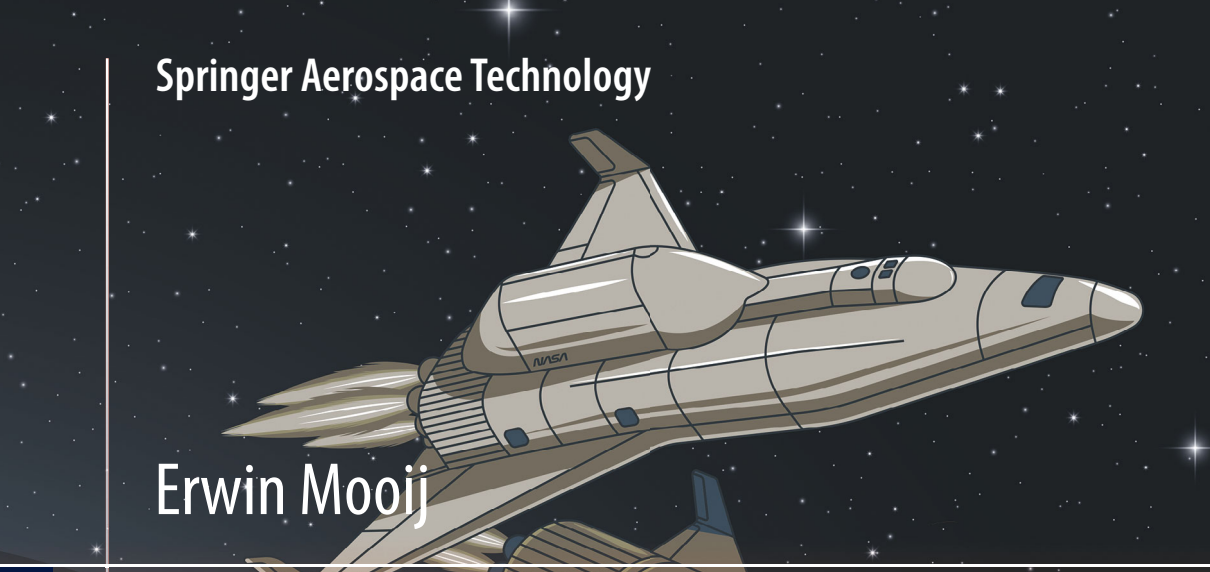
Springer Aerospace Technology



Erwin Mooij

Re-entry Systems

Springer

Springer Aerospace Technology

Series Editors

Sergio De Rosa, DII, University of Naples Federico II, Napoli, Italy

Yao Zheng, School of Aeronautics and Astronautics, Zhejiang University, Hangzhou, Zhejiang, China

Elena Popova, Air Navigation Bridge Russia, Chelyabinsk, Russia

The series explores the technology and the science related to the aircraft and spacecraft including concept, design, assembly, control and maintenance. The topics cover aircraft, missiles, space vehicles, aircraft engines and propulsion units. The volumes of the series present the fundamentals, the applications and the advances in all the fields related to aerospace engineering, including:

- structural analysis,
- aerodynamics,
- aeroelasticity,
- aeroacoustics,
- flight mechanics and dynamics
- orbital maneuvers,
- avionics,
- systems design,
- materials technology,
- launch technology,
- payload and satellite technology,
- space industry, medicine and biology.

The series' scope includes monographs, professional books, advanced textbooks, as well as selected contributions from specialized conferences and workshops.

The volumes of the series are single-blind peer-reviewed.

To submit a proposal or request further information, please contact: Mr. Pierpaolo Riva at pierpaolo.riva@springer.com (Europe and Americas) Mr. Mengchu Huang at mengchu.huang@springer.com (China)

The series is indexed in Scopus and Compendex

Erwin Mooij

Re-entry Systems

Springer

Erwin Mooij
Faculty of Aerospace Engineering
Delft University of Technology
Delft, Zuid-Holland, The Netherlands

ISSN 1869-1730 ISSN 1869-1749 (electronic)
Springer Aerospace Technology
ISBN 978-3-031-62173-4 ISBN 978-3-031-62174-1 (eBook)
https://doi.org/10.1007/978-3-031-62174-1

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

This book has evolved over the years that passed since 2007 during which I have been teaching the M.Sc. course "Re-entry Systems" at Delft University of Technology. Initially, I primarily discussed the motion of vehicles entering the atmosphere, thereby focussing on the three entry mechanisms of ballistic, gliding, and skipping flight. Aeroheating is, of course, a major constraint on entry flight, so this topic was included right from the beginning. In the years after, I added lectures on guidance, navigation, and control, as well as entry, descent, and landing systems. Each year a single lecture was dedicated to a "special topic", which included the terminal-area energy management of winged vehicles, like the Space Shuttle, advanced guidance systems that enable precision landing, and a brief overview of reusable thermal-protection systems.

The layout of the book follows the current lecture series, and only serves as background material for the students. It is obvious that the material is far too much for a single course, but from experience, the lecture notes have served many students that have done their thesis in one of the covered fields. The chapters on aeromanoeuvring and thermal-protection systems deal with topics that are not taught in class. Some of the intermezzos and sections in the book have an asterisk "*" as a prefix. The material discussed there goes beyond the standard level, and provides additional information. In the text, I have tried to be consistent in the use of British English and SI units. Where deemed necessary, I have also included the US units of measure.

The enthusiasm and motivation of the 70+ students annually in class, as well as the many M.Sc. students doing their final thesis on related topics, have kept me going all these years, to formalise the lecture notes into something more substantial. I am very grateful to all of them, but want to mention a few in particular who provided me with input for several parts of the book. In no particular order, I want to acknowledge Sven de Ridder (terminal area energy management), Alexander Ronse (orbital decay and entry of rocket stages), Michele Facchinelli (aerobraking), Enrico Zucchelli and Jorrik Engelsma (aerocapture), Jaimy Hess (aerogravity assist), Ingo Gerth (E-guidance), and Svenja Woicke (she wrote the first version of the section on advanced landing systems). Svenja was also a guest lecturer and, during her Ph.D. studies, taught me a lot about this special topic, simply because "I know this,

because I studied this, and you didn't". And she was very right, of course! Last, but definitely not the least, I want to mention my friend and long-time collaborator Kees Sudmeijer, who sadly passed away recently: *Kees, the stimulating and sometimes 36 heated discussions we have had during the 30 years that we have known each other and worked together, have certainly left their mark on the book*. Thank you!

I am also indebted to my colleagues at university, especially Ron Noomen and Dominic Dirkx. Ron was also my M.Sc. thesis supervisor when I was struggling at ESTEC to get my Huygens simulations running. *I will never forget our meetings of only a few seconds at the bottom of the stairs in Leiden station. In passing by you handed me an envelope with copied chapters from mathematics books that I could not get otherwise.* I always felt like a spy receiving top-secret information. His generous donation of a few of his exam questions that appear as questions in this book is greatly appreciated. Dominic, you were first my student, and later we were teaching classes on astrodynamics and trajectory optimisation together. How time flies. Despite your best efforts to tell me "no, don't write that book", after we had successfully (but after a painful process) published your M.Sc. thesis as a book, also with Springer, I chose to ignore you and make fun of you, which became a mutual habit of ours, to great joy of our students in class.

I am very much indebted to my (former) students Michaela Brchnelova for her valuable input to the chapter on aerothermodynamics, and Chloé Carer, who read through all problems and answers ("I still have nightmares from these derivations"), and my good friend and former colleague Dr. Derek Gransden, currently at Carleton University (Ottawa, Canada). Derek proofread the whole book and saved me from some embarrassing errors and typos. You did a fantastic job, and, of course, all remaining errors and typos are my own. You even managed to spot the autocorrect typo "dust brains" when I was talking about storing the solar-dust grains in the sample container of the Genesis spacecraft. The simple annotation "Zombie?" said it all! But, seriously, thank you for all you did.

Finally, I want to thank my daughter Jeanine, who made many of the Illustrator drawings (including the cover), and, along the way, taught me a few things such that I did not have to bother her so much. Her creative input has made this book look a lot better, and, yes, you were right with "You *really* have to change that font to make the figures more consistent with the text". Thank you for all the help, and for enduring my grumpy complaining about the lack of progress at dinner time.

Delft, The Netherlands Erwin Mooij
April 2024

Contents

1 Introduction 1
1.1 Early History 1
1.2 Key Persons 8
1.3 The Space Age 33
1.3.1 Setting the Stage 33
1.3.2 The Capsules 37
1.3.3 Gliding Flight 50
1.3.4 Visiting Alien Worlds 65
1.4 Layout of the Book 76
References 80

2 Planetary Entry Environment 83
2.1 Planetary Shape 85
2.1.1 Regular Bodies 85
2.1.2 Surface Models 88
2.1.3 Irregular Bodies 96
2.2 Gravity Field 99
2.2.1 Spheroid Bodies 100
2.2.2 Irregular Bodies 113
2.3 Atmospheres 115
2.3.1 Exponential Model 121
2.3.2 Earth 125
2.3.3 Mars 137
2.3.4 Titan 145
2.3.5 Venus 148
References 155

3 Aerothermodynamics 157
3.1 Basics 159
3.1.1 Flow-Field Parameters 159
3.1.2 Similarity Parameters 164
3.1.3 Conservation Laws 168

3.1.4 Forces and Moments . . . 168
3.1.5 Flow-Field Description . . . 171
3.1.6 Numerical and Experimental Techniques . . . 176
3.2 Low-Speed Aerodynamics . . . 192
3.2.1 Subsonic Flow . . . 192
3.2.2 Transonic Flow . . . 194
3.3 Supersonic Flow . . . 195
3.3.1 Introduction . . . 196
3.3.2 Normal Shock . . . 199
3.3.3 Oblique Shock . . . 206
3.3.4 Expansion Wave . . . 212
3.3.5 Boundary Layer . . . 217
3.4 Hypersonic Flow . . . 226
3.4.1 Flow Properties . . . 227
3.4.2 Boundary Layer . . . 230
3.4.3 Flow Complications . . . 234
3.4.4 Local Inclination Methods . . . 249
3.5 Free-Molecular Flow . . . 271
3.6 Aerodynamic Heating . . . 283
3.6.1 Boundary-Layer Dynamics . . . 286
3.6.2 Convection . . . 292
3.6.3 Radiation . . . 300
3.6.4 Heat Balance . . . 307
3.7 *Navier-Stokes Equations . . . 307
References . . . 317

4 Fundamentals of Motion . . . 321
4.1 Introduction . . . 321
4.1.1 Classical Mechanics . . . 321
4.1.2 Reference Frames . . . 323
4.2 State-Variable Definitions . . . 327
4.2.1 Position and Velocity . . . 327
4.2.2 Attitude and Angular Rates . . . 331
4.3 Frame Transformations . . . 335
4.3.1 Unit-Axis Rotations . . . 335
4.3.2 *Euler's Rotation Theorem . . . 337
4.3.3 Standard Frame Transformations . . . 341
4.4 Coordinate Transformations . . . 353
4.4.1 Position and Velocity . . . 353
4.4.2 Attitude . . . 356

4.5 General Formulation of the Equations of Motion 358
4.6 *Equations of Motion in Spherical Coordinates 363
4.6.1 General 363
4.6.2 Wind Equations 365
References 373

5 Planar Motion 375
5.1 Introduction 375
5.2 Equations of Motion 378
5.3 Ballistic Entry 383
5.3.1 Flight Mechanics 385
5.3.2 Thermal Loads 396
5.3.3 Summary 402
5.3.4 Case Study: Huygens Entry 403
5.3.5 Case Study: Impact of Atmosphere-Model Choice on Mission Design 411
5.4 Gliding Entry 422
5.4.1 Flight Mechanics 424
5.4.2 Thermo-Mechanical Loads 433
5.4.3 Entry Corridor 440
5.4.4 Footprint 446
5.4.5 Case Study: Horus Entry 458
5.5 Skipping Entry 465
5.5.1 Introduction 465
5.5.2 Flight Mechanics 470
5.5.3 Thermo-Mechanical Loads 489
5.5.4 Numerical Verification of an Aerocapture Mission 498
5.5.5 Case Study: Apollo Entry 502
5.6 *Second-Order Approximation 508
5.6.1 Equations of Motion 508
5.6.2 Verification of an Aerocapture Mission—Revisited 519
References 530

6 Characteristic Motion of Entry Vehicles 533
6.1 Non-linear Equations of Motion 534
6.2 Eigenvalues and Eigenmotion 538
6.3 Linearisation of Equations of Motion 555
6.4 Eigenmotion of Winged Entry Vehicles 563
6.4.1 State-Space Model 563
6.4.2 Nominal Mission 578
6.4.3 HORUS Characteristic Motion 582
6.5 Eigenmotion of Entry Capsules 606

6.5.1 Apollo Nominal Mission . . . 606
6.5.2 State-Space Model . . . 611
6.5.3 Apollo Characteristic Motion . . . 613
6.5.4 Huygens Characteristic Motion . . . 625
References . . . 632

7 Guidance, Navigation, and Control . . . 633
7.1 Introduction . . . 633
7.2 Sensors . . . 638
7.2.1 Inertial Measurement Unit . . . 639
7.2.2 Global Positioning System . . . 643
7.2.3 Other Sensors . . . 648
7.3 Navigation . . . 651
7.3.1 Linear Kalman Filter . . . 652
7.3.2 Extended Kalman Filter . . . 657
7.3.3 Integrated Navigation System . . . 662
7.3.4 Case Study: HORUS Navigation System . . . 664
7.4 Actuators . . . 683
7.4.1 Control Surfaces . . . 684
7.4.2 Reaction Control System . . . 688
7.4.3 Retro Rockets . . . 690
7.4.4 Other Actuators . . . 693
7.5 Control . . . 697
7.5.1 Introduction . . . 697
7.5.2 State and Output Feedback Control . . . 703
7.5.3 PID Control . . . 708
7.5.4 Optimal Control . . . 713
7.5.5 Incremental Non-linear Dynamics Inversion . . . 735
7.5.6 Simple Adaptive Control . . . 738
7.5.7 Case Study: HORUS Attitude Controller Design . . . 741
7.6 Guidance . . . 756
7.6.1 Introduction . . . 756
7.6.2 Guidance System Space Shuttle . . . 759
7.6.3 Other Guidance Principles . . . 766
7.6.4 Case Study: Heat-Flux Tracking . . . 774
References . . . 794

8 Terminal Area Energy Management . . . 799
8.1 Introduction . . . 799
8.2 Flight Mechanics . . . 802
8.2.1 Non-linear Equations of Motion . . . 803
8.2.2 Flat, Non-rotating Earth . . . 805
8.2.3 Steady-State Approximation . . . 807
8.2.4 Diving Flight in the Vertical Plane . . . 812
8.2.5 Energy-State Approximation . . . 813

8.2.6 Steady Coordinated Turn 816
8.3 Terminal Area Guidance Algorithms 820
8.3.1 Introduction 820
8.3.2 Space Shuttle 822
8.3.3 Buran 826
8.4 Case Study: Terminal-Area Flight of HORUS 829
8.4.1 Vehicle Characteristics 831
8.4.2 Steady-State Approximation 835
8.4.3 Energy-State Approximation 837
8.4.4 Optimal Trajectories in the Vertical Plane 839
8.4.5 Maximum-Dive Trajectories in the Vertical Plane 844
8.4.6 Energy-Tube Concept 846
8.4.7 Influence of a Turn on the Optimal Trajectories 849
8.4.8 Conclusions 852
References 855

9 Entry, Descent, and Landing Systems 857
9.1 History 859
9.2 Parachute Systems 873
9.2.1 Introduction 873
9.2.2 Hardware Elements 876
9.2.3 Parachute Types 877
9.2.4 Materials 885
9.2.5 Operational Aspects 889
9.2.6 Parachute Dynamics Simulation 904
9.3 Powered Descent 922
9.3.1 Introduction 922
9.3.2 Gravity Turn 927
9.3.3 E-Guidance 941
9.3.4 Terminal Descent 950
9.4 Other Decelerators 957
9.4.1 Drag Skirt 958
9.4.2 Ballute 958
9.4.3 Inflatable Aerodynamic Decelerator 962
9.5 Advanced Descent and Landing Systems 971
9.5.1 Landing Site Selection 972
9.5.2 Integrated Navigation Approach 980
9.6 Landing Systems 992
9.6.1 Sea Landing 994
9.6.2 Shock-Attenuation Techniques 1001
9.7 Missions 1025
9.7.1 Surveyor and Apollo Lunar Lander 1025

9.7.2 Viking 1034
9.7.3 Galileo 1039
9.7.4 Stardust 1044
9.7.5 MER-A and B 1049
9.7.6 Mars Science Laboratory 1054
9.8 Case Study: Huygens 1060
9.8.1 Mission and System 1061
9.8.2 Simulation Model 1063
9.8.3 Results 1064
References 1076

10 Thermal Protection Systems 1085
10.1 Historical Development 1086
10.1.1 The Early Days 1086
10.1.2 Winged Vehicles 1090
10.1.3 Planetary Probes 1103
10.2 Design Aspects 1112
10.2.1 TPS Requirements 1112
10.2.2 Aerothermodynamics 1114
10.3 Materials 1124
10.3.1 Material Types 1126
10.3.2 Characteristics 1140
10.4 Design Options 1158
10.4.1 Characterisation 1158
10.4.2 Passive Systems 1159
10.4.3 Semi-Passive Systems 1164
10.4.4 Active Systems 1168
10.5 TPS Modelling 1176
10.5.1 General Approach 1176
10.5.2 TPS Mass Estimation Relationships 1180
10.5.3 Ablation Modelling 1181
10.5.4 Active-Cooling Modelling 1195
10.5.5 Finite-Difference Model 1199
10.6 Case Study 1221
References 1251

11 Aero-assisted Manoeuvres 1259
11.1 Aerobraking 1260
11.1.1 Concept and Missions 1260
11.1.2 Theory 1268
11.1.3 Guidance 1275
11.2 Aerocapture 1284
11.2.1 Concepts and Missions 1284

11.2.2 Characteristic Trajectories 1288
11.2.3 Analytic Predictor-Corrector Guidance 1293
11.2.4 Optimal Aerocapture 1298
11.2.5 Entry Corridor 1311
11.2.6 Entry-Corridor Variations 1320
11.3 Aerogravity Assist 1324
11.3.1 Introduction 1324
11.3.2 Gravity Assist 1327
11.3.3 Atmospheric Passage at Constant Altitude 1344
11.3.4 Feedback Guidance 1364
11.3.5 Case Study: Aerogravity Assist at Mars 1375
References 1390

Answers 1395

Appendix A: General Formulation of the Equations of Motion 1467

Appendix B: Reference Vehicles 1483

Appendix C: Mathematical Tools and Techniques 1515

Index 1541

Chapter 1
Introduction



1.1 Early History

To write about the (early) history of entry vehicles is by no means as rewarding as writing about the history of flight, or in general terms, flying. Who has not heard the tale of Daedalus and Icarus in Greek mythology? Daedalus was a skilful craftsman and artisan, who had constructed the famous Labyrinth of Knossos in Crete to contain the gruesome minotaur. To prevent his knowledge of his Labyrinth from spreading to the public, King Minos had locked Daedalus in a tower. Since Minos controlled the land and sea routes, Daedalus could not escape from Crete and began to construct wings for himself and his young son, Icarus. As discussed in detail in the Metamorphoses, one of the works of the Roman poet Ovid (born 43 B.C.), Daedalus tied feathers together, from smallest to largest so as to form an increasing surface. He secured the feathers at their midpoints with string and at their bases with wax, and gave the whole a gentle curvature like the wings of a bird (Fig. 1.1).

After successfully trying the wings he fixed the wings also to his son and taught him to fly, thereby warning him not to fly too high, because the heat of the Sun would melt the wax, nor too low, because the sea foam would soak the feathers. Well on their way to escape King Minos, Icarus became overly confident and began to soar upward toward the Sun. The blazing Sun softened the wax which held the feathers together and they came off. Icarus fell into the sea and drowned.

The myth of Icarus seems to be quite farfetched, although at a more Earthly scale a young Benedictine monk named Eilmer also fastened wings to his arms and his feet, and climbed to the top of the tower of Malmesbury Abbey, England, with the idea of flying like a bird (White 1961). It was the year 1010. Eilmer had been fascinated by the flight of jackdaws for many years, and his bird-like wings were constructed of ash or willow-wand, covered with a light cloth, and attached to pivots on either side of a back-brace, with handholds so he could flap the wings (Fig. 1.2). He jumped from the tower, and using the southwest wind he glided down into the small valley by the river Avon. Towards the end of the flight he panicked, lost control and came

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_1



Fig. 1.1 Sculpted panel on the base of Giotto's Bell Tower in Florence, Italy, depicting the myth of Daedalus. *Image source* Archivi Alinari, Florence

down quite hard, breaking both his legs. He had covered a distance of about 600 feet, 150 feet lower than his starting point. Soon after he recovered he realised the shortcomings of his design—the absence of a tail—and was planning a second flight. However, the abbot forbade him to continue this dangerous endeavour, and that was the end of Eilmer's flying career.

It is stories like these that emphasise man's desire to fly, initially only through the skies to be free as a bird, but later for more practical reasons as transportation and military advantage. Discussing the history of entry vehicles immediately raises the question of how this entry vehicle got outside of the atmosphere in the first place. Therefore, in this historical context, it is logical to talk a little about the history of rockets as well. And in that case one is right to talk about the flying monk, since trying to fly like a bird inspired even the great Leonardo da Vinci, who in turn inspired many many others who all contributed to the state of aerospace engineering as it is now. So, back to where we left of, the history of flying.

Evliya Çelebi, an Ottoman Turkish traveler and writer, wrote about two brothers, Hezârfen Ahmed and Lagari Hasan Çelebi, who were both legendary Ottoman aviators of 17th-century Istanbul. Hezârfen Ahmed is said to have achieved sustained unpowered flight with a winged apparatus circa 1630–1632. He had practiced his

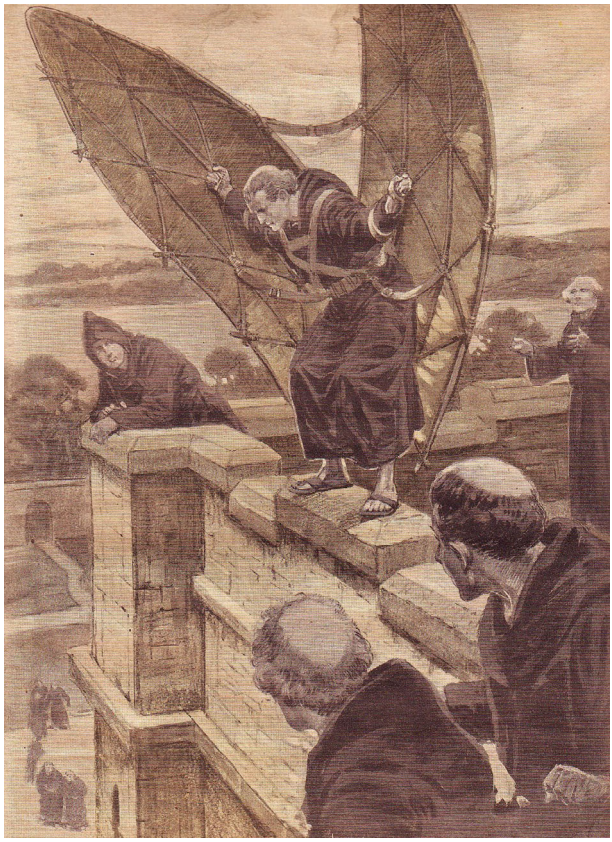
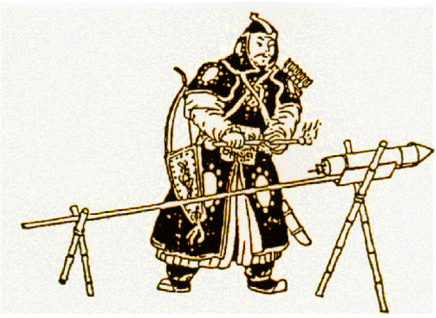


Fig. 1.2 Artist impression of Eilmer, the flying monk, just before his pioneering flight. *Image source* Unknown

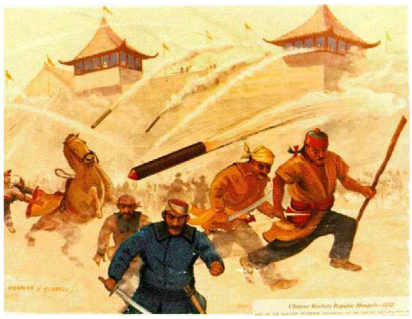
flying several times over the Okmeydani neighborhood of Istanbul, before he flew, with the help of a south-west wind, from the very top of the Galata Tower and landed in the Doğancilar square in ÜskÜdar. He covered a distance of 3,358 m over an elevation change of 85.6 m, which took him right over the Bosphorus. The sultan (on his own accord or forced by his advisors), initially impressed by the feat, later on saw him as a threat and exiled him to Algeria. There, he died heartbroken at the age of 31 (Harding 2006).

In 1633, Hezârfen's brother Lagari Hasan launched in a 7-winged rocket using 50 okka (1 okka is 1282 g) of gunpowder from Sarayburnu, the point below Topkapi Palace in Istanbul. He lifted off and rose high in the sky, before landing in the sea using a simple parachute. The sultan, mightily impressed, rewarded Lagari with silver and the rank of Sipahi (member of the cavalry) in the Ottoman army.

Even though Lagari was the first official "rocket man", the idea to use of gunpowder to propel objects through the sky was a very old one. The discovery of gun powder



(a) Drawing of an early Chinese soldier lighting a rocket. Drawing courtesy of NASA.



(b) Battle of Kai-Fung-Fu. Image by Charles Hubbell.

Fig. 1.3 Examples of Chinese rocketry

dates back to ninth-century China, attributed to Chinese alchemists searching for an elixir of immortality. This discovery led to the invention of fireworks and the earliest gunpowder weapons in China. The Chinese history in the use of primitive rockets spans centuries, notably armies fired flaming rockets at enemies, see Fig. 1.3a. In the Battle of Kai-fung-fu in 1232. Chinese troops used bamboo casings loaded with gun powder mixed with iron shrapnel, and aimed them as projectiles against an invading Mongol army, see Fig. 1.3b. Eyewitnesses mentioned the enormous thunderous sound that could be heard for about 25 km, and upon impact it devastated everything in all directions for 600 m.

Even though they were defeated, the Mongols learned from this and the warriors quickly included the Chinese fire arrow in their arsenal. They employed these rockets in various battles during invasions of India, the Middle East, and portions of Europe in the 13 and 14th centuries, e.g., on February 15, 1258, when Mongol invaders used rockets to capture Baghdad. It is commonly believed that these actions were the primary cause for the spread of rocket technology out of China.

The first use of gun powder in a space application also originates from China, although it is more legend than fact. Around 1500, a Chinese official called Wan Hu was so much captured by the beauty of the Moon that he decided to go there, taking advantage of China's advanced rocket and fireworks technology. To achieve his dream he fitted 47 powder rockets to a bamboo framework connecting a pair of large kites (some sources claim that there was a second set of rockets that would ignite at a later stage, making Wan Hu's design the first multi-stage rocket). On the day of lift-off, Wan, splendidly attired, climbed into his rocket chair. His instructions to his 47 servants were very clear: to simultaneously lite the fuses and, if they cared for their own lives, hastily run for cover. And so it happened (Fig. 1.4). When the rockets were ignited there was a huge explosion, and a thick smoke obscured Wan Hu and his chair from view. After the smoke had cleared, Wan and the chair were gone, and were said never to have been seen again.



Fig. 1.4 Chinese magistrate Wan Hu on his chair, propelled by 47 rockets. *Image* Civil Air Patrol

Whether this story is indeed more fiction than fact does not really matter. In Chinese culture Wan Hu is still extremely popular and is credited with being the first to attempt a manned rocket flight (and thus being the first astronaut). In honour of this attempt a crater on the far side of the Moon has been named Wan-Hoo.

Kazimierz Siemienowicz (born c. 1600–c. 1651), was a Polish-Lithuanian general of artillery, gunsmith, military engineer, artillery specialist and pioneer of rocketry. After contributing his expertise to several battles, Siemienowicz published *Artis Magnae Artilleriae pars prima* (Great Art of Artillery, the First Part) in 1650. This treatise, which included discussions on rocketry and pyrotechnics, remained a standard work in these fields for two centuries. The book provided the standard designs for creating rockets, fireballs, and other pyrotechnic devices. It discussed for the first time the idea of applying a reactive technique to artillery. It contains a large chapter on caliber, construction, production and properties of rockets (for both military and civil purposes), including multistage rockets (see Fig. 1.5), batteries of rockets, and rockets with delta wing stabilisers. According to rumour, Siemienowicz was killed by members of the metallurgy/gunsmith/pyrotechnics guilds, who were opposed to him publishing a book about their secrets, and that they hid or destroyed the manuscript of the second part.

Of course, military applications dominated the development of rockets. In 1780 the British began to annex the territories of the Sultanate of Mysore in southern India, during the Second Anglo-Mysore War. Although rocket technology had made

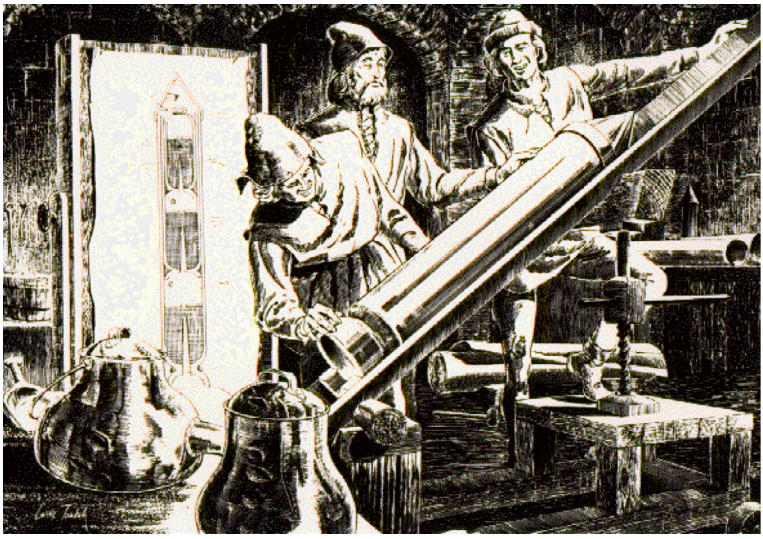


Fig. 1.5 Artillery expert, Kazimierz Siemienowicz, published a series of drawings for a staged rocket. *Image* Motorola Inc.

its way to India and Europe by the 13th century, development of accurate cannons had sidelined rockets as a military technology in Europe. But not in India ... Hyder Ali, the *de facto* ruler of Mysore, built on the experience of his father in the use of rockets, improved them and significantly expanded their use. Technological innovations included the use of high-quality iron casing (better than what was then available in Europe) for the combustion chamber, enabling the use of higher-powered explosive charges. By effectively utilising Mysorean rockets and rocket artillery against the closely massed British forces, the British were defeated during the Battle of Guntur.

Learning from this adverse experience the British military engineers, led by Sir William Congreve, improved the design of captured Indian rockets and developed a series of more efficient bombardment rockets. They were used against the very Indians that had provided them with the basic technology. Hyder Ali's son Tipu Sultan greatly expanded the use of rockets after Hyder's death, deploying as many as 5,000 rocketeers at a time. These rockets deployed by Tipu during the Battle of Pollilur were much more advanced than the British East India Company had previously seen, chiefly because of the use of iron tubes for holding the propellant; this enabled higher thrust and longer range for the missile (up to 2 km range). But, as British accounts describe the use of the rockets during the third and fourth Anglo-Mysorean wars, during the final confrontation at the Siege of Seringapatam (1799) the British breached the walls of the fortress, killed Tipu Sultan and conquered the kingdom.

Perhaps the most famous application of Congreve's rockets was the British bombardment of the American Fort McHenry (Baltimore) during the War of 1812 (which



Fig. 1.6 The bombardment of Fort McHenry in 1814. *Image* public domain

lasted until 1815), Fig. 1.6. In 1814, British ships, which had withdrawn to just beyond the range of Fort McHenry's cannons, bombed the fort for 25 hours. Despite the large number of rockets and cannon balls launched at the fort, damage was light due to recent fortification that had been completed prior to the battle. Also the fact that Congreve's rockets were not very accurate contributed to the partially failed bombardment. In fact, this battle was the turning point in the war in favour of the Americans.

William Hale (21 October 1797–30 March 1870), born in Colchester, England, was a self-taught British inventor and rocket pioneer. Building on the earlier Congreve rocket design, Hale patented a new form of rotary rocket in 1844. To improve its stability in flight, he eliminated the guide stick from the design and instead vectored part of the thrust through canted exhaust holes to provide rotation of the rocket, a technique called spin stabilisation. Hale's rockets could have a mass of up to 60 pounds and were noted for their noise and glare on ignition. Soon after he had patented his rocket design, he sold the manufacturing rights to the United States government. They were subsequently used by the US army during the Mexican-American war (1846–1848) and later also during the US Civil War, albeit with mixed results. During the Civil War, the Confederate States of America also did their own development. They came with a 12-foot long rocket, with 10 pounds of gunpowder and the mark C.S.A. on its brass case. This rocket intended to be first ballistic missile. It was launched from Richmond, Virginia, with Washington, D.C., as target, and seen to roar rapidly up and out of sight. Alas (well, fortunately for those in Washington), no one ever saw the rocket land.

Of course, the history of rocket development does not stop here. And, in fact, we have not even touched upon re-entry vehicles yet. However, the progress to come is mainly made due to a limited number of pioneering giants: the Russian Konstantin Eduardovich Tsiolkovsky (1857–1935), the American Robert Hutchings Goddard (1882–1945), and the German Hermann Julius Oberth (1894–1989), each one considered to be the father of modern rocketry. Be that as it may, their developments would not have been possible without someone like, for instance, Isaac Newton. Therefore, in the next section we will introduce a number of (personally selected) key persons who have shaped the aerospace field in general, and the development of (manned) spaceflight in particular.

1.2 Key Persons

The development of rockets and re-entry systems has been a long process, both from a theoretical point of view and from the technological side. Over the centuries many people have contributed to this development, but without a limited number of key persons this development would not have been possible. This section introduces key people and discusses their major contributions to the field. Obviously, it is impossible to include everybody that in one way or the other contributed to the aerospace field. Therefore, this section is based on a (possibly biased) personal selection.

The discussion begins with one of the most brilliant inventors of all time, i.e., Leonardo da Vinci (Fig. 1.7). Born on April 15, 1452, in or around Vinci, a small town some 40 km west of Florence, Italy, Leonardo was the illegitimate son of a notary, Piero da Vinci, and a peasant woman named Caterina. Even though Leonardo was and is renowned primarily as a painter, our interest is, of course, his ideas and designs of human-powered flight, and his flying machines.

As a young boy, in 1466 he left Vinci and moved to Florence, where he started working in the workshop of Andrea del Verrocchio. This workshop produced many of the theatrical machineries used to astonish the public. Among the special effects were all kinds of flying contraptions. Due to the volume and variety of the work, Leonardo evolved many skills, including drafting, chemistry, metallurgy, metal working, plaster casting, leather working, mechanics and carpentry as well as the artistic skills of drawing, painting, sculpting and modelling. At the same time, he also developed a passion for the animal kingdom, most notably the flight of birds. He applied this passion to work on winged dragons and other types of fantastic creatures, which were often portrayed in Florentine art. Leonardo's drive to work on flying machines was further stimulated by the early work of Siennese engineers, who were, amongst others, one of the first to deal with man's movement through air on at least two occasions, proposing a parachute and two very unlikely wings (Laurenza 2004).

In 1483, he moved to Milan to work for Ludovico il Moro, the Duke of Milan and patron to many artists (he presided over the final and most productive stage of the Milanese Renaissance, and is probably best known as the man who commissioned the The Last Supper). Da Vinci strengthened and developed his theoretical studies,



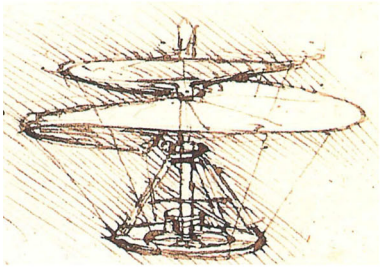
Fig. 1.7 Leonardo da Vinci (1452–1519), undated self portrait

especially in the fields of anatomy and mechanics. This proved in particular useful for the study on human flight. About ten years after he reached Milan, Leonardo did an immense volume of research on flying machines (e.g., Fig. 1.8a, c), culminating in the design of his so-called "flying ship", also known as the ornithopter (Fig. 1.8b). One of the conclusions of his theoretical ponderings stated that the force that an object exerts upon the air is the same as the force exerted by the air on that object. Later, Newtonian "aerodynamic reciprocity" confirmed that, i.e., "*flight is a purely mechanical phenomenon, a result of the dynamic action between wing and air*" (Laurenza 2004).

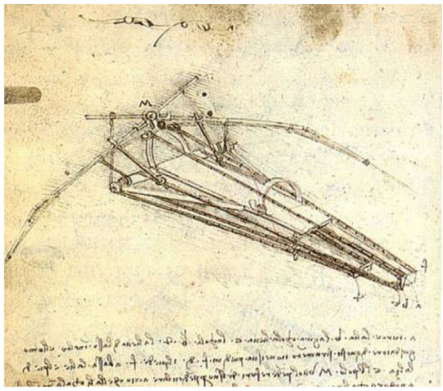
In 1500, Leonardo returned to Florence where he continued to work on his flying machines. However, studies of the animal world dominated this period. In fact, his observations on the flight of birds took the foreground. His studies on human-powered flight were more and more directed towards reproducing natural flight. Moreover, at that time he called his flying machine "bird"! Even during the last years of his life he did not abandon the idea of human flight, but completely focussed on the theoretical aspects; he did not try to build something (anymore), even though some simple experiments were all related to flight and flying. In 1516, he entered the service of the French king, and he moved to France where he spent the last three years of his life. Leonardo died on May 2, 1519. He was buried in the Chapel of Saint-Hubert in Château d'Amboise.

Relatively few of Leonardo's designs were constructed or were even feasible during his lifetime. Although he made important discoveries in anatomy, civil engineering, optics, and hydrodynamics, he did not publish his findings and they had no direct influence on later science. His conceptual drawings, though, have inspired many generations in the past and will undoubtedly continue to do so in the future.

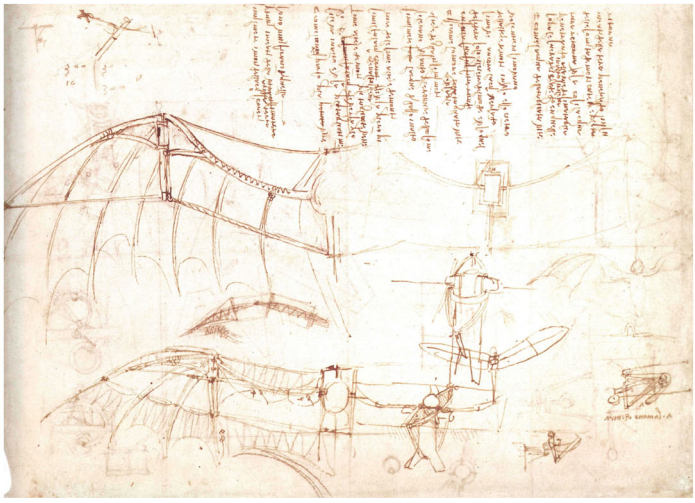
Nicolaus Copernicus (1473–1543), Fig. 1.9a, was born in Poland and studied mathematics and optics in Italy. In 1530, he published *De Revolutionibus*, in which



(a) The 'aerial screw'.



(b) Ornithopter.

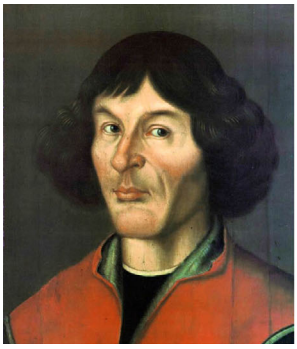


(c) Flying machine.

Fig. 1.8 Several of Leonardo da Vinci's designs. *Images* public domain

he stated that the Sun, not the Earth, was the centre of our Solar System.[1] This was contrary the Ptolomaic view on the world that put the Earth in the centre of the universe. The Greek-Roman citizen of Egypt, Ptolemy (c. AD 90–c. AD 168), mathematician and astronomer, amongst others, explained the motions in terms of orbits (epicycles) carried on a larger orbit (deferent). Even though this system was

[1] Copernicus was not the first one to propose the heliocentric model. In as early as the third century BC, Aristarchus of Samos, an ancient Greek astronomer and mathematician, presented the first known model that placed the Sun at the centre of the known universe with the Earth revolving around it. He also put the other planets in their correct order of distance around the Sun.



(a) Nicolaus Copernicus, portrait from Town Hall in Thorn (1580)



(b) Portrait of Galileo Galilei by Justus Sustermans painted in 1636



(c) Portrait of Johannes Kepler by an unknown artist (1610)

Fig. 1.9 Nikolaus Copernicus, Galileo Galilei, and Johannes Kepler. *Images* public domain

conceived many centuries ago, the epicycle deferent ratios were very close to modern values of planet/Earth orbit ratios, and even though the planets appeared to reverse motions at times, it worked very well. In fact, contrary to popular myths, Ptolemy's system was not overly cumbersome, and it accounted for subtleties like the uneven motion of the Sun. And imagine, Ptolemy did such a good job that it took 1500 years to improve on him!

Copernicus claimed that the Earth rotated about its axis once per day, and traveled around the Sun once each year. The accuracy of his model was no better than that of Ptolemy, though, because both were using the same observational data. It is not known how he arrived at this controversial heliocentric idea, and he died just as the theory was published. Nevertheless, it was the start of a change in the way the world was viewed, and Copernicus became known as the initiator of the Scientific Revolution.

One of the earliest supporters of so-called Copernicanism was the Italian physicist, mathematician, astronomer, and philosopher Galileo Galilei (1564–1642). Galileo was born in Pisa, Italy, and as a young man he seriously considered to become a priest. His father urged him to enroll at the University of Pisa for a medical degree instead. When he was studying medicine, once upon a day in 1581 he noticed a swinging chandelier moving under the influence of air currents. It seemed that, no matter the amplitudes of the arcs, the swinging back and forth would always take the same amount of time. Once back at home, he set up two pendula of equal length and swung them with different sweeps, and indeed confirmed that they kept time together.

Because his father really wanted him to become a physician (who could make a good deal of money), he deliberately kept Galileo away from mathematics. However, Galileo accidentally attended a lecture on geometry and decided he wanted to study mathematics and natural philosophy. His father, though reluctant, finally let him.

After successful completion of his studies, in 1589, he was appointed to the chair of mathematics in Pisa and stayed there until 1592. He then moved to the University of Padua, to teach geometry, astronomy and mechanics until 1610. During this period Galileo made significant discoveries in both pure fundamental science and the practical applied science. Examples of the former covered the areas of kinematics of motion (e.g., the path of a projectile is a parabola) and astronomy. Galileo discovered not only the four largest moons of Jupiter, i.e., Io, Europa, Ganymede, and Callisto, in 1610, but also the rings of Saturn and the Milky Way. The practical applied science included strength of materials and improvements to the telescope. Note that contrary to what some people believe, he did *not* invent the telescope.

It was already in 1597 when Galileo indicated his belief in the Copernican System, and did so in a letter to Johannes Kepler, the famous German astronomer. This belief (and his public expression of it) would lead to an almost lifelong battle with the Catholic church, which claimed it was bordering heresy. In 1616, Galileo was officially warned by the Church not to hold or defend the Copernican System, and his book *De revolutionibus orbium coelestium* was placed on the List of Prohibited Books. However, Galileo remained true to his beliefs, although at his trial later he denied that he had done so. He also "mocked" and contradicted Aristotelian physics, which the Church believed to be true. In 1633, the Inquisition sentenced Galileo to imprisonment for vehement suspicion of heresy, but this sentence was commuted to house arrest. He continued to work, but all his writings were put on the List of Prohibited Books. This included the *Discorsi e dimostrazioni matematiche, intorno à due nuove scienze* (Discourses and Mathematical Demonstrations Relating to Two New Sciences), which was Galileo's final book (1638) and a sort of scientific testament covering much of his work in physics over the preceding thirty years. It was published in The Netherlands, where the Inquisition could not cause havoc. Galileo remained a prisoner until his death in 1642.

Galileo's most original contributions to science were in mechanics: he helped clarify concepts of acceleration, velocity, and instantaneous motion, and thus became one of the "giants" that Newton refers to when claiming that he could only have seen further, because he stood on the shoulders of giants. However, Galileo was not always right. For instance, he dismissed the idea, held by his contemporary Johannes Kepler, that the Moon caused the tides as "useless fiction": he claimed the tides were caused by the sloshing back and forth of water in the seas, because of the Earth's rotation on its axis and revolution around the Sun. He also refused to accept Kepler's elliptical orbits of the planets, considering the circle the "perfect" shape for planetary orbits.

Johannes Kepler (1571–1630), Fig. 1.9c, born in Germany, started his career as a mathematics teacher at a seminary school in Graz, Austria, where he became an associate of Prince Hans Ulrich von Eggenberg. From 1600 onwards, he studied with the Danish astronomer Tycho Brahe in Prague (until Brahe's unexpected death in 1601), and used Brahe's data on observations of Mars to determine its orbit. He did so because the Ptolomaic data were very old and not so accurate. Initially, Kepler was reluctant to abandon perfectly circular motion (but despite his mystic tendencies, he did when the evidence required it). Kepler determined that the orbits of planets could be drawn as ellipses with the Sun always at one of the foci (and the other one empty).

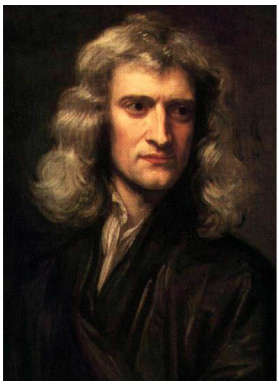
This postulate became known as Kepler's First Law. Together with the Second Law, a line connecting a planet and the Sun will sweep over equal areas in equal times as the planet moves about its orbit, the First Law was published in *Astronomia nova* and Kepler was the first to demonstrate that the Copernican model is more accurate than the Ptolemaic one for uses, such as navigation and prediction.

In 1610, Kepler wrote to Galileo requesting him to provide hime with one of his telescopes or lenses. However, Galileo claimed he was too busy to build one and replied that he did not have a spare to send. As a result, Kepler started a theoretical and experimental investigation of telescopic optics using a telescope borrowed from Duke Ernest of Cologne. He described his findings in the published manuscript *Dioptrice* (1611). There, he also described an improved telescope (known as the Keplerian telescope) in which two convex lenses can produce higher magnification than Galileo's combination of convex and concave lenses.

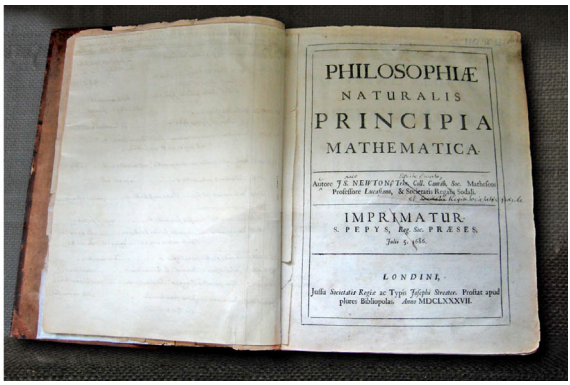
Among the many numerological relationships among the planets, his Laws are the three that have proven to have a physical basis. Kepler's Third Law states that the square of the orbital period of a planet is directly proportional to the cube of the semi-major axis of its orbit. This law was published in 1619, in *Harmonices Mundi*. At the time, he was a mathematics professor in Linz, Austria, after having been the imperial mathematician to Emperor Rudolf II and his two successors Matthias and Ferdinand II, in Prague (until 1612). In 1626, Kepler moved to Ulm and later became the advisor of General Wallenstein. Kepler provided astronomical calculations for Wallenstein's astrologers and occasionally wrote horoscopes himself. In those years until his death in 1630, Kepler spent much of his time travelling, from the imperial court in Prague to Linz and Ulm to a temporary home in Sagan, and finally to Regensburg.

Shortly after he reached Regensburg, he fell ill and died on November 15, 1630. A key figure in the 17th century scientific revolution was no more. Although Kepler contributed to many fields of science, mathematics and astronomy, he is best remembered for the Laws of Planetary Motion that carry his name. The Laws were codified by later astronomers using his major works *Astronomia nova*, *Harmonices Mundi*, and *Epitome of Copernican Astronomy*. These works also provided one of the foundations for Isaac Newton's theory of universal gravitation.

Isaac Newton (1642–1727) is often referred to as the greatest physicist of all time, but it may be clear that he had illustrious predecessors. In 1642, the year Galileo died, Newton was born in Woolsthorpe, Lincolnshire, England, on Christmas Day (Christianson 1996). His father had died three months earlier, and as Isaac was born prematurely, he was also not expected to survive. Isaac's mother remarried a wealthy elderly clergyman from the next village, and went to live there. She left Isaac, only three years old, in the care of his grandmother. When the clergyman died, Isaac's mother came back after eight years. Two years later, Newton went to Grantham to study at the Grammar School. He stayed with the local apothecary, and was fascinated by the chemicals. The plan was that at age seventeen he would come home and look after the farm. He turned out to be a total failure as a farmer: he'd rather sit under a tree reading books or crafting wood models than watching the sheep that went astray on the neighbour's property.



(a) Godfrey Kneller's 1689 portrait of Isaac Newton



(b) Cover of the *Principia*. Copyright Andrew Dunn

Fig. 1.10 Isaac Newton

In 1661, Isaac continued his studies at university, Cambridge Trinity College, partially the doing of his uncle William, who convinced his mother his talents were better spent there than in the fields. Isaac paid his way through the first three years of college by waiting tables and cleaning rooms for the fellows (faculty) and the wealthier students. In 1664, he obtained a scholarship that financially secured him for four years. Due to the spreading plague, the university was closed Summer of 1665 and Newton returned home, where he spent two years concentrating on problems in mathematics and physics. He wrote later that during this time he first understood the theory of gravitation, which we shall discuss below, and the theory of optics (he was the first to realise that white light is made up of the colours of the rainbow), and much mathematics, both integral and differential calculus,[2] and infinite series.

The credit for developing calculus of variations (1665–1666) has not been without any dispute. Building on Galileo's and others' concepts of instantaneous motion, as well as the method of infinitesimals of Kepler (1616) and Cavalieri (1635), Newton developed the theory at the age of 23, but did not publish about it. This was not so uncommon for him, because he was secretive and rarely published. German polymath and philosopher in the history of mathematics and philosophy, Gottfried Leibniz (1646–1716), began working on his variant of calculus in 1674, and published his first paper employing it in 1684. This led immediately to heated discussions and accusations of Newton that Leibniz had stolen from his work. Years later, a weary Leibniz appealed to the *Royal Society* to pass verdict on the matter of credit. Not realising his blunder, a committee, handpicked by Newton, gave full credit to Newton and put everything Leibniz had done in a very bad light. The full story is known, now, and credit goes to both having independently developed the calculus. Leibniz

[2] Using his calculus, Newton determined the mass of the Sun and each of the planets, Earth included. He estimated the Earth's density between five and six times that of water. The current estimate is 5.5.

established a superior system of notation in his equations, though, which ultimately replaced Newton's more inconvenient method, and is still in use today (Christianson 1996).

In his 1632 published book *Dialogo sopra i due massimi sistemi del mondo* (Dialogue Concerning the Two Chief World Systems), Galileo had written down the basic principle of relativity. This principle shows that the laws of physics are the same in any system that is moving at a constant speed in a straight line, regardless of its particular speed or direction. This means that there is no absolute motion or absolute rest. Newton took that principle and thus developed his Laws of Motion (this principle is also central to Einstein's special theory of relativity). He published his famous *Philosophiae Naturalis Principia Mathematica* in 1687 (Fig. 1.10), in which he covers his three Laws of Motion that form the foundation of classical mechanics, his Law of Universal Gravitation, and a derivation of Kepler's Laws of Planetary Motion.[3]

As the name implies, Kepler had obtained the Laws of Planetary Motion first, of course, but only empirically. Newton, however, demonstrated that Kepler's three Laws were but special cases of his three Laws of Motion, if a force of a particular kind, i.e., the gravitational force, was assumed to exist between all objects in the universe having mass. The expression of this force is popularly known as the Inverse-Square Law, derived from the fact that its magnitude is inversely proportional with the square of the distance between the two attracting bodies.[4] Once this law is known, one can use calculus to derive Kepler's Laws. But Newton went beyond the fact of simply proving this. He showed that Kepler's Laws of Planetary Motion were only approximately correct, and supplied the quantitative corrections that with careful observations proved to be valid.

For a minute, let us turn to the central topic of the *Principia*, the universality of the gravitational force. The story goes that Newton was sitting under an apple tree in his garden in Lincolnshire, when an apple fell on his head. Newton himself, though, said that he was staring out the window in his house when he saw the apple fall.[5] In any case, the falling apple triggered something in him and he thought of it in terms of an attractive gravitational force towards the Earth, and realised that the same force might extend as far as the Moon. He was familiar with Galileo's work on

[3] Newton appears to have been a stern man. Humphrey Babington, who had joined the ranks of the senior fellows at Trinity College, states that he saw Newton laugh only once in five years time. An acquaintance had borrowed a copy of Euclid (Euclid of Alexandria (300 B.C.), was a Greek mathematician, often referred to as the "Father of Geometry"). The acquaintance asked what use its study would be to him. "Upon which Sir Isaac was very merry" (White 1993). This "outburst" of joy is better understood if one realises that Newton proved many of his laws and theories without using complicated mathematics, but only Euclidean geometry.

[4] Understanding this, it easily follows that planets that are closer to the Sun will move faster, as for a given mass the acceleration is proportional to the force acting on the body.

[5] Common to either version is an apple falling from a tree. One cannot help but wonder what happened with this apple tree? Is it still there? The King's School in Grantham, Linconshire, England, where Newton went to school, claims to have purchased the tree and moved it to its garden. This is a heated disagreement with the Woolsthorpe-Manor people who are currently in charge of the upkeep of Newton's home, and claim that the tree is in their garden.

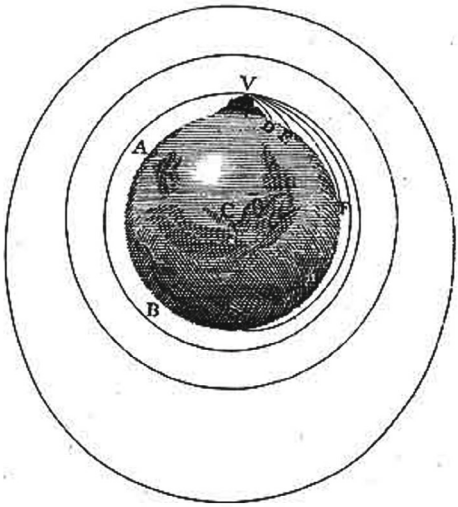


Fig. 1.11 Newton's Mountain illustrating orbital flight, from the *Principia*

projectiles, and suggested that the Moon's motion in orbit could be understood as a natural extension of that theory.

To get a grasp on this, consider a cannon placed on top of a high mountain and horizontally shooting a projectile. When using more and more powder in successive shots it will propel the cannonball faster and faster (Fig. 1.11). As the cannonball's speed increases, it travels farther from the mountain before hitting the Earth. One can easily imagine that eventually, the cannonball moves so quickly that the curved Earth drops away beneath it and it never hits the Earth at all: the cannonball then orbits the Earth. Doug Adams, in his third volume of *The Hitchhiker's Guide to the Galaxy* put it so eloquently: "There is an art to flying, or rather, a knack. The knack lies in learning how to throw yourself at the ground and miss. [...] Clearly, it is this second part, the missing, that presents the difficulties."

Also Newton's work on the gravitational force did not pass without dispute. Robert Hooke[6] almost found experimental proof that gravity follows an inverse square law, and hypothesised that such a relation governs the motions of the planets. When Newton, working independently on this, published his work in the first volume of the *Principia*, Hooke raised his all too familiar cry of theft. He claimed that he had already proven this six years before and had communicated this in a letter to Newton. Even though Hooke had corrected a rare blunder Newton made regarding the path of a falling body, this was, of course, nowhere near the prove that the greater the distance between a planet and the Sun, the less intense the gravitational attraction will be between them (Christianson 1996). Newton was obviously outraged and removed any reference to Hooke from the manuscript of the remaining volumes of the *Principia*.

[6] Robert Hooke (28 July 1635–3 March 1703), was an English natural philosopher, architect and polymath, known, amongst others, for his law ("Hooke's Law") that states that the force needed to extend or compress a spring by some distance scales linearly with respect to that distance.

But it was not just gravity that kept the two men apart. Both Newton and Hooke worked on optics, and did experiments on dispersion, the nature of colour, and the wave nature of light, based on the observation that a prism decomposes white light into the colours of the visible spectrum (published in *Opticks*, in 1704 after Hooke's death). Newton was one embracing the idea of a light beam being composed of minute particles, or corpuscles, whereas Hooke leaned towards the theory that light is composed of waves. Whereas we know now that light behaves like both particle and wave, it led to heated discussions, insults and accusations, which would last a lifetime. On another occasion, and even though Newton had built the first practical reflecting telescope, Hooke claimed he had developed a smaller and more accurate instrument.

However, no matter how difficult Newton may have been as a person, it is without doubt that he was one of the greatest scientists of all time and a key figure in the scientific revolution. His book *Philosophiae Naturalis Principia Mathematica* laid the foundations of classical mechanics, and we will draw heavily from it in the upcoming chapters.

Before moving on to the "giants of rocketry" someone you may not expect in this overview of the great people of science has to be introduced to you. But actually, when you learn about how he influenced certain people one has to agree that he has deserved his place here. The person we are talking about is the French author from Brittany, Jules Verne (1828–1905), see Fig. 1.12. Jules Verne pioneered the Science Fiction genre, and wrote about air, space and underwater travel, before any of this had ever materialised. He became famous for his books *Twenty Thousand Leagues Under the Sea*, *Journey to the Centre of the Earth*, *Around the World in Eighty Days*, and, of course, *From the Earth to the Moon*, which was published in 1865.

From the Earth to the Moon tells the story of the Baltimore Gun Club and their attempts to build an enormous Columbiad space gun and launch three people in a



Fig. 1.12 Jules Verne around 1878, photographed by Félix Nadar

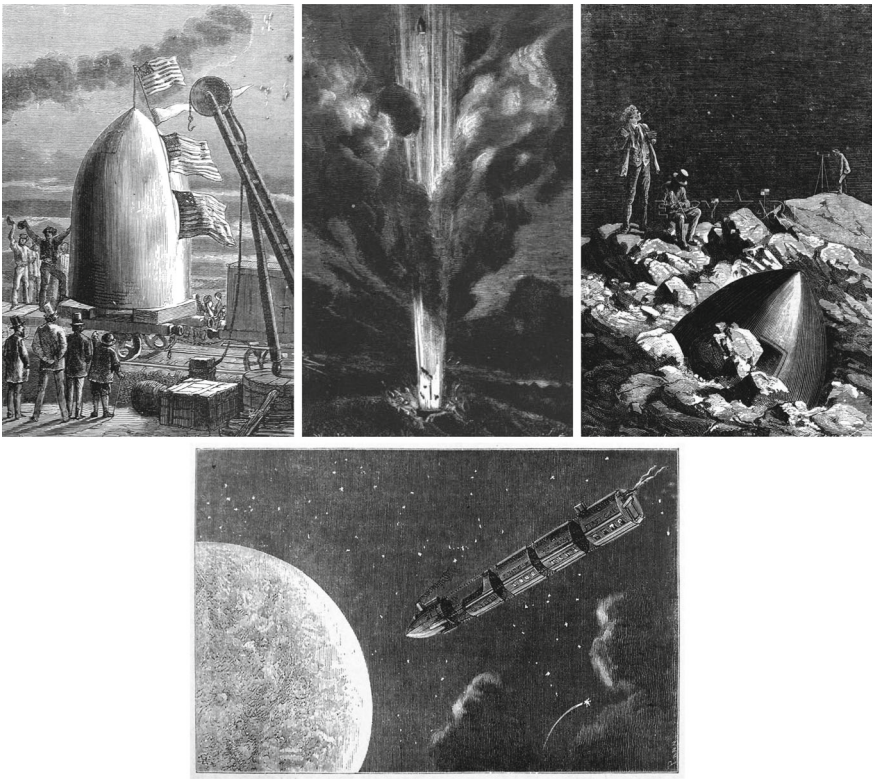


Fig. 1.13 Jules Verne's *From the Earth to the Moon*

projectile with the goal to land on the Moon (Fig. 1.13). Before publication of his book, rockets, or gun-launched projectiles, were only seen as a means of destruction, without much thought put into the operation: launch it and let us hope it does not return. Jules Verne changed all that, because he saw rockets also as a means to travel. He did some rough calculations on, amongst others, the requirements for the cannon. Given the comparative lack of empirical data on this at that time, it is quite impressive that some of his results were remarkably accurate. Of course, the scenario of a cannon launch is impractical for safe manned space travel, because a much longer muzzle than what Verne conceived would have been required to reach escape velocity, while at the same time limiting the accelerations acting on the passengers to limits that they can survive. Nonetheless, this book was—and still is—a great inspiration for generations of rocket scientists.

Konstantin Tsiolkovsky (1857–1935) was only an eight-year old boy when Jules Verne's book was published, but it captured him all the same and would eventually drive him to develop the theory of spaceflight. Generally considered to be one of the three founding fathers of astronautics, alongside Robert Goddard and Hermann

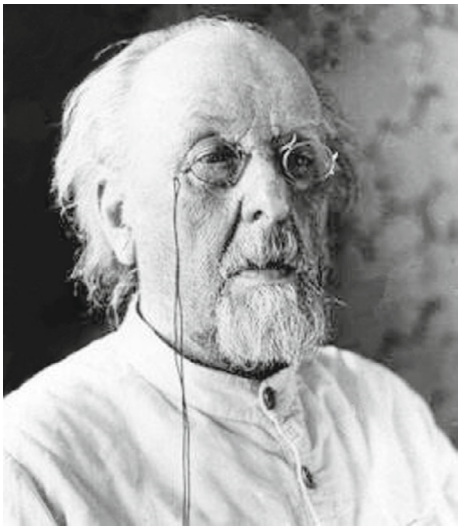


Fig. 1.14 Konstantin Eduardovich Tsiolkovsky, Russian rocket scientist and pioneer of cosmonautics

Oberth,[7] Tsiolkovsky (Fig. 1.14), a nearly deaf Russian schoolteacher, was a self-taught theoretical rocket expert and space-travel pioneer, "light-years" ahead of his time (Angelo 2006). In the 1950 and 1960s, his works inspired leading Soviet rocket engineers, such as Sergei Korolev and Valentin Glushko, and thus contributed to the success of the Soviet space programme. However, it was not an easy path that Tsiolkovsky had to walk. Even though he has a statue in Moscow to honour his achievements, he was considered mad at the time, partially fuelled by his reclusive nature and unusual habits (Gainor 2008).

As early as 1894, Tsiolkovsky designed a monoplane and also worked on airships. However, too far ahead of his time, his work on the airplane and the airship did not receive much recognition from the official representatives of Russian science, and he did not get any monetary or moral support. His descriptions and drawings of the monoplane led to a design, similar in appearance and aerodynamics, that would be constructed 15–18 years later. In his apartment, Tsiolkovsky developed the first aerodynamics laboratory in Russia, and in 1897 he built the first Russian wind tunnel with an open test section and developed an experimentation method to use it. Around 1900, executing numerous experiments on simple shapes, he determined the drag coefficient of spheres, flat plates, cylinders, cones, and other bodies. In contrast,

[7] Sometimes a fourth "founding father" is mentioned. Robert Albert Charles Esnault-Pelterie (November 8, 1881–December 6, 1957) was a pioneering French aircraft designer and spaceflight theorist. Unaware of Tsiolkovsky's 1903 work, in 1913 he published a paper on the rocket equation, and calculated the energy that is necessary to reach the Moon and nearby planets. In 1929, Esnault-Pelterie proposed the idea of the ballistic missile for military bombardment, a study that was funded by the French War Department. In 1931, he began to experiment with various types of rocket propulsion systems, including liquid propellants. Unfortunately, in the end the work failed to create an interest in rocketry within France. Yet another one of his contributions was the idea of manoeuvring rockets by vectored thrust.

even though he was interested in the theories of space flight, he never built a rocket or motor himself.

During his life Tsiolkovsky pioneered in a variety of fields, and published his findings in over 500 papers and books. He discussed the design of liquid-fuelled rockets, rockets with steering thrusters, multistage boosters, space stations, airlocks for exiting a spacecraft into the vacuum of space, and closed-cycle biological systems to provide food and oxygen for space colonies. He was also the first person to conceive the space elevator, becoming inspired in 1895 by the newly constructed Eiffel Tower in Paris. In 1903 he published his famous "Rocket Equation" in *Exploration of Outer Space by Means of Rocket Devices*, which relates the change in the rocket's speed to the exhaust velocity of the engine, and the initial and final mass of the rocket. He contradicted Jules Verne's idea of using a cannon for space travel and concluded that a gun would have to be impossibly long. The gun used in the original story would subject the payload to about 22,000 g.

In 1926 Tsiolkovsky defined his "Plan of Space Exploration", consisting of sixteen steps for human expansion into space:

1. Formation of rocket airplanes with wings. In 1947, Chuck Yeager was the first pilot to exceed the speed of sound in the Bell X-1 rocket plane.
2. Progressively increasing the velocity and altitude of these rocket planes. The North American rocket-powered aircraft, X-15, holds the speed record of manned, powered aircraft, set in October 1967 (Mach 6.72 at 31,120 m altitude).
3. Design and construction of actual rockets without wings. Modern rockets originated in 1926, when Goddard attached a supersonic nozzle to the combustion chamber of a liquid-propellant rocket. In 1947 the US Military launched a modified V2 into space, which became the first man-made object to re-enter at hypersonic speeds. Development of rockets for satellite launch and manned spaceflight continued throughout the 1950 and 1960s.
4. Establish the capability to land on the surface of the sea. The splashdown method of landing was used for the US Mercury, Gemini and Apollo missions, starting with the first (sub-orbital) Mercury mission on May 5, 1961.
5. Reaching a velocity large enough to conquer Earth's gravity (about 8 km/s), and the first flight into Earth orbit. On October 4, 1957, just after the 100th anniversary of Tsiolkovsky's birthday, the Soviet Union successfully launched the first artificial satellite, Sputnik, into Earth orbit. In 1961, this followed suit with the launch of Vostok-1, carrying Yuri Gagarin.
6. Lengthening rocket flight times in space. After the first manned spaceflight, duration of rocket flight extended to the point that the USA could successfully fly astronauts to the Moon and land them on its surface, in 1969.
7. Experimental use of plants to make an artificial atmosphere in spacecraft. In December, 1996, US Astronaut John Blaha, aboard the Russian space station MIR, harvested the first wheat crop completely grown in space. Since 2002, the Lada greenhouse has been used to perform almost continuous plant growth experiments on the station.

8. Using pressurised space suits for activity outside of spaceships. The first spacewalk took place on 18 March 1965, when Alexey Leonov performed a 12 min extravehicular activity (EVA), exiting the Voskhod-2 capsule.
9. Making orbiting greenhouses for plants. In August 2015, astronauts on the International Space Station ate the first vegetables grown in space
10. Constructing large orbital habitats in orbits around the Earth. Even though we have a permanently inhabited space station orbiting Earth, this station is not of the scale that Tsiolkovsky envisioned.
11. Using solar radiation to grow food, to heat space quarters, and for transport throughout the Solar System.
12. Colonisation of the asteroid belt.
13. Colonisation of the entire Solar System and beyond.
14. Achievement of individual and social perfection.
15. Overcrowding of the Solar System and the colonisation of the Milky Way (the Galaxy).
16. The Sun begins to die and the people remaining in the Solar System's population go to other suns.

Apart from many official tributes, such as naming the most prominent crater on the far side of the Moon, Tsiolkovskiy, after him (Fig. 1.15), and naming asteroid 1590 Tsiolkovskaja after his wife, Tsiolkovsky is also often commemorated in popular culture. For instance, in the science fiction series *Star Trek: The Next Generation*, episode *The Naked Now*, the crew of the starship USS Enterprise responds to messages received from the USS Tsiolkovsky, a science vessel of the Oberth Class monitoring the collapse of a supergiant star. The USS Tsiolkovsky is listed in the Federation Starfleet Class Database as NCC-53911; the dedication plaque for the ship bore a quote from Tsiolkovsky: "The Earth is the cradle of the mind, but one cannot remain in the cradle forever". Other recognitions are, for instance, the name

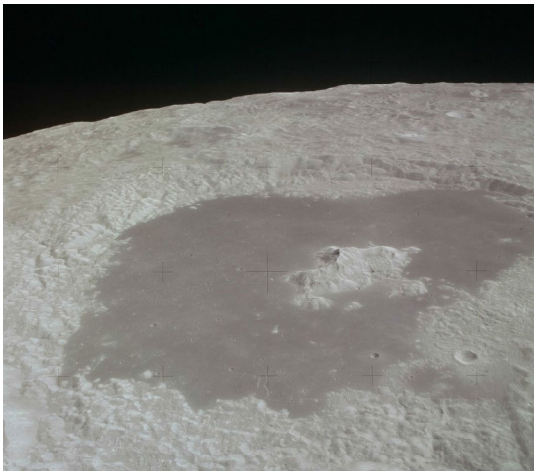


Fig. 1.15 The Tsiolkovskiy crater on the far side of the Moon taken on the Apollo-15 mission in 1971. *Image* NASA



Fig. 1.16 Robert Hutchings Goddard, the father of modern US rocketry. *Image* NASA

"Tsiolkovsky Towers" for the Mars-based space elevators in the Horus Heresy novel *Mechanicum* by Graham McNeill, set in the Warhammer 40 k universe. The space elevators are also mentioned in the book *Limit*, written by Frank Schätzing in 2009, where its theory is applied in a hypothetical future, and the video-games *Sid Meier's Alpha Centauri* and *Sid Meier's Alien Crossfire* use one of Tsiolkovsky's quotes when the technology of the space elevator is developed.

The American engineer, professor, physicist, and inventor Robert Goddard (1882–1945), see Fig. 1.16, was inspired by works of science fiction, primarily *War of the Worlds* by H. G. Wells and *From the Earth to the Moon* by Jules Verne (Gainor 2008). As a kid, he was an avid reader, going often to the local public library to borrow books on physical sciences. Around 1901, Goddard read Newton's *Principia*, and concluded that Newton's Third Law of Motion applied to motion in space. In 1904, Goddard enrolled at Worcester Polytechnic Institute and obtained his BSc in physics in 1908. He then changed to Clark University where he got M.A. and PhD degrees in physics in 1910 and 1911, respectively. He stayed on for another year as an honorary fellow in physics, but accepted a research fellowship at Princeton University's Palmer Physical Laboratory in 1912.

Suffering from a poor health, Goddard lost quite some time before he could finish his education. It did not stop him from putting his innovative ideas to paper, and trying to have them published. As a high-school student, he wrote down his ideas on space travel in a draft article, "The Navigation of Space", which was rejected by *Popular Science News*. *Scientific American*, on the other hand, published his article proposing a method for balancing aeroplanes using gyro-stabilisation in 1907. His first article on a potentially liquid-fuelled rocket saw the light on February 2, 1909. Earlier, he had studied different options to increase a rocket's efficiency using methods differing from conventional solid-fuel rockets, i.e., using liquid hydrogen as a fuel with liquid oxygen as the oxidiser. In 1912, he explored mathematically the practicality of using rocket propulsion to reach high altitudes and even the Moon.

His health kept plaguing him and in early 1913 Goddard became seriously ill with tuberculosis and had to leave his position at Princeton. Despite the dark prognosis of

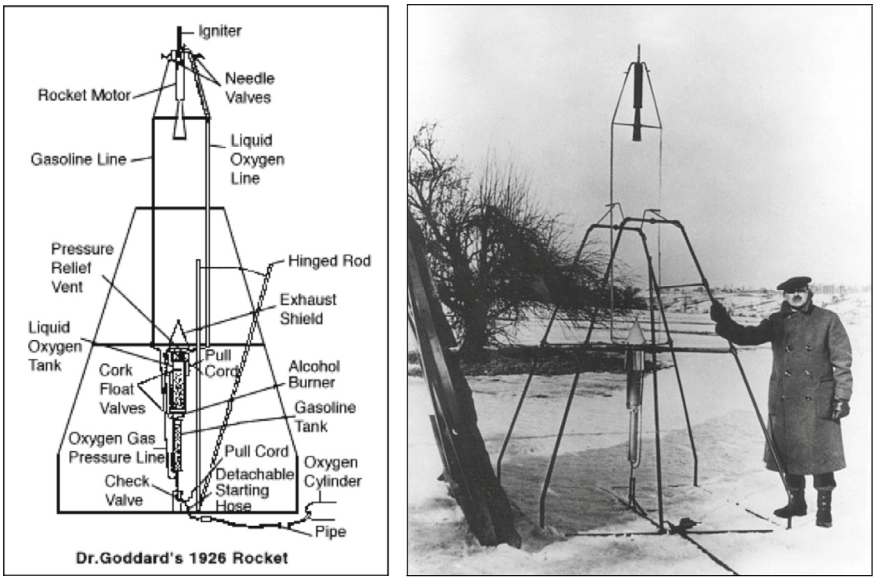
his doctors that he would not recover, Goddard kept exercising and his health slowly improved. It was in this period of recovery that he began to produce some of his most important work, and also started to realise the importance to keep the intellectual property of his ideas. Goddard's first rocket patent application was submitted in October 1913, and over the years many would follow. In 1914, two of his patents that would become crucial milestones in the history of rocketry were accepted and registered: the first one described a multi-stage rocket powered with a solid fuel, and the second one dealt with a rocket fuelled with a solid fuel (explosive material) or with liquid propellants (gasoline and liquid nitrous oxide). Overall, he published 214 patents, many (131 of them) were only awarded after his death.

At the Clark physics lab, Goddard started testing his rocket designs. He confirmed that solid-fuel rockets were converting only about 2% of their fuel into thrust. By using De Laval nozzles he could improve the efficiency, and by mid 1915, he had increased the average efficiency up to 40%, with a nozzle exit velocity of 2,051 m/s. Later tests had even shown an efficiency of 63%. Even though not fully recognised at the time, the experiments suggested that rockets could be made powerful enough to escape Earth and travel into space, but Goddard realised that it would require the more efficient liquid propellants to do so. Later in 1915, he designed an elaborate (static) test, and proved that a rocket will work in vacuum and does not need the air to push against. In fact, the performance of rockets decreases while operating in the atmosphere.

When the United States entered World War I in 1917, Goddard went to work for the Army with the goal of designing rockets for military applications. In late 1919, the Smithsonian published Goddard's groundbreaking work that he had already prepared in 1916, *A Method of Reaching Extreme Altitudes*. It includes Goddard's mathematical theories of rocket flight, the experiments with solid-fuel rockets, and the possibilities he saw of exploring Earth's atmosphere and beyond. Along with Konstantin Tsiolkovsky's earlier work, this book is considered to be one of the pioneering works of the physics of rocketry.

In 1926, despite the many difficulties he encountered, Goddard launched the first liquid propellant rocket (Fig. 1.17), using gasoline and liquid oxygen as fuel. The rocket flew for only two and a half seconds, climbed 12.5 m, and landed 56 m away. This might seem very unimpressive—and for current standards it is—but this rocket test defined a whole new era in rocket flight. Goddard's experiments in liquid-propellant rockets continued for many years, refining and enlarging his rockets every time, and between 1930 and 1945, 31 rockets were launched. In 1932, he was the first one to use vanes in the rocket motor blast for guidance and developed a gyroscope system for flight control. He was the first one to develop pumps suitable for rocket fuels, and designed a payload compartment for scientific instruments. Parachute recovery systems were employed to return rockets and instruments safely. In 1937, he successfully launched a rocket with a motor pivoted on gimbals under the influence of a gyro mechanism.

Even though the US government never really developed an interest in his work (the press even ridiculed him by saying that he did not have the knowledge of a high-school student, simply because they failed to see his brilliance), it is without a



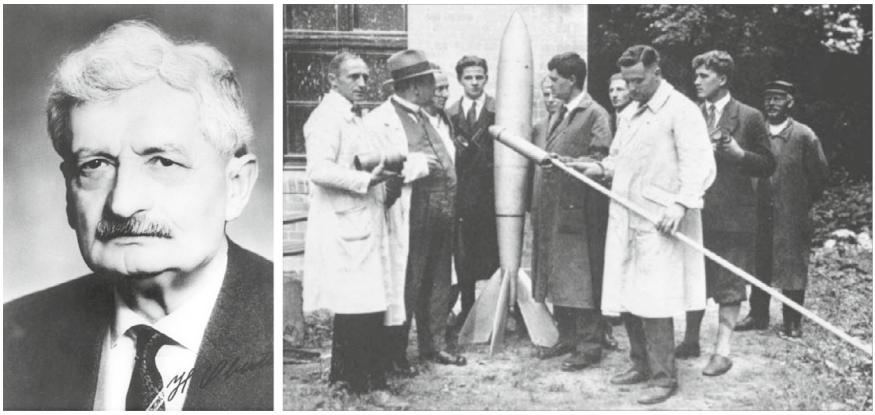
(a) Goddard's rocket design (b) Experimental setup

Fig. 1.17 Flight testing the first liquid-fuelled rocket (1926). *Images* public domain

shred of doubt that Goddard's influence on rocket development has been enormous. Even today, rocket scientists use at least some of his ideas or devices when they design, construct, or launch a modern liquid-propellant rocket. Therefore, Goddard is generally considered to be the founding father of American rocketry. He died of throat cancer in 1945, having worked on rockets until his last breath.

Besides having been an inspiration for many people, amongst others astronauts Buzz Aldrin and Jim Lovell, and NASA flight controller Gene Kranz, Goddard was honoured in many different ways. The Goddard Space Flight Center, a NASA facility in Greenbelt, Maryland, was established in 1959, and there is a crater Goddard on the Moon. Further references include release 13 of the Linux distribution Fedora, which is named after Goddard, and in the television series Star Trek: The Next Generation, a shuttlecraft is called Goddard. Finally, the experimental Blue Origin Goddard reusable vertical launch and landing rocket is named in honour of him.

The third founding father of rocketry and astronautics is Austro-Hungarian born Hermann Oberth (Fig. 1.18a). Born on June 25, 1894, in Hermannstadt, Transylvania (currently Sibiu, Romania), as a young boy Oberth grew up in Romania. When he developed scarlet fever at the age of 11 he was sent to Italy to recover, and while there (as did his illustrious companions Tsiolkovsky and Goddard), he read Verne's *From the Earth to the Moon* and *Around the Moon*. Reading and re-reading these books to the point of memorisation, he became fascinated by the concept of space travel. Intrigued by the cannon launch (Fig. 1.13), he calculated the necessary velocity with which the capsule had to leave the gun to overcome Earth's gravity and reach the



(a) Hermann Oberth (b) Oberth, right beside the rocket. Second from right: young Wernher von Braun.

Fig. 1.18 Hermann Oberth, the *Father of Space Travel*. *Images* NASA

Moon, and confirmed that Verne had been correct. However, he also computed the acceleration required, and concluded that it would not have ended well for the space travellers: "*Forty-seven thousand times the Earth's gravity. They would be flattened into pancakes. A cannon is not good for space flight. It must be done with a rocket.*"[8] As a school student of only 14, he constructed his first model rocket. Talking to people who understood gunpowder, he knew that this would not work for rockets, and he reached at the same conclusion as Goddard, namely that the rockets should use liquid propellants. In 1912, he finished his design based on the most powerful liquid propellant combination he could think of, liquid hydrogen and liquid oxygen. He also realised that using a single stage would lead to unrealistically large rockets, so he arrived independently at the concept of the multistage rocket, albeit on paper only. He was very correct to conclude that "*If there is a small rocket on top of a big one, and if the big one is jettisoned and the small one is ignited, then their speeds are added.*"

In 1912, Hermann and his parents moved to Germany and he began his studies medicine at the University of Munich. However, this was interrupted by the outbreak of World War I. He was drafted in the Austro-Hungarian army and he served in medical units on the battlefield. He realised practising medicine was not his passion, and after being wounded in combat, Oberth pursued his old dream of flight. In 1917, he had submitted his liquid-propellant rocket design to the German government, but it was too far ahead of its time and, therefore, rejected. Without faltering, though, Oberth started to work on a Ph.D. degree in physics from the University of Heidelberg. He presented his doctoral thesis on rocketry in 1922, but it was received with scepticism

[8] https://www.meaus.com/spacePioneer.html, accessed July 11, 2019. Speech given by rocket scientist and long-time colleague and friend of Oberth, Prof. Ernst Stuhlinger, at the International Space Hall of Fame, Alamo Gordo, New Mexico, in October, 1976.

and he was even ridiculed, and Oberth gave up on the idea: "*I refrained from writing another one, thinking to myself: Never mind, I will prove that I am able to become a greater scientist than some of you, even without the title of doctor.*" He and his wife used their savings to partially cover the publishing cost, and his dissertation appeared in 1923 as the book *Die Rakete zu den Planetenräumen*, which has been published as *The Rocket into Interplanetary Space* (Oberth 2014). In it, he explained mathematically, how rockets could achieve a speed that would allow them to escape Earth's gravitational pull and he proved that flight beyond the atmosphere is possible. The book was received with great enthusiasm and finally gave him the widespread recognition that he deserved.

Until 1922, Oberth was unaware of Goddard's work in the United States and until 1925, he was unfamiliar with that of Tsiolkovsky in the Soviet Union as well. He started corresponding with both and he openly acknowledged their precedence in deriving the equations linked to space flight. In 1929, Oberth's *Wege zur Raumschiffahrt*, translated as *Ways to Spaceflight* (Oberth 1970), won the first annual Robert Esnault-Pelterie-André Hirsch Prize, which allowed him to continue his research on liquid-propellant rocket motors. To further support himself and his family, from 1924 through 1938, Oberth taught physics and mathematics at the Stephan Ludwig Roth High School in Mediaş, Romania. In 1931, he submitted a patent for his rocket design to the Romanian Patent Office, which got awarded. Just a while earlier he became a member of the *Verein fur Raumschiffahrt* or VfR ("Spaceflight Society"), which was founded in Germany in 1927 to popularise the idea of space travel and to perform serious experiments with rocket engines. Within two years, the VfR had 870 members and had become the largest rocket society in the world. Oberth was soon voted its president. He was a mentor to many student members, and he even took one of them as an assistant: 18-year old Wernher Von Braun (see also Fig. 1.18b), who would later become one of the leading scientists in rocketry research for the German and the United States governments.

The founders of the VfR were impressed by Oberth's mathematically-sound theories and in 1930–1931 the society launched many small rockets, designated *Mirak*. The highlight was undoubtedly the launch of Oberth's patented design, which was first launched on May 7, 1931, near Berlin. By 1932, VfR rockets had a range of 5 km and could reach an altitude of 1,500 m. Unfortunately, by that time the society had fallen on financially hard times, and when Hitler rose to power, he forbade all rocket-related experimentation. After having worked at universities in Vienna and Dresden, Oberth became a German citizen in 1940 and he worked under von Braun at the German rocket development centre at Peenemünde during World War II. There, they developed the world's first modern-war rocket, the V-2 or "vengeance weapon", for the German Army. After the war, when Von Braun had been relocated to the United States to work for the US Army's Ballistic Missile Agency in Huntsville, Alabama, they worked together again. However, three years later Oberth retired and returned to Germany, where he continued to write books on rocketry and space travel.

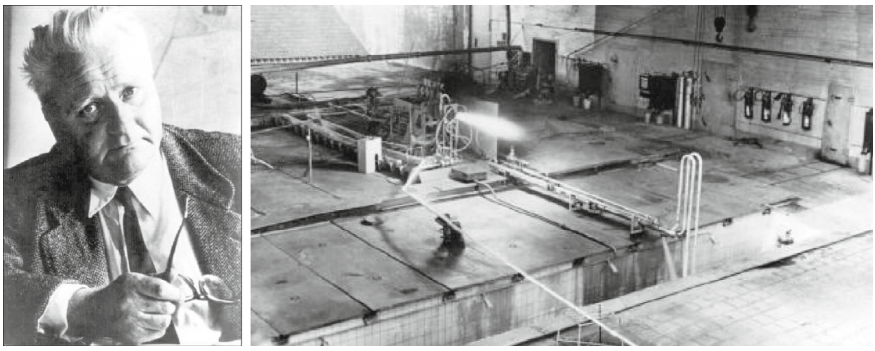
On December 28, 1989, Hermann Oberth died at the age of 95, in Nuremberg, West Germany. To honour this giant of rocket science, there is a crater on the Moon named Oberth (following in the footsteps of Wan Hu and Konstantin Tsiolkovsky),

Asteroid 9253 is designated Oberth, as is the Hermann Oberth Raumfahrt Museum in Feucht. In the Star Trek movie *The Search for Spock* the Oberth-class of starships, hypothetically to be in his honour, is mentioned, with two of the star ships in that class being the USS Oberth and USS Tsiolkovsky. On a more serious note, the Oberth effect is named after him, which states that a rocket engine when travelling at high speed generates more useful energy than one at travelling at low speed.[9]

Another member of the VfR was Austrian rocket pioneer Eugen Sänger (1905–1964). Born on 22 September 1905 in Preßnitz (Bohemia), part of the former Austro-Hungarian Empire, Sänger studied civil engineering at the Technical Universities of Graz and Vienna, Austria. As a student, he came in contact with Hermann Oberth's book *Die Rakete zu den Planetenräumen*, which inspired him to change from studying civil engineering to aeronautics and to join the VfR. He was convinced that the way into space was by using a rocket plane rather than a rocket, and focussed his initial research on rocket propulsion. However, his original Ph.D. thesis on rocket-powered flight was rejected, because the university found it too fictional. He was only allowed to submit a thesis on the far more "down-to-Earth" topic of the statics of wing trusses. In 1933, he published his rejected thesis as the book *Raketenflugtechnik*, i.e., Rocket Flight Engineering, which was considered to be a quantum leap ahead.

Sänger's rocketplane design was a horizontal takeoff-and-landing, hypersonic glider, and in 1935 and 1936 he published several articles about this. These attracted the attention of the *Reichsluftfahrtministerium* (RLM, the German Aviation Ministry), which had been focussed on developing a bomber that could strike the United States from Germany (the so-called Amerika Bomber project). They saw Sänger's ideas as a possible way to achieve this. The ministry offered him a research institute in Trauen, in the Lüneburger Heide region, and also built a liquid oxygen plant and a test stand for a 100-tonne thrust engine (Fig. 1.19b). Indeed, Sänger agreed to lead a rocket development team. From 1936 till 1944, he worked for the Luftwaffe (German airforce) on rocket and ramjet development. In that time, he conceived the concept of gliding and skipping flight, while developing the Silverbird Bomber, see also Sect. 5.5. He did this together with Irene Bredt, initially his PhD student who later became his wife. By 1942, however, the RLM canceled the Silverbird project, due to its complexity and high cost involved. Sänger was forced to shift his focus to ramjet-technology development, to which he made important contributions. He was largely responsible for designing the Skoda-Kauba SK P.14 ramjet fighter.

[9] The so-called Oberth manoeuvre has been described by Adams and Richardson (2010), who found that the most energy-efficient method for a spacecraft to burn its engine is at the lowest possible orbital periapsis, when its orbital velocity (and thus its kinetic energy) is the largest. Since the vehicle is only for a short time near periapsis, the Oberth manoeuvre will be most effective when the vehicle can generate as much impulse as possible in the shortest possible time. This means that the Oberth manoeuvre will be much more useful for high-thrust rocket engines compared to low-thrust reaction engines (e.g., ion drives), which take a long time to gain speed. The Oberth effect also aids in understanding the behaviour of multi-stage rockets: the upper stage can generate much more usable kinetic energy than the total chemical energy of the propellants it carries (Oberth 1970).



(a) Sänger (b) Trauen

Fig. 1.19 Eugen Sänger, and the rocket-engine test stand in Trauen, Germany. *Images* public domain

After the war, Sänger and Bredt worked for the French from 1946 to 1954, as consultant engineers. Amongst others, Sänger worked on a number of different jet, ramjet and rocket propulsion systems, and contributed to the experimental Griffon turbo-ramjet aircraft (maiden flight in 1957). In April 1952, the former general Dornberger and director of Peenemünde, at the time employed by the Bell Aircraft company, which was studying the skip-glide technique, travelled to Paris and offered the Sängers to come to the US and work for him. However, Sänger rejected the offer, because he and his wife "liked Paris better". The Russian approach was slightly different. After World War II, much of the research on the Silverbird had fallen into Soviet hands, along with hardware that may have included an uncompleted mock-up of the spaceplane. Stalin was so impressed with the Sänger-Bredt spaceplane that he considered to have them kidnapped. The Soviets continued to work on the design and improved it in several ways, but in the end it proved unworkable and the Russians eventually abandoned this improved antipodal bomber (Rose 2008). And there was no more need to kidnap the Sängers.

After returning to Germany, his research from 1954–1963 included winged shuttle designs. In 1957, he was directing a jet propulsion research institute in Stuttgart, and between 1961 and 1963, he was a consultant for the aircraft company Junkers in designing a ramjet-powered space-plane that unfortunately never left the drawing board. Eugen Sänger died in Berlin, on 10 February 1964. His legacy, fundamental work on gliding and skipping flight, as well as his Silverbird design, would prove important to the development of the X-15 and X-20 *Dyna-Soar*, and ultimately the Space Shuttle.

Oberth's pupil, Wernher von Braun (1912–1977), Fig. 1.20a, was born in Wirsitz, part of the German empire (now Wyrzysk, Poland), but soon after (1915) the family moved to Berlin (Spangenburg and Moser 2008). Twelve-year-old Wernher, inspired by speed records established by Max Valier and Fritz von Opel in rocket-propelled cars, was taken into custody by the local police, because he had attached fireworks

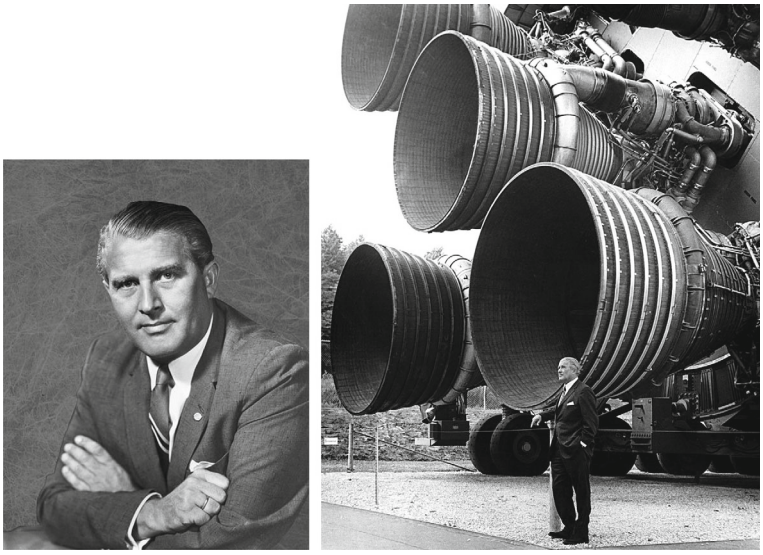
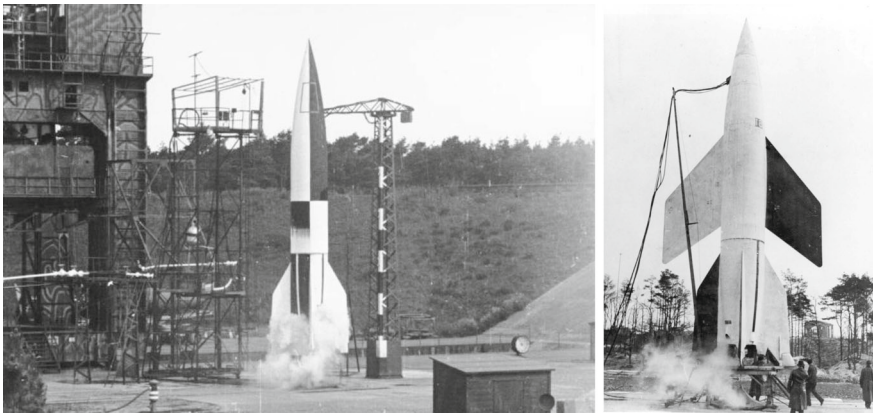


Fig. 1.20 Wernher Von Braun in 1960 (left), and in front of the engines of a Saturn-V rocket (right). *Images* NASA

to a toy car and detonated it in a crowded street. It was only a year later that he got hold of a copy of Oberth's book *Die Rakete zu den Planetenräumen*, which amplified his fascination for space travel. From then on, he committed himself to learning physics and mathematics—his weakest subjects!—to pursue his interest in rocket engineering. In 1930, he attended the Technische Hochschule, Berlin, and also became a member of the VfR, where he assisted Willy Ley in his liquid-fuelled rocket-motor tests in collaboration with Oberth. After getting his degree in mechanical engineering, von Braun obtained a doctorate in physics in 1934 for a thesis with the public title *About Combustion Tests*.[10]

In 1937, the German government established a rocket development and test facility near Peenemünde, the Peenemünde Rocket Centre. It covered 45 km^2 and employed as many as 18,000 people. Von Braun, at the age of only 25, became technical director of the centre, and oversaw the development of liquid-fuel rocket engines for aircraft and jet-assisted takeoffs, in collaboration with the German airforce. They also designed his creation, the long-range A-4 ballistic missile, which became known as the much dreaded V-2 rocket (Fig. 1.21a). This rocket seems to have been based on Goddard's work, and included gyroscopic control, steering by means of motor exhaust, gimbal steering, and exhaust-powered fuel pumps. However, even though von Braun was certainly inspired by his work, it was later confirmed that Goddard's publications did not cover enough detail to verify that claim. When V-2 rocket MW

[10] Von Braun's complete thesis was kept classified by the German army, and was not published until 1960. Its title: *Construction, Theoretical, and Experimental Solution to the Problem of the Liquid Propellant Rocket*, dated April 16, 1934.



(a) A-4 (b) A-4b

Fig. 1.21 The standard V2 rocket (A-4) and its winged variant (A-4b). *Images* Deutsches Bundesarchiv and US Air Force

18014 was vertically launched on 20 June 1944, it crossed the so-called Kármán line (equal to 80 or 100 km, depending on the organisation that defines it) to obtain a maximum altitude of 176 km, and so it became the first man-made object to travel into space. It also became the first man-made object to return into the atmosphere and impact on Earth, as it was in a sub-orbital trajectory only. In 1944, the team in Peenemünde started to work on a winged version to extend the A-4's flight range, the A-9. Flight tests of two hastily constructed A-4b rockets (Fig. 1.21b) failed (partially), and the advance of the Red Army put a stop to further development (Zaloga 2003).

Following World War II, along with about 1,600 other German scientists, engineers, and technicians, Von Braun (secretly) came to the United States in September 1945, as part of Operation *Paperclip*.[11] Von Braun began working for the US Army, testing ballistic missiles, based on the designs of his original brainchild, the V-2. Even though working with missiles his research focussed more and more on launching missiles for space travel, rather than warheads. Under army supervision, Von Braun helped create test launch sites for the Redstone and Jupiter ballistic missiles that formed the basis for launcher development, most notably the Jupiter C, Juno II and Saturn I launch vehicles. He and his team were responsible for developing the Jupiter C Re-entry Test Missile and launching the western world's first scientific Earth satellite, Explorer I, on January 31, 1958, which event marked the birth of America's space programme.

The launch of Soviet satellite Sputnik on 4 October, 1957, forced the Americans to speed up their launcher development, if they wanted to be a player in the Space Race. Some of von Braun's earlier proposals for new launchers that had been denied

[11] The Soviet Union forcibly recruited more than 2,200 German specialists with their Operation *Osoaviakhim* during one night on October 22, 1946. Also Great Britain was securing rocket technology and know-how, through their Operation *Backfire*.

earlier, were now approved. As director of the Marshall Space Flight Center, he led the development of the Saturn V, which took the Americans to the Moon, thereby claiming victory of the Space Race. Just a year after Neil Armstrong, Buzz Aldrin, and Michael Collins successfully used his technology to land on the lunar surface, Von Braun became NASA's Deputy Associate Administrator for Planning. This position allowed him to carry out his visions and plans to bring men into space, but these plans turned out to be too big, even for NASA. Von Braun retired in 1972 and died five years later, leaving behind a legacy far more important than he ever realised.

Before we come to an end of discussing those "giants" that made the space age possible, two more names should be mentioned, namely Sergei Korolev (1907–1966) and Hsue-Shen Tsien (1911–2009). Sergei Pavlovich Korolev (Fig. 1.22a) was born on 12 January 1907, in Zhytomyr, located in present-day Ukraine. He was the lead Soviet rocket engineer and spacecraft designer in the Space Race between the United States and the Soviet Union during the 1950 and 1960s. However, Korolev started out as an aircraft engineer, fascinated by those flying machines ever since being a kid. By 1929, when he graduated from Moscow's Bauman High Technical School (Russia's MIT) his interest in rocketry and the possibilities of space travel had developed, and he founded an organisation like the VfR in Germany, the *Gruppa Isutcheniya Reaktivnovo Dvisheniya*, GIRD, or Group for Investigation of Reactive Motion. The group was recognised by the state and transformed into the RNII (Jet Propulsion Research Institute), with Korolev as Deputy Chief. With his team, he developed missiles and a manned rocket-powered glider.

Before the latter could be made to fly, in 1938 Korolev was arrested and imprisoned in several Siberian prison camps for almost six years, which took a toll on his health (Gerovitch 2008). The arrest was typical for the Stalinistic times, some say for being a saboteur of military technology, others say for alleged mismanagement of funds (he spent the money on yet unsuccessful experiments with rocket devices). While some of his colleagues from the RNII were executed, in late 1940, Korolev and other arrested engineers in Soviet prisons were made to work on military projects from their cells. In 1944, Korolev was released and had to investigate the German military rockets, and as part of *Operation Osoaviakhim* he was involved in collecting captured enemy rocket hardware and technicians to aid Soviet rocket research. Derived from the V-2 (A-4c)—renamed to R-1—Korolev developed the R-7 that became the world's first Intercontinental Ballistic Missile (ICBM) with a maiden flight on 15 May 1957. Derivatives of the R-7 include the Vostok, Voskhod and Soyuz rockets, which, as of 2019 with over 2,000 launches, have been used for all Soviet (and later Russian) manned spaceflights.

Korolev was appointed to lead the Soviet space programme, made Member of Soviet Academy of Sciences, overseeing the early successes of the Sputnik and Vostok projects. He and his team designed the Soyuz spacecraft, and its derivatives are still carrying people into space, after all these years since its first flight in the 1960s. It can safely be said that Korolev's impact on the Russian space programme has been enormous, including the launch of the first man (Yuri Gagarin) into space, and he is considered by many as the father of practical astronautics. On 14 January 1966, Korolev died under somewhat mysterious circumstances after having under-



Fig. 1.22 Sergei Korolev (left) and Hsue-Shen Tsien (right), with from left to right: Ludwig Prandtl (von Kármán's doctoral adviser), Hsue-Shen Tsien and Theodor von Kármán's (Tsien's doctoral adviser). *Images* public domain

gone a minor surgery, although his poor health as a result of his imprisonment, in combination with his extreme working hours, had made him physically weak.

Hsue-Shen Tsien (western name Qian Xuesen), born on 11 December 1911, was an outstanding Chinese aerodynamicist and cyberneticist (Fig. 1.22b), who made important contributions to rocket science and he established engineering cybernetics. Starting his academic career at National Chiao Tung University, in August 1931, he moved to the US and continued his studies at the Massachusetts Institute of Technology (MIT). After a single year he received an MSc in mechanical engineering. Drafted by Theodore von Kármán,[12] who recognised him as an outstanding mathematician and "undisputed genius", Tsien moved to the California Institute of Technology (CalTech) and obtained his PhD in 1939. While being at CalTech Tsien got inspired by rockets and became a member of the university's so-called Suicide Squad of rocket experimenters (known, because of their often loud rocket experiments), who laid the groundwork for testing done by the Jet Propulsion Laboratory (JPL).[13] During his early career, Tsien was a huge catalyst to advances in high-speed aerodynamics and jet propulsion. He worked on designing an intercontinental space plane, which would later inspire the X-20 Dyna-Soar that in turn stimulated

[12] Theodore von Kármán (11 May 1881–6 May 1963) was a Hungarian-American mathematician, aerospace engineer, and physicist, who was mainly working in aeronautics and astronautics. To his credit are many key advances in aerodynamics, particularly his work on supersonic and hypersonic airflow characterisation, and he is one of the most distinguished aerodynamic theoreticians of the twentieth century. His name is associated with the von Kármán line that defines the boundary between atmosphere and space. He was the first person to calculate at which altitude the atmosphere becomes too thin to support aeronautical flight and arrived at 83.6 km.

[13] The Jet Propulsion Center found its roots in CalTech's Guggenheim Aeronautical Laboratory (headed by Von Kármán), where Tsien and the Suicide Squad performed their rocket experiments. The JPL was formally established in 1943 by Tsien and a few others.

the development of the Space Shuttle. In 1949, with the recommendation of von Kármán, Tsien became Robert H. Goddard Professor of Jet Propulsion at Caltech.

However, his success story did not last. In 1950, the infamous Los Angeles Police Department's "Red Squad" arrested him and accused him of being a communist. Neither proof of that nor him being a spy were admittedly never found, but the damage was done and the loss for the US space programme enormous. After a house arrest of five years, and, despite protests by his colleagues, stripped of his security clearance, Tsien was finally allowed to leave the US (it was rumoured that he was exchanged for 11 US airmen that were held captive by China). Back in China, Tsien became director of its rocket research programme, and under his leadership it made large strides forward. Disillusioned by his American experience, he never returned to the US, and, after his retirement in 1991, he refused to talk to westerners. On October 31, 2009, Hsue-Shen Tsien, the Father of the Chinese Space Program, died at the age of 97 in Beijing.

Now that we have introduced the actors, it is time to set the stage. In the next section, the discussion will start with the early struggles of both the United States and the Soviet Union to reach space. To describe the early development of launch system is, of course, required to reach that point in time that what we launch into space may also re-enter again. Crucial, of course, for manned spaceflight. After the heroic flight of Yuri Gagarin, a brief overview of both entry capsules and winged entry vehicles performing gliding flights will be given.

1.3 The Space Age

1.3.1 Setting the Stage

After the war, both in the US and the Soviet Union, the initial development of rockets was based on V-2 heritage. Many train loads with V-2 engines, fuselages, propellant tanks, gyroscopes, and associated equipment were transported to the White Sands Proving Grounds in New Mexico. With the aid of the German rocket scientists led by von Braun, the V-2s were assembled and equipped with different experiments. The first firing of a V-2 was a static test for 57 s, and took place on March 15, 1946. Just a month later (April 16, 1946), the first flight test took the V-2 to an altitude of about 6 km. Between 1946 and 1952, a total of 67 V-2 rockets were assembled and tested, and even though the success rate was under 70% it provided the US with valuable experience in the assembly, pre-flight testing, handling, fuelling, launching, and tracking of large missiles.

A major advance was made in 1947, though. The V-2 (as a first stage) was combined with the 7.3 m long WAC Corporal sounding rocket (as a second stage), and this combination, called Bumper, became the first multi-stage rocket after the vision of Tsiolkovsky. The first four launches failed, but when Bumper 5 was launched on February 24, 1949 (Fig. 1.23), it accelerated to 8,290 km/h to become the first flight

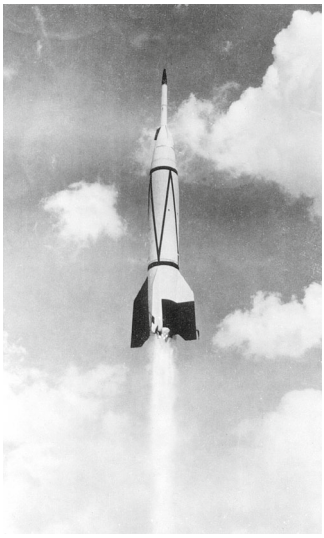


Fig. 1.23 Launch of Bumper 5 from white sands proving grounds on February 24, 1949. *Image* NASA

of more than five times the speed of sound. When it re-entered it also became the first manmade object to re-enter at hypersonic speeds. On July 24, 1950, Bumper WAC Round B-8 became the first rocket to fly from Florida's Cape Canaveral Air Force Station.

In the Soviet Union a similar development took place. In October 1947, the USSR supported by German scientists (indirectly monitored by Korolev) launched rebuilt V-2s in Kapustin Yar. The first Soviet missile completely manufactured in Russia, albeit a duplicate of the V-2, was the R-1 that was first launched in September 1948. In all secrecy, and completely underestimated by the western world, the Soviets continued to work on larger missiles, the R-2 and R-5, and finally the R-7. It came as a shock to the US when on 4 October, 1957, they picked up a soft beep coming from space. It was the sound of the transmitter of Sputnik-1, the first manmade object orbiting the Earth, and launched by the USSR with their R-7 derived Sputnik-PS launcher that could carry 500 kg into low Earth orbit (LEO). It was the first major blow in the *Space Race* between the Sovjet Union and the United States that had begun in earnest in 1955, when both countries announced to launch an artificial satellite into space. In November of the same year, the Soviet Union launched Sputnik-2, the first satellite to carry a living animal, a dog named Laika. Eager to catch up, the US attempted to launch their first satellite, but unfortunately the Vanguard launcher exploded on the launchpad. But then, on January 31, 1958, the US Explorer-1 was successfully launched. With its instrument package it detected high-altitude radiation surrounding Earth, which was the first major scientific discovery of the space age.

Many (frequently failed) launches followed from both sides, each trying to beat the other. Successes were there for the USSR, i.e., the Luna-2 was the first probe to successfully impact on the Moon (13 September, 1959), and Luna-3 probe returned the first pictures of the far side of the Moon with 70% photo coverage (launched on

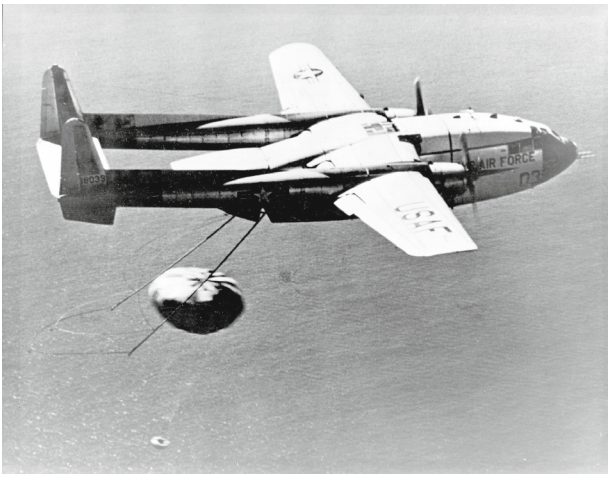


Fig. 1.24 Mid-air recovery of Discoverer-14 on August 18, 1960. *Image* USAF

4 October 1959). Then, on March 11, 1960, the US launched the first deep-space probe, Pioneer-5, that returned data from a distance of 36.2 million km. Another first for the Americans was the recovery of the first object returned from Earth orbit, i.e., the re-entry capsule of Discoverer-13 on August 10, 1960. The Discoverer was part of the Corona programme, a series of spy satellites taking pictures of the Soviet Union and China. The US Air Force (USAF) did not want to risk having the photo canister fall in the wrong hands upon return to Earth, so the capsule was intended to be caught in mid-air by an airborne claw mounted on a passing airplane towing such that it could be winched aboard (Fig. 1.24), or it could land at sea. A total of nine successful mid-air retrievals were carried out.

This first retrieval was just nine days before the recovery of Soviet Korabl Sputnik 2, a biosatellite, which took the two space dogs, Belka and Strelka, into orbit and then safely returned them to Earth. This satellite was the third test flight of the Vostok spacecraft, and part of USSR's manned spaceflight programme aiming to launch the first human into space. It had been preceded by the Korabl-Sputnik 1 that was launched empty into a near-Earth orbit, and Korabl Sputnik that also carried two dogs, but unfortunately, the launch vehicle exploded 28.5 s after ignition. Korabl-Sputnik 4, also known as Vostok-3KA No.1, the first of the Vostok spacecraft intended to put the first cosmonaut into space, was launched on 9 March, 1961, atop a Vostok-K rocket (an upgrade of the Sputnik PS launcher). It carried a mannequin that was named Ivan Ivanovich, a dog named Chernushka, some mice and a guinea pig. It successfully reached its orbit, and was deorbited again after a single orbit. It reentered on its first pass over the Soviet Union and was successfully recovered, with all animals alive. To test its ejection seat, during descent the mannequin was ejected from the spacecraft and came down separately under its own parachute. It was a bit charred when it landed, but it survived and was "fit" enough to be launched again on March



Fig. 1.25 The Vostok-3KA spacecraft (Image: NASA), and Gagarin's entry capsule after landing. *Image* ESA

25, together with dog Zvezdochka in Korabl-Sputnik 5. It all ended well, Ivan was ejected and safely descended under parachute, and Zvezdochka was recovered after landing with the capsule. Considered to be a dress rehearsal, everything was now ready for a manned flight.

And then the day was there ... the launch of the first human into space ... April 12, 1961. Atop the Vostok-K rocket, cosmonaut Yuri Gagarin was sitting in the Vostok-1 (Vostok-3KA No. 3[14]). Designed by Soviet rocket scientists under supervision of Sergei Korolev, the Vostok-3KA consisted of a spherical descent module (with a mass of m = 2,460 kg, and a diameter of d = 2.3 m) and a conical instrument module (a mass of 2,270 kg, a length of 2.25 m, and a width of 2.43 m), which contained propellant and the engine system (Fig. 1.25a). All three stages the Vostok-K functioned nominally and delivered Gagarin and Vostok 1 into Earth orbit, and ten minutes after launch, Gagarin officially became the first person to achieve orbital velocity: his historic trip around the Earth was a fact.

After the planned, single orbit the retro rockets fired and deorbited the spacecraft to re-enter the atmosphere. The two halves of the spacecraft were meant to separate before entry but a bunch of wires kept them connected and the combination entered. Due to the off-nominal configuration, the draft experienced strong gyrations, but fortunately the service module broke free before it got worse. The re-entry capsule came into its nominal orientation, and during the descent Gagarin experienced a maximum *g-load* of eight times the gravitational pull (8 g).[15]

In the final phase of the descent, Gagarin ejected from the craft (at about 7 km), like Ivan before him, and descended with his own parachute. The capsule landed separately, also by parachute—but not so smoothly, as it hit the ground hard enough to create an impact crater and bounced a few times before coming to rest (Fig. 1.25b). At the time, the separated landing raised an interesting question: was Gagarin really

[14] Starting at 28 July 1960, a total of eight Vostok 3KA spacecraft were flown, six of them with a human crew. The last flight, Vostok-6, took place on June 16, 1963, and took the first woman to space, Valentina Tereshkova.

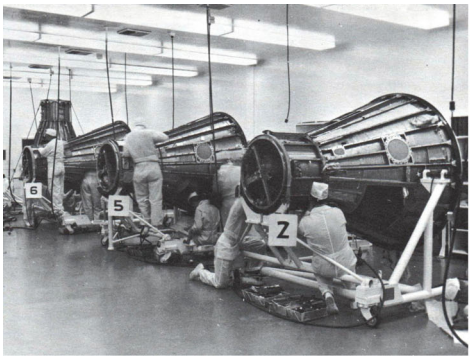
[15] With a spherically-shaped entry capsule—symmetric for any cross-section—only ballistic flight is possible, which typically gives rise to the largest thermo-mechanical loads.

the first cosmonaut? The Fédération Aéronautique Internationale (FAI) rules in 1961 stated that a pilot must land with the spacecraft so that for the FAI record books it is officially a spaceflight (Siddiqi 2000). The Soviet Union went to great length to hide the truth and officially insisted that Gagarin had landed with the Vostok, and forced him to lie in press conferences. And so, the FAI accepted Gagarin as the first cosmonaut. It was not until 1971 that the Soviet Union admitted that Gagarin had ejected and landed separately from the Vostok descent module, but by then it was well established that Gagarin was the first man in space.

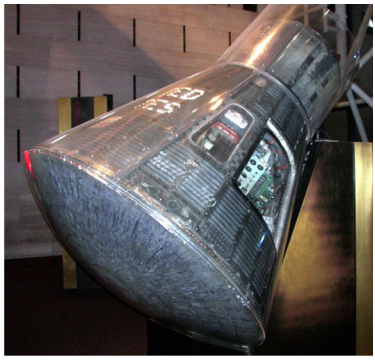
1.3.2 The Capsules

Project Mercury was the first human spaceflight programme of the United States, initiated in 1958 and continuing until 1963. The goal was to put a man into Earth orbit and return him safely, ideally before the Soviet Union, although that did not quite work out. The programme began under supervision of the USAF, but was taken over by the newly formed civil National Aeronautics and Space Administration (NASA). Seven astronauts were selected to fly the Mercury spacecraft, although in the end only six flew. They became known as the Mercury Seven and each spacecraft was given a name ending with a "7" by its pilot. Under NASA leadership, twenty unmanned developmental flights were conducted, some using animals, before moving on to the manned flights. On May 5, 1961, just under a month after Gagarin's orbital flight, Mercury flight MR-3 (Mercury Redstone-3, also known as Freedom-7) was launched with a Redstone MRLV MR-7. Its sub-orbital flight only lasted 15 minutes, but made Alan B. Shepard Jr. the first American in space. The primary objective of the mission was to demonstrate that the astronaut could endure the high *g–loads* during launch and re-entry.[16] The second sub-orbital flight followed on 21 July, 1961, when Virgil "Gus" Grissom completed his sub-orbital flight in his Liberty Bell-7. The third flight, finally, brought the Americans alongside the Russians: the Friendship-7 (MA-6), with on-board John Glenn Jr., was launched with an Atlas rocket. The Liberty Bell-7 achieved orbital flight, and John Glenn achieved what Yuri Gagarin had done three months earlier (Fig. 1.26). However, the Mercury re-entered the Earth's atmosphere after three orbits instead of one, and subsequently splashed down in the North Atlantic Ocean. Total flight duration: 4 h, 55 min, 23 s. Even though only for a short while, the US held the record of longest flight! On August 6, 1961, Vostok-2 was launched with on-board cosmonaut Gherman Titov. He stayed in orbit for a full day—and orbited Earth over 17 times—to study the effects of a more prolonged period of weightlessness on the human body. This new flight record was not broken by Aurora-7 (May 24, 1962) nor Sigma-7 (October 3,

[16] The *g–load* felt by the astronauts during the ballistic entry was 8 g, the same as endured by Gagarin, despite the different shape of the Vostok-1. We will see later, in Chap. 5, that the maximum *g–load* is independent from the so-called *ballistic parameter*, which is determined by the mass and the shape of the vehicle. Only the initial conditions at the beginning of re-entry will change this value.



(a) Manufacturing of the Mercury spacecraft at the McDonnell Aircraft plant, St. Louis (Image: NASA).



(b) John Glenn's Mercury MA-6, at display in the National Air and Space Museum, Washington.

Fig. 1.26 The Mercury spacecraft

1962), launched by Atlas and carrying M. Scott Carpenter and Walter M. Schirra, Jr., respectively. Eventually, it was the last Mercury flight on May 15, 1963 (Faith-7, with L. Gordon Cooper on-board) that orbited the Earth 22 times before entry and landing.

The Mercury was a one-man capsule (Fig. 1.27), 3.3 m long, 1.9 m in diameter, with an escape tower on top. The launch mass of the entry capsule for the orbital missions was about 1350 kg (with the exception of Sigma-7, which had a launch mass of 1,964 kg and a landing mass of 1,110 kg). The pressurised cabin was about the same size as that of a telephone booth; it was made of titanium, while the capsule's outer shell consisted of a nickel alloy, i.e., René 41, able to withstand high temperatures. The blunt end was covered with an ablative heat shield made out of an aluminium honeycomb covered with multiple layers of fiberglass to protect against heat of re-entry by slowly burning away.[17] Prior to landing the heat shield was detached to deploy a landing skirt to reduce the impact load. At the top of the capsule, the parachutes were stored: a drogue to stabilise free fall and two main chutes, a primary and reserve. Attitude control about all three axes was achieved by 18 small thrusters that were operated by the astronaut. The deorbitation engines, attached to the heat shield, were released prior to re-entry.

In the early days of Mercury, the second human spaceflight programme of NASA was announced. President Kennedy, in his State of the Union of May 25, 1961, said to the US Congress that the US "*should commit itself to achieving the goal, before this decade is out, of landing a man on the Moon and returning him safely to the Earth.*" One of the accommodations to prepare for Moon landing was Project Gemini that started in 1961 and concluded in 1966. Its objective was to develop techniques for

[17] The regression was about one centimetre, with a weight loss of about 7 kg (Launius and Jenkins 2011).

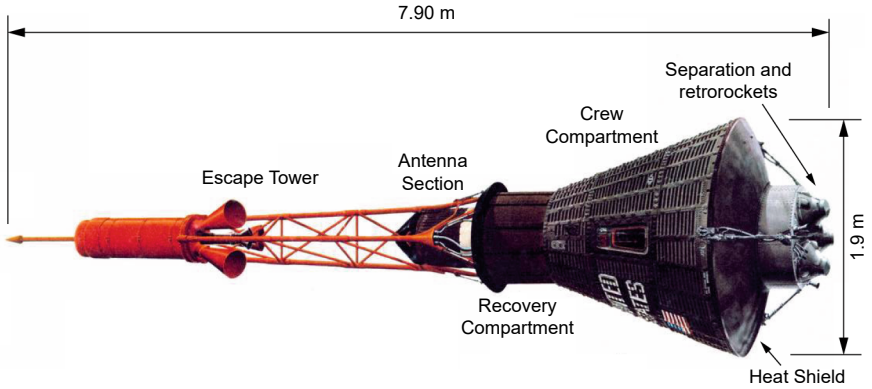
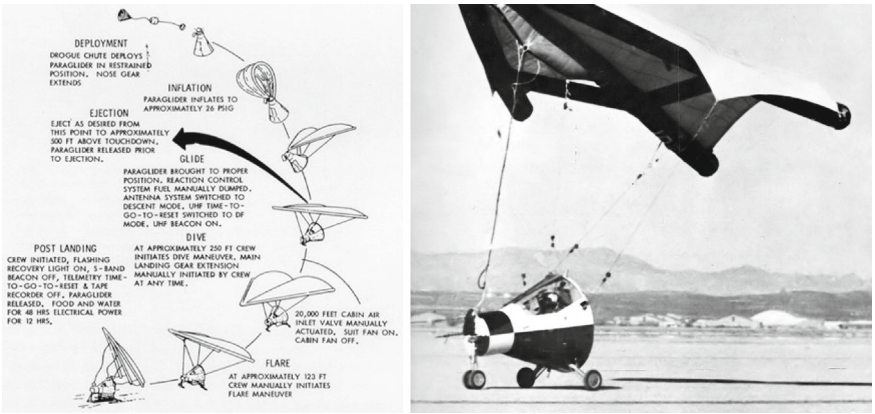


Fig. 1.27 The different components of the Mercury spacecraft. Adapted from NASA image



(a) Sequence of events (b) Tests at Edwards Air Force Base in August 1964

Fig. 1.28 Gemini paraglider. *Images* NASA

advanced space travel, consisting of (i) a two-person crew, (ii) missions long enough for a trip to the Moon and back, (iii) first American spacewalks, (iv) new orbital manoeuvres including rendezvous and docking, and (v) perfection of techniques for atmospheric re-entry and landing at a pre-selected location. Landing at a pre-selected location meant that the concept of sea landing would be reviewed, and in the original design the Gemini was supposed to land horizontally on land. To accommodate that and to avoid drifting away during parachute descent, the idea was to use an inflatable paraglider wing. Several low-altitude drop tests were done with success (Figs. 1.28 and 1.29), but ultimately, the design was dropped, and parachutes were used to make a sea landing.

Meanwhile, on the other side of the world, in the ongoing Space Race also the Soviet Union was working on its Moon Programme and had defined the Soyuz



Fig. 1.29 The Gemini paraglider on display at the Smithsonian Air and Space Museum Steven F. Udvar-Hazy Center; insert right: test payload for the paraglider wing, right insert: paraglider research vehicle 1A. Note the almost "casual" display in between an Apollo boilerplate Command Module, and two (flight models of) Gemini capsules. *Images* author

project. It was, of course, of prime importance that they would remain ahead in the race, and when NASA announced its Gemini programme the Soviets needed to do something. The Soyuz project was facing delays, and the progress in Gemini indicated that this would allow the United States to set new space records. To prevent his from happening, the Voskhod spacecraft series was conceived, which was derived from Vostok and much simpler than Soyuz, which was put on hold. Thus, the Voskhod programme became the second Soviet human spaceflight project. It began with two unmanned missions, one of which failed and another one successfully tested several hardware elements. Then, two one-day manned missions were flown. The Voskhod-1 was launched on 12 October, 1964, and with the three cosmonauts V. Komarov, K. Feoktistov and B. Yegorov on-board it became the first multi-man spacecraft. The descent module had a length (and diameter) of 2.3 m, and weighed 2,900 kg, of which the heat shield took 837 kg and the recovery equipment 151 kg. The landing parachute opened at an altitude of 2.5 km, and contrary to Vostok-1, the cosmonauts were landing inside the spacecraft. To soften the land impact, a landing rocket was used. The maximum deceleration during ballistic entry was the same as for Vostok, i.e., 8 g.

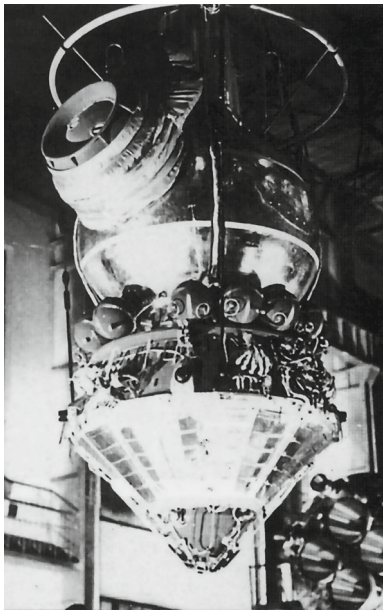


Fig. 1.30 Voskhod-2, with descent and service module, and an inflatable airlock. *Image* Roscosmos

Another first came with the Voskhod-2 (Fig. 1.30), launched on 18 March, 1965, when cosmonaut Alexei Leonov made the first-ever spacewalk of 12 minutes. To leave the capsule, an airlock attached to the capsule was inflated. This airlock was required to protect the avionics from a temperature drop and to keep the oxygen inside the capsule, as there was no reserve. A third, unmanned mission was a 22-day long flight with two dogs on-board in 1966, but before the effect of "long-duration" flight on the human body could be studied, the programme got cancelled and the four remaining manned missions scrapped. A change in policy of Soviet leadership allowed the designers to shift their focus back to the third manned spaceflight programme, the Soyuz.

In the meantime, Gemini progressed as well as could be. Launched by the Titan II launch vehicle, the Gemini spacecraft always carried a two-astronaut crew. The design had resulted in a scaled version of the Mercury capsule, 5.61 m long and 3.0 m wide (Fig. 1.31), with a launch weight varying from 3,220 to 3,790 kg (Rumerman 2007). Gemini weighed more than twice as Mercury, but with its 50% more volume it seemed to be tighter for double the crew. Contrary to Mercury, the retrorockets, electrical power, propulsion systems, oxygen, and water were located in a separate Adapter Module, connected to the Re-entry Module, and would be detached prior to re-entry. To perform rendezvous and docking manoeuvres (and changing orbit altitude and inclination), Gemini was equipped with an Orbit Attitude and Manoeuvring System, containing sixteen thrusters for translation control in three perpendicular axes. Like Mercury, it could also change its attitude in roll, pitch and yaw.



(a) Gemini-7 in orbit, as photographed from Gemini-6, with right the black entry capsule and left the white adapter module.



(b) Recovery of Gemini-8 astronauts David Scott and Neil Armstrong.

Fig. 1.31 The Gemini spacecraft. *Images* NASA

Instead of having an escape tower like Mercury, Gemini would have ejection seats, similar to the early Russian capsules. The ejection-seat system suffered from development problems, and extensive testing was performed. When NASA conducted an unmanned test of the ejection system, they found that the ejection seat performed perfectly... Unfortunately, the spacecraft's clam-shell hatch failed to open and the rocket-powered seat blasted right through the 2-inch thick hull plating. Watching nearby, astronaut John Young, destined to fly the first Gemini mission with Gus Grissom, was horrified, but maintained his sense of humour when he dryly remarked: "That's one hell of a headache, but a short one!" (Evans 2009).

Prior to the manned flights, in 1964 and 1965 two unmanned Gemini missions were flown to test out systems and the heat shield. During 1965 and 1966, when the Soviets had only launched two Voskhods, ten Gemini crews flew LEO missions. Although they were not the first with two-man crews, the experience they gained was significant, and in that respect it put the United States in the lead. In their future Moon programme, rendezvous and docking would be a crucial element, and this was (for the first time) tested with Gemini-6 and -7 that flew simultaneously in the same orbit. Even though they did not actually dock, the proximity operations taught NASA a lot (Fig. 1.31). Other major achievements were that (i) Ed White became the first American to make a spacewalk on June 3, 1965 (Gemini-4), (ii) Gemini-5 was in orbit for eight days, which showed that lunar flight (and back) was possible, enabled by the (first) use of new fuel cells to generate its electrical power, (iii) Gemini-7 broke the endurance record and stayed in orbit for 14 days, (iv) Gemini-8, by the hand of pilot Neil Armstrong, accomplished the first space docking with the pre-launched unmanned Agena target vehicle (derived from Lockheed Aircraft's Agena-D upper stage rocket), (v) Gemini-11 set a manned Earth orbital altitude record of 1,369.0 km in September

1966, using the Agena target vehicle's propulsion system, and (vi) Edwin "Buzz" Aldrin (Gemini-12) proved during a 2 h and 9 m tethered spacewalk that useful work could be done outside a spacecraft without life-threatening exhaustion (Rumerman 2007).

After cancellation of the Voskhod programme, the Soviet Union moved back to the development of Soyuz, their third manned space programme. Originally intended for their Moon programme, the Soyuz is a series of spacecraft designed by the Korolev Design Bureau (now RKK Energia) in the 1960s that is still in service today. Of course, it has seen gradual change over the years, with many different (flight) versions, the major ones being (i) Soyuz-A, cancelled to move on with Voskhod (1963), (ii) Soyuz-7K-OK (1967–1971), (iii) Soyuz-7K-L3/LOK (1970–1972), (iv) Soyuz-7K-T (1973–1981), (v) Soyuz-7K-TM (1975), (vi) Soyuz-T (1976–1986), (vii) Soyuz-TM (1986–2002), (viii) Soyuz-TMA (2003–2012), ix) Soyuz-TMA-M (2010–2016), and (x) Soyuz-MS (2016–current). As of mid-2019, the Soyuz is the only means of bringing astronauts to the International Space Station.[18]

A Soyuz spacecraft consists of three parts (see also Fig. 1.32a), i.e., a spheroid orbital module, which provides accommodation for the crew during their mission, a small aerodynamic re-entry module to return the crew to Earth (with entry masses varying between 2,500 and 3,000 kg), and a cylindrical service module with solar panels attached, which contains the instruments and engines. Of these three, only the re-entry capsule returns to Earth (Fig. 1.32b), whereas the orbital and service modules will burn up upon re-entry in the atmosphere. Entry is semi-ballistic, where the attitude-control system keeps a small positive attitude to create some lift that reduces the maximum deceleration load. However, in case of emergency the crew can be faced with a purely ballistic entry, thereby pulling around 8 g, as has been common with all Russian entry capsules.

The first Soyuz flight (version Soyuz-7K-OK) was unmanned and started on 28 November, 1966, whereas the first manned mission, Soyuz 1, launched on 23 April 1967. The latter ended with a crash due to a parachute failure, and it killed cosmonaut Vladimir Komarov (Yuri Gagarin was the backup crew for that mission). To date, the Soyuz-7K-OK was among the very few fatalities of the Soyuz programme, with in addition only Soyuz-11, where the crew died due to depressurisation of the capsule during re-entry (after they had made the first successful transfer of cosmonauts to the first space station in history, Salyut-1, on 7 June, 1971). Despite these tragedies, the Soyuz-7K-OK has also seen many (first) achievements, i.e., the first unmanned automated docking between two 7K-OK spacecraft in 1967, the first docking between two manned spacecraft (Soyuz-4 and Soyuz-5) in January 1969, and the longest manned flight involving a single spacecraft (Soyuz-9, in 1970, with an 18-day flight).

[18] The Chinese Shenzhou spacecraft is using Soyuz technology to support China's manned space programme. Even though its design resembles the Soyuz, it is larger in size (9.25 by 2.8 m, with a volume of 14 m^3). The first launch was on November 19, 1999, and the first manned launch, Shenzou-5, on October 15, 2003. Astronaut Yang Liwei was in space for more than 21 hours, and made China the third country with a successful manned space programme. The 11th and last launch was on 17 October 2016, and Shenzhou-11 rendezvoused with the Chinese space station Tiangong-2.



(a) Soyuz TMA-7, with the three distinct elements orbital, descent and service module (Image: Public Domain)



(b) Soyuz TMA-6 descent module (Image: S. Vigovskiy, through NASA)

Fig. 1.32 The Soyuz TMA, with 22 spacecraft launched between 2002 and 2012

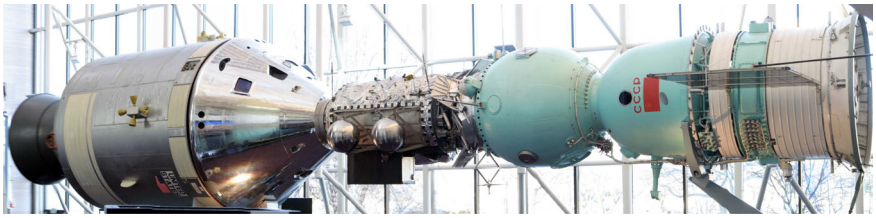


Fig. 1.33 Apollo-Soyuz in the National Air and Space Museum, Washington D.C. *Image* Toytoy at English Wikipedia

After the accident with Soyuz-11, design changes led to the second-generation of Soyuz spacecraft, the Soyuz-7K-T (Soyuz-12 through Soyuz-40). The main change was to reduce the crew size to two, such that the cosmonauts could now wear their space suits and to accommodate extra life-support systems, which was not the case in the earlier version. The Soyuz-7K-T was intended to only be a Salyut space-station ferry and had therefore no solar panels. The Soyuz-T was the third generation of the family, and was in service for seven years, with a total of 16 flights. It was a major upgrade from previous Soyuz, with improved systems, and it could carry three cosmonauts again, all wearing pressure suits. Many of the experiences, obtained during the Apollo-Soyuz test project (July 1975), where the US and USSR in a joint mission docked their Apollo and Soyuz capsules, were incorporated in the new Soyuz design. In Fig. 1.33 the docked configuration of the Apollo and Soyuz spacecraft is shown.

The Soyuz-TM crew transport is the fourth-generation Soyuz spacecraft (1986–2002, with a total of 34 flights), used to transport cosmonauts and astronauts to both the Russian Mir space station and the ISS. Improvements were related to systems for rendezvous and docking, communications, propulsion, and an integrated parachute/landing engine system. The Soyuz-TMA (2002–2012, 22 flights) was adapted to NASA requirements, amongst others to accommodate taller astro-

nauts, and it was followed by the upgraded Soyuz-TMA-M (2010–2016, 20 flights) with improved systems that also led to a mass saving of 70 kg. Finally, the latest generation of the Soyuz is the Soyuz-MS, with changes mostly in the communications and navigation subsystems. With a first launch on 7 July, 2016, the spacecraft is still in service, with Soyuz MS-13 launched on 20 July 2019.[19] It is surprising, though, that despite the many hardware improvements made over more than 55 years, the entry procedures have not changed much: close to a ballistic entry, concluded by a parachute descent and a landing assisted by a short burst of the landing rockets.

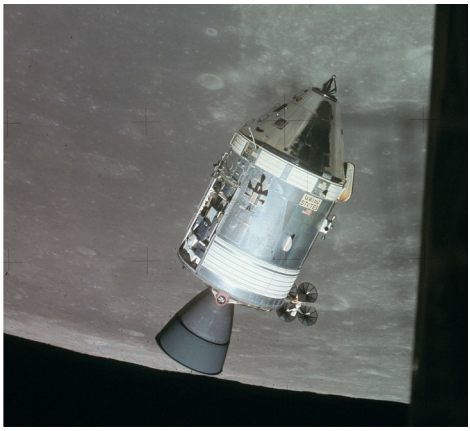
The third part of the US manned space programme during the Space Race was the Apollo programme, and had the objective to execute the actual Moon landing (Rumerman 2007; Woods 2008). Apollo ran from 1961 to 1972, and until 1966 all went well. The focus was not only on the three Apollo components, i.e., the Apollo Command Module (CM), the Service Module (SM), and the Lunar Module (LM), see Fig. 1.34. It also required the development of a new launcher, the Saturn V, which would be strong enough to launch 45 tonnes into a transfer orbit to the Moon, as well as a new flight-control centre in Houston, two new launch pads at Cape Canaveral, the 160 m high Vehicle Assembly Building (VAB) to integrate the 110 m tall Saturn-V, and the so-called crawler that had to transport the launcher configuration to the launch pad. The Apollo CM was the conic-shaped crew cabin (3.91 m wide, 3.23 m high, and having a launch mass of about 5,500 kg), with an ablative heat shield made of phenolic formaldehyde resin of maximum 5.1 cm thickness. It was equipped with a reaction control system (RCS) for attitude control both outside and inside the atmosphere, and its descent and landing system consisted of two drogue parachutes with mortars, three main parachutes, three pilot parachutes to deploy the mains, and three inflation bags for uprighting the capsule after landing in the ocean, if necessary. The SM was a 7.5 m long cylinder with a diameter of almost 4 m, containing the orbital engine, the fuel and oxidiser tanks, as well as the oxygen tanks for the CM. The LM was the vehicle that performed the actual Moon landing, and would bring two astronauts to the lunar surface. It consisted of two parts, the bottom part with the descent engine would also serve as launch platform for the top part, which had an ascent engine to bring the astronauts back to the CSM (CM+SM) that remained in orbit around the Moon. The total LM was about 7 m high, 9.4 m wide and 9.4 m deep, and had a launch mass of nominally 15,200 kg.

Disaster struck on 27 January, 1967, when a fire broke out in the Apollo-1 CM during a prelaunch test, killing the three astronauts Virgil Grissom, Roger Chaffee and Edward White. Main reasons were that the escape hatch could not be opened from the inside and the atmosphere in the cabin was 100% oxygen, burning extremely fast. The redesign caused a delay, of course, and it was not until 9 November, 1967, that the redesigned Apollo-4 (CSM) was launched with the Saturn-V to test the CM heat shield. The translunar entry velocity was obtained by boosting the CM with the SM's main engine. Apollo-5 carried the (unmanned) LM to test its manoeuvrability in LEO. Apollo-6 did further tests of the CM and the Saturn-V.

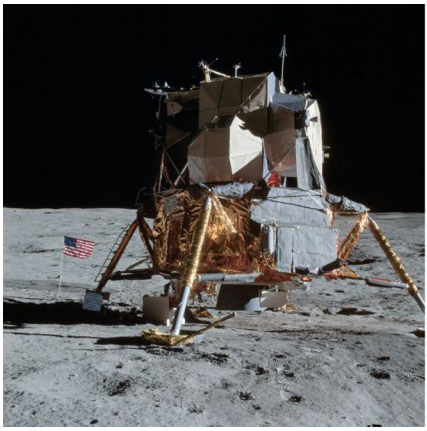
[19] The launch of Soyuz MS-10 was aborted shortly after launch due to booster failure (11 October 2018), but the crew landed safely 20 km east of Jezkazgan, Kazakhstan.



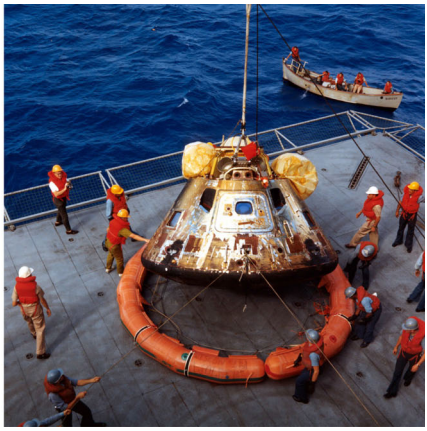
(a) Launch with Saturn V



(b) Apollo-15 Command and Service Module, seen from the Lunar Module



(c) Lunar Module of Apollo 14 on the Moon



(d) Recovery of Apollo-11 Command Module

Fig. 1.34 The main elements of the Apollo Programme. *Images* NASA

On 11 October, 1968, the Apollo programme started in all earnest with the manned flights (see Table 1.1 for an overview of the manned missions with some characteristics). We will not go into much detail here, as that could easily take up the rest of the book, but focus on some aspects relevant for entry, descent, and landing.[20] With Apollo-7 the first manned entry into the Earth's atmosphere was performed, albeit at a lower speed than when returning from the Moon. Nonetheless, the CM and its heat shield performed well. The entry was no longer a ballistic flight, as

[20] In Chap. 9, the Apollo mission will be discussed in far more detail, in terms of descent trajectory to the lunar surface, as well as the entry and parachute descent into the Earth's atmosphere.

Table 1.1 Manned apollo flights

Name	Crew	Launch	Landing	Time on moon	Remark
Apollo-1	Grissom Chaffee White	27-01-1967			Fire during ground test, astronauts died
Apollo-7	Schirra Eisele Cunningham	11-10-1968	22-10-1968		Ten-day test flight of Apollo CM (168 Earth orbits)
Apollo-8	Borman Lovell Anders	21-12-1968	27-12-1968		First manned flight around Moon
Apollo-9	McDivitt Scott Schweickart	03-03-1969	13-03-1969		First manned test Apollo CM and Lunar Module Earth orbit
Apollo-10	Stafford Cernan Young	18-05-1969	26-05-1969		First test Lunar Module in lunar orbit
Apollo-11	Armstrong Aldrin Collins	16-07-1969	24-07-1969	21 h 36 m	First manned Moon landing
Apollo-12	Conrad Bean Gordon	14-11-1969	24-11-1969	31 h 31 m	Visit to Surveyor-3
Apollo-13	Lovell Haise Swigert	11-04-1970	17-04-1970		Explosion in Service Module; safe return crew
Apollo-14	Shepard Mitchell Roosa	31-01-1971	09-02-1971	33 h 30 m	Landing near hills; use of wheelbarrows
Apollo-15	Scott Irwin Worden	26-07-1971	07-08-1971	66 h 55 m	Landing near mountains; first lunar rover
Apollo-16	Young Duke Mattingly	16-04-1972	27-04-1972	71 h 00 m	Second lunar rover
Apollo-17	Cernan Schmitt Evans	07-12-1972	19-12-1972	74 h 59 m	Sixth/last manned Moon landing; third lunar rover

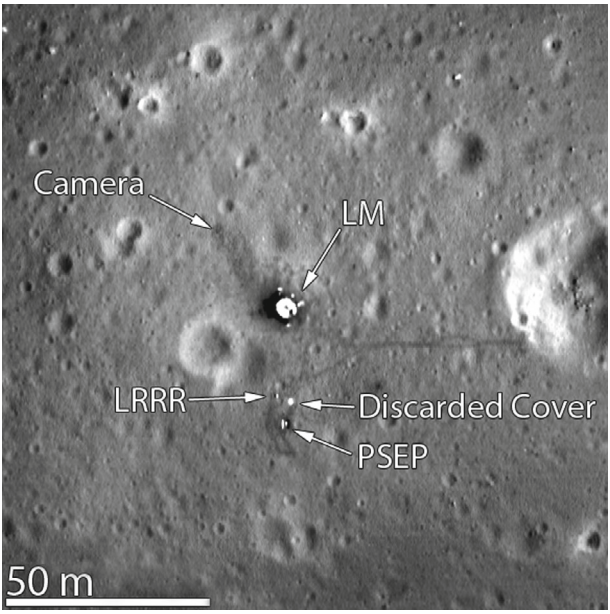


Fig. 1.35 Tranquility Base, imaged in March 2012 by the Lunar Reconnaissance Orbiter. *Image* NASA/GSFC/Arizona State University

was the case for Mercury and Gemini, but due to a clever location of the capsule's centre of mass a natural angle of attack was obtained, providing lift that reduced the thermo-mechanical loads.

Apollo-10 was the dress rehearsal for the actual Moon landing of Apollo-11. The LM descended towards the lunar surface, and the crew took many pictures of *Mare Tranquillitatis* (Sea of Tranquility) to prepare the mission planning for Apollo-11. The LM was under-fuelled, i.e., it had the amount of fuel that a LM would have when it would be ascending from the Moon. This was done as a precaution, so that the Apollo-10 crew was not tempted to actually land. Had they done that, they would not have been able to take off again. Further worth mentioning about the return flight of Apollo-10 is that it set the record for the highest speed attained by a crewed vehicle: 39,897 km/h (= 11.08 km/s, just under the Earth's escape velocity of 11.2 km/s).

And then the day of a man setting foot on another celestial body came: "*Houston, Tranquility Base here. The Eagle has landed.*", said Neil Armstrong after landing (at 20:17:40 UTC on 20 July, 1969), a couple of hours later (at 02:56 UTC, 21 July) followed by: "*That's one small step for [a] man, one giant leap for mankind*". Kennedy's goal was accomplished with some months to spare. In Fig. 1.35, the landing site as observed by the Lunar Reconnaissance Orbiter (LRO) is shown, with the launch platform of the LM, the Passive Seismic Experiment Package (PSEP), and the Lunar Ranging Retro Reflector (LRRR), as well as the trails that the astronauts walked.

Five subsequent Apollo missions also landed astronauts on the Moon (Apollo-13 was supposed to be the third mission to land, but an explosion in the SM forced the astronauts to abort the mission and return to Earth), the last in December 1972. In these six spaceflights, twelve men walked on the Moon, while three missions carried lunar rovers that in total drove 90.2 km. A total of 382 kg Moon rock was returned to Earth. In total, the Apollo programme had cost $25.4 billion (or approximately $153 billion in 2018 dollars when adjusted for inflation via the GDP deflator index[21]). This was the main reason that the last three remaining missions, Apollo-18 to –20, got scrapped and the programme was cancelled.

The first Moon landing was, of course, a major win in the Space Race. It was not that the Russians were not doing anything in that respect, although at the time they denied taking part in the race to the Moon. They secretly pursued two programmes in the 1960s: manned lunar flyby missions using the Soyuz-7K-L1 spacecraft (also known as Zond), launched with the UR-500K (Proton) rocket, and a manned lunar landing using Soyuz-7K-L3 and LK Lander spacecraft launched with the N1 rocket. Unmanned Zond were launched into a slingshot trajectory around the Moon to test the return flight into the atmosphere and make use of the skip-flight principle (bouncing off the atmosphere) to reduce the entry velocity, conceived by Eugen Sänger. Zond-5 (September 1968), with two turtles on board, was the first spacecraft to circle the Moon and come back to Earth, but failed to do the skip and came in ballistic—but the turtles survived! The Americans believed that the "success" of Zond-5 would soon be repeated with human cosmonauts, and this gave a push to the Apollo programme. Zond-6 performed the first successful skip entry on 16 November 1968, and also Zond-7 and -8 managed to do this. A manned flyby was never performed, though. Major setbacks in the Russian Moon programme originated from the lack of success with the N1 development. This launcher, equivalent to the Saturn-V, was required to bring the cosmonauts to the Moon and back.

Two main reasons made the Russians decide to stop their Moon programme: (1) the dual American successes of the first manned lunar orbit on December 24–25, 1968 (Apollo 8) and the first Moon landing on July 21, 1969 (Apollo 11), and (2) a series of catastrophic N1 failures. And maybe there was a reason #3 ... Korolev died in 1966, and his deputy, Vasily Mishin (1917–2001), did not have Korolev's political astuteness or power; this led to the eventual downfall of the N1. All four unmanned N1 launches (1969–1972) out of 12 planned tests ended in failure, each before first-stage separation. The longest flight lasted 107 s, just before first-stage separation. The two flight-ready N1Fs were scrapped and their remains were unceremoniously used as shelters and storage sheds. The boosters were deliberately broken up in an effort to cover up the USSRs failed Moon attempts, and for a long time the Russians claimed that they were never planning to send men to the Moon.

[21] United States Congress House Committee on Science and Astronautics (1973). 1974 NASA authorization: hearings, Ninety-third Congress, first session, on H. R. 4567. Page 1271. Washington: US Government Printing Office.

1.3.3 Gliding Flight

Ever since the onset of the Space Age it was clear for the engineers that to lower the loads on astronauts and vehicles, as well as to increase flight range and manoeuvrability, the vehicles' lifting capabilities would have to be increased. Experience with the traditional capsule-like spacecraft from the Mercury, Gemini, and Apollo series had shown that they had very little control over where they landed. In the 1950 and 1960s, a lot of research was carried out and many flight-test vehicles were designed, built, and flown. Crucial for the development of the US Space Shuttle were the so-called lifting bodies, so rather than adding wings to the vehicle to produce lift, the body itself would be the main contributor.

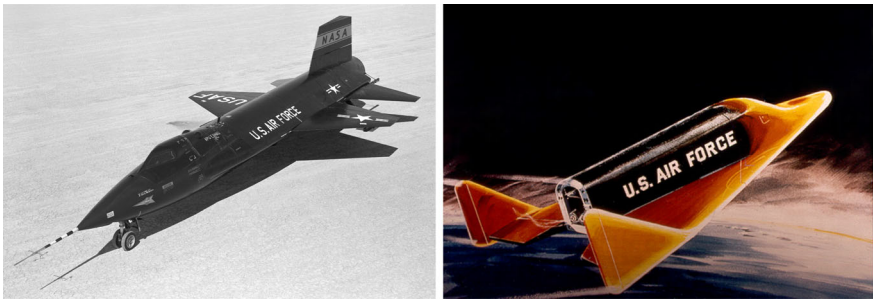
These lifting bodies did not stand by themselves. A whole series of experimental aircraft, so-called X-planes, had been conceived before to support the development of advanced aircaft and spacecraft alike (Miller 2001). On October 14, 1947 it was Chuck Yeager to fly the Bell X-1 (Fig. 1.36), an experimental rocket-powered aircraft, at Mach 1 at an altitude of 13,700 m, and to become the first one to "break the sound barrier", although the only barrier to "break" was to overcome the steep increase in drag prior to reaching Mach 1. The rocket-propelled X-15 (Fig. 1.37a) was designed to be a hypersonic research aircraft, with programme initiation traced back to June 24, 1952, and its maiden flight taking place on June 8, 1959.[22] The three constructed X-15s flew a total of 199 test flights, with the last one on 24 October 1968, and during those flights the X-15 set speed and altitude records. The official world record for the highest speed ever recorded by a manned, powered aircraft was achieved on October 17, 1967, by an X-15: William J. Knight tested the boost guidance, and during descent reached Mach 6.70 at 31,120 m (equivalent to a velocity of 2,021 m/s). As of 2019, this record remains to be broken.

Like many of the X-planes, the X-15 was designed to be carried aloft and drop launched from under the wing of a B-52 mother ship. It had a thick wedge tail to provide hypersonic directional stability, although this significantly increased the drag at lower speeds. Because the X-15 flew to the edge of space, where the air was too thin to support the use of control surfaces, it was also equipped with an RCS that used small rocket thrusters, and it showed a smooth transition between aerodynamic and reaction control, and back. In the course of the flight programme, eight pilots received their "astronaut wings", because during a total of 13 flights they exceeded the altitude of 80 km. From the many technical achievements of the X-15 programme, only a few are mentioned here. It had the first reusable superalloy structure capable of withstanding the temperatures and gradients of hypersonic entry, it led to improved material processing techniques for Inconel-X and titanium, and the discoveries that the hypersonic boundary layer is turbulent and not laminar, that turbulent heating rates are significantly lower than predicted, and that "hot spots" are generated by

[22] Before 1958, USAF and NACA officials discussed an orbital X-15 spaceplane, the X-15B, that would launch into outer space from atop an SM-64 Navaho missile. This project was cancelled in favour of the Mercury programme.



Fig. 1.36 The Bell X-1 rocket-powered aircraft on display at the Smithsonian Air and Space Museum Steven F. Udvar-Hazy Center. *Images* author



(a) X-15 (b) X-20

Fig. 1.37 The X-15 #1 on a lakebed, 1960, and an artist impression of the X-20 *DynaSoar*. *Images* NASA

surface irregularities, and finally the first direct measurement of hypersonic skin friction, and that it was lower than predicted (Miller 2001).

The Boeing X-20 (Fig. 1.37b), nicknamed *DynaSoar*, from the flight characteristic of "dynamic soaring", was a USAF programme to develop a piloted manoeuvrable spaceplane that could be used for a variety of military missions, although it was also intended for conducting hypersonic and orbital experiments (Miller 2001). The military missions would include aerial reconnaissance, bombing, space rescue, satellite

maintenance, and possibly to sabotage enemy satellites, although clear decisions on what the X-20 should be capable of doing were lacking. In a way, the development of the X-20 followed up on the ideas of Sänger's hypersonic bomber and its US counterparts of the 1950s, the Bomber Missile "BoMi", Hypersonic Weapons Research and Development Supporting system (Hywards), the Brass Bell reconnaissance vehicle, and the Rocket Bomber "Robo" (Rose 2008). The programmed kicked-off on 24 October 1957, and by 1959, the X-20 had become the USAF's preferred means for launching military manned spacecraft into orbit. Towards the end of 1963 (December 10), however, the programme was cancelled, despite the fact that construction of the X-20 had started and that about 5.5 billion current-day dollars had already been spent. Main reasons for the cancellation were, apart from the obvious budgetary problems, the uncertainty over the booster's ability to launch the vehicle into orbit and a lack of a clear project goal.

The X-20 was to be launched on top of a rocket (only late in the project the Titan III was chosen), and due to its delta-wing shape it would fly more like an aircraft than a space capsule. Nominally it would be launched into a sub-orbital trajectory, but it would use a skip-glide technique to extend its flight range, although there were plans for orbital flights as well. The fuselage would have been constructed of the René-41 super alloy, as were the upper surface panels. The bottom surface was to be made from molybdenum sheets placed over insulated René-41, while the nose-cone was to be made from graphite with zirconia rods. Despite its cancellation, the X-20 has served as a testbed for many advanced technologies, including aerodynamics, structures and re-entry heating data obtained from its materials programme, and provided a very important input to the Space Shuttle development. In addition, guidance-system development found its way into the X-15 and other programmes.

One crucial element in the development of gliding entry vehicles consisted of the so-called lifting bodies (Fig. 1.38), although in the 1960s tens of spaceplane concepts were studied as well, these never left the drawing boards (Jenkins 2008). Even though the concept of lifting bodies had been known for a long time, NASA's refinements started in the early 1960s with the M2-F1, followed by the M2-F2, X-23 PRIME, HL-10, M2-F3, X-24A & B, and Hyper III[23] (see for an overview Table 1.2, and for more details on the X-planes the excellent book by Miller (2001)). The lifting-body concept was conceived to allow for horizontal landing after atmospheric re-entry, thereby reducing the aerothermodynamic heating due to the lifting capabilities as compared to the ballistic capsules. The absence of the wings would further reduce the risk of extreme heating damaging the vehicle. Additionally, having lift would increase the manoeuvrability and thus the landing accuracy.

All of the lifting bodies were test flown from NASA's Dryden Flight Research Center (currently the Neil A. Armstrong Flight Research Center), except for the X-23 PRIME that was launched from Vandenburg Airforce Base. It all began with

[23] Hyper III was developed to support the M2 programme, and was a full-scale, radio-controlled lifting body with a flat bottom and sides, and a simple straight wing with no control surfaces. It was designed to simulate a pop-out wing that had been proposed for a re-entry vehicle. However, after a single three-minute test flight, i.e., launched from a helicopter at 3,300 m it glided 5 km, turned round, came back and landed, the programme got cancelled.



Fig. 1.38 Lifting bodies X-24A, M2-F3 and HL-10. *Image* NASA

the M2-F1, nicknamed the "flying bathtub", which was a lightweight, unpowered prototype aircraft, developed to flight-test the wingless lifting body concept. Initially towed by a car, and further accelerated by its solid-fuel rocket, it was subsequently towed by a C-47 aircraft. It could reach a maximum velocity of 240 km/h and had a flight range of 16 km. After more than 400 ground tows and 77 air tows, the last flight of the M2-F1 took place on 16 August, 1966, and by then it had proved the lifting-body concept and paved the way for more heavyweight designs, Northrop's M2-F2 and HL-10, and the USAF X-24 programme.

The M2-F2 had an empty weight of 2,095 kg, was 6.76 m long and had a "wing span" of 2.94 m. It was equipped with a Reaction Motors upgraded XLR-11 four-chamber rocket engine, with 36 kN of thrust (although the engine was never used). Air dropped by a modified B-52, it would perform unpowered gliding flights from about 13.7 km altitude, and reach a maximum speed close to 750 km/h. Most tests were done to establish the performance, longitudinal and lateral stability characteristics, and the controllability. A crash at the end of the 16th flight ended the programme. The HL-10 flew more successfully, with unpowered glides and powered flights, both subsonically and supersonically. The highest Mach number obtained was 1.86, and after 37 flights the very successful programme came to an end. The lifting-body results were so valuable that early designs of the Space Shuttle were derived from the HL-10 shape.

Because the pilots found that the M2-F2 had lateral-stability problems, its successor M2-F3 was rebuilt from the crashed vehicle, but with an additional third vertical fin. After three years of redesign and rebuilding, the first test flight of the M2-F3 took place on 2 June, 1970. Lateral stability had indeed improved significantly. Many aspects of controllability were investigated, amongst others the use of a rate-command augmentation control system, and a side-arm control stick similar to those still in use in many aircraft. A reaction-control system was installed, similar to those used in the low-density regimes of the atmosphere to test its effectiveness for vehicle control.

Table 1.2 Lifting-body research in the 1960–1970s (Images: NASA; X-23 Image: USAF)

	Vehicle	Year[a]	# Flights	Characteristics
	M2-F1	1963–1966 (16-08-63)	77	• Unpowered • Proof-of-concept of wingless lifting body concept of atmospheric entry vehicles • Demonstrated low-cost prototyping and flight testing
	M2-F2	1966–1967 (12-07-66)	16	• Heavyweight lifting body • Unpowered glides • Testing of stability and control characteristics • Crash landing
	X-23 PRIME	1966–1967 (21-12-66)	3	• Launched with Atlas into sub-orbital trajectory • Aerodynamic and reaction control • Structural and heating research • Mid-air retrieval by JC-130B Hercules after Mach 2 ballute and subsequent parachute deployment • Two vehicles lost during recovery, one successful

(continued)

Table 1.2 (continued)

	Vehicle	Year[a]	# Flights	Characteristics
	HL-10	1966–1970 (22-12-66)	37	• Heavyweight lifting body • Powered and unpowered flight • Validate concept of safely manoeuvring and landing a low lift-over-drag vehicle • Fastest (1,976 km/h) and highest (27.524 km) lifting-body flight
	Hyper-III	1969 (12-12-69)	1	• Support of the M2 lifting-body programme • Programme cancelled after a single three-minute flight
	M2-F3	1970–1972 (02-06-70)	27	• See M2-F2 • Redesign with third fin for better lateral stability and control characteristics
	X-24A	1969–1975 (17-04-69)	28	• Powered and unpowered • Dropped from B-52 • Rocket engine used to ascend to higher altitudes • Unpowered reentry and landing

(continued)

Table 1.2 (continued)

	Vehicle	Year[a]	# Flights	Characteristics
	X-24B	1973–1975 (01-08-73)	36	• See under X-24A • Higher lift-to-drag ratio • Max. velocity 1873 km/h, max. altitude 22.59 km • Operational feasibility of accurate unpowered reentry vehicle landings

[a] The time range indicates the operational years only. First flights are indicated in brackets

With the Martin-Marietta X-23 PRIME, the USAF took a new direction of testing lifting bodies. They had gained experience with its ASSET flight-test vehicle, six of which were launched between 1963 and 1965, reaching a maximum altitude between 50 and 70 km, and a velocity ranging from 4 to 6 km/s. The USAF had focussed mostly on verification of the superalloy heat shield of the X-20. ASSET's biconic shape and low delta wing were intended to represent DynaSoar's forward nose section, where the aerodynamic heating would be the most severe (about 2200 °C). After the X-20 got cancelled, the remaining vehicles were used in re-entry heating and structural investigations.[24] With X-23 PRIME, the focus was put on manoeuvrability during re-entry into the Earth's atmosphere, with cross-range manoeuvres up to about 1,150 km. The X-23 was constructed from titanium, beryllium, stainless steel, and aluminium, and fully covered with an ablative heat shield varying between 2 and 7 cm thickness. The nose cap was made out of a carbon phenolic material.

Launched by Atlas, the first vehicle entered the atmosphere at Mach 25. A ballute was deployed at Mach 2, but the parachute system failed to open and the vehicle crashed in the ocean.[25] The second flight was not so much more successful, in terms of recovery. Although the parachute had deployed, its reefing line cutters had not worked and the parachute was not fully opened. This made it impossible for the JC-130B Hercules aircraft to perform the mid-air retrieval recovery procedure. The third flight, however, was a complete success. It executed bank manoeuvres at hypersonic speeds to induce cross-range, controlled by its pair of lower flaps (whereas outside the atmosphere a nitrogen-gas RCS was used), the ballute and parachute systems worked well, and the vehicle was recovered mid-air.

[24] Out of the six missions, one failed due to launcher malfunctioning, one was recovered, two were not planned to be recovered, but all mission goals were met, and two had failed to be recovered.

[25] Ballutes and parachute systems will be discussed extensively in Chap. 9.

The X-24A was derived from the X-23 PRIME, and had the same fat, short teardrop shape. Contrary to the X-23, its vertical fins were movable and used for control. Launched from under a B-52, the X-24A would either glide down, or use its rocket engine to accelerate and climb to a higher altitude. The X-24A unquestionably demonstrated that lifting-bodies could consistently make precision landings onto a hard runway, thereby confirming the concept for the future Space Shuttle. The X-24B had a completely different shape than the X-24A, because the engineers at Edwards Airforce base wanted to perform similar tests on an even more advanced lifting-body shape. Most of the 36 flights were used to demonstrate precision-landing techniques that would be used on the forthcoming Space Shuttle. The X-24B was the last aircraft to fly in Dryden's lifting-body programme, in which well over 200 flight tests were carried out with six different vehicles.

The path the USSR followed was in a way similar to that of the US, albeit way more secretive. The Soviet Union had also seen a long history in rocket-plane development and they had (and still have) their share of experimental aircraft and spacecraft (Gordon and Gunston 2001). For space-limitation reasons we will focus on the Mikoyan-Gurevich MiG-105 "Spiral", the associated series of BOR test vehicles, and eventually the Russian space shuttle, Buran (Hendrickx and Vis 2007). Starting 1965, the Russian lifting-body Mikoyan-Gurevich MiG-105 or EPOS (Russian acronym for Experimental Passenger Orbital Aircraft) was developed, an air-launched spaceplane stemming from similar military requirements as the X-20. The Spiral spaceplane had the shape of a conventional delta wing, a large upturned nose to reduce the aerothermodynamic heating, and included innovative variable-dihedral wings. During launch and reentry, these were folded upward at 60°, defining the vehicle to be a lifting body, but at subsonic speeds the pilot lowered the wings into the horizontal position to improve flying characteristics. Spiral was equipped with an engine that could be used for a powered landing and a retry in case the landing approach was missed. The idea was that the spaceplane, combined with a liquid-fuel booster stage, would be launched at high altitude from the back of a custom-built hypersonic jet. The latter was never built, but several test vehicles were built to test the spaceplane.

The (scale) models of the spaceplane were known as Unmanned Orbital Rocket Planes (Bespilotnye orbitalnye raketoplany or BOR). BOR-1 was a (flight-tested) wooden mock-up, launched on 15 July, 1969, atop a Kosmos-3M launch vehicle. Deployed at 100 km altitude, it burned in the atmosphere at a height of about 60–70 km at a speed of 13,000 km/h, as planned, and transmitting data on its trajectory until the last moment. BOR-2 (Fig. 1.40a), a 1:3 scale model of Spiral, saw four sub-orbital launches between 1969 and 1972. Equipped with an ablative heat shield, it further tested the lifting-body shape. BOR-3 was the larger 1:2 scale model (1973–1974), but both vehicles were lost: the first one during launch and for the second one the parachute failed to open, after an otherwise successful mission.

A full-scale prototype for subsonic flights (known as MiG 105-11) had been constructed by the mid-1970s (Fig. 1.39). Between June 1976 and September 1978, it made several manned autonomous flights, using its own turbojet engine to take off and land 10–15 s later at the same runway, as well as piloted drop tests from a



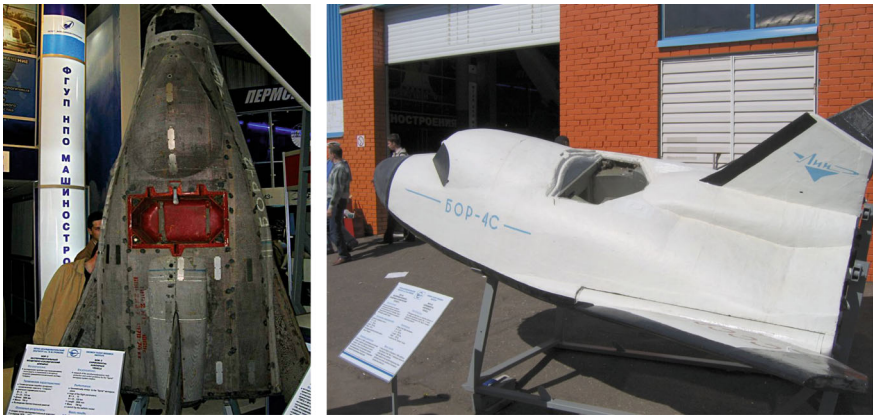
Fig. 1.39 The Russian lifting-body Mikoyan-Gurevich MiG-105 "Spiral", at the Monino Central Air Force Museum (Moscow). *Image* Bernhard Gröhl

Tupolev Tu-95K. This modified bomber aircraft was used as missile platform in a similar manner as the Americans used the B-52. A model for supersonic tests was built, but never flown, and a model for hypersonic tests was only partially built.

Interest in the reusable spaceplane started to diminish in the mid-seventies, when the Russian military became more and more worried about the US Space Shuttle development. In their opinion, the Shuttle was never going to be economically viable, unless it had a strong (not publicly announced) military focus. This idea was used to convince president Brezhnev to put (a lot of) money in the development "to get what the Americans have", and to stop development of Spiral, because "this is not at all what the Americans are doing". Approval for the development of the Buran space shuttle came, but the Spiral test vehicles were not immediately discarded. The engineers realised that to test, for instance, heat-shield technology, the shape of the vehicle is less important.

So, BOR-4 had survived (Fig. 1.40b), and with it the move from sub-orbital to orbital missions was made. Initially developed to support the Spiral spaceplane programme, after cancellation its purpose was redirected to test the heat-shield tiles and reinforced carbon-carbon for Buran, then under development. It would also provide critical data on the vehicle's aerodynamic characteristics through the different flight regimes. A total of seven vehicles was built, with two (unconfirmed) sub-orbital launches and four confirmed orbital launches between 1982 and 1984, of which three were recovered. The first flight may have been the first try of BOR-4 for the orbital flight, on 5 December, 1980. After BOR-4, there was BOR-5 and BOR-6, but these were dedicated to the Buran programme and will be discussed shortly.

In the 1960s, there were many discussions and studies going on about developing space stations, and to reduce the launch cost to support construction of these stations. Many different configurations were studied, but *the* reusable space-launcher concept was not easily established (Jenkins 2008). The cost of the project was a major issue,



(a) BOR-2 (b) BOR-4S

Fig. 1.40 Spiral spaceplane test vehicles, BOR-2 and BOR-4S, at MAKS airshow (Zhukovskiy, Russia), 2005. *Images* WikiCommons

and to spread out the cost an effort was made to interest the US Airforce. In the late 1950s, the then military projects had been combined in the X-20 programme, and after cancellation in 1963, the US military were looking for solutions to launch large payloads and have sufficient manoeuvrability to guarantee a quick return from orbit and reach military airfields. Also a quick turn-around time, and the recovery of payloads from Earth orbit and return them to Earth, became sought-after capabilities. The USAF demanded a much larger vehicle, and to lower the cost, several allowances on reusability were made. As a feasible compromise, the Space Shuttle programme originated in the late 1960s, and the final agreed-upon design would include a reusable spaceplane (the orbiter) with three liquid-fuelled engines, an expandable external tank (ET), which would hold the fuel for those engines, and two reusable Solid-fuel Rocket Boosters (SRBs) that were chosen over liquid-fuelled boosters for the lower development cost.

The first orbiter built was the *Enterprise*, which did not have any orbital capabilities (it was built without engines and functional heat shield), and was meant to do approach and landing tests. Originally named *Constitution*, "Star Trek" fans started a petition asking the White House to instead select the name of the starship that James T. Kirk captained in the original TV series. President Johnson was not against this idea ... The crew of the Star-Trek series was invited to the roll-out of the *Enterprise* on 17 September, 1976 (Fig. 1.41).

The approach and landing tests were a series of taxi and flight trials, conducted in 1977 to test the vehicle's flight characteristics both on its own and when mated to the Shuttle Carrier Aircraft, a modified Boeing 747. A total of 16 tests were carried out, i.e., three taxi tests, five captive-inert flights (the (passive) orbiter remained mated to the aircraft to test the flying characteristic of the mated configuration), three captive-active flights (manned and active orbiter mated to the aircraft, to determine the



Fig. 1.41 Crew members of the *United Space Ship (USS) Enterprise (NCC-1701)*, popular culture icon from the television series *Star Trek*, attend the roll-out of America's first Space Shuttle, named *Enterprise*, in Palmdale, California, on 17 September, 1976. *Image* NASA (From left to right they are: Dr. James C. Fletcher (NASA Administrator), DeForest Kelley (Dr. "Bones" McCoy), George Takei (Mr. Sulu), James Doohan (Chief Engineer Montgomery "Scotty" Scott), Nichelle Nichols (Lt. Uhura), Leonard Nimoy (logically, Mr. Spock), Gene Roddenberry (*Star Trek* creator), democratic congressman Don Fuqua, and Walter Koenig (Ensign Pavel Chekov))

optimum separation conditions to conduct future free flights), and five free flights to do the actual approach and landing tests. Other important tests carried out were the so-called mated vertical ground vibration tests (4 October, 1978, and onwards), where *Enterprise* was mated with an external tank and two inert solid rocket boosters to do vibration tests representative of different (launch) phases, and the transport of the mated configuration to Launchpad 39-A (1 May, 1979) (Jenkins 2008).

After the roll-out of *Enterprise*, it took four more years of development, and on 12 April, 1981, Space Shuttle *Columbia* (STS-1) lifted off from Kennedy Space Center, with commander John Young and pilot Robert Crippen onboard the first orbital flight (Fig. 1.42). There were, of course, many (minor) anomalies detected, which was to be expected with the first flight of such a complex system, of which not all systems can be adequately tested on-ground. The end result was a huge success, though, when after 53 h in space doing 36 orbits *Columbia* landed safely at Edwards Airforce Base, California. Major anomalies concerned a significant mis-prediction of the orbiter's aerodynamics at high Mach numbers during re-entry, resulting in an excessive body-flap deflection to trim the vehicle, and a stronger roll-yaw coupling during the bank



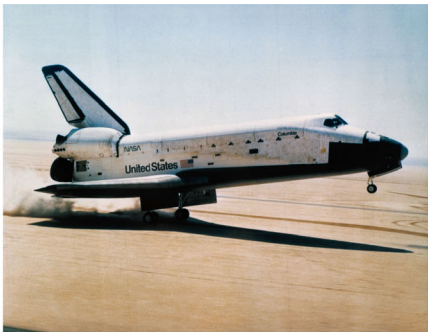
(a) Launch (Image: NASA).



(b) In orbit (Image: NASA).



(c) Gliding descent (Image: NASA).



(d) Landing (Image: Edwards AFB History Office).

Fig. 1.42 The first Space-Shuttle mission, Columbia (STS-1), with launch on April 12, 1981

manoeuvres. Also, some damages in the heat shield were found. All anomalies were studied and after modifications *Columbia* could fly the next four Shuttle missions. Referring to Fig. 1.42, it can be seen that the external tank of STS-1 (as was the case for STS-2) was painted white. From the third Shuttle flight onwards, the tank was no longer painted and got the familiar brown appearance. This saved 272 kg in mass.

A typical mission of the Space Shuttle was conducted as follows. After assembly of the Space Shuttle (and possibly its payload, although that could be done at the launch pad as well), the ET and the SRBs, the mated configuration was transferred by Crawler Transporter to the launch pad. At launch the three Space Shuttle main engines were ignited, followed by the solid rocket boosters. Due to the huge thrust levels, the Shuttle rose quickly, and after about seven seconds, it rolled over to a heads-down attitude to align the configuration's velocity vector with the desired

orbital plane.[26] Approximately 120 s into the descent, the SRBs were burnt out and separated, and fell back to Earth to splash in the ocean to be recovered and refurbished. After about eight an a half minute, the main engines were cut off, and the ET was jettisoned, which subsequently re-entered and disintegrated in the atmosphere. The Orbital Manoeuvring System (OMS) put the Space Shuttle in the correct orbit, and its orbital mission could begin.

After completion of the mission, the orbiter was rotated into a tail-first attitude such that the OMS could perform a deorbitation manoeuvre. The RCS then turned the nose back into flight direction, and gave the orbiter a large (positive) angle-of-attack orientation, such that the encountered maximum heat load would be minimal during entry. The Shuttle was targeted towards the so-called terminal area, a circular area around the runway of the landing site, and to keep its heading error within limits, several so-called bank reversals were executed.[27] Once close to the terminal area, a dedicated guidance system brought the orbiter in line with the runway, and the vehicle could land. The typical characteristics of this so-called gliding entry flight will be discussed in detail in Chap. 5.

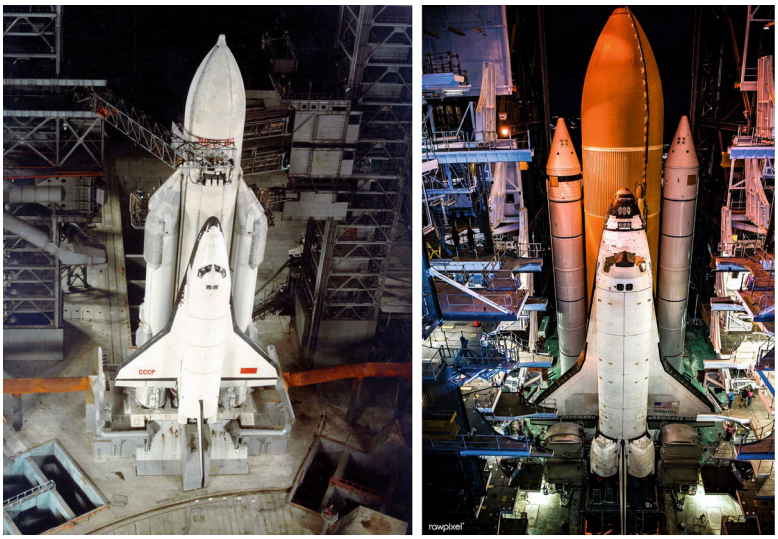
Four fully operational orbiters were initially built: *Columbia*, *Challenger*, *Discovery*, and *Atlantis*. Of these, two were lost in mission accidents, *Challenger* in 1986 during launch and *Columbia* in 2003 during re-entry. A fifth operational orbiter, *Endeavour*, was built in 1991 to replace *Challenger*, and mainly constructed from spare parts already available. The last Space Shuttle, *Atlantis*, was launched in July 2011. Between 1981 and 2011, 135 missions were flown, and the total in-orbit mission time was 1322 days, 19 h, 21 min and 23 s. Despite its many successes one cannot help but wonder, whether the huge investment in this programme was justified. The total cost of the Space Shuttle programme, including the orbiters and other hardware, buildings, facilities, training, salaries, etc., is calculated to be $209.1 billion (in 2010 dollars).[28] Divided over 135 flights this is about $1.5 billion per flight, which challenges the original goal of reducing launch cost.

The decision to develop the Buran, the "Russian space shuttle", came in 1976, at the expense of the Spiral spaceplane. Before the go-ahead, many different options were considered, but the main uncertainty was that Soviet military did not know what military requirements the Space Shuttle was based on. As the design of the Shuttle had been frozen two years before, it would be better to stick to the American design as much as possible, such that they would have the same capability in case an unknown American space threat would appear. The Russians denied the claim that they copied the design of the US, and said that the Buran design was driven by the

[26] The heads-down orientation facilitated to maintain a negative angle of attack that reduced the wing loading during peak dynamic pressure, typically between 30 and 60s after lift-off.

[27] Banking is the rotation about the velocity vector, which, at zero angle of attack is the same a rolling about the vehicle's longitudinal axis. At large angle of attack, however, this manoeuvre is dynamically different. A reversal is a manoeuvre to switch the bank angle from positive to an equivalent negative value, or *vice versa*. This manoeuvre and the associated guidance is discussed in more detail in Chap. 7.

[28] Logsdon, J., "Was the Space Shuttle a Mistake?", *MIT Technology Review*, July 6, 2011. (https://www.technologyreview.com/s/424586/was-the-space-shuttle-a-mistake/, accessed 23 July, 2019).



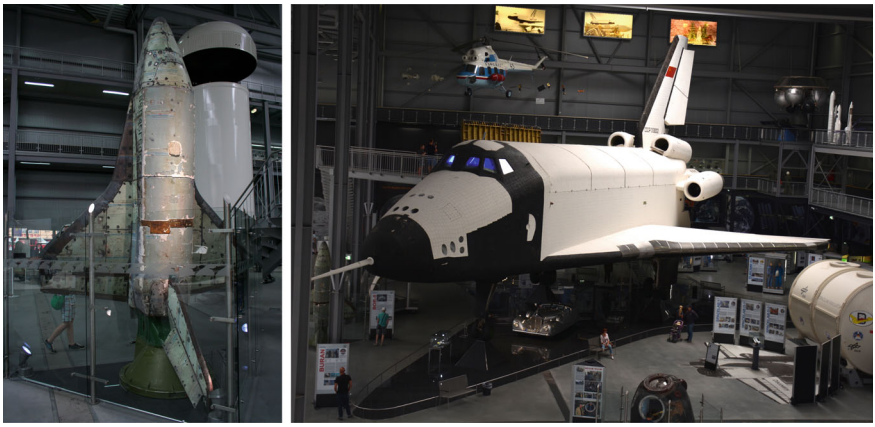
(a) Buran (b) Space Shuttle

Fig. 1.43 Comparison between the Buran (Image: Roskosmos) and the Space Shuttle. *Image* NASA

same aerodynamics, leading to the delta shape. However, the US spaceplane studies of the 1960s—as well as their own studies—showed many different configurations (Jenkins 2008). It was the US military requirements that drove the final design of the Space Shuttle, and thus also that of the Buran. But despite a similar configuration (geometry) and payload capacity, there were many differences as well (Fig. 1.43).

The Buran did not have any main engines (and had due to the weight saving a 30 tonnes payload capacity versus the 26 tonnes of the Space Shuttle), and relied for the launch on the specially developed Energiya rocket, the most powerful rocket ever built. The Energiya core stage had liquid-propellant engines, and so did its four boosters, contrary to the SRBs. An advantage of this was that the engines could be throttled down (and even switched off) in case of an emergency, such that Buran could make it safely back to the surface. Moreover, from the beginning the Buran was meant to be fully autonomous, which would greatly enhance its operability. But still, the Russians thankfully made use of the American decision to make the Space Shuttle an *unclassified* programme, so all technology development was open to the world. For the Russians it meant that they had a management problem: how to manage all the gathered information. However, this did not diminish the achievement of the Russian engineers. The design was frozen after just four years, and construction began well before the first Shuttle flight, so the flight experience of the Shuttle was not included in the final design. By 1984 the first full-scale orbiter was rolled out.

Of course, a lot of testing took place in parallel. We already mentioned BOR-4 that was meant to support the Spiral programme. In 1977, though, it was decided to re-use the design of BOR-4 and flight-test Buran's heat-shield materials on orbital missions



(a) BOR-5 (b) Buran Analog BST-02

Fig. 1.44 Test models for the Buran, exhibited in Technik Museum Speyer, Germany *Images* author

five times. The aerodynamic behaviour of Buran was wind-tunnel tested using 85 different scale models, varying from 1:3 to 1:550, in 25 different wind tunnels with a Mach range of 0.1 to 2.0 (Hendrickx and Vis 2007). In addition, BOR-5 was defined and built (Fig. 1.44a), a 1:8 scale model of Buran to test its aerodynamic characteristics on suborbital missions, also to be flown five times (four flights were successful, flown from mid 1985 till end of 1986). Besides 1:1 scale models for structural and environmental testing, as well as many other models, the so-called Buran Analog BST-02 was constructed in 1984, and to be used for approach and landing tests. Equipped with four jet engines, it could take off under its own power. It was used for nine taxi tests and twenty-five test flights between 1985 and 1988 before being retired. It is now an exhibit in the Technik Museum Speyer, Germany (Fig. 1.44b).

On 15 November, 1988, twelve years after the approval for the Energiya-Buran development was given, the first (and which appeared to be only one) orbital flight of the Buran took place. The unmanned vehicle orbited the Earth twice and came back, fully automated. Despite the strong crosswind, it landed within a few meters from the calculated location on the runway. To date, it remains the only automated flight of a winged spaceplane, capable of carrying astronauts to and from space.[29] The Buran was supposed to be flown again after five years, and in the meantime a second vehicle was 95–97% finished, and the construction of two flight models of the second series had started, when the Soviet Union collapsed and the Russian space programme faced serious budget problems. In 2002, the Buran, mounted on an

[29] The American X-37 orbital spaceplane has achieved a similar result, by returning unmanned from space four times. However, this spacecraft is much smaller (and weighs only about 5,000 kg), and cannot carry astronauts. There were plans, though, to enlarge the vehicle (leading to the X-37C), such that it could carry up to six astronauts in its pressurised cargo bay.

Energiya rocket, was stored away at the Baikonur Cosmodrome, when negotiations with NASA were underway for a possibly US-funded mission. On 12 May, a severe storm hit the spaceport and, as a result of poor maintenance, the hangar that housed the Buran collapsed. The collapse killed several workers and destroyed the Buran-Energiya combination. This was the very sad ending of a potentially very successful programme.

The *H-2 Orbiting Plane, Experimental* (HOPE-X) was part of Japan's effort to build and fly its own shuttle, HOPE, which would fly a crew of four and was to be launched on top of the H2-launcher. Development started in the 1980s, and it showed a steady progress over the years, with a number of dedicated and progressively more involved flight tests. This began on 3 February 1994, with the launch of the Orbital Re-entry Experiment Vehicle (OREX), Fig. 1.45a. OREX, with a launch mass of 865 kg, was launched into a ballistic, sub-orbital trajectory and tested various communication systems (amongst other GPS, as well as aspects of the black-out phase), measured the aerothermodynamic heating profile and tested heat-shield components. The capsule had a blunt-cone shape, and the heat shield was constructed of carbon/carbon material and ceramic tiles. After a single orbit it re-entered, and before it splashed into the Pacific Ocean it transmitted measurement data to ground control. The second vehicle tested was the *Hypersonic Flight Experiment* (HYFLEX), a lifting body with similar shape as HOPE (Fig. 1.45b). On February 11, 1996 it was launched into a sub-orbital trajectory, in this case with the solid-propellant rocket J-I. HYFLEX tested the carbon-carbon heat-shielding tiles to be used on HOPE and gathered data on controlled hypersonic lifting flight. Unfortunately the recovery system failed, and, even though the descent flight and splash down were successful, it sank and recovery was abandoned. The third test vehicle was the *Automatic Landing Flight EXperiment*, ALFLEX (Fig. 1.45c), a 6.1 m long vehicle with a wing span of 3.78 m and weighing 768 kg. It was built to develop the basic technology of automatic landing, and for that it was carried by helicopter to an altitude of 1500 m, then released to automatically glide down to the designated landing site, navigating by means of onboard computers. ALFLEX completed thirteen drop tests from July to August 1996, all of which were successful. The fourth and final test that was done before the HOPE-X programme was cancelled in 2003, was a set of balloon drop tests of the *High Speed Flight Demonstration*, HSFD (Fig. 1.45c). The two individual vehicles (25% scale models of HOPE) were dropped at 5 and 30 km, respectively, and verified autonomous navigation technology, estimated transonic aerodynamic characteristics, and tested the guidance and control design technology for autonomous heading alignment and landing.

1.3.4 Visiting Alien Worlds

In previous sections, the rise of the space nations was discussed, each with their individual development of launchers, satellite launches, and the progress in their manned space programme, which would require dedicated entry systems to return



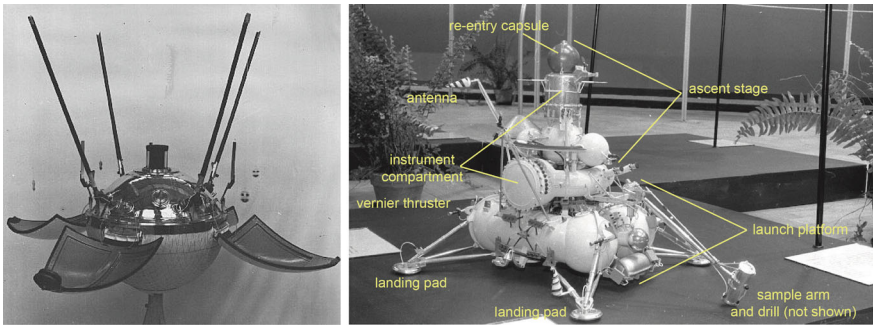
(a) OREX (b) HYFLEX

(c) ALFLEX (d) HSFD-1

Fig. 1.45 NASDA/JAXA flight-test vehicles in preparation for HOPE spaceplane development. *Images* JAXA

the cosmonauts and astronauts safely back to Earth. The events of launching animals into space, the first cosmonaut, the first spacewalk, and the momentous landing on the Moon of Apollo-11 often diverts the attention from the struggles of launching robotic explorers to other celestial bodies. In Sect. 1.3.1 the Russian Luna and American Pioneer were briefly mentioned. These spacecraft were part of a much larger effort of sending spacecraft to the Moon or into deep space. In the current section, these and other programmes will be elaborated upon, but with the focus on landing missions. Specific details on all exploration missions between 1958 and 2016 are provided by Siddiqi (2018).

As was the case with many aspects of spaceflight in the early years, the Russians took the lead in the robotic exploration of the Moon. However, it started with a number of failures: in a period of about 2.5 months (September-December, 1958), the three impactor missions Luna E-1 No.1/2/3 all failed due to launcher failures. Then Luna came along, a programme of a series of robotic spacecraft missions sent to the Moon between 1959 and 1976. Fifteen were successful, each designed as either an orbiter or lander. After yet another launch failure for Luna-1, the Luna-2 was the first probe to successfully impact on the Moon (13 September, 1959), and the Luna-3 probe returned the first pictures of the far side of the Moon with 70% photo coverage (launched on 4 October, 1959). After several other failures of different nature, on



(a) Luna-9 (Image: NASA) (b) Luna-24 (Image: WikiCommons/User Svobodat; labels added)

Fig. 1.46 Models of Luna-9 landing vehicle and Luna-24

3 February, 1966, Luna-9 made the first successful soft landing on the Moon (Fig. 1.46a), using a combination of retro rockets and landing bags.

The historical importance of Luna-9 was great. It was the first man-made object to softly land on another celestial body, even though it was the 12th attempt at a landing by the Soviet Union. Pictures of the lunar surface were transmitted back to Earth, but they were not released immediately by the Soviet authorities. However, the Jodrell Bank Observatory in England picked up the signal as well and noticed that the signal format used was identical to the internationally-agreed system used by newspapers. After contact with the *Daily Express*, newspaper employees rushed with a suitable receiver to the Observatory and the pictures from Luna-9 were decoded and published worldwide. It was curious why the Soviets had not encrypted their signal, and BBC News speculated that the spacecraft's designers deliberately fitted the probe with equipment conforming to the standard, to enable reception of the pictures by Jodrell Bank and in that way obtaining higher-resolution pictures, or that the Russian scientists wanted to avoid that the Soviet government would use the pictures for unrightful political gain.[30]

The next lander mission was Luna-13 (24 December, 1966), it landed successfully and transmitted TV pictures of the lunar landscape, and did soil measurements. After these successful landings the nature of the missions changed, and the Russians sent rovers and tried to get samples from the Moon, all with robotic missions. Due to a launcher failure, Luna 1969A (19 February, 1969) did not deliver the rover (Lunokhod 0) to the Moon, but Luna-17 did: it successfully landed and deployed the first robotic rover ever, Lunokhod-1. The rover exceeded its lifetime by far (17 November, 1970 until 4 October, 1971, instead of just three months) and drove over 10 km. The second rover, Lunokhod-2, was successfully delivered by Luna-21, and was active from 15 January until May, 1973, although the actual final date is not known. The rover

[30] http://news.bbc.co.uk/onthisday/hi/dates/stories/february/3/newsid_4063000/4063471.stm, accessed July 24, 2019.

travelled a distance of 37 km. In terms of sample return, the first six missions failed, either due to launcher failure or the spacecraft did not manage to leave Earth orbit. It was Luna-16 that became the first robotic sample return mission (20 September, 1970). This feat was repeated by Luna-20 (1972) and Luna-24 (1976), the latter took samples from the depths of 118, 143, and 184 cm of the lunar regolith. In total, the three missions returned $101+30+170.1=301.1$ g of sample mass in a container stored in a small entry capsule (Fig. 1.46b). The entry capsule was launched from the lunar surface with a smaller portion of the spacecraft, the ascent stage. In February 1978, Russian scientists published a paper in which they claimed that they had found traces of water in the lunar samples (Akhmanova et al. 1978). NASA's Moon Mineralogy Mapper (M^3) spectrometer onboard India's ISRO Chandrayaan-1 probe has confirmed the presence of water on the Moon in 2009 and 2018. Luna-24 marked the end of Russian activities to, on, and around the Moon, and to date remains the most recent lunar-sample-return mission.

The Americans started their lunar-lander programme with a series of failures. Ranger-3 though -5 (launched within a few months from each other in 1962) all missed their target. Ranger-6, launched on 2 February, 1964, was only a partial success: it had impacted, but due to a power failure it could not transmit any pictures. Fortunately, Ranger-7, -8, and -9 (1964–1965) faired better, and they returned pictures (and even a TV broadcast) until impact. The American Surveyor program was a series of seven unmanned missions, designed to reach the surface of the Moon by a soft landing. The purpose of the program was to develop enough technical knowledge to support the Apollo programme. Although Surveyor-1 soft-landed on the Moon about four months after the Russian Luna-9, it was a milestone for the Americans. Five of the spacecraft were successful (Table 1.3). During the Apollo programme, the Surveyor landing sites *Mare Tranquillitatus*, *Sinus Medii*, and *Oceanus Procellarum* were designated Apollo Landing Sites (ALS), i.e., ALS2, ALS3, and ALS5. Surveyor-2 suffered from a mid-course correction failure on the way to the Moon, and it lost control. Tumbling, it crashed on the Moon. Surveyor-4 also failed to land, after an otherwise flawless mission: two and a half minute before landing, telemetry was lost. Both failed Surveyor missions were targeted to land in the *Sinus Medii* region. The launch masses of the Surveyor spacecraft varied between 995 and 1039 kg, the landing mass between 294 and 306 kg. Surveyor-6 became the first monitored rocket launch from the lunar surface by using its liquid-fuelled vernier engines, and to land again some 3.5 m away. However, Surveyor-3 was the unintentional spacecraft "taking off" from the Moon; its engines failed to shut down, because its lunar-descent radar gave off the wrong signals. As a consequence, the spacecraft bounced twice (to 10 m and 3.4 m altitude, respectively) before performing a soft landing. Surveyor-3 became even more famous, when its landing site was chosen for the Apollo-12 mission to land (Fig. 1.47). Parts of the spacecraft were returned to Earth to study the effect of the harsh lunar environment on the degradation of materials.

The Russian Venera programme ran from 1961 to 1984 with the objective to gather information about the planet Venus. It was a highly successful programme, in which ten probes successfully landed on the surface of the planet, while thirteen probes successfully entered the Venusian atmosphere. It did not start out so well,

Table 1.3 Landing parameters for Surveyor missions (Siddiqi 2018)

	Time of landing	Landing site	Longitude (°)	Latitude (°)
Surveyor-1	02-06-1966, 06:17:36 UT	*Oceanus Procellarum*	43.34 W	2.47 S
Surveyor-3	20-04-1967, 00:04:53 UT	*Oceanus Procellarum*	23.34 W	2.94 S
Surveyor-5	11-09-1967, 00:46:44 UT	*Mare Tranquillitatus*	23.18 E	1.41 N
Surveyor-6	10-11-1967, 01:01:06 UT	*Sinus Medii*	1.40 W	0.49 N
Surveyor-7	10-01-1968, 01:05:36 UT	*Tycho Crater* north rim	11.41 W	41.01 S

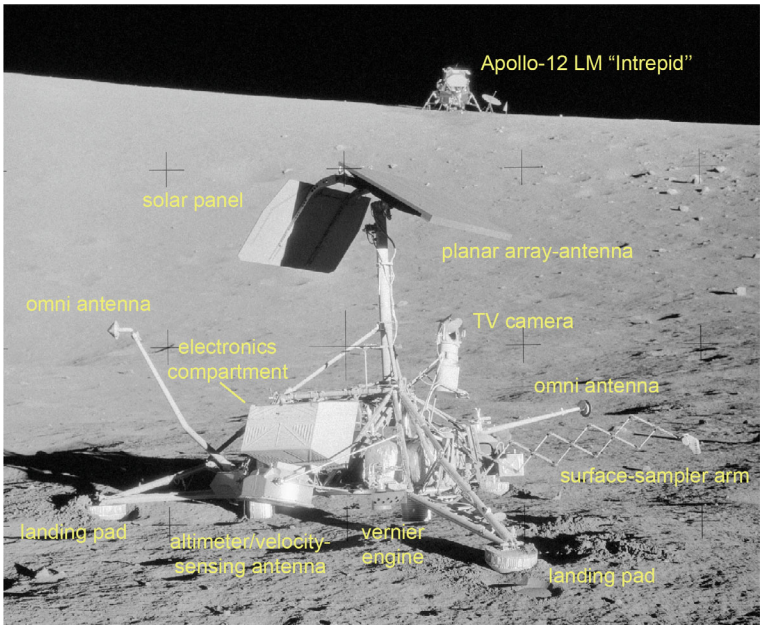
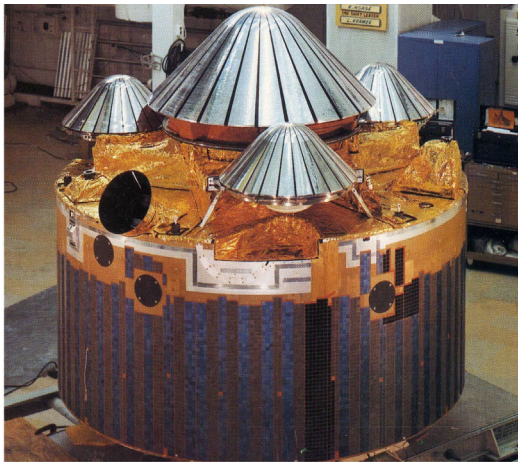


Fig. 1.47 The Surveyor-3, as photographed by the Apollo-12 astronauts. *Image* NASA, with added text labels

though. Venera-1 and -2, launched in February, 1961, failed to leave Earth orbit, and even though Venera-3 became the first human-made object to impact another planet's surface (1 March, 1966), due to data-probe malfunction, no data from the entry and descent was transmitted to Earth. Venera-4 (18 October, 1967) became the first successful planetary entry probe to send data from within the Venusian



(a) Venera-4 (Image: NASA/RSA)



(b) Pioneer Venus Multiprobe (Image: NASA)

Fig. 1.48 Russian and American Venus missions

atmosphere back to Earth.[31] The entry capsule of Venera-4 was a small, almost spherical capsule of about 1 m diameter (see Fig. 1.48a, with the capsule at the bottom of the spacecraft). The capsule had an improved heat shield compared to Venera-1 and -2, and was pressurised at 25 bar, as the atmospheric pressure was unknown at the time. During entry, the heat-shield temperature rose to 11,000 °C and the maximum deceleration clocked at 300 g (Ulivi 2007), but this was below the maximum centrifuge-test loads of 450 g. At 52 km altitude the probe's parachute deployed, and supported the descent for 26 km. Signal transmission stopped, when the conditions became to harsh for the capsule to sustain; Venera-4 was not designed for landing.

The first successful landing on Venus was achieved by Venera-7. With a similar capsule as Venera-4, it landed on Venusian surface on 15 December, 1970. The parachute opened at a height of 60 km, but seemed to have failed along its descent path, such that the landing velocity was higher than predicted (around 16.5 m/s). The thick atmosphere contributed to surviving the landing, although the probed appeared to have rolled onto its side. Therefore, the signal received was very weak, but could be analysed nonetheless. It provided valuable information about the surface that could not be seen through the thick layer of the atmosphere, and confirmed that humans cannot survive on the surface of Venus, nor that there is any liquid water. The subsequent Veneras (9 through 14), all consisted of an orbiter and a lander, and continued with a string of successes. In 1975, Venera-9 was the first to enter

[31] Venus has an extremely thick atmosphere that completely hides the surface from the visible eye. The conditions for landers are very harsh: the atmosphere is 50 times as dense compared to Earth's surface density, its (surface) temperatures rises to 460 °C, and it has a staggering surface pressure of over 90 bar! These conditions provide a challenge to any landing-mission design.



(a) Vega spacecraft (Image: WikiCommons, User Daderot)



(b) Descent module (Image: WikiCommons, G. Landis)

Fig. 1.49 The Vega Venus spacecraft and balloon, exhibited in Smithsonian's Udvar-Hazy Center (Dulles International Airport, USA)

an orbit around Venus, and its lander was the first to return TV pictures from the surface of another planet. It operated for 53 min on the surface, whereas Venera-10's lander transmitted for 65 min. The Venera-11 and -12 landers conducted the first soil analysis and sent the first colour pictures from the surface (December 1978). Despite some instrument failures, they continued to operate for 95 and 110 min, respectively. Two more lander missions followed with Venera-13 and -14 (1982), and both landers worked well beyond their designed lifetime. With Venera-15 and -16, two radar surface mapping missions that covered about 25% of the surface, concluded the Venera programme.

But it did not end the Soviet effort to land on Venus. Vega was a redesigned Russian-French cooperation mission to study Halley's comet, and consisting of two spacecraft. While flying by Venus, each spacecraft would release a descent module, carrying a balloon and lander mission (Fig. 1.49; note the large spherical descent module on top). The two spacecraft were launched six days apart in December 1984. Vega-1 flew past Venus on June 11, 1985, followed by Vega-2 four days later. The two aerobots, named Vega-1 and Vega-2, were designed to float at about 54 km from the surface, in the most active layer of the Venusian cloud system. At this altitude, pressure and temperature conditions of Venus are similar to those of Earth, though the planet's winds move at hurricane velocity. The instrument pack had enough battery power for sixty hours of operation, and measured temperature, pressure, vertical wind speed and aerosol density. Both balloon missions were operational for 46 h at an altitude of 54 km and returned a wealth of information. The instruments of the Vega-1 lander activated too early, and despite the soft landing hardly any data was transmitted. The Vega-2 lander transmitted data from the surface for 56 min.

At first sight, Venus did not seem to hold the same interest for the Americans as it did for the Russians. Still, in spite of never having landed on Venus, there are some interesting achievements. Mariner-2, launched 14 December, 1962, performed the first successful Venus flyby, at a minimum distance of 34,773 km. This was repeated by Mariner-5 that flew by Venus on 19 October, 1967 at a minimum distance 5,000 km, and Mariner-10 on 5 February, 1974. The latter flew by Venus at a minimum distance of 5,768 km while being on its way to Mercury, and therefore this flyby became known as the first gravity assist by an interplanetary spacecraft.

The Pioneer Venus Orbiter (also known as Pioneer Venus-1 or Pioneer-12), was the orbital component of the Pioneer-Venus project. Launched in May 1978, it entered Venus orbit on 4 December 1978 to conduct atmospheric and magnetic studies, and it kept transmitting data until contact was lost on 8 October 1992. Uncontrolled, it entered the atmosphere on 22 October 1992 and disintegrated soon thereafter. The atmospheric component of the Pioneer-Venus project was the Pioneer Venus Multiprobe (Pioneer Venus-2 or Pioneer-13), see also Fig. 1.48b. The main spacecraft bus was a transporter for four atmospheric probes, that were released before entering the atmosphere at different locations. The bus descended in the atmosphere as well, and burnt up, but continued to transmit data until 110 km altitude. All four probes were successful, and throughout descent kept transmitting data about atmospheric properties and composition, as well as radiation levels. Neither of the probes had photographic instruments and were not designed to survive landing, but would transmit data until that moment. Only the larger probe had a parachute, deployed at 67 km, but that did not prevent loss of signal after impact. However, one of the smaller probes survived impact and continued to transmit data from the surface for over an hour.

One of the later US missions that is worth mentioning is Magellan, launched from the Space Shuttle *Atlantis* payload bay in 1990. Being an orbiter, it was inserted into Venus orbit on 7 August 1990, from where it mapped 98% of the planet to 100-meter resolution with its synthetic aperture radar (SAR). As an end-of-life experiment, Magellan was using the upper atmosphere of Venus to circularise its orbit, a concept known as aerobraking. At the very end, it was deliberately deorbited into Venus's atmosphere. Due to differential drag acting on the solar panels the spacecraft started rotating ("windmilling"), which motion was countered by the RCS thrusters. Together with the data obtained from the aerobraking manoeuvre it provided valuable information on the upper-atmosphere force mechanism. Contact with Magellan was lost on 13 October, 1994.

Contrary to their successful Venus programme, the Soviet Union has been very unlucky with its Mars programme. The Soviets already started in 1960 to launch their first interplanetary probe to Mars, but third-stage launcher failure prevented the spacecraft reaching its parking orbit and it never got beyond 120 km altitude and re-entered again. Mars-1 was launched in 1963, but that spacecraft closest distance to Mars was 193,000 km when it failed. The US did a better job, and the Mariner-4 (launched 1965) performed the first successful fly-by of Mars, returning 21 pictures showing a cratered surface. It was a close call for who would be orbiting Mars first, but it was the US with Mariner-9 that completed the task (14-11-1971), just

13 days before Mars-2 succeeded. Mariner-9 stayed operational for almost a year, and mapped 100% of the surface. Both spacecraft are still in orbit, and Mariner-9 is expected to enter Martian atmosphere after 2022. Mars-2 also carried a lander, but that one failed and it crashed, although it did receive the dubious honour of becoming the first man-made object on Mars.

The Soviet Mars-3 lander was more successful, in the sense that it survived the (soft) touchdown (2 December 1971). Unfortunately no data was returned, although a signal was received for 20s. Also Mars-6 landed (1974), but contact was lost at landing. Since then, the track record with Mars has been extremely bad. Out of the 22 missions sent to Mars (flyby, orbiter, or lander), only two were completely successful (the orbiters of Mars-2 and Mars-3). Three were partially successful (Mars-3 lander, Mars-5 orbiter and Phobos-2,[32]) and the rest of the missions failed, either due to launcher failure, or spacecraft failure.

The Americans were way more successful. In 1975 the US launched the two Viking missions, consisting of an orbiter and a lander (Fig. 1.50). The Viking-1 orbiter entered Mars orbit on 19 June 1976 and began observing the Red Planet below in search for a good landing spot for its lander. The Viking-1 descent module was separated from its mothership on 20 July, and after deorbitation entered the thin Martian atmosphere. At 6 km altitude, still flying at supersonic speeds the 16-m parachute deployed, and just over 1.5 min later the lander softly touched down with the help of its retro rockets. It performed the first successful Mars landing that returned pictures and data. Viking-2 repeated Viking-1's success, but landed on the opposite side of Mars (3 September 1976). The Viking-1 lander operated for over six years until November 11, 1982. Viking-2's life was a bit shorter, it shut down on 12 April 1980, when its batteries failed. Both missions were extremely successful, not only in terms of science, but its technology heritage was the basis of many Mars missions to come.

The Mars-Pathfinder mission brought a change in Mars exploration. By airbag landing it delivered a small 11.5 kg rover, Sojourner, to the surface (1997). Designed to operate for seven sols (the standard Mars day, comparable in length to an Earth day), it was functional for 83 sols before contact was lost with both rover and Pathfinder base station. Following this success, in 2004 the Mars Exploration Rovers A and B (Spirit and Opportunity) were put softly on Mars's surface. Designed for a lifetime of about 92 Earth days, contact with Spirit was lost 22 March 2010, and Opportunity continued to operate till 10 June 2018 (!). After the soft landing of Phoenix in the north polar region (2008), on 26 November, 2011, an Atlas-V took off from Cape Canaveral with on-board Mars Science Laboratory (MSL) and the rover Curiosity. Reaching Mars more than eight months later, the descent stage was separated from its cruise stage and entered the Martian atmosphere. Commonly executed as a ballistic entry, MSL performed the first-ever guided flight in Martian atmosphere to reduce the landing ellipse to 7 by 20 km (for comparison: the Viking landing ellipse was about 100 by 300 km). It carried out the most accurate Martian landing of any known

[32] Designed as lander on Mars' moon Phobos it entered into Mars orbit and did some observations, but contact was lost before landing.



(a) Viking Descent Module (Image: NASA)



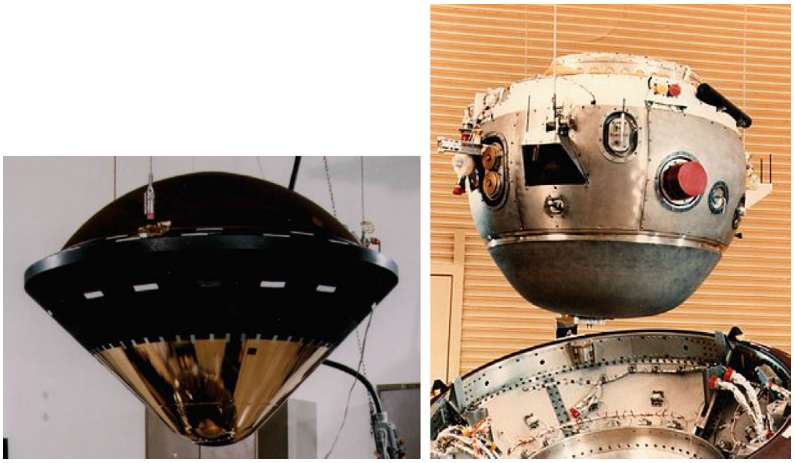
(b) Lander (Image: NASA/JPL-Caltech/University of Arizona)

Fig. 1.50 The Viking lander mission to Mars

spacecraft, the final landing place being off by a mere 2.4 km. As of July 2019, the rover is still operational and has been so for 2477 sols (2544 total days), thereby hugely exceeding its two-year lifetime.

Before bringing this section to an end, a few more words on the landing missions to other celestial bodies, each with their own challenges. The Galileo mission to Jupiter not only brought an orbiter to study the Jovian system, but also a descent module that was dropped into Jupiter's atmosphere (Fig. 1.51). Launched on October 18, 1989, from the payload bay of Space Shuttle *Atlantis*, it was piggybacking with the Galileo orbiter until five months before it would rendezvous with Jupiter. Targeted on a direct-entry course, the Galileo probe entered Jupiter's atmosphere on 7 December 1995, obviously not to land on this gas giant, but to conduct atmospheric measurements. After the initial phase of the entry was over, a stabilising drogue parachute was deployed and the heat shield released, such that the descent module could begin its mission. For almost an hour it transmitted its findings to the Galileo orbiter that transmitted the data back to Earth. After the probe's mission the orbiter stayed in the Jovian system for eight years. Apart from all the discoveries it made, it observed comet Shoemaker-Levy 9's collision with Jupiter in 1994. On September 21, 2003, Galileo's mission was terminated by sending it into Jupiter's atmosphere, thereby avoiding that by chance, Galileo could contaminate local moons with terrestrial bacteria.

Concerning the small bodies in the Solar System, i.e., asteroids and comets, several missions have been sent out to observe them, others have flown past them and also observed them, but only four missions have resulted in a landing on those bodies. It was in 2001, that NASA's NEAR-Shoemaker successfully landed on asteroid 433 Eros. Not designed as a lander, at the end of its mission the complete spacecraft made a slow, controlled descent to the surface on February 12, 2001, which ended with a soft touchdown. To everybody's surprise, the spacecraft had not been damaged and



(a) Entry configuration. (b) Descent module.

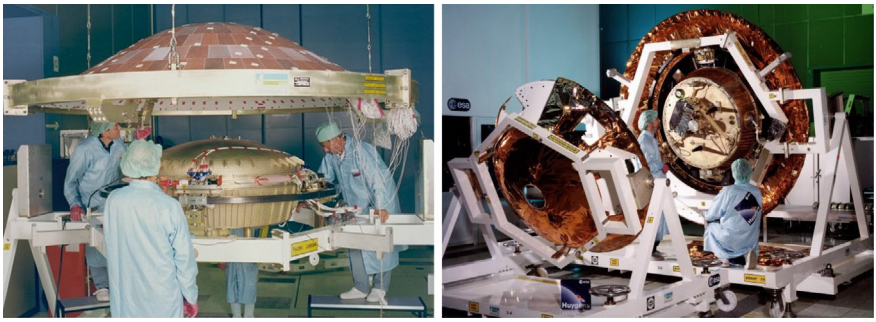
Fig. 1.51 The Galileo probe. *Images* NASA

kept on transmitting. Amongst others, data on Eros's composition were sent back to Earth. Sixteen days later contact was lost, and the mission was officially over.

Stardust was a mission to collect dust samples from the coma of comet Wild 2, as well as samples of cosmic dust, and return these to Earth for analysis. In 1999, Stardust was launched into a Sun-centred (heliocentric) trajectory and indeed collected the samples. While passing Earth on 15 January, 2006, it released an entry probe with the sample canister that subsequently made a ballistic entry and parachuted descent to Earth's surface. It became the first sample-return mission of its kind (not counting the Luna sample return missions), and also a typical example of an aerocapture mission where a spacecraft enters a planetary atmosphere directly from its interplanetary trajectory. Stardust's primary mission was completed after delivery of the entry capsule, but it continued to operate until 2011.

The Japanese mission Hayabusa (launched 2003), landed (briefly) on asteroid Itokawa and collected samples in the form of tiny grains of asteroidal material. These grains were stored in a small entry capsule that was released from the main spacecraft three hours prior to entry. After a ballistic entry, the capsule was successfully recovered on on 13 June, 2010. Its follow-on mission Hayabusa-2, launched in 2014 and rendezvoused with near-Earth asteroid 162173 Ryugu on 27 June, 2018, is still ongoing (until at least 2031), and has returned asteroid samples to Earth on 5 December, 2020.

The American/European Cassini-Huygens mission was launched on 5 October, 1997 by Titan-IV. Cassini, the orbiter, spent 13 years and 76 days at Saturn, and the discoveries it made and data it has transmitted will keep the scientists busy for many years to come. On 15 September, 2017, Cassini plunged into Saturn's atmosphere on a deliberate "crash" course, a worthy end of a worthy mission. Huygens (Fig. 1.52),



(a) Entry configuration. (b) Combining descent module and heat shield

Fig. 1.52 The Huygens probe. *Images* ESA

the European contribution to the mission, was a probe that entered the atmosphere of Saturn's largest moon, Titan, on January 14, 2005, and landed on Titan near the Adiri region after a parachuted descent of almost 1.5 hours. The probe was designed to gather data during the atmospheric descent, and was to survive a surface mission for three minutes only. However, until 90 minutes after landing the lander kept on transmitting data to Cassini.

In Chap. 9, several of the missions just briefly touched upon in this introductory chapter, will be discussed in a lot more detail. Discussions there will, of course, focus on the descent systems and the particular entry missions. Included are Surveyor and Apollo, Galileo, Huygens, and several of the Mars missions, i.e., Viking, Mars Exploration Rover, and Mars Science Laboratory.

1.4 Layout of the Book

While discussing the different entry missions in the previous section, several aspects of entry, descent and landing already came forward. Divided into capsules and winged vehicles, the particularities of these vehicles will lead to either ballistic, semi-ballistic, or gliding flight. And, if the conditions are right, possibly skipping flight, irrespective of vehicle shape. A second distinction that can be made, drawing from the examples of landing on Venus, Earth, the Moon, Mars, or the smaller bodies of our Solar System, like asteroids and comets, is the type of supporting equipment required to survive entry, descent and landing. Typically, for Earth landing of a capsule, a parachute system is required to lower the velocity to acceptable values. In the case of manned missions, one might lower the impact velocity even more by using a landing engine. Such a landing engine or retro rocket is a necessity for landing on Mars, since the atmosphere is so thin that parachutes alone will not sufficiently lower the landing velocity. Of course, when there is no atmosphere to begin with, powered descent is mostly the only means to guarantee a safe landing.

To begin the discussion on entry flight, it serves the purpose of the discussion knowing what the entry environment is. Chapter 2 introduces the planetary flight environment. The main (passive) force components acting on the re-entry system are typically of gravitational and aerodynamic origin. For landing in an atmosphere-less environment like the Moon, it is, of course, just gravity; and to counter that an active propulsion system is required. However, it is not just this part of the environment that one needs to know. Landing sites on the Moon in the 1960s were selected in the large maria, flat planes without too many craters and hills or mountains. Especially for relatively simple, unmanned systems a large safe so-called landing ellipse is required. However, with a pilot on board, or an advanced navigation system, evasive manoeuvres can be executed when encountering a hazardous situation. But to design (and simulate) such a system, detailed knowledge of the shape of the target body is required. Therefore, the three main elements that will be discussed are the shape of the target body, also called primary or central body, as well as its gravity field and atmosphere, when it is present. The focus will not be so much on the physical aspects of these models, but how one uses a simplified representation that can be used in a design and/or simulation environment.

In terms of external loads acting on the system, landing on an atmosphere-less body is an easy system to analyse. After all, there is no aerodynamic friction and the vehicle does not heat up because of that. Thus, the structural design can be simplified due to the absence of a heat shield, and neither does a parachute system have to be carried along and operated. On the other hand, without atmosphere fuel-efficient manoeuvres may be required, and a robust guidance system must ensure that the vehicle can land before it runs out of fuel. Most planetary missions to date were targeted to Venus and Mars, both planets with an atmosphere, but even the return flight of the Apollo missions ended with a flight through Earth's atmosphere. To understand the interaction of the entry vehicle with the surrounding atmosphere, Chap. 3 will cover the fundamental concepts of aerothermodynamics. Starting with a general flow-field description, different phenomena are described that play a part in dedicated velocity regimes. For each of these velocity regimes, e.g., the supersonic and hypersonic regime, specific problems that have to be dealt with by the vehicle designers will arise. Some engineering models to calculate the aerodynamic properties of entry vehicles, as well as some rules of thumb for assessing the thermal loads will be provided, such that the engineer can quickly do a first assessment of these aspects of re-entry flight.

The following Chap. 4 will lay the groundwork for studying all trajectory aspects of re-entry. In its general form, the motion of a(n entry) vehicle is governed by Newton's Laws of Motion, and in particular his second law, the famous $\mathbf{F} = m\mathbf{a}$: force equals mass times acceleration. In principle, this law is only valid in a so-called *inertial frame of reference*. If one wants to express the motion with respect to a rotating frame, e.g., one that is fixed to the Earth, special care should be taken, and at the very least transformations between the rotating and inertial frame are needed. Solving the motion from Newton's second law, requires an expression for the acceleration, which can be done in so-called Cartesian coordinates. Through the time dependency of acceleration-velocity-position, also velocity and position will be

expressed in Cartesian components, unless coordinate transformations are applied. Talking about position and velocity, they describe the *state* of the system, in other words, they are state variables. By choosing a different set of state variables, e.g., spherical coordinates for either position, velocity or both, will affect the formulation of the equations of motion, derived from $\mathbf{F} = m\mathbf{a}$.

All of the above mentioned aspects need to be clarified first, before one can start describing (and calculating!) the motion of re-entry vehicles. So, having covered reference frames, frame transformations, state variables, and the general formulation of the equations of motion in Chap. 4, the following Chap. 5, will deal with three types of entry flight: ballistic, gliding, and skipping flight. For a basic understanding, several analytical expressions of fundamental properties, such as maximum deceleration and heat load, flight range and flight time, will be derived. Since the general three-dimensional picture is too complicated (not to say impossible) to derive those expressions for, the discussion is restricted to two-dimensional space, which represents the so-called planar motion. The application due to this assumption does not limit us: the predictions with the analytical expressions are shown to be quite accurate, and moreover, they are often used as the basis for guidance-system design, as we will find out in Chap. 7.

Leading to the design of guidance, navigation and control (GNC) systems, the natural motion or eigenmotion should be well understood before moving on to the more complicated aspects of GNC-system design. Eigenmotion analysis, as discussed in Chap. 6, is by definition related to linear systems. Re-entry flight is a non-linear process and has to be made suited for analysis by linearising the equations of motion, and to formulate the resulting equations in matrix format, the so-called state-space form. As this form is also used in linear state-feedback control-system design, this linearisation serves therefore two purposes. From the system part of the state-space model one can extract the eigenvalues that gives information about the uncontrolled natural motion of the system. This will be done for both a winged vehicle and a capsule, to understand the difference in behaviour. In combination with the system part, if also the controls part of the state-space model is used, a linear state-feedback controller can be designed. This will be illustrated in the next Chap. 7.

Control of spacecraft is typically automated, unless for manned missions when there is a pilot on-board. But even in that case, parts of the flight will be automatically controlled, even if it were only to increase the safety. A complete GNC system includes, as the name implies, three major components. From a physical point of view, the GNC system requires information from the outside world, provided by a set of sensors, such as accelerometers and gyroscopes. Due to instrument and measurement errors, these sensor measurements are usually not directly used, so estimation algorithms are in place to provide the estimated state. The combination of sensors and estimators is called the navigation system in this book, and should lead to answer the question: *where am I?* Comparing the estimated state with what it should be, e.g., a to-be-specified reference state, the guidance logic will calculate corrective actions to remove any difference between the two states, and answers the question: *where should I go?* Finally, the vehicle needs some hardware to make this happen, the actuators, for instance, control surfaces or reaction-control thrusters. These actuators

cannot be activated directly, and the translation between the guidance and actuator commands is taken care of by the control system. This system basically answers the question: *how should I get there?*

For winged vehicles, the guidance during the trajectory is usually split into two parts, the hypersonic-descent guidance and the terminal-area guidance. Chapter 8 focusses on just the terminal-area energy-management system, which is a specific guidance system balancing the potential and kinetic energy to guarantee that the vehicle reaches the runway with the right heading and velocity. The terminal area is typically a circular area around the runway of about 60–80 km radius. If the hypersonic guidance can intersect the trajectory with the terminal-area interface at the right altitude-velocity combination, then the terminal-area guidance should be able to guarantee a safe landing. Since the re-entry vehicle is unpowered it typically boils down to a one-shot approach and landing procedure.

So far, the focus has been on the flight-mechanics aspects of controlled re-entry, without examining the sub-systems that enable it. Although this book is not about re-entry vehicle design, it is important to highlight two important sub-systems, i.e., the so-called entry, descent, and landing system, to be discussed in Chap. 9, and the thermal-protection system, covered in Chap. 10. In case a vehicle does not provide a large enough aerodynamic force to reduce the velocity to acceptable values, such that impact can be survived, additional means are required. An atmosphere-less descent will require a powered propulsion system, but in an atmospheric descent passive means can be deployed (in addition). The combination of the two systems is also an option, when the atmosphere is too thin to provide sufficient deceleration, as is the case on Mars. However, the entry and descent system can be more extensive than just adding these two sub-systems. For instance, probes sent to Venus, Jupiter and Titan had small so-called spin vanes attached to the outer wall, to guarantee a certain minimum spin rate. Finally, as landing system to absorb some of the terminal energy, one could think of using (ordinary) landing legs, crushable structures, or landing bags.

Looking at the orbital energy of a vehicle that has to go (or return) to the surface of the planetary body below, it does not take a lot of math that all that energy has to go somewhere, if one wants to end up with zero energy at landing. During atmospheric entry, the only mechanism to dissipate the energy is through friction, which means that this will be converted to heat. One can imagine that the amount of heat is substantial, so the vehicle has to be protected with a thermal protection system (TPS). The older capsules all used ablative heat shields, which absorb the heat by burning away the shield material (the process known as ablation). Unfortunately, this means that the heat shield can be used only once. More advanced designs can actively cool the vehicle wall using a cooling liquid, and these active TPSs will enable reusability of the spacecraft.

The Magellan spacecraft, at the end of its mission of mapping Venus, lowered its orbital apoapsis, i.e., the farthest orbital point from the planet, by passing through the atmosphere at its lowest point, the periapsis. The resulting drag resulted in dissipation of the orbital energy, with the apoapsis lowering as a result. This process is called aerobraking, and is one of more so-called aero-assisted manoeuvres. The

Stardust capsule, carrying the cosmic-dust samples, entered the Earth atmosphere directly from its interplanetary trajectory, which is an example of aerocapture. In Sect. 1.3.4, Mariner-10 was programmed to fly by Venus on its way to Mercury, and thus performing the first gravity assist. However, if the spacecraft had flown lower it could have used the atmospheric flight to its advantage, and made the flyby more efficient. Such a manoeuvre, the aerogravity assist, is, of course, not without risk: the slightest offset may result in loss of the vehicle. Each of the above-mentioned manoeuvres relies on an atmospheric-flight segment, similar to the hypersonic flight of an entry vehicle. Hence, the last Chap. 11 will discuss these manoeuvres in more detail.

Problems

Judge whether the following statements are TRUE or FALSE (Problems 1.1–1.10).

1.1. John Glenn was the first American in space.
1.2. Sergei Korolev was the inventor of the basic rocket equation.
1.3. Explorer-1 was the first successful U.S. satellite.
1.4. Spirit, Opportunity, and Curiosity are NASA rovers that roam(ed) the surface of Mars.
1.5. The first rocket to reach hypersonic speed at entry into the atmosphere was an unmodified V-2, captured after World War II.
1.6. Johannes Kepler defined the physical principles of force, motion, and gravitational attraction.
1.7. Mid-air recovery is a concept developed in the late 1950s and used in the 1960s.
1.8. Even though Apollo-10 was the dress rehearsal for the actual Moon landing of Apollo-11, it could have landed on the Moon.
1.9. Wernher von Braun was the chief designer of the Space Shuttle Transportation System.
1.10. Robert Goddard was an American rocket pioneer who dramatically improved the efficiency of the solid-fuelled rocket engines.

References

Adams RB, Richardson GA (2010) Using the two-burn escape maneuver for fast transfers in the solar system and beyond, AIAA-2010-6595. In: 46th AIAA/ASME/SAE/ASEE joint propulsion conference and exhibit. Nashville, TN

Akhmanova M, Dement'ev B, Markov M (1978) Possible water in Luna 24 regolith from the Sea of Crises. Geochem Int 15:Nr. 166

Angelo Jr. JA (2006) Rockets. Facts on File, Inc

Evans B (2009) Escaping the bonds of Earth: the fifties and the sixties. Springer-Praxis Books in Space Exploration

Christianson GE (1996) Isaac Newton and the scientific revolution. Oxford University Press
Gainor C (2008) To a distant day. The rocket pioneers. University of Nebraska Press
Gerovitch S (2008) Stalin's Rocket Designers' leap into space: the technical intelligentsia faces the thaw. Osiris, Second Ser 23:189–209
Gordon Y, Gunston B (2001) Soviet X-planes. Midland Publishing
Harding J (2006) Flying's strangest moments: extraordinary but true stories from over 1100 years of aviation history. Anova Books, Strangest Series
Harvey B, Zakutnyaya O (2011) Russian space probes. Scientific Discoveries and Future Missions. Springer-Praxis Books in Space Exploration
Hendrickx B, Vis B (2007) Energiya-Buran. The Soviet Space Shuttle. Springer-Praxis Books in Space Exploration
Jenkins DJ (2008) Space shuttle. The History of the national space transportation system. The first 100 missions, 3rd edn. Fifth Printing, Midland Publishing
Launius, RD, Jenkins DR (2011) Coming home—re-entry and recovery from space. NASA-SP-2011-593
Laurenza D (2004) Leonardo on flight. Giunti Editore, Florence-Milan
Miller J (2001) The X-planes, X-1 to X-45. Midland Counties Publications
Miller R (1993) The dream machines: an illustrated history of the spaceship in art, science, and literature. Krieger Publishing Company
Oberth H, Sorensen TC, Kehr J, Ciancone ML, Englert PAJ (2014) The rocket into planetary space. Walter De Gruyter Inc
Oberth H (1970) Ways to spaceflight, NASA TT F-622, 1970. Translation of Wege zur Raumschiffahrt, R. Oldenbourg Verlag , Munich-Berlin, 1929
Rose B (2008) Secret projects: military space technology. Midland Publishing
Rumerman JA, Gamble C, Okolski G (2007) U.S. human spaceflight. A record of achievement 1961–2006', NASA-SP-2007-4541. Monographs Aerosp History (41)
Siddiqi AA (2000) Challenge to Apollo: the Soviet Union and the space race, 1945–1974. NASA SP-2000-4408
Siddiqi AA (2018) Beyond Earth. A chronicle of deep space exploration, 1958–2016. NASA SP-2018-4041
Spangenburg R, Moser DK (2008) Wernher von Braun. Rocket visionary. Revised Edition, Chelsea House Publishers
Ulivi P, Harland DM (2007) Robotic exploration of the solar system: the golden age 1957–1982. Springer-Praxis Books in Space Exploration
White Jr, L (1961) Eilmer of Malmesbury, an eleventh century aviator: a case study of technological innovation, its context and tradition. Technol Culture 2(nr. 2):97–111
Westfall RS (1993) The life of Isaac Newton. Cambridge University Press
Woods WD (2008) How Apollo flew to the Moon. Springer-Praxis Books in Space Exploration
Zaloga SJ (2003) V-2 Ballistic missile. 1942–1952. Osprey Publishing

Chapter 2
Planetary Entry Environment



Our Solar System consists of 9 planets, hundreds of moons and thousands of smaller bodies, such as asteroids and comets (Figs. 2.1 and 2.2). If we look at past and present missions it seems that at first sight most of these bodies are of no interest for studying the subject of re-entry. Russia and the United States sent their probes to Venus and Mars, the Americans landed on the Moon, and a Japanese spacecraft on an asteroid in November 2005. And let us not forget about the joint NASA/ESA mission Cassini/Huygens: the European Huygens probe made a perfect entry, descent and landing in the atmosphere and on the surface of Titan, the largest moon of Saturn.

However, digging a little deeper we find that the American Galileo probe made an entry in the atmosphere of Jupiter in 2003, ESA has sent Rosetta towards the comet 67P/Churyumov–Gerasimenko and reached there in 2014, after which it send the Philae lander down to its surface. Moreover, there are several plans to visit some of the smaller bodies in our Solar System, i.e., the asteroids, and, finally, the concept of aerogravity assist (see Chap. 11) makes the gas giants Jupiter, Saturn, Uranus and Neptune potential candidates for an atmospheric flyby.

The choice of the central body to which the re-entry vehicle is returning from space, has a significant influence on the trajectory. The Moon, for instance, does not have an atmosphere, so therefore there will be no aerodynamic forces exerted on the re-entry vehicle. Also the mass of the central body is of importance, because it is directly related to the gravitational forces: the gravitational pull of Jupiter will be tremendous compared to the one by Mars, for instance. Asteroids, on the other hand, may have a very low mass compared to planets, but since their shape can be extremely irregular, landing on such a body may pose a serious challenge for the navigation and hazard-avoidance system.

In the following sections several characteristics of the planetary environment that is relevant for all aspects of entry, descent and landing are covered. In Sect. 2.1 the planetary shape of potential target bodies is discussed. We will not restrict to the few planets and moons of the past, but also include asteroids in the discussion. Comets, on the other hand, will not be part of any discussion in this chapter, because the dedicated

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_2



Fig. 2.1 Bodies of the Solar System: the Sun, planets, moons, asteroids and comets. Image: NASA

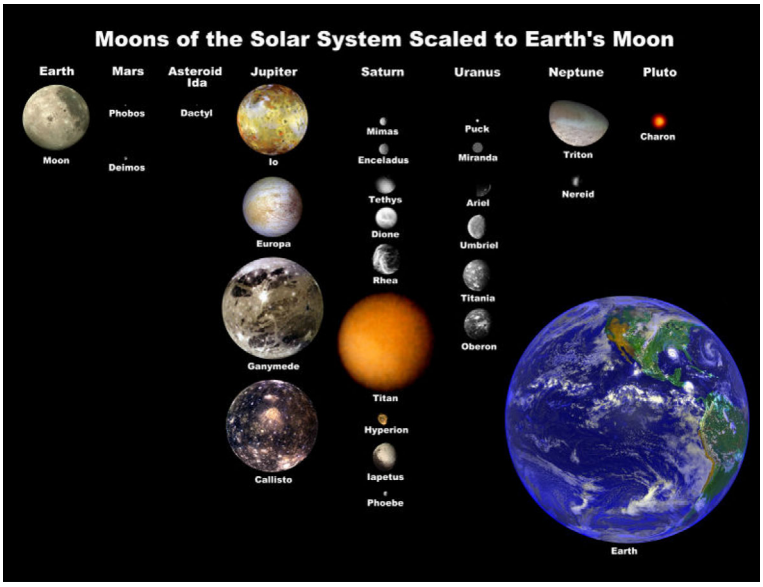


Fig. 2.2 Moons of the Solar System, in size compared to Earth. Image: NASA

outgassing models when comets are getting closer to the Sun are considered to be outside the scope of this text. Section 2.2 gives an overview of the different ways to model the gravity field of both regular and irregular bodies. The so-called spherical harmonics model can in principle be used for all bodies of interest in the field of entry and landing. In the last section of this chapter, Sect. 2.3, the atmospheres of planets and moons will be discussed. Of course, the focus will be on the Earth, Venus and Mars as potential target planets for entry, but also Titan will pass review because of the breakthrough Huygens mission. The gas giants, even though relevant for aerogravity-assist missions, will not be covered in detail, although a simple model for the upper layers of the atmosphere will be provided.

2.1 Planetary Shape

A common confusion is that the shape of a celestial body is the only determinant of the gravitational field. However, in the end it is the combination of shape *and* mass distribution that determines the gravitational irregularities. For gaseous bodies, such as Jupiter and Saturn, the rotational rate of the planet has a significant impact on the shape. For example, Saturn is a gas giant and due to a combination of its lower density, rapid rotation and fluid state, Saturn is an oblate spheroid; that is, it is flattened at the poles and bulges at the equator. Its equatorial and polar radii differ by almost 10%. Therefore, while discussing the shape of celestial bodies we will also include the rotational rate of these bodies. An additional reason is that while converting inertial position and velocity to relative coordinates, i.e., with respect to the rotating central body, we need the rotational rate to do so.

This section is divided into three parts. In Sect. 2.1.1 regular bodies are discussed. These bodies can be modelled as spheroids or ellipsoids. However, when a more accurate surface description is required, e.g., for precision-landing studies or the development of hazard-avoidance algorithms, detailed surface models are needed. An overview of these models is given in Sect. 2.1.2. Finally, Sect. 2.1.3 covers irregular bodies. These bodies require a specific modelling that can also be used for the calculation of gravity fields.

2.1.1 Regular Bodies

Celestial bodies, which are considered to be regular and of importance for entry studies (or atmospheric flyby) are the Earth-like planets Earth, Venus, Mars, Titan and the Moon, and the gas planets Jupiter, Saturn, Neptune and Uranus. Parameters, which characterise these bodies deal with the shape, size and rotational rate. Even though the shape of, for instance, the Earth is determined by many factors, e.g., lunisolar gravitational tides, winds, Earthquakes, post-glacial rebound, plate motion, melting of ice and atmospheric pressure, an approximation, which is adequate for

the purpose of re-entry, is to consider the shape of the regular bodies as an ellipsoid of revolution. The parameter defining this ellipsoid is called the ellipticity f, or alternatively oblateness or flattening, and is defined as

$$f = \frac{R_e - R_p}{R_e} = 1 - \frac{R_p}{R_e} \tag{2.1}$$

with

R_p = mean radius at the pole (m)
R_e = mean radius at the equator (m).

For the Earth, R_p is about 21 km smaller than R_e. We can use the ellipticity to derive an expression for the radius at an arbitrary point along the surface, R_s. Starting with the general expression of an ellipse and using polar coordinates, the following general expression can be derived:

$$R_s = R_e \left[1 - \frac{f}{2}(1 - \cos 2\delta^*) + \frac{5}{16} f^2 (1 - \cos 4\delta^*) - \ldots \right] \tag{2.2}$$

Here, δ^* is the geographic latitude which is for the ellipsoid not the same as the geocentric latitude δ (see Fig. 2.3). However, since the difference is very small it is justified to approximate δ^* with δ. In addition, as R_e is not known with infinite accuracy, for first-order analyses we may approximate Eq. (2.2) by

$$R_s \approx R_e \left[1 - \frac{f}{2}(1 - \cos 2\delta) \right] = R_e (1 - f \sin^2 \delta) \tag{2.3}$$

We can use this expression to define the height above the planetary surface. Let the position of the re-entry vehicle be specified by the distance, R, to the origin of the central body, the longitude, τ, and the latitude, δ. The height, h, is therefore given by

$$h = R - R_s = R - R_e (1 - f \sin^2 \delta) \tag{2.4}$$

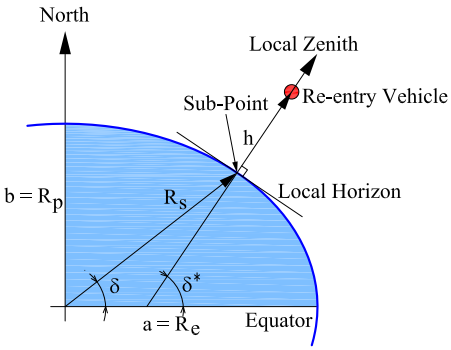


Fig. 2.3 The re-entry vehicle above the reference ellipsoid; δ is the geocentric latitude for a spherical central body, δ^* the geographic latitude for the ellipsoid

Table 2.1 Some characteristics of celestial bodies (Yoder 1995)

Celestial body	R_e (km)	R_p (km)	e (–)	ω_{cb} (rad/s)
Earth	6378.136	6356.751	$3.3528{\cdot}10^{-3}$	$7.29220{\cdot}10^{-5}$
Moon	1738.1	1736.0	$1.20{\cdot}10^{-3}$	$2.662{\cdot}10^{-6}$
Mercury	2440.0	–	–	$1.240{\cdot}10^{-6}$
Venus	6051.8	–	–	$-2.992{\cdot}10^{-7}$
Mars	3397.0	3375.0	$6.4763{\cdot}10^{-3}$	$7.08822{\cdot}10^{-5}$
Jupiter	71492.0	66854.0	0.06487	$1.75853{\cdot}10^{-4}$
Saturn	60268.0	54364.0	0.09796	$1.65543{\cdot}10^{-4}$
Uranus	24766.0	24342.0	0.0171	$-1.01237{\cdot}10^{-4}$
Neptune	25559.0	24973.0	0.02293	$1.08338{\cdot}10^{-4}$
Titan	2575.0	–	–	$4.56069{\cdot}10^{-6}$

For the rotation of the central body, we assume that it is rotating with a constant angular velocity $\boldsymbol{\omega}_{cb}$ directed along the Z-axis of the rotating planetocentric frame,[1] so

$$\boldsymbol{\omega}_{cb} = (0, 0, \omega_{cb})^{\mathrm{T}} \tag{2.5}$$

The default values for the several parameters introduced above for several planets and moons are listed in Table 2.1.

In some situations it is possible to simplify the shape of the central body, for instance, when we want to facillitate the derivation of the equations of motion to obtain analytical expressions, or when approach and landing studies of Shuttle-like vehicles are performed. The first simplification is to consider the central body as a (non-)rotating spheroid, see also Fig. 2.4. The surface radius, R_s, is constant and equal to the equatorial radius, R_e. The height above the surface simply becomes $h = R - R_e$. A second simplification, commonly used in aircraft studies, but also a justified simplification when studying the landing of the Space Shuttle, is to assume a flat-Earth approximation. As a consequence, this also means that the Earth (or in general, the central body) is non-rotating, as the rotation has lost its meaning. The equatorial radius has become infinite, and latitude and longitude have lost their meaning as well. The coordinate system used is a Cartesian one, where the XY-plane may define the surface and the Z-axis the altitude direction. This simplification is not commonly used in entry studies, contrary to the spheroid approximation.

[1] The reader is referred to Chap. 4 for a definition of the rotating planetocentric frame, as well as other reference frames.

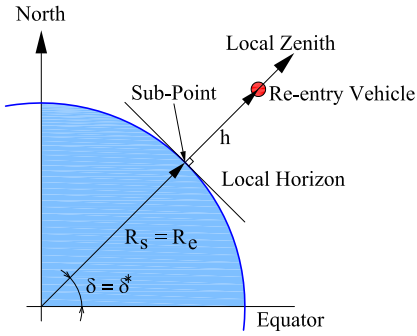


Fig. 2.4 The re-entry vehicle above the reference spheroid; the geodetic and geocentric latitude are identical

2.1.2 Surface Models

In some situations more knowledge about the surface (i.e., accuracy) is required, and should therefore also be modelled. Small, irregular bodies like asteroids will require a detailed model for proximity operations (and serves as input for gravity-field models, which will be discussed later in Sect. 2.1.3). Large, regular bodies may also require a detailed (local) model for precision landing and hazard avoidance, e.g., in case of robotic missions to the Moon and Mars with autonomous navigation systems.

There are several ways to describe a planetary surface. A contoured topographic map is perhaps the most familiar way of representing terrain, although essentially all information is projected on a two-dimensional plane (Li et al. 2004). A digital elevation model (DEM) is a digital model or 3-D representation of a terrain's surface. It is commonly used for a planet (including Earth), moon, or asteroid, and is created from terrain elevation data. Since there is no common usage of the terms digital elevation model, digital terrain model (DTM) and digital surface model (DSM) in scientific literature, we will keep on using the term DEM when discussing surface models. A DEM can be represented as a raster (a grid of squares, also known as a height map when representing elevation). DEMs are commonly built using remote sensing techniques, but in case of Earth they may also be built from land surveying. The quality of a DEM is a measure of how accurate the elevation is at each pixel (absolute accuracy) and how accurately the morphology is presented (relative accuracy). (Space-based) Methods for obtaining elevation data used to create DEMs include lidar (light detection and ranging), SAR (synthetic aperture radar), Doppler radar and real time kinematic GPS (global positioning system). It may be clear that the more satellites have been orbiting a celestial body the more detailed these models become. Amongst others, detailed models are available for Earth, Moon, Mars, and Venus, and are discussed next.

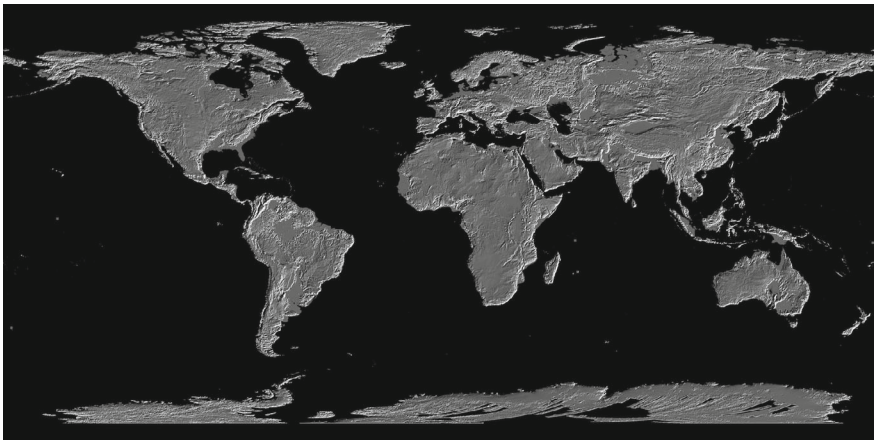


Fig. 2.5 Global Earth-map derived from ASTER GDEM globe map data sets. Image: NASA/JPL

2.1.2.1 Earth

The Earth has been observed in many ways and for hundreds of years. Very detailed DEMs are available that are used for many purposes, including modelling water flow or mass movement (for example avalanches and landslides), creation of relief maps, 3D flight planning, creation of physical models (including raised-relief maps), reduction (terrain correction) of gravity measurements (gravimetry, physical geodesy), GPS and flight simulation (Li et al. 2004).

A high-quality DEM from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) instrument of the Terra satellite is freely available for 99% of the globe, and represents elevation at a 30-m resolution. A similarly high resolution was previously only available for the United States territory with the Shuttle Radar Topography Mission (SRTM) data, while most of the rest of the planet was only covered in a 3 arc-second resolution (around 90 m). In Fig. 2.5 an image derived from ASTER GDEM globe map data sets is shown. The complete data sets can be freely downloaded.[2] A novel global DEM has been generated by the TanDEM-X satellite mission, which started in July 2010. TanDEM-X is a German Earth observation satellite using SAR. The so-called WorldDEM, available since 2014, features a vertical accuracy of 2 m (relative) and 4 m (absolute), within a horizontal raster of approximately $12 \times 12\,\mathrm{m}^2$.

A less accurate, but more easily accessible model has been developed by Hirt et al. (2012). The developed model is a spherical-harmonics model,[3] and is based on SRTM V4.1, SRTM30_PLUS and ETOPO data. The normalised coefficients for a

[2] http://asterweb.jpl.nasa.gov/gdem-wist.asp. Last accessed: October 31, 2011.

[3] Spherical harmonics models are discussed in more detail in Sect. 2.2.1, which deals with gravity models for spheroidal bodies.

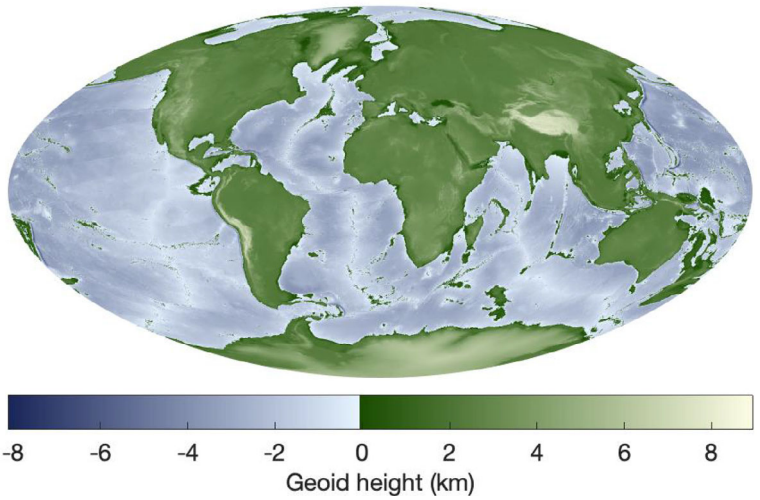


Fig. 2.6 Global topography and bathymetry of the Earth (land plus ice masses) using a spherical-harmonics expansion (order/degree = 2160), referenced to mean sea level. Data source provided in Footnote 4 and discussed by Hirt et al. (2012)

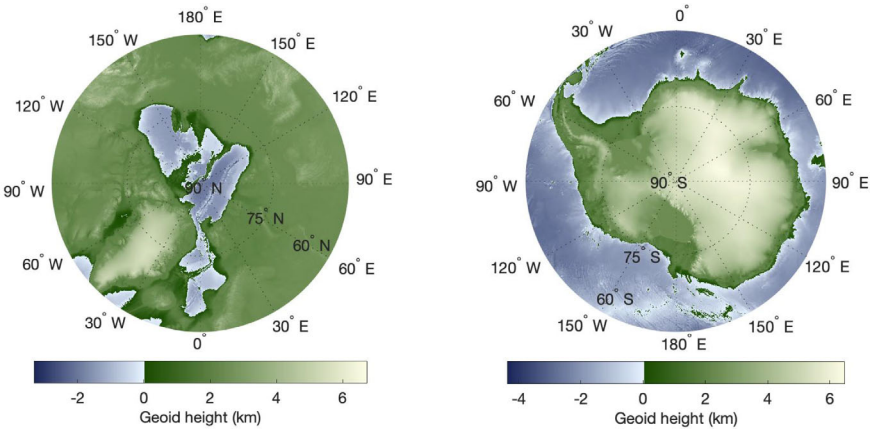
2160° model of either the complete Earth or just the solid portion is freely available.[4] Details on how the models have been derived can be found in the quoted reference. In Figs. 2.6 and 2.7 this topography is shown, including the land and ice mass.

2.1.2.2 Moon

One of the best topographic map of the Moon that is currently available is derived from measurements obtained with the Japanese SELENE spacecraft, and was the first model that covers the poles. SELENE began its nominal mapping phase on December 30, 2007 with its laser altimeter LALT. The total radial topographic error is ±4.1 m, whereas the total horizontal topographic error amounts to ±77 m. The topographic data have been made available to everyone by the Japan Aerospace Exploration Agency, JAXA (Araki et al. 2009). The Lunar Reconnaissance Orbiter (LRO), launched by NASA on June 18, 2009, is currently still orbiting the Moon in an eccentric polar mapping orbit. The spacecraft has made a 3-D map of the Moon's surface at 100-m resolution and 98.2% coverage, whereas the Apollo landing sites have even been observed at 0.5-m resolution. Also the LRO data are freely available for download,[5] and are provided in different formats, amongst others as spherical-harmonics coefficients.

[4] http://ddfe.curtin.edu.au/gravitymodels/Earth2012/topo_shape_to2160/. Date of access April 10, 2020.

[5] https://pds-geosciences.wustl.edu/missions/lro/default.htm. Last accessed April 10, 2020.



(a) North pole (65°-90° N) (b) South pole (65°-90° S)

Fig. 2.7 Polar topographic maps of the Earth (land plus ice masses) using a spherical-harmonics expansion (order/degree = 2160). Data source provided in Footnote 4 and discussed by Hirt et al. (2012)

Such an alternative surface model is described by, amongst others, Wieczorek (2007). The spherical-harmonics model for the lunar shape is provided as radii with respect to the centre of mass, and can be converted to height above (or under) the lunar geoid by subtracting the reference radius of $R_e = 1{,}737.1513$ km. Figures 2.8 and 2.9 show the topographic map of the entire Moon and the poles, respectively. The maximum degree and order is 2050, suitable for a 0.1° map with a spatial resolution of 3 km.

2.1.2.3 Mars

The Mars Orbiter Laser Altimeter (MOLA) is an instrument on board of Mars Global Surveyor, which measured the topography and surface roughness of Mars surface in the period between 1997 and 2001. The altimeter measurements provide global topological datasets with vertical accuracy of approximately 1 m. The topographic resolution grid size is 1/64° in latitude and 1/32° in longitude, which corresponds to $1 \times 2\,\text{km}^2$ at the equator (Smith et al. 2000). The raw data from MOLA Experiment Data Records (EDRs) are transferred into Precision Experiment Data Records (PEDRs), which contain data obtained from measured and derived MOLA profiles as well as the areoid values (gravitational equipotential surface). The profile date are put into Experiment Gridded Data Products (EGDP) with various resolutions depending on density of measurement data. The available resolutions are 4, 16, 32, 64 and 128

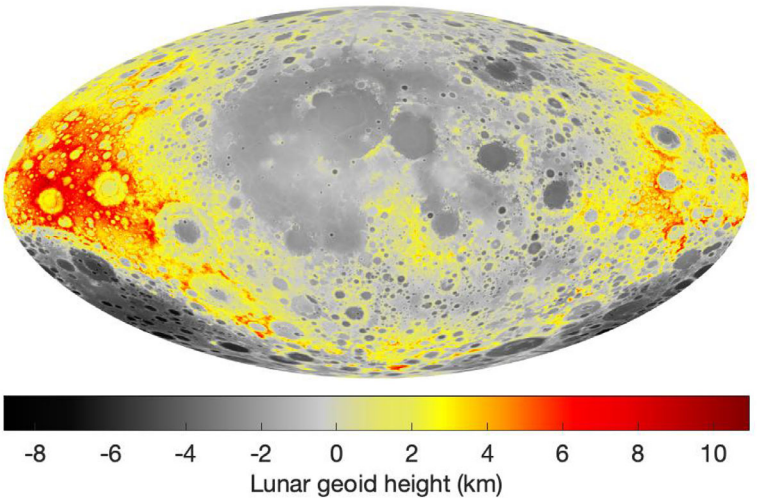
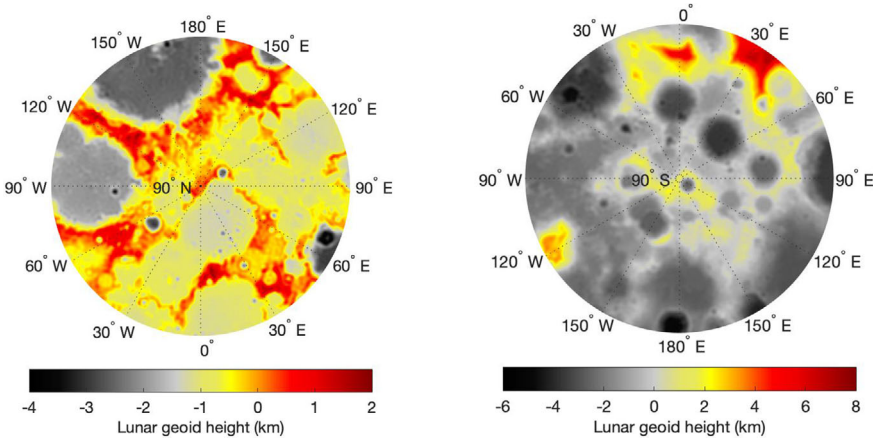


Fig. 2.8 Lunar global topographic map obtained using a spherical-harmonics expansion (order/degree = 2160). Reference of the height is a sphere whose radius is 1737.15 km and whose origin is set to the centre of mass. Full range of the topography is about 19.81 km. Data source provided in Footnote 5



(a) North pole (85°-90° N) (b) South pole (85°-90° S)

Fig. 2.9 Polar topographic maps of the Moon using a spherical-harmonics expansion (order/degree = 2160). Data source provided in Footnote 5

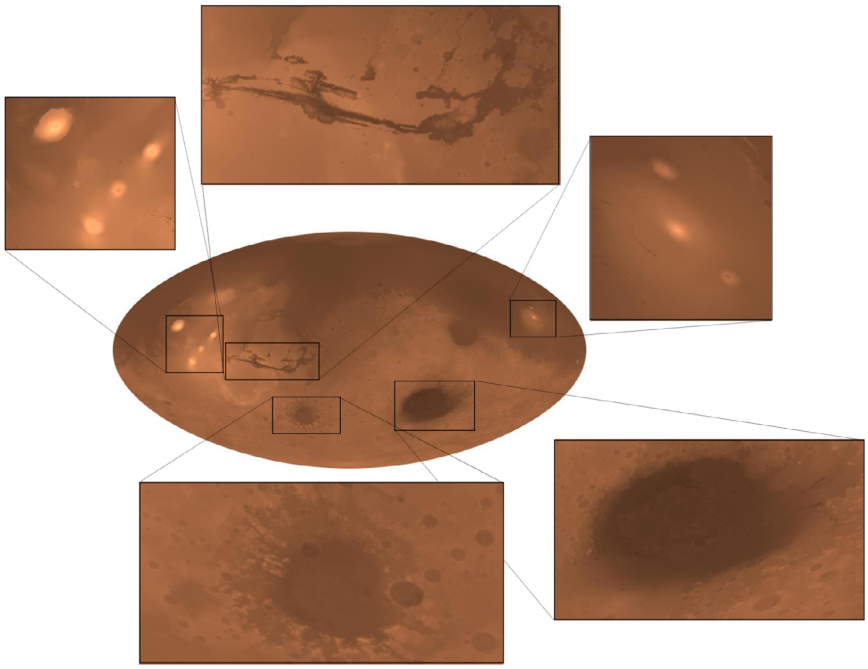


Fig. 2.10 Medium-resolution (0.03125 (1/32) by 0.03125°) MOLA image of Mars. Data obtained from NASA, see Footnote 6

pixel per degree. An example surface map is shown in Fig. 2.10, generated with data obtained from NASA.[6]

A practical implementation of the data is yet again described by Wieczorek (2007), where, as before for the Earth and the Moon, a spherical-harmonics model for the topography of Mars is presented. This model is a degree and order 2600 shape model of Mars derived from the PDS 1/32° resolution gridded data set. Data can be downloaded from Wieczorek's GitHub repository,[7] and has been used to generate the topography shown in Figs. 2.11 and 2.12.

2.1.2.4 Venus

To conclude the discussion on the "regular" bodies, in Figs. 2.13 and 2.14 the topography of the Venus surface, modelled with a spherical-harmonics model is shown. Also in this case, the model has been developed by Wieczorek (2007); it is a degree

[6] https://pds-geosciences.wustl.edu/missions/mgs/mola.html. Date accessed: April 10, 2020.

[7] https://github.com/MarkWieczorek/web/tree/master/spherical-harmonic-models-topography. Date accessed: April 10, 2020.

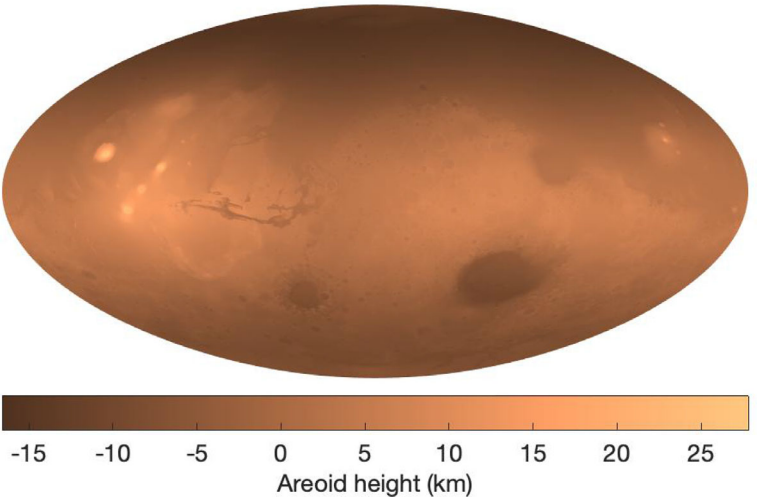
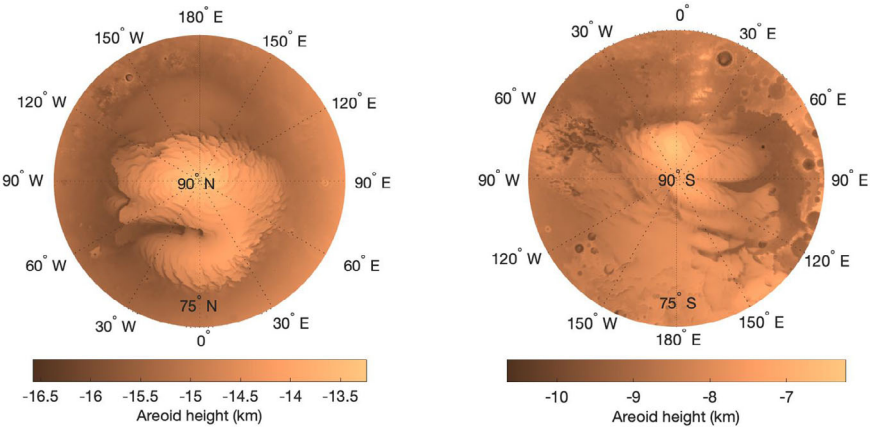


Fig. 2.11 Mars global topographic map obtained using a spherical-harmonics expansion (order/degree = 2600). Reference of the height is a sphere whose radius is 3,389.50 km and whose origin is set to the centre of mass. Full range of the topography is about 44.40 km. Data source provided in Footnote 7



(a) North pole (75°-90° N) (b) South pole (75°-90° S)

Fig. 2.12 Polar topographic maps of Mars using a spherical-harmonics expansion (degree/order = 2600). Data source provided in Footnote 7

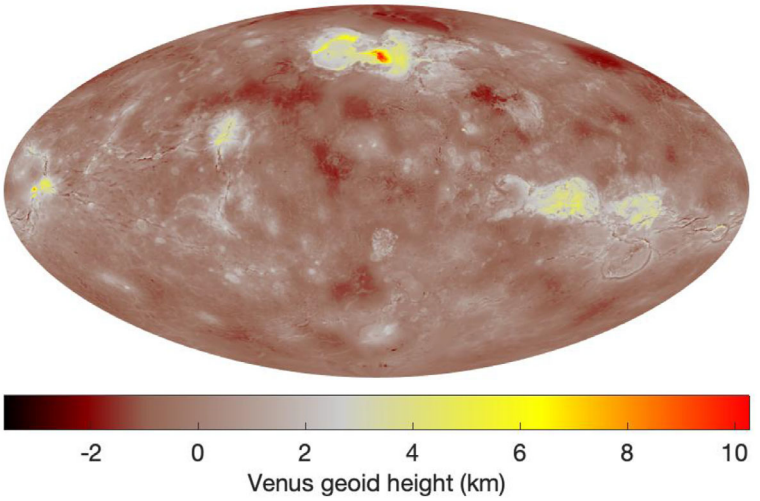
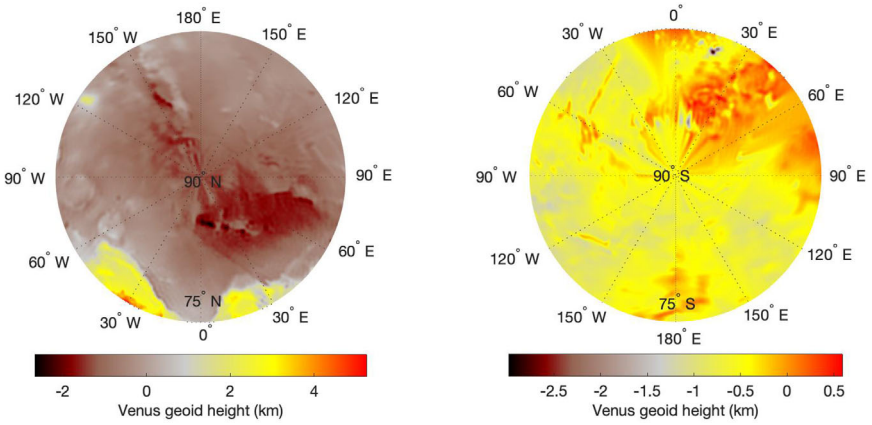


Fig. 2.13 Global topography of Venus, derived with spherical-harmonic model (order/degree = 719). Reference of the height is a sphere whose radius is 6,051.88 km and whose origin is set to the centre of mass. Full range of the topography is about 13.89 km. Data source provided in Footnote 7



(a) North pole (75°-90° N) (b) South pole (75°-90° S)

Fig. 2.14 Polar topographic maps of Venus using a spherical-harmonics expansion (order/degree = 719). Data source provided in Footnote 7

and order 719 shape model derived from Magellan (GTDR3.2), Pioneer Venus, and Venera 15/16 altimetry data.

2.1.3 Irregular Bodies

One method that is used to model the gravity of a body is by assuming it is a constant-density polyhedron, see, for instance, Werner and Scheeres (1997), who applied it to model asteroid 4769 Castalia. This approach is very versatile, as it can easily incorporate valleys, ridges, craters all the way through the asteroid, etc. A polyhedron is a three-dimensional solid body which surface consists of planar faces that are polygons. The polygons meet each other along straight edges, and each connect two vertices. The vertex coordinates alone are insufficient to describe the polyhedron; the connective topology must also be described: for each face its vertices must be defined, which all have to lie in the same plane. This polyhedron can also be used as a surface model and used for landing studies. In this section we will look at polyhedra models of three different asteroids, i.e., Šteins, Kleopatra and Itokawa. These asteroids were (quite arbitrarily) selected because of their irregularity and, of course, availability of data.

2867 Šteins is a small main-belt asteroid that was discovered on 4 November, 1969, by N.S. Chernykh. It was visited by Rosetta in 2008 (Fig. 2.15). The ESA spacecraft Rosetta, launched on 2 March, 2004, on an Ariane 5, has been on a mission to study the comet 67P/Churyumov–Gerasimenko, where it arrived in 2014. From light-curve data a basic triaxial ellipsoid description of the asteroid's shape has been obtained (Sierks, in Melman et al. (2009)). Its dimensions are $3.35 \times 2.90 \times 2.25$ km. Moreover, by a process called limb profiling, a full polyhedron model is obtained with a total of 14,536 faces. It is constructed by Rosetta's OSIRIS Team (MPS/UPD/LAM/IAA/RSSD/INTA/UPM/DASP/IDA), and shown in Fig. 2.16, including the top, side and front views. The crater is not visible in this model, since it does not come out by limb profiling. More advanced techniques (shading and photoclinometry) are needed to get to that level of detail. Nonetheless,

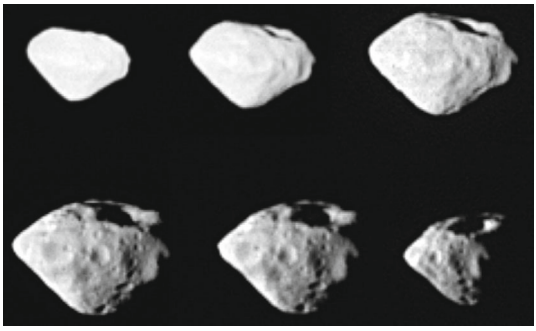


Fig. 2.15 Asteroid 25143 Šteins as seen by Rosetta, taken at different distances. Image: ESA

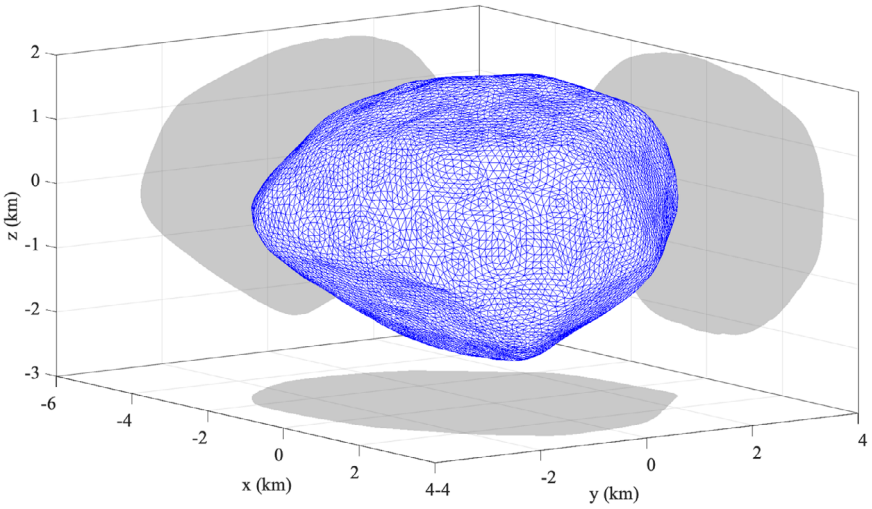


Fig. 2.16 Steins polyhedron model with 14,536 faces. Datasource: Jorda et al. (2012)

this is the most accurate model that is currently available, and it is assumed that it represents the true shape of the asteroid.

Asteroid 216 Kleopatra is a trinary main-belt asteroid that was discovered by Johann Palisa on April 10, 1880, in Pola. It is named after Cleopatra, Queen of Ancient Egypt. Kleopatra has an unusual shape that has been compared to a dog's bone (Fig. 2.17). It is a so-called rubble-pile asteroid, which means that it does not consist of a single body, but of numerous pieces of rock that have coalesced under the influence of gravity. Rubble piles have low density, because there are large cavities between the various "chunks" that make them up. By bouncing radar signals off the asteroid astronomers were able to develop a more detailed computer model of its shape. It is a relatively large asteroid, measuring $217 \times 94 \times 81$ km ($\pm 25\%$) (Ostro et al. 2000). From the computer model a polyhedra model has been derived (Fig. 2.18) that can serve as surface model and used to set up a gravity model.

Itokawa is an S-type asteroid (Figs. 2.19 and 2.20). Radar imaging by Goldstone in 2001 observed an ellipsoid 630 ± 60 m long and 250 ± 30 m wide, which was confirmed by the Japanese Hayabusa spacecraft. This probe arrived in the vicinity of Itokawa on September 12, 2005, and initially "parked" in an asteroid−Sun line at 20 km, and later 7 km, from the asteroid (Itokawa's gravity was too weak to provide an orbit, so the spacecraft adjusted its orbit around the Sun until it matched the asteroid). Hayabusa landed on November 20 for thirty minutes, but failed to operate a device designed to collect soil samples. On November 25, a second landing and sampling sequence was attempted. The sample capsule was returned to Earth and landed at Woomera, South Australia on June 13, 2010, around 13:51 UTC. On November 16, 2010, JAXA reported that dust collected during Hayabusa's voyage was indeed from the asteroid.

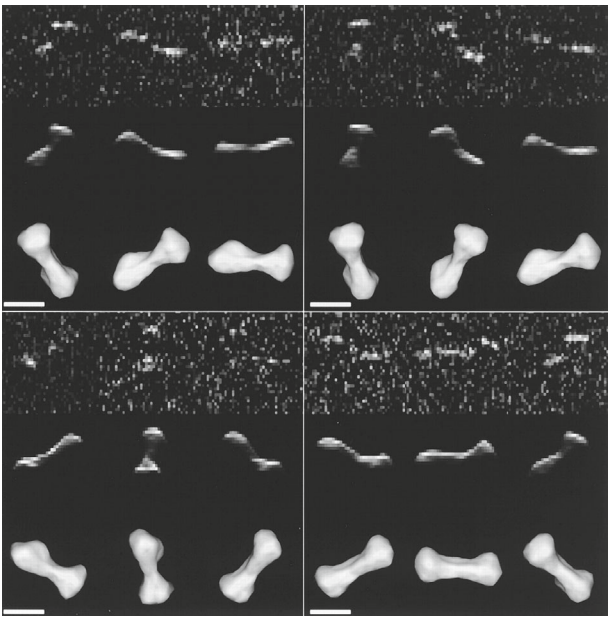


Fig. 2.17 Shape reconstruction results. Each quadrant shows Arecibo radar images from one date (top), corresponding images calculated from the shape model (middle), and corresponding plane-of-sky views of the model (bottom) (Ostro et al. 2000)

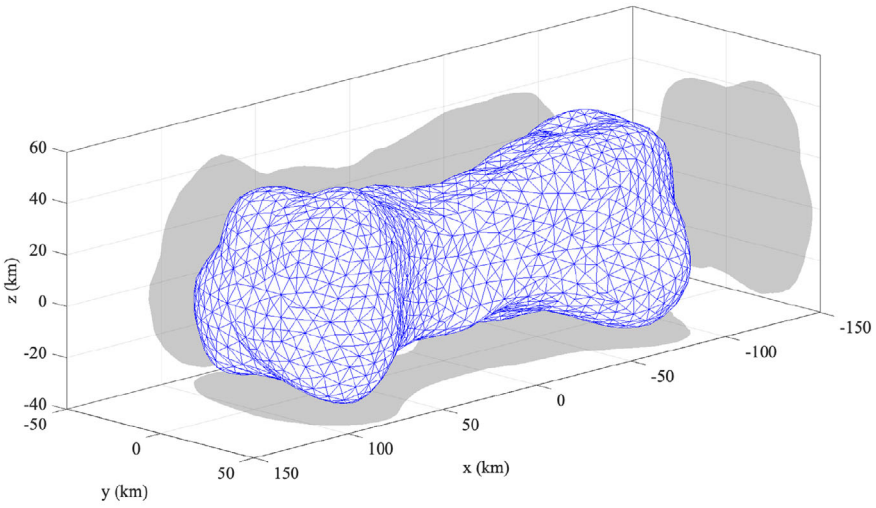


Fig. 2.18 Kleopatra polyhedron model, with 4,092 faces. Datasource: https://echo.jpl.nasa.gov/asteroids/shapes/shapes.html, accessed May 14, 2022

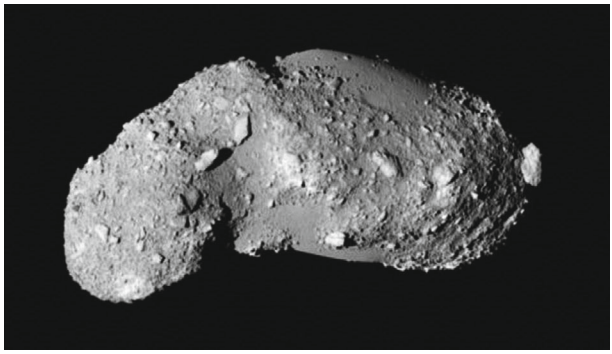


Fig. 2.19 Asteroid 25143 Itokawa, as seen by Hayabusa. Image: JAXA

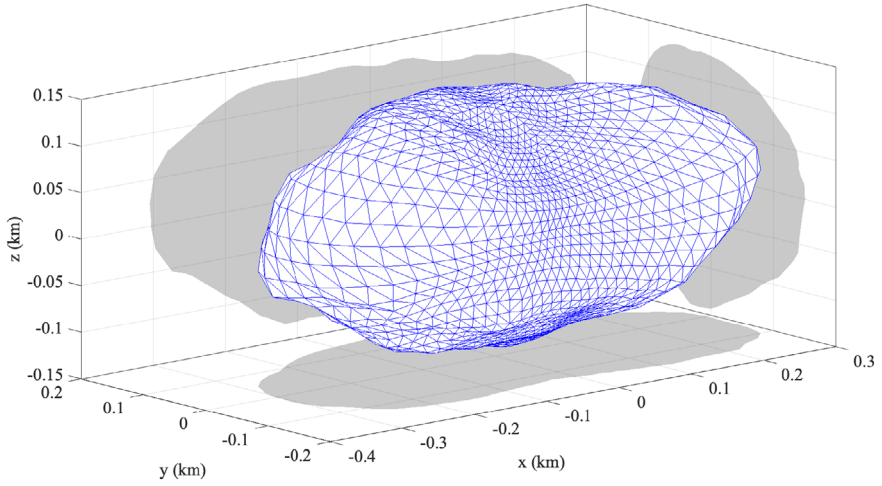


Fig. 2.20 Itokawa polyhedron model, with 3,688 faces. Datasource: https://echo.jpl.nasa.gov/asteroids/shapes/shapes.html, accessed May 14, 2022

2.2 Gravity Field

The gravitational force is one of the external forces acting on a re-entry vehicle. Depending on the application one can model this force in different ways, ranging from constant to a position-dependent quantity. In this section, the various ways of modelling the gravity field are discussed, starting the discussion with the one for regular bodies, i.e., spheroids like the Earth. However, this does not mean that the gravity field is regular too, since that depends on the internal mass distribution of the body. Mountains and seas will, of course, affect this regularity. On the other end of the spectrum, irregular bodies, such as asteroids, are found, which may have equally irregular gravity fields. To accurately model such a field is important for

satellites orbiting at close distance from the surface, as the orbital stability is directly related to the gravitational perturbations. When landing on an asteroid, usually an active guidance system is employed that can counteract these perturbations, so that it seems that they become less important. However, when precision landing is required, then, because also the surface is very irregular (and partially unknown), an accurate knowledge of the gravity field will aid in meeting the landing requirements.

In this section we will discuss the following. In Sect. 2.2.1 the gravity fields of spheroid bodies is introduced, beginning with the so-called "inverse square" law that assumes that the body is a point mass. From there on, the step will be made to modelling the gravity field as a series of spherical harmonics. In the next Sect. 2.2.2 the spherical-harmonics model will also be used as an approximation for the gravity field of irregular bodies.

2.2.1 Spheroid Bodies

Two point masses, M and m, separated by a vector distance R, attract each other with a force given by Newton's law of gravitation as

$$\mathbf{F_G} = \frac{GMm}{R^2}\hat{\mathbf{r}} \tag{2.6}$$

where $\hat{\mathbf{r}}$ is the normalised radius vector. In the above equation, G is the universal gravity constant ($= 6.668 \cdot 10^{-11}$ m^3/kg s^2). Let M be the mass of the central body, e.g., the Earth, then it is convenient to define the geocentric gravitational constant μ as GM (for the Earth, $\mu = 3.9860047 \cdot 10^{14}$ m^3/s^2; values for other celestial bodies can be found in Table 2.2). Equation (2.6) is then rewritten as

$$\mathbf{F_G} = m\frac{\mu}{R^2}\hat{\mathbf{r}} \tag{2.7}$$

which is commonly known as the central-field model.

Moreover, if the gravitational potential U is defined as

$$U = -\frac{\mu}{R} \tag{2.8}$$

then the gravitational force may be rewritten as the gradient of the scalar U:

$$\mathbf{F_G} = m\frac{\mu}{R^2}\hat{\mathbf{r}} = -m\nabla U \tag{2.9}$$

with ∇ as the nabla or gradient operator.

However, planetary bodies are usually only considered as point masses when their gravitational pull is considered to be a so-called third-body perturbation for satellites.

Table 2.2 Some gravitational characteristics of celestial bodies. g_0 is the value at the equator; μ has been calculated from the mass of the body with $G = 6.668{\cdot}10^{-11}$ $m^3/kg\ s^2$. Data taken from Yoder (1995)

Celestial body	μ (m^3/s^2)	g_0 (m/s^2)	g_0-ratio (–)
Earth	$3.98319648{\cdot}10^{14}$	9.78	1.00
Moon	$4.90031320{\cdot}10^{12}$	1.62	0.17
Mercury	$2.20177360{\cdot}10^{13}$	3.70	0.38
Venus	$3.24631580{\cdot}10^{14}$	8.87	0.91
Mars	$4.27985580{\cdot}10^{13}$	3.69	0.38
Jupiter	$1.26598648{\cdot}10^{17}$	23.12	2.36
Saturn	$3.79049128{\cdot}10^{16}$	8.96	0.92
Uranus	$5.78995776{\cdot}10^{15}$	8.69	0.89
Neptune	$6.83003240{\cdot}10^{15}$	11.00	1.12
Titan	$8.97312760{\cdot}10^{9}$	1.35	0.14

Once in close proximity to a planetary body this assumption does not commonly hold, and this main attracting body has to be treated as a body of finite dimensions with an inhomogeneous mass distribution. Especially for Earth (and otherwise Earth-like shapes), U may be expressed as the summation of a central-field term, representing a mass-symmetric body, and a correction term for the Earth's non-symmetric mass distribution in the form of an appropriate spherical harmonic expansion:

$$U = -\frac{\mu}{R} + U_c(R, \tau, \theta) \tag{2.10}$$

where τ is the geocentric longitude and θ is the co-latitude (θ ranges from 0° for $\delta = 90°$ to, 180° for $\delta = -90°$). Assuming that the centre of mass of the planetary body is coinciding with the origin of the reference frame, U_c can be written as (Vallado 2001)[8]:

$$U_c(R, \tau, \theta) = \frac{\mu}{R}\left\{\sum_{n=1}^{\infty}\left[\left(\frac{R_e}{R}\right)^n J_n P_{n,0}(\cos\theta) + \sum_{m=1}^{n}\left(C_{n,m}\cos m\tau + S_{n,m}\sin m\tau\right) P_{n,m}(\cos\theta)\right]\right\} \tag{2.11}$$

with R_e being the radius of the Earth, J_n the zonal harmonic coefficients (J_n can also be written as $C_{n,0}$), $P_{n,m}$ the Legendre polynomials, and $C_{n,m}$ and $S_{n,m}$ the tesseral harmonic coefficients for $n \neq m$ and sectoral harmonic coefficients for

[8] Often one sees in literature the assumption of coinciding centre of mass, such that the summation in Eq. (2.11) and following begins from $n = 2$, rather than 1. At the end of this section, this assumption is revisited.

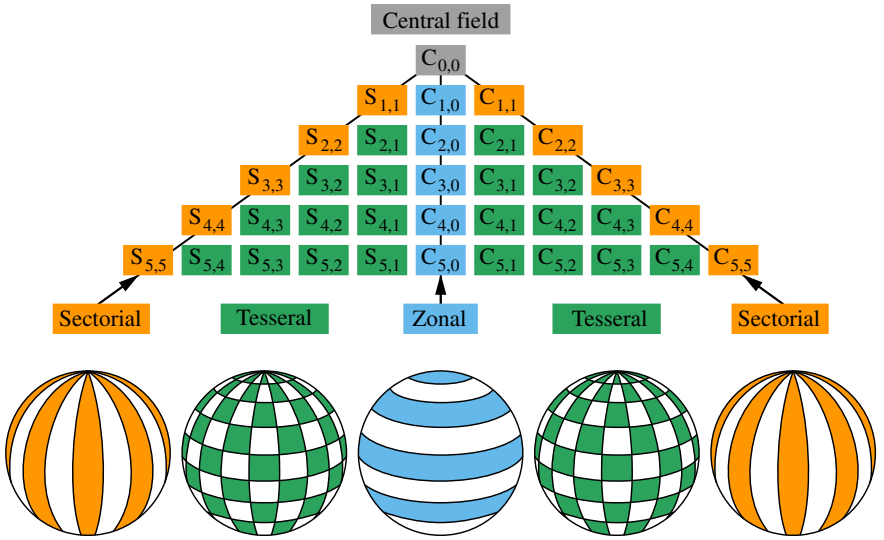


Fig. 2.21 Graphical representation of spherical-harmonics coefficients and their relation to mass offsets in the (spherical) body of interest

$n = m$; the indices n and m are the degree and order of the expansion, respectively.[9] Figure 2.21 provides a visual representation of the different types of coefficients for low degree and order, separating the coefficients in the three categories of zonal, sectorial, and tesseral.

By definition, the coefficients $C_{n,m}$ and $S_{n,m}$ have a relation with the mass distribution of the attracting body. Let us assume that the mass of this body is given by

$$M = \iiint_{\mathcal{V}} \tilde{\mathbf{r}} \rho(\tilde{\mathbf{r}}) \mathrm{d}x \mathrm{d}y \mathrm{d}z \tag{2.12}$$

where $\tilde{\mathbf{r}}$ is the location of a mass element, $\mathrm{d}m$. This (Cartesian) location can also be expressed in spherical coordinates, to link it with the longitude and latitude:

$$\tilde{\mathbf{r}} = \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} \tilde{r} \cos\delta \cos\tau \\ \tilde{r} \cos\delta \sin\tau \\ \tilde{r} \sin\delta \end{pmatrix} \tag{2.13}$$

in a similar way that defines the location on, for instance, the Earth's surface. For the body with mass M, the centre of mass is defined to be

[9] The coefficients with n have a special meaning, which will be discussed in more detail at the end of the section.

$$\mathbf{r_{cm}} = \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} = \frac{1}{M} \iiint_{\mathcal{V}} \tilde{\mathbf{r}} \rho(\tilde{\mathbf{r}}) \mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{2.14}$$

which will be used later, to link $\mathbf{r_{cm}}$ to the coefficients of degree 1.

Returning to the coefficients, they can be written as

$$C_{n,m} = \frac{2-\delta_{0m}}{M} \frac{(n-m)!}{(n+m)!} \iiint_{\mathcal{V}} \frac{\tilde{r}^n}{R^n} P_{n,m}(\sin\delta) \cos(m\tau) \rho(\tilde{\mathbf{r}}) \mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{2.15}$$

and

$$S_{n,m} = \frac{2-\delta_{0m}}{M} \frac{(n-m)!}{(n+m)!} \iiint_{\mathcal{V}} \frac{\tilde{r}^n}{R^n} P_{n,m}(\sin\delta) \sin(m\tau) \rho(\tilde{\mathbf{r}}) \mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{2.16}$$

where the argument to the Legendre polynomials has been changed from co-latitude to latitude; for the sake of brevity the argument $\sin\delta$ to P will be dropped, so $P_{n,m} = P_{n,m}(\sin\delta)$ from now on. In these equations, δ_{0m} is the Kronecker delta, i.e., $\delta_{ji} = 1$ if $i = j$ and 0 otherwise.

In each of the expressions introduced so far, the coefficients, $C_{n,m}$ and $S_{n,m}$, as well as the Legendre polynomials, $P_{n,m}$, are denormalised. Since normalised coefficients ($\bar{C}_{n,m}$ and $\bar{S}_{n,m}$) appear in literature, the relationships between these and the denormalised coefficients are as follows:

$$C_{n,m} = \Pi_{n,m} \bar{C}_{n,m} \quad \text{and} \quad S_{n,m} = \Pi_{n,m} \bar{S}_{n,m} \tag{2.17}$$

where $\Pi_{n,m}$ is given by:

$$\Pi_{n,m} = \sqrt{\frac{(n-m)!(2n+1)(2-\delta_{0m})}{(n+m)!}} \tag{2.18}$$

Note that when the normalised coefficients are used (in Eq. (2.11) and also in the related equations below), the Legendre polynomials should also be normalised. The relation between $P_{n,m}$ and $\bar{P}_{n,m}$ is

$$P_{n,m} = \frac{\bar{P}_{n,m}}{\Pi_{n,m}} \tag{2.19}$$

Since the example models in the tables below provide normalised coefficients, from now on the the combination of $\bar{C}_{n,m}$, $\bar{S}_{n,m}$, and $\bar{P}_{n,m}$ will be used in the description of the spherical-harmonics model.

In case the spherical-harmonics model has to be implemented in a computer program, for implementation purposes Eqs. (2.10) and (2.11) are rewritten as:

$$U(R,\tau,\delta)=\frac{\mu}{R}\left\{1+\sum_{n=1}^{n_{max}}\left[\left(\frac{R_e}{R}\right)^n\sum_{m=0}^{n}\left(\bar{C}_{n,m}\cos m\tau+\bar{S}_{n,m}\sin m\tau\right)\bar{P}_{n,m}\right]\right\} \tag{2.20}$$

The gravitational acceleration **g** is defined to be

$$\mathbf{g}=\begin{pmatrix}g_n & g_e & g_d\end{pmatrix}^{\mathrm{T}}=\begin{pmatrix}-\frac{1}{R}\frac{\partial U}{\partial\delta} & -\frac{1}{R\cos\delta}\frac{\partial U}{\partial\tau} & -\frac{\partial U}{\partial R}\end{pmatrix}^{\mathrm{T}} \tag{2.21}$$

with

$$\frac{\partial U}{\partial\delta}=\frac{\mu}{R}\sum_{n=1}^{n_{max}}\left(\frac{R_e}{R}\right)^n\sum_{m=0}^{n}\left(\bar{C}_{n,m}\cos m\tau+\bar{S}_{n,m}\sin m\tau\right)\frac{\partial\bar{P}_{n,m}}{\partial\delta} \tag{2.22}$$

$$\frac{\partial U}{\partial\tau}=\frac{\mu}{R}\sum_{n=1}^{n_{max}}\left(\frac{R_e}{R}\right)^n\sum_{m=0}^{n}m\left(\bar{S}_{n,m}\cos m\tau-\bar{C}_{n,m}\sin m\tau\right)\bar{P}_{n,m} \tag{2.23}$$

$$\frac{\partial U}{\partial R}=\frac{\mu}{R^2}\left\{1+\sum_{n=1}^{n_{max}}\left[\left(\frac{R_e}{R}\right)^n\sum_{m=0}^{n}\left(\bar{C}_{n,m}\cos m\tau+\bar{S}_{n,m}\sin m\tau\right)(n+1)\bar{P}_{n,m}\right]\right\} \tag{2.24}$$

The above definition provides the components of **g** in the so-called vertical frame, i.e., in north, east, and down direction.

The (normalised) Legendre polynomials are computed via recursion formulae:

Zonals: $m = 0$

$$\bar{P}_{0,0}=1 \tag{2.25}$$

$$\bar{P}_{1,0}=\sqrt{3}\sin\delta \tag{2.26}$$

$$\bar{P}_{n,0}=\frac{1}{n}\left[(2n-1)\sin\delta\sqrt{\frac{2n+1}{2n-1}}\bar{P}_{n-1,0}+(n-1)\sqrt{\frac{2n+1}{2n-3}}\bar{P}_{n-2,0}\right] \tag{2.27}$$

Tesserals: $m \neq 0$ and $m < n$

$$\begin{aligned}\bar{P}_{n,m}=&\sqrt{\frac{(n-m)(n-1-m)(2n+1)}{(n+m)(n-1+m)(2n-3)}}\bar{P}_{n-2,m}\\&-(2n-1)\cos\delta\sqrt{\frac{(2n+1)(1+\delta_{1m})}{(n+m-1)(n+m)(2n-1)}}\bar{P}_{n-1,m-1}\end{aligned} \tag{2.28}$$

Table 2.3 Normalised spherical harmonics coefficients for the Earth: the ITG-Grace2010 gravity-field model, $R_{e,Earth} = 6378.1366\,\text{km}$ and $\mu_{Earth} = 3.986004415{\cdot}10^5\ \text{km}^3/\text{s}^2$ (Mayer-Gürr et al. 2010)

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
2	0	$-0.48416922{\cdot}10^{-3}$	0.00000000
2	1	$-0.23496643{\cdot}10^{-9}$	$0.14670562{\cdot}10^{-8}$
2	2	$0.24393389{\cdot}10^{-5}$	$-0.14002759{\cdot}10^{-5}$
3	0	$0.95719820{\cdot}10^{-6}$	0.00000000
3	1	$0.20304650{\cdot}10^{-5}$	$0.24819221{\cdot}10^{-6}$
3	2	$0.90478857{\cdot}10^{-6}$	$-0.61899625{\cdot}10^{-6}$
3	3	$0.72128184{\cdot}10^{-6}$	$0.14143769{\cdot}10^{-5}$
4	0	$0.54000342{\cdot}10^{-6}$	0.00000000
4	1	$-0.53614922{\cdot}10^{-6}$	$-0.47356758{\cdot}10^{-6}$
4	2	$0.35049863{\cdot}10^{-6}$	$0.66248219{\cdot}10^{-6}$
4	3	$0.99086430{\cdot}10^{-6}$	$-0.20095555{\cdot}10^{-6}$
4	4	$-0.18848951{\cdot}10^{-6}$	$0.30882588{\cdot}10^{-6}$

Sectorials: $m = n$

$$\bar{P}_{1,1} = \sqrt{3}\cos\delta \tag{2.29}$$

$$\bar{P}_{n,m} = -(2n-1)\cos\delta\sqrt{\frac{(2n+1)}{(n+m-1)(n+m)(2n-1)}}\,\bar{P}_{n-1,m-1} \tag{2.30}$$

The derivative relationship, finally, is given by

$$\frac{\partial\bar{P}_{n,m}}{\partial\delta} = \frac{\partial\bar{P}_{n,m}(\sin\delta)}{\partial\delta} = \bar{P}_{n,m+1} - m\tan\delta\,\bar{P}_{n,m} \tag{2.31}$$

To serve as reference, for each of the celestial bodies that are potential targets for re-entry studies, i.e., Earth, the Moon, Venus, and Mars the coefficients of a low-order spherical harmonics model are provided. It is noted that depending on the application more or fewer terms, i.e., higher or lower degree and/or order of the model, should be included. The reader should also realise that whenever more data from recent missions become available, the coefficients may change and their uncertainty reduce. However, this usually applies to higher-degree and/or order coefficients.

To begin with, for the Earth, in Table 2.3 the coefficients of ITG-Grace2010 gravity-field model are listed. For Mars, Venus and the Moon a subset of the coefficients can be found in Tables 2.4, 2.5, and 2.6.[10] It is noted that the Moon's grav-

[10] Complete datasets for the gravity-field models, including the uncertainties in the coefficients, can be downloaded from https://pds-geosciences.wustl.edu/dataserv/gravity_models.htm. Date accessed: April 10, 2020.

Table 2.4 Normalised spherical harmonics coefficients for Mars: the GMM-3 gravity-field model, $R_{e,Mars} = 3{,}396.0\,\text{km}$ and $\mu_{Mars} = 4.282837285418775 \cdot 10^4\ \text{km}^3/\text{s}^2$. Complete model is $n = m = 120$. Data source: Mars Reconnaissance Orbiter, see Footnote 10

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
2	0	$-0.87502113 \cdot 10^{-3}$	0.00000000
2	1	$0.59031496 \cdot 10^{-9}$	$-0.49433617 \cdot 10^{-10}$
2	2	$-0.84635904 \cdot 10^{-4}$	$0.48934626 \cdot 10^{-4}$
3	0	$-0.11896035 \cdot 10^{-4}$	0.00000000
3	1	$0.38042981 \cdot 10^{-5}$	$0.25177322 \cdot 10^{-4}$
3	2	$-0.15948009 \cdot 10^{-4}$	$0.83615176 \cdot 10^{-5}$
3	3	$0.35054186 \cdot 10^{-4}$	$0.25574571 \cdot 10^{-4}$
4	0	$0.51294073 \cdot 10^{-5}$	0.00000000
4	1	$0.42162333 \cdot 10^{-5}$	$0.37633848 \cdot 10^{-5}$
4	2	$-0.95253939 \cdot 10^{-6}$	$-0.89808128 \cdot 10^{-5}$
4	3	$0.64568895 \cdot 10^{-5}$	$-0.19319621 \cdot 10^{-6}$
4	4	$0.30971854 \cdot 10^{-6}$	$-0.12873047 \cdot 10^{-4}$

Table 2.5 Normalised spherical harmonics coefficients for Venus: MGN-V-RSS-5-GRAVITY-L2-V1.0 gravity-field model, $R_{e,Venus} = 6{,}051.0\,\text{km}$ and $\mu_{Venus} = 3.24858592 \cdot 10^5\ \text{km}^3/\text{s}^2$. Complete model is $n = m = 120$. Data source: Magellan radio science and tracking data, see Footnote 10

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
2	0	$-0.196972 \cdot 10^{-5}$	0.000000
2	1	$0.268000 \cdot 10^{-7}$	$0.132400 \cdot 10^{-7}$
2	2	$0.857770 \cdot 10^{-6}$	$-0.955300 \cdot 10^{-7}$
3	0	$0.796820 \cdot 10^{-6}$	0.000000
3	1	$0.234830 \cdot 10^{-5}$	$0.541620 \cdot 10^{-6}$
3	2	$-0.853000 \cdot 10^{-8}$	$0.809060 \cdot 10^{-6}$
3	3	$-0.188010 \cdot 10^{-6}$	$0.213480 \cdot 10^{-6}$
4	0	$0.715800 \cdot 10^{-6}$	0.000000
4	1	$-0.457420 \cdot 10^{-6}$	$0.491600 \cdot 10^{-6}$
4	2	$0.126320 \cdot 10^{-6}$	$0.483700 \cdot 10^{-6}$
4	3	$-0.174660 \cdot 10^{-6}$	$-0.116490 \cdot 10^{-6}$
4	4	$0.172510 \cdot 10^{-6}$	$0.137661 \cdot 10^{-5}$

ity field is more non-linear than the others, and coefficients up to higher degree and order should typically be included for precision-landing studies. It is advised to determine this on a case-by-case basis. Irrespective, at any time the validity of spherical-harmonics models should be considered. Because of their definition, the models are only valid for locations above the reference geoid, i.e., for radial distances larger than the reference radius, defining the so-called *Brillouin sphere*. Depending

Table 2.6 Normalised spherical harmonics coefficients for the Moon, $R_{e,Moon} = 1{,}738$ km and $\mu_{Moon} = 4.90279980693169 \cdot 10^3$ km^3/s^2. Complete model is $n = m = 660$. Data source: Lunar Reconnaissance Orbiter, see Footnote 10

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
2	0	$-0.90882924 \cdot 10^{-4}$	0.0000000
2	1	$0.84954065 \cdot 10^{-10}$	$0.9772702 \cdot 10^{-9}$
2	2	$0.34670944 \cdot 10^{-4}$	$-0.2406424 \cdot 10^{-9}$
3	0	$-0.31974039 \cdot 10^{-5}$	0.0000000
3	1	$0.26367949 \cdot 10^{-4}$	$0.5454562 \cdot 10^{-5}$
3	2	$0.14171538 \cdot 10^{-4}$	$0.4877944 \cdot 10^{-5}$
3	3	$0.12275202 \cdot 10^{-4}$	$-0.1774375 \cdot 10^{-5}$
4	0	$0.32347925 \cdot 10^{-5}$	0.0000000
4	1	$-0.60134612 \cdot 10^{-5}$	$0.1664333 \cdot 10^{-5}$
4	2	$-0.71161737 \cdot 10^{-5}$	$-0.6777023 \cdot 10^{-5}$
4	3	$-0.13499174 \cdot 10^{-5}$	$-0.1344499 \cdot 10^{-4}$
4	4	$-0.60069115 \cdot 10^{-5}$	$0.3926611 \cdot 10^{-5}$
5	0	$-0.22378187 \cdot 10^{-6}$	0.00000000
5	1	$-0.10116123 \cdot 10^{-5}$	$-0.41189034 \cdot 10^{-5}$
5	2	$0.43995351 \cdot 10^{-5}$	$0.10571040 \cdot 10^{-5}$
5	3	$0.46619443 \cdot 10^{-6}$	$0.86988723 \cdot 10^{-5}$
5	4	$0.27542005 \cdot 10^{-5}$	$0.67624332 \cdot 10^{-7}$
5	5	$0.31107298 \cdot 10^{-5}$	$-0.27545431 \cdot 10^{-5}$
6	0	$0.38184191 \cdot 10^{-5}$	0.00000000
6	1	$0.15282769 \cdot 10^{-5}$	$-0.25995935 \cdot 10^{-5}$
6	2	$-0.43973006 \cdot 10^{-5}$	$-0.21676869 \cdot 10^{-5}$
6	3	$-0.33175420 \cdot 10^{-5}$	$-0.34274198 \cdot 10^{-5}$
6	4	$0.34122058 \cdot 10^{-6}$	$-0.40580194 \cdot 10^{-5}$
6	5	$0.14543742 \cdot 10^{-5}$	$-0.10341804 \cdot 10^{-4}$
6	6	$-0.46842414 \cdot 10^{-5}$	$0.72298542 \cdot 10^{-5}$
7	0	$0.55933882 \cdot 10^{-5}$	0.00000000
7	1	$0.74716915 \cdot 10^{-5}$	$-0.11974837 \cdot 10^{-6}$
7	2	$-0.65012294 \cdot 10^{-6}$	$0.24111006 \cdot 10^{-5}$
7	3	$0.59942806 \cdot 10^{-6}$	$0.23573269 \cdot 10^{-5}$
7	4	$-0.84369455 \cdot 10^{-6}$	$0.75651532 \cdot 10^{-6}$
7	5	$-0.20682822 \cdot 10^{-6}$	$0.10693088 \cdot 10^{-5}$
7	6	$-0.10653046 \cdot 10^{-5}$	$0.11004282 \cdot 10^{-5}$
7	7	$-0.18202646 \cdot 10^{-5}$	$-0.15999966 \cdot 10^{-5}$
8	0	$0.23468301 \cdot 10^{-5}$	0.00000000
8	1	$0.41735905 \cdot 10^{-8}$	$0.10980291 \cdot 10^{-5}$
8	2	$0.30093221 \cdot 10^{-5}$	$0.19305656 \cdot 10^{-5}$
8	3	$-0.18890483 \cdot 10^{-5}$	$0.95448280 \cdot 10^{-6}$
8	4	$0.34086972 \cdot 10^{-5}$	$-0.52824948 \cdot 10^{-6}$
8	5	$-0.12480443 \cdot 10^{-5}$	$0.29185713 \cdot 10^{-5}$
8	6	$-0.16604389 \cdot 10^{-5}$	$-0.21146631 \cdot 10^{-5}$
8	7	$-0.15096536 \cdot 10^{-5}$	$0.32688521 \cdot 10^{-5}$
8	8	$-0.24856047 \cdot 10^{-5}$	$0.21164274 \cdot 10^{-5}$

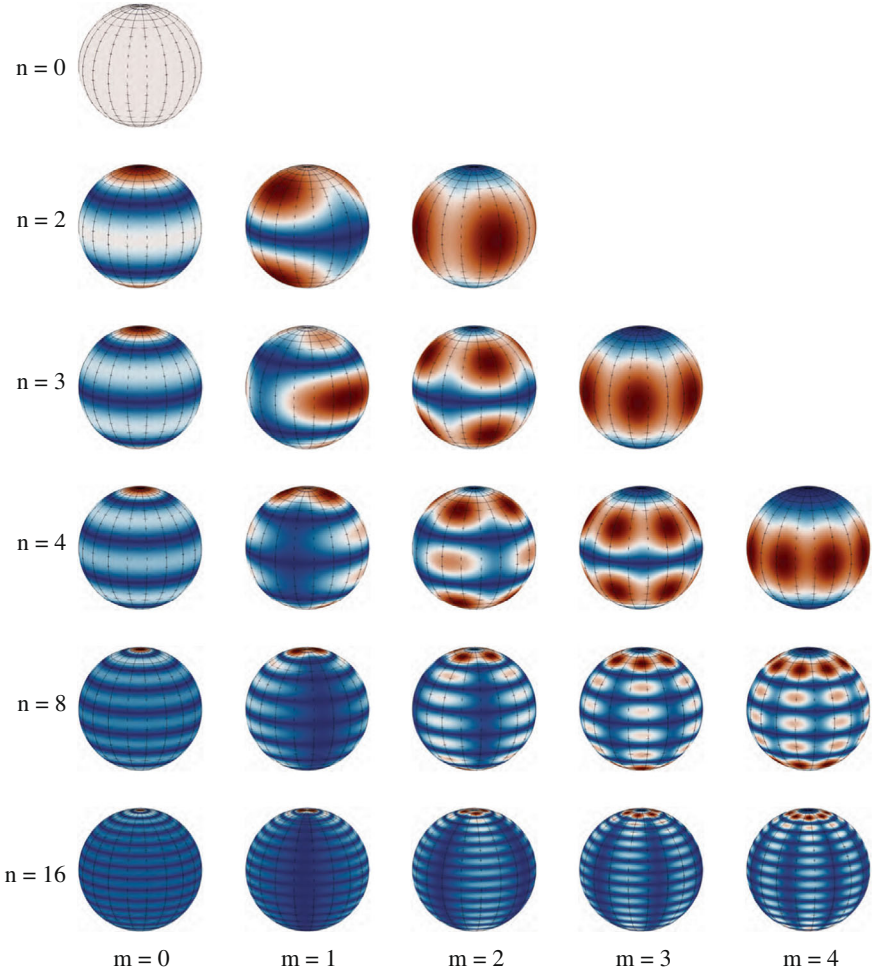


Fig. 2.22 The effect of individual degree-and-order terms, n, m, to the gravitational acceleration (based on the ITG-Grace2010 gravity-field coefficients). Note that the colour variations indicate relative fluctuations in between the minimum (= dark blue) and maximum (= dark red) acceleration values of each combination of n and m

on the elevation, some surface locations may have a lower radius and in principle the model can no longer be used.

To illustrate the contribution of different degree and order to the gravitational acceleration, Fig. 2.22 shows the individual components for the Earth's gravity field (using the ITG-Grace2010 coefficients listed in Table 2.3). For each degree, n, and order, m, the figure shows the combined effect of $\bar{C}_{n,m}$ and $\bar{S}_{n,m}$, but only for the specific combination of n and m (not the summation up to that degree and order). Degree and order 0 represent the central-field term, hence the uniform colour. For

degree 2 and higher (with order 0), the zonal bands are clearly visible, and with increasing order the "checker" pattern of the tesseral terms starts appearing, albeit merged with the sectorial bands.

As discussed before, for a typical mission analysis a high accuracy of the gravitational acceleration is not always required. We already discussed the so-called central-field model. This model can also be obtained from the general definition by spherical harmonics, i.e., it simply follows from Eqs. (2.22)–(2.24) by realising that $n = m = 0$. In that case the summations, which start at $n = 1$ do not contribute such that $\frac{\partial U}{\partial R} = \frac{\mu}{R^2}$ and $\frac{\partial U}{\partial \tau} = \frac{\partial U}{\partial \delta} = 0$.

In other cases a slightly more accurate model of g is required, for instance, for guidance and navigation in the upper layers of the atmosphere where gravity is the dominating force. Then, the central body is considered to be an ellipsoid with mass symmetry about the polar axis, but otherwise mass inhomogeneity in lateral direction. This means that only the zonal terms contribute to the gravitational acceleration ($m = 0$) and that the Legendre polynomials are given by Eqs. (2.25) through (2.27). From Eq. (2.23) it can be seen that $\frac{\partial U}{\partial \tau} = 0$, because of the $m = 0$ right after the summation term. For re-entry simulations, depending on the required accuracy, it usually suffices to include only a few terms, mostly only the term for $n = 2$, but in some situations one also finds $n = 3$ and $n = 4$ terms. Concluding, in that case g is defined as

$$g = \begin{pmatrix} g_n & 0 & g_d \end{pmatrix}^{\mathrm{T}} \tag{2.32}$$

where, because of the assumptions that were made concerning the rotational symmetry of the central body around the polar axis, z_R, there is no component of the gravity vector out of the meridian plane ($m = 0$). The coefficients J_2, J_3, and J_4 are determined experimentally, and are called harmonics. The even harmonics give rise to the oblate terms, which are symmetric about the equatorial radius; the odd harmonics are antisymmetric about the equatorial radius and are the so-called "pear-shape" terms. Evaluating the Legendre polynomials for $n = 2$, 3, and 4 with $m = 0$, yields

$$\begin{aligned} g_n = -3\tfrac{\mu}{R^2}\left(\tfrac{R_e}{R}\right)^2 \sin\delta\cos\delta\Big[J_2 + \tfrac{1}{2}J_3\left(\tfrac{R_e}{R}\right)\operatorname{cosec}\delta\,(5\sin^2\delta - 1) \\ +\tfrac{5}{6}J_4\left(\tfrac{R_e}{R}\right)^2(7\sin^2\delta - 3)\Big] \end{aligned} \tag{2.33}$$

$$\begin{aligned} g_d = \tfrac{\mu}{R^2}\Big[1 + \tfrac{3}{2}J_2\left(\tfrac{R_e}{R}\right)^2(3\sin^2\delta - 1) - 2J_3\left(\tfrac{R_e}{R}\right)^3\sin\delta\,(5\sin^2\delta - 3) \\ - \tfrac{5}{8}J_4\left(\tfrac{R_e}{R}\right)^4(35\sin^4\delta - 30\sin^2\delta + 3)\Big] \end{aligned} \tag{2.34}$$

In Fig. 2.23 the error that is made when including only the J_2-effect and compared to a 80×80 spherical-harmonics model using the ITG-Grace2010 gravity-field model. The difference varies between -1.5 and 1 mm/s^2 for the Earth's surface (i.e., $R = R_e$). Note that even though the difference is not large in absolute sense, the

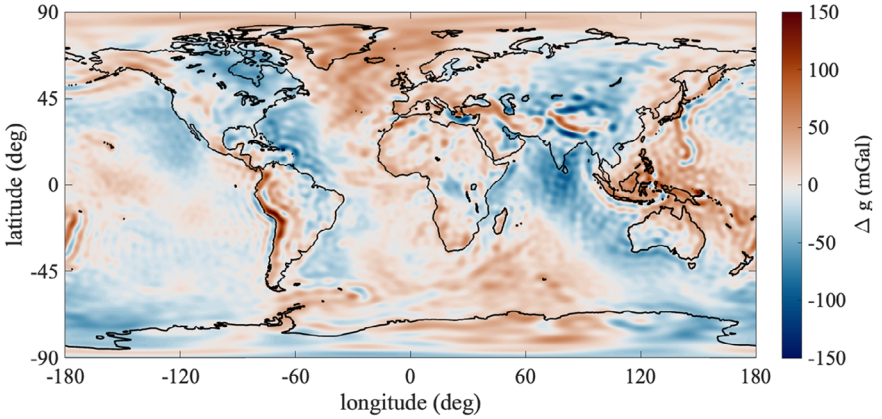


Fig. 2.23 Difference in gravitational acceleration at the Earth's surface between the model including only the J_2-term and the 80×80 spherical-harmonics model, both using the ITG-Grace2010 gravity-field coefficients

induced error is integrated over time and can accumulated to much larger differences, depending on the trajectory and flight duration. For an altitude of $h = 120\,\text{km}$, the error pattern changes a bit, with maximum differences between -0.8 and $0.5\,\text{mm/s}^2$.

Example 2.1 (*Low-order gravity model with J_2-term*) To derive a low-order gravity model from the complete spherical-harmonics model we first establish that $m = 0$, retaining only the zonal harmonics. For re-entry simulations in the Earth's sphere of influence it is usually sufficient to include the first-order effect due to the oblateness of the Earth. The two non-zero components of the gravitational acceleration, i.e., g_d and g_n are, according to Eqs. (2.21), (2.22) and (2.24):

$$g_d = -\frac{\partial U}{\partial R} = \frac{\mu}{R^2}\left\{1 + 3\left(\frac{R_e}{R}\right)^2 \bar{J}_2 \bar{P}_{2,0}\right\} \tag{2.35}$$

$$g_n = -\frac{1}{R}\frac{\partial U}{\partial \delta} = \frac{\mu}{R^2}\left(\frac{R_e}{R}\right)^2 \bar{J}_2 \frac{\partial \bar{P}_{2,0}}{\partial \delta} \tag{2.36}$$

With $\bar{P}_{0,0} = 1$ and $\bar{P}_{1,0}(\sin\delta) = \sqrt{3}\sin\delta$ the normalised Legendre polynomial for $n = 2$, i.e., $\bar{P}_{2,0}$, can easily be calculated. Using Eqs. (2.25)–(2.27) it is found that

$$\begin{aligned}\bar{P}_{2,0} &= \frac{1}{2}\left(3\sin\delta\sqrt{\frac{5}{3}}\bar{P}_{1,0} - \sqrt{\frac{5}{1}}\bar{P}_{0,0}\right) \\ &= \frac{1}{2}\left(3\sin\delta\sqrt{\frac{5}{3}}\sqrt{3}\sin\delta - \sqrt{\frac{5}{1}}\right) = \frac{1}{2}\sqrt{5}\left(3\sin^2\delta - 1\right)\end{aligned} \tag{2.37}$$

Since the low-order models that appear in literature are commonly denormalised, $\bar{P}_{2,0}$ is denormalised with Eq. (2.19):

$$P_{2,0} = \frac{\frac{1}{2}\sqrt{5}\left(3\sin^2\delta - 1\right)}{\Pi_{2,0}} = \frac{1}{2}\left(3\sin^2\delta - 1\right) \tag{2.38}$$

with $\Pi_{2,0} = \sqrt{5}$, following from Eq. (2.18). $P_{2,0}$ substituted into Eq. (2.35) gives

$$g_d = \frac{\mu}{R^2}\left\{1 + \frac{3}{2}\left(\frac{R_e}{R}\right)^2 J_2\left(3\sin^2\delta - 1\right)\right\} \tag{2.39}$$

Inspecting Eq. (2.36) shows that $\frac{\partial \bar{P}_{2,0}(\sin\delta)}{\partial\delta}$ is needed to determine g_n. This derivative is evaluated with Eq. (2.31), yielding

$$\frac{\partial \bar{P}_{2,0}}{\partial\delta} = \bar{P}_{2,1} \tag{2.40}$$

From the above equation it is clear that $\bar{P}_{2,1}$ has to be evaluated. This polynomial follows from Eq. (2.28) as

$$\bar{P}_{2,1} = -\sqrt{5}\cos\delta\,\bar{P}_{1,0} = -\sqrt{15}\cos\delta\sin\delta \tag{2.41}$$

With a "denormalisation" coefficient, Eq. (2.18), of $\sqrt{\frac{10}{6}}$ (note that $n = 2$ and $m = 1$ in this case) one gets:

$$P_{2,1} = -3\cos\delta\sin\delta \tag{2.42}$$

$P_{2,1}$ substituted in Eq. (2.36), finally, gives:

$$g_n = -3\frac{\mu}{R^2}\left(\frac{R_e}{R}\right)^2 J_2\cos\delta\sin\delta \tag{2.43}$$

The denormalised zonal coefficient J_2 follows from the fact that $J_2 = C_{2,0} = \Pi_{2,0}\bar{C}_{2,0}$. From Table 2.3, $\bar{C}_{2,0} = -0.48416922\cdot10^{-3}$, so $J_2 = -1.08263529\cdot10^{-3}$. ■

Degree 0 and 1 terms

The gravitational potential Eq. (2.10) separates the perturbing terms from the central-field potential, with the perturbing terms given by Eq. (2.11). For that reason, the coefficients listed in Tables 2.3, 2.4, 2.5 and 2.6 represent the perturbing harmonics only. If included in these sets, the central-field term is represented by $\bar{C}_{0,0} = 1.0$ (with $\bar{S}_{0,0} = 0.0$). This can be easily shown the definition of $C_{0,0}$, following from Eq. (2.15):

$$C_{0,0} = \frac{1}{M}\frac{0!}{0!}\iiint_{\mathcal{V}} \frac{\tilde{r}^0}{R^0} P_{0,0}(\sin\delta)\cos 0\rho(\tilde{\mathbf{r}})\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{2.44}$$

With $P_{0,0} = \bar{P}_{0,0} = 1$, and knowing that the volume integral over all mass elements is equal to the total mass:

$$C_{0,0} = \frac{1}{M}\iiint_{\mathcal{V}} \rho(\tilde{\mathbf{r}})\mathrm{d}x\mathrm{d}y\mathrm{d}z = \frac{M}{M} = 1 \tag{2.45}$$

But there is more to it. When the body is not an ideal rotational ellipsoid, it is not guaranteed that the coefficients of degree 1 are all zero. Ignoring the coefficients $S_{n,0}$, as they are zero by definition (following from Eq. (2.16), with $m = 0$), there are three coefficients that require special attention: $C_{1,0}$, $C_{1,1}$, and $S_{1,1}$. All three are related to the centre-of-mass location and the choice of reference frame.

The combination of $n = 1$ and $m = 0$ substituted into Eq. (2.15) gives

$$C_{1,0} = \frac{1}{M}\iiint_{\mathcal{V}} \frac{\tilde{r}}{R} P_{1,0}\rho(\tilde{\mathbf{r}})\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{2.46}$$

with Eq. (2.26) providing $\bar{P}_{1,0} = \sqrt{3}\sin\delta$, and thus, using Eqs. (2.18) and (2.19), $P_{1,0} = \frac{\bar{P}_{1,0}}{\sqrt{3}} = \sin\delta$. Therefore,

$$C_{1,0} = \frac{1}{M}\iiint_{\mathcal{V}} \frac{\tilde{r}}{R}\sin\delta\rho(\tilde{\mathbf{r}})\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{2.47}$$

With $\tilde{z} = \tilde{r}\sin\delta$ this yields

$$C_{1,0} = \frac{1}{MR}\iiint_{\mathcal{V}} \tilde{z}\rho(\tilde{\mathbf{r}})\mathrm{d}x\mathrm{d}y\mathrm{d}z = \frac{\bar{z}}{R} \tag{2.48}$$

For the combination $n = m = 1$, $C_{1,1}$ is given by

$$C_{1,1} = \frac{2}{M}\frac{0!}{2!}\iiint_{\mathcal{V}} \frac{\tilde{r}^1}{R^1} P_{1,1}\cos\tau\rho(\tilde{\mathbf{r}})\mathrm{d}x\mathrm{d}y\mathrm{d}z = \frac{1}{MR}\iiint_{\mathcal{V}} \tilde{r}\cos\delta\cos\tau\rho(\tilde{\mathbf{r}})\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{2.49}$$

where $P_{1,1} = \cos\delta$ has been substituted. With $\tilde{x} = \tilde{r}\cos\delta\cos\tau$, it follows that

$$C_{1,1} = \frac{\bar{x}}{R} \tag{2.50}$$

In a similar fashion it follows that

$$S_{1,1} = \frac{\bar{y}}{R} \tag{2.51}$$

Normalisation of the above coefficients follows from the inverse relation of Eq. (2.17).

As mentioned, in case the centre of mass is located in the origin of the reference frame, these three coefficients are zero. If not, they have non-zero values and should be included in the series expansion. Finally, it is noted that $C_{2,1}$ and $S_{2,1}$ are both zero, when the z-axis of the body is aligned with its main axis of inertia (Montebruck and Gill 2000).

2.2.2 Irregular Bodies

It is not hard to imagine that highly irregular shapes, such as asteroids, will cause significantly different orbits than in the case of a pure spherical potential. As discussed in the previous section, a common way to describe the gravity field of spheroid bodies is with a spherical-harmonics model. However, the more irregular the shape the more the describing potential will deviate from a "pure" spheroid. The deviations will become more apparent the closer one gets to the surface of the body. The gravity of a body can also be modelled by assuming it is a constant-density polyhedron. This approach is very versatile, as it can easily incorporate valleys, ridges, craters all the way through the asteroid, etc.

Werner and Scheeres (1997) have derived a method that describes the gravity field exactly for the given shape and density, up to the asteroid's surface. This method is also applied by Melman et al. (2009) to derive a model for the gravity field of Šteins, and similarly for Kleopatra and Itokawa. These models were subsequently converted to spherical-harmonics models and used to compare the accuracy of either model. They found that for Šteins the spherical harmonics diverge close to the surface. Hence, for accuracy reasons the use of polyhedra is advised when close to the small body. However, the computations with the polyhedron for the same accuracy level are a factor of 20 slower than the computations with spherical harmonics. Therefore, when further away from the body, the use of spherical harmonics will give a higher efficiency, while retaining the required accuracy.

To conclude the discussion about gravity fields, in Tables 2.7, 2.8, and 2.9 the spherical-harmonics gravity models of order and degree 4 are listed for Šteins, Kleopatra and Itokawa. These models have been derived using the polyhedra shown in Figs. 2.16, 2.18, and 2.20, under the assumption that the reference frame has its origin in the asteroid's centre of mass. It is clear from the given coefficients that, for instance, the J_2 effect, which for the Earth is four orders of magnitude smaller than the central term, is "only" about 23 times smaller for Šteins, a factor 15 for Kleopatra and 28 for Itokawa. Also the other coefficients are relatively much larger, clearly reflecting the irregularity of the asteroids. It is noted that for the implementation of

Table 2.7 Normalised spherical harmonics coefficients of Šteins. The reference radius is 3,350 m

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
2	0	$-4.37{\cdot}10^{-2}$	0.00
2	1	0.00	0.00
2	2	$2.04{\cdot}10^{-2}$	0.00
3	0	$5.18{\cdot}10^{-3}$	0.00
3	1	$1.84{\cdot}10^{-3}$	$1.13{\cdot}10^{-3}$
3	2	$2.10{\cdot}10^{-3}$	$-3.17{\cdot}10^{-4}$
3	3	$-2.54{\cdot}10^{-3}$	$-7.48{\cdot}10^{-3}$
4	0	$8.39{\cdot}10^{-3}$	0.00
4	1	$-3.12{\cdot}10^{-4}$	$-2.13{\cdot}10^{-4}$
4	2	$-3.82{\cdot}10^{-3}$	$-5.67{\cdot}10^{-4}$
4	3	$-2.99{\cdot}10^{-4}$	$-1.27{\cdot}10^{-4}$
4	4	$-4.66{\cdot}10^{-4}$	$-6.41{\cdot}10^{-4}$

Table 2.8 Normalised spherical harmonics coefficients of Kleopatra. The reference radius is 114.166 km

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
2	0	$-6.69{\cdot}10^{-2}$	0.00
2	1	0.00	0.00
2	2	$1.14{\cdot}10^{-1}$	0.00
3	0	$-8.98{\cdot}10^{-4}$	0.00
3	1	$-3.58{\cdot}10^{-4}$	$-3.04{\cdot}10^{-4}$
3	2	$8.43{\cdot}10^{-4}$	$2.97{\cdot}10^{-4}$
3	3	$-2.42{\cdot}10^{-3}$	$3.66{\cdot}10^{-3}$
4	0	$1.33{\cdot}10^{-2}$	0.00
4	1	$-7.69{\cdot}10^{-4}$	$-2.93{\cdot}10^{-4}$
4	2	$-1.96{\cdot}10^{-2}$	$6.62{\cdot}10^{-4}$
4	3	$-7.95{\cdot}10^{-4}$	$1.19{\cdot}10^{-3}$
4	4	$2.45{\cdot}10^{-3}$	$-1.87{\cdot}10^{-3}$

these gravitational models, also the gravitational parameter μ is required (it is used as a scaling factor). This parameter is dependent on the assumed average density of the asteroids. The actual value of μ will *not* change the spherical-harmonics coefficients. Currently calculated values for μ are $7.70 \cdot 10^{-6}$ km^3/s^2 (with an assumed average density of $\rho = 1.5$ g/cm^3) for Šteins, $\mu = 0.310$ km^3/s^2 ($\rho = 6.55$ g/cm^3) for Kleopatra and $\mu = 2.34{\cdot}10^{-9}$ km$^3/s^2$ ($\rho = 1.98$ g/cm^3) for Itokawa. The respective volumes are 77, 708,868 and 0.0178 km^3.

The assumption of a coinciding centre-of-mass location with the origin of the reference frame can, of course, always be done when one (artificially) generates

Table 2.9 Normalised spherical harmonics coefficients of Itokawa. The reference radius is 313 m

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
2	0	$-3.59{\cdot}10^{-2}$	0.00
2	1	0.00	0.00
2	2	$6.03{\cdot}10^{-2}$	0.00
3	0	$3.87{\cdot}10^{-3}$	0.00
3	1	$4.38{\cdot}10^{-3}$	$2.06{\cdot}10^{-3}$
3	2	$-4.60{\cdot}10^{-3}$	$-2.19{\cdot}10^{-3}$
3	3	$-5.67{\cdot}10^{-3}$	$-8.13{\cdot}10^{-3}$
4	0	$5.84{\cdot}10^{-3}$	0.00
4	1	$-1.04{\cdot}10^{-3}$	$-4.28{\cdot}10^{-4}$
4	2	$-8.35{\cdot}10^{-3}$	$-1.31{\cdot}10^{-4}$
4	3	$5.54{\cdot}10^{-4}$	$1.21{\cdot}10^{-3}$
4	4	$1.05{\cdot}10^{-2}$	$-2.63{\cdot}10^{-6}$

the gravity-field model. For consistency reasons, when a model has been generated using a shape model and there is an option to compare it with a measurements-derived model this assumption may not always hold. A typical example is the model for asteroid Eros-433. The spherical-harmonics coefficients are available in two sets: one derived from radio-science data (the so-called NEAR15A gravity model), where besides the coefficients also mass, density, inertia tensor, and the location of the centre of mass have been estimated, and one that has been calculated using the shape model. The measurement-based model has thus a zero offset of the centre-of-mass location, implying that $\bar{C}_{1,0}$, $\bar{C}_{1,1}$, and $\bar{S}_{1,1}$ are all zero. The shape-derived model, on the other hand, assumes an average, yet constant density, so its centre of mass has a slightly different location. For that reason, the aforementioned coefficients are non-zero. For completeness, the shape-derived spherical harmonics model is listed in Table 2.10 the corresponding centre-of-mass offset is $\Delta\mathbf{r}_{cm} = (-9.656, 2.429, 32.584)^{\mathrm{T}}$ m.

2.3 Atmospheres

The atmosphere plays an important role in the design of entry vehicles. It is responsible for reducing the entry vehicle to acceptable landing speeds by converting kinetic energy to heat due to friction. At the same time, this heat poses a serious problem for entry-vehicle designers who have to come up with smart solutions for thermal protection. Due to the severe heating environment, the thermal-protection system (TPS) is loaded at its limit and an accurate knowledge about the interaction of the vehicle with the atmosphere is required to support the design process. In this section, the atmospheres of the bodies in the Solar System are discussed, together with some commonly used models for the analysis of re-entry vehicles.

Table 2.10 Normalised spherical harmonics coefficients of Eros-433, generated from the shape model. The reference radius is 16.0 km. https://sbn.psi.edu/pds/resource/nearbrowse.html, date accessed: 25 September, 2023. Note that the given accuracy is the same as provided by NASA's planetary data system

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
0	0	1.000000000000	0.000000000000
1	0	$1.175785831520 \cdot 10^{-03}$	0.000000000000
1	1	$-3.484427594460 \cdot 10^{-04}$	$8.766452698130 \cdot 10^{-05}$
2	0	$-5.285148878740 \cdot 10^{-02}$	0.000000000000
2	1	$1.021293512930 \cdot 10^{-04}$	$8.314827416250 \cdot 10^{-02}$
2	2	$1.171641181310 \cdot 10^{-05}$	$-2.819769459150 \cdot 10^{-02}$
3	0	$-1.746658679040 \cdot 10^{-03}$	0.000000000000
3	1	$4.086789748400 \cdot 10^{-03}$	$-8.401928754980 \cdot 10^{-04}$
3	2	$3.400820184730 \cdot 10^{-03}$	$-1.049252521580 \cdot 10^{-02}$
3	3	$2.127432677370 \cdot 10^{-03}$	$-1.221642051680 \cdot 10^{-02}$
4	0	$1.307711276510 \cdot 10^{-02}$	0.000000000000
4	1	$-1.449369221770 \cdot 10^{-04}$	$-3.130233427200 \cdot 10^{-04}$
4	2	$1.647971980510 \cdot 10^{-04}$	$-1.946510011750 \cdot 10^{-04}$
4	3	$-1.764730392110 \cdot 10^{-02}$	$1.769372863880 \cdot 10^{-02}$
4	4	$4.623965128100 \cdot 10^{-03}$	$-9.118275274890 \cdot 10^{-03}$
5	0	$8.040269731590 \cdot 10^{-04}$	0.000000000000
5	1	$-2.791765238870 \cdot 10^{-03}$	$3.601526439980 \cdot 10^{-03}$
5	2	$-1.230910225660 \cdot 10^{-03}$	$5.901588590910 \cdot 10^{-04}$
5	3	$-1.003790649660 \cdot 10^{-03}$	$-7.741235325160 \cdot 10^{-04}$
5	4	$4.497119857680 \cdot 10^{-04}$	$-1.034639930470 \cdot 10^{-02}$
5	5	$4.645819165710 \cdot 10^{-03}$	$-5.986302199180 \cdot 10^{-03}$
6	0	$-4.958280336910 \cdot 10^{-03}$	0.000000000000
6	1	$-5.848388650290 \cdot 10^{-05}$	$-5.656243124620 \cdot 10^{-03}$
6	2	$-8.779840037550 \cdot 10^{-05}$	$1.798612884550 \cdot 10^{-03}$
6	3	$6.600756525270 \cdot 10^{-03}$	$-5.628527194580 \cdot 10^{-04}$
6	4	$-1.188912547480 \cdot 10^{-03}$	$-1.202848311210 \cdot 10^{-04}$
6	5	$2.893725631600 \cdot 10^{-04}$	$5.187143981030 \cdot 10^{-03}$
6	6	$1.860196354340 \cdot 10^{-04}$	$-1.562296163960 \cdot 10^{-03}$
7	0	$3.923399201210 \cdot 10^{-04}$	0.000000000000
7	1	$1.811938547760 \cdot 10^{-03}$	$4.075777475970 \cdot 10^{-04}$
7	2	$5.261744188700 \cdot 10^{-04}$	$3.843931902200 \cdot 10^{-03}$
7	3	$4.969228567990 \cdot 10^{-04}$	$2.035935441050 \cdot 10^{-03}$
7	4	$-2.347349608900 \cdot 10^{-04}$	$1.502211084770 \cdot 10^{-04}$
7	5	$-2.476609205240 \cdot 10^{-03}$	$-4.499749592270 \cdot 10^{-04}$
7	6	$-1.458303912990 \cdot 10^{-03}$	$-6.790061431970 \cdot 10^{-03}$
7	7	$-3.417036641460 \cdot 10^{-04}$	$-2.244364637360 \cdot 10^{-03}$

(continued)

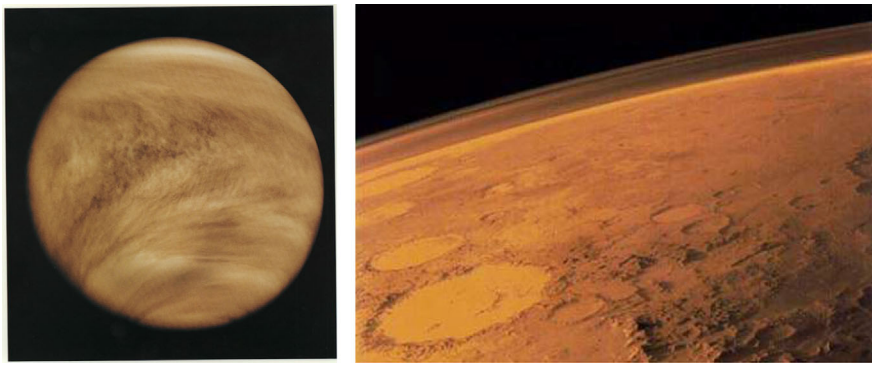
Table 2.10 (continued)

Degree	Order	$\bar{C}_{n,m}$	$\bar{S}_{n,m}$
8	0	$2.332698296190 \cdot 10^{-03}$	0.000000000000
8	1	$1.176328666590 \cdot 10^{-04}$	$3.886108808310 \cdot 10^{-04}$
8	2	$3.362497962950 \cdot 10^{-05}$	$3.491867096780 \cdot 10^{-05}$
8	3	$-3.122792657720 \cdot 10^{-03}$	$-2.236538379940 \cdot 10^{-03}$
8	4	$3.431369303920 \cdot 10^{-04}$	$4.701611646030 \cdot 10^{-06}$
8	5	$-2.243887071000 \cdot 10^{-04}$	$-5.127049850290 \cdot 10^{-04}$
8	6	$-7.063776582180 \cdot 10^{-05}$	$7.874768865870 \cdot 10^{-05}$
8	7	$2.686346867920 \cdot 10^{-03}$	$2.381956825870 \cdot 10^{-03}$
8	8	$-4.162893880770 \cdot 10^{-04}$	$1.330995240810 \cdot 10^{-03}$

An atmosphere is a layer of gases that may surround a material body of sufficient mass (a planet or moon), held in place by the gravity of this body. They are formed by heating, outgassing (releasing), and reprocessing volatiles (volatiles are substances that readily vaporise at relatively low temperature). For all planets, the escape of some gases and the retention of others due to temperature and surface gravity played an important role in how their atmospheres evolved. Several planets and moons in the Solar System have a substantial atmosphere. The atmospheres of the planets Venus and Mars are primarily composed of carbon dioxide, with small quantities of nitrogen, argon, oxygen and traces of other gases (Fig. 2.24). The climate on Venus is widely known to be unpleasant at the surface, the planet roasts at temperatures close to 500°C under a suffocating blanket of sulphuric acid clouds and a crushing atmosphere more than 90 times the pressure of Earth's. The Martian atmosphere is much thinner, having a much lower pressure and temperature compared to Earth. Despite the occasional, and sometimes planet-wide dust storms, the conditions are such that life could be supported, and looking at Fig. 2.25 one might forget not being on Earth.

The atmospheric composition on Earth is largely governed by the by-products of the very life that it sustains. Earth's atmosphere contains roughly (by molar content/volume) 78.09% nitrogen, 20.95% oxygen, a variable amount (average around 0.247%) water vapour, 0.93% argon, 0.038% carbon dioxide, and traces of hydrogen, helium, and other "noble" gases. Even though our atmosphere may seem turbulent at times (Fig. 2.26), compared to most atmospheres in the Solar System it is quite "peaceful".

The low temperatures and higher gravity of the gas giants, i.e., Jupiter, Saturn, Uranus, and Neptune, allows them to more readily retain gases with low molecular masses. These planets have hydrogen-helium atmospheres, with trace amounts of more complex compounds. The atmospheres are known for their highly active phenomena. Jupiter, for instance, shows a wide range of these phenomena, including band instabilities, vortices (cyclones and anticyclones), storms and lightning. Famous, of course, is the Great Red Spot, a vortex large enough to contain two or



(a) Venus' thick atmosphere

(b) The (very thin) Martian atmosphere as seen by NASA's Mars Reconnaissance Orbiter.

Fig. 2.24 Atmospheres of Venus and Mars. Images credit: NASA

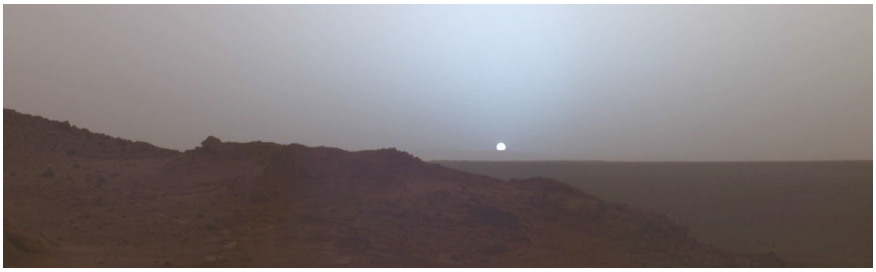
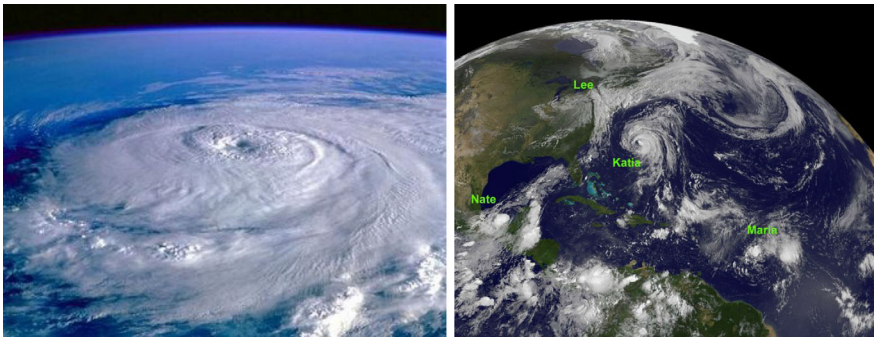


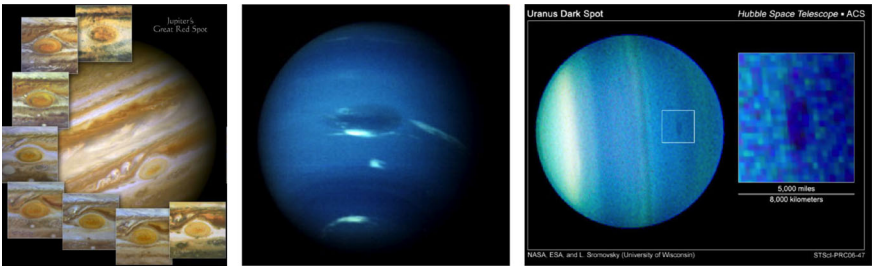
Fig. 2.25 Photograph of Martian Sunset taken from the Gusev crater, taken by NASA's Mars Exploration Rover Spirit on 19 May 2005. Image: NASA



(a) Earth's cloud system from space. Image: NASA.

(b) GOES-13 satellite image of four tropical systems in the Atlantic on Sept. 8, 2011. Image: NASA/NOAA GOES Project.

Fig. 2.26 The Earth's atmosphere, rich with cloud systems and turbulent phenomena

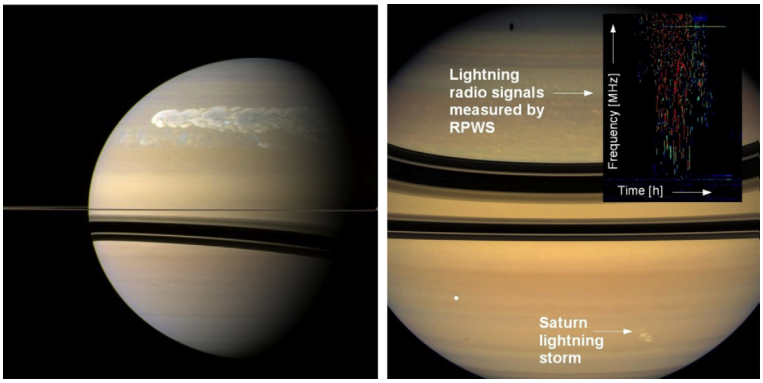


(a) Jupiter with the famous Great Red Spot.

(b) Neptune, with its Great Dark Spot.

(c) Uranus, with its Dark Spot.

Fig. 2.27 The gas giants Jupiter, Neptune and Uranus. Images: NASA



(a) Storm in the atmosphere of Saturn.

(b) Saturn lightning storm.

Fig. 2.28 Saturn. Images: NASA

three planets the size of Earth (Fig. 2.27a). Similarly, Neptune has a Great Dark Spot (Fig. 2.27b). These weather patterns are driven by the strongest sustained winds of any planet in the Solar System, with recorded wind speeds as high as 2,100 km/h. Initially, Uranus appeared to be quite featureless. However, it has a similar cloud pattern as Neptune, and the Hubble Space Telescope observed the first Dark Spot back in 2006 (de Pater and Littauer 2010), see Fig. 2.27c. Also Saturn has similar atmospheric features as the other three. Closely observed by the Cassini spacecraft, on February 25, 2011, an approximately true-colour view of Saturn with an enormous storm was captured (Fig. 2.28). The northern storm is obvious, but below it there is a subtler band of whirling clouds.

Two satellites of the outer planets possess non-negligible atmospheres: Titan (Fig. 2.29), a moon of Saturn, and Triton, a moon of Neptune, which contain mainly nitrogen. Pluto, in the nearer part of its orbit, has an atmosphere of nitrogen and methane similar to Triton's, but these gases are frozen when farther from the Sun. Titan served as the scene for the ESA Huygens probe's entry, descent and landing,

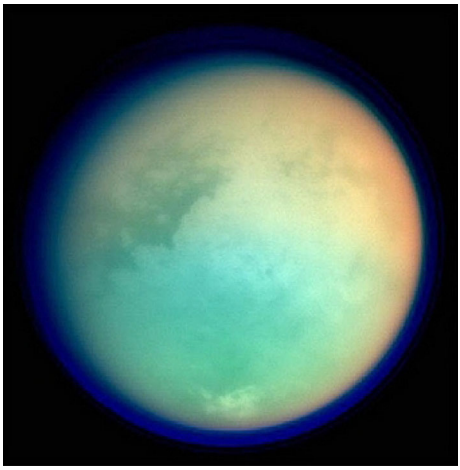


Fig. 2.29 Titan, with its thick nitrogen atmosphere. Image: NASA

and is—apart from our own moon—the first moon visited by a robotic lander. Other bodies within the Solar System have extremely thin atmospheres not in equilibrium. These include the Moon (sodium gas), Mercury (sodium gas), Europa (oxygen), Io (sulfur), and Enceladus (water vapour) (de Pater and Littauer 2010). The atmospheric composition of an extra-solar planet was first determined using the Hubble Space Telescope. Planet HD 209458b is a gas giant with a close orbit around a star in the constellation Pegasus. The atmosphere is heated to temperatures over 1,000 K, and is steadily escaping into space. Hydrogen, oxygen, carbon and sulfur have been detected in the planet's inflated atmosphere.

To compute the aerodynamic forces and moments, we need the atmospheric density as a function of altitude. Also temperature and pressure are needed to compute, for instance, Mach number and Reynolds number. The atmosphere is not a static medium, primarily due to solar radiation, and it is partly for this reason that it is impossible to develop a complete model, which gives the proper atmospheric conditions at every location for each time of the year. Therefore, recognise that no matter how complicated and extended the atmosphere model is, it will always be an approximation of the real atmosphere, especially when the model is based on a limited amount of experimental data. In Table 2.11 some basic properties of planetary atmospheres are listed.

Table 2.11 Properties of some planetary atmospheres (de Pater and Littauer 2010)

Planet	Composition (mole fractions)	Average surface temperature (K)	Surface pressure (bar)
Earth	78.08% N_2, 20.95% O_2	288	1.013
Moon	He, Ar, Na, K	277	$3{\cdot}10^{-15}$
Venus	0.035% N_2, 96.5% CO_2	737	92
Mars	0.027% N_2, 95.3% CO_2, 0.016% Ar	215	0.00636
Jupiter	86.4% H_2, 13.6% He	165[a]	1.013[a]
Saturn	88% H_2, 11.9% He	134.8[a]	1.013[a]
Titan	95% N_2, 0.014-0.049% CH_4, 0.004% H_2	93.7	1.47

[a] The surface for the gas planets is defined as the altitude where the pressure is 1.013 bar

2.3.1 Exponential Model

To accurately describe the atmosphere with all its (chaotic) phenomena is a complex (and probably impossible) task, and would not help in understanding the fundamentals of re-entry mechanics. Even though more complicated models when (numerically) simulating the flight of re-entry vehicles are used, for an analytical approach one would need a simplified atmosphere model. The first simplification that we make is to consider the atmosphere to be an *ideal gas*. An ideal gas is a theoretical gas composed of a set of randomly-moving, non-interacting point particles. The ideal gas concept is useful, because it obeys the ideal gas law, a simplified equation of state.[11] The ideal gas law is given by

$$p = \rho RT = \rho \frac{R^*}{M} T \tag{2.52}$$

with

$R = \frac{R^*}{M}$ = gas constant for an ideal gas in J/(kg K)
R^* = universal gas constant = 8314.32 J/(kmol K)
M = molecular mass of the atmosphere under consideration (kg/kmol)

This equation relates the atmospheric pressure, p, with the atmospheric density, ρ, and temperature, T, through a proportionality constant, i.e., the gas constant R. In this simple model, R is assumed to be constant, e.g., $R = 287$ J/(kg K) for air, although in reality M is varying with h. For Earth and the simple exponential model, the molecular weight of the mixture of atmospheric gases can be considered to be constant up to 100 km and is equal to $M = 28.9644$ kg/kg-mol.

To find a relation for the density as a function of altitude, assume that the atmosphere is in hydrostatic equilibrium. The hydrostatic equation for the atmosphere is

[11] An equation of state is a thermodynamic equation describing the state of matter under a given set of physical conditions.

$$\mathrm{d}p = -\rho g \mathrm{d}h \tag{2.53}$$

Integrating Eq. (2.53) requires knowledge of both ρ and g as a function of h. The latter problem can be avoided by introducing the geopotential altitude, z, which is related to the geometric altitude h by the relation:

$$g_0 \mathrm{d}z = g \mathrm{d}h \tag{2.54}$$

with g_0 is the gravitational acceleration at the surface of the Earth, $h = 0$. For a spheroid, Newton's law of gravitation may be written as

$$g = \frac{g_0 R_e^2}{(R_e + h)^2} \Rightarrow g = \frac{g_0}{\left(1 + \frac{h}{R_e}\right)^2} \tag{2.55}$$

The geopotential altitude is derived by substituting Eq. (2.55) in Eq. (2.54):

$$g_0 \mathrm{d}z = g \mathrm{d}h \Rightarrow g_0 \mathrm{d}z = \frac{g_0}{\left(1 + \frac{h}{R_e}\right)^2} \mathrm{d}h \tag{2.56}$$

Integrating both sides leads to

$$\int_0^z \mathrm{d}z^* = \int_0^h \frac{1}{\left(1 + \frac{h^*}{R_e}\right)^2} \mathrm{d}h^* \Rightarrow \tag{2.57}$$

$$\frac{z}{R_e} = \int_0^h \frac{R_e}{(R_e + h^*)^2} \mathrm{d}h^* \Rightarrow \tag{2.58}$$

$$\frac{z}{R_e} = 1 - \frac{R_e}{R_e + h} = \frac{h}{R_e + h} \Rightarrow \tag{2.59}$$

$$\frac{z}{R_e} = \frac{h}{R_e}\left(\frac{1}{1 + \frac{h}{R_e}}\right) \approx \frac{h}{R_e}\left(1 - \frac{h}{R_e}\right) \tag{2.60}$$

For low Earth orbits (LEO) the ratio $h/R_e \approx 0.06$, which means that $z \approx h$ within 6.3% ($h = 400\,\mathrm{km}$).

Using the state and hydrostatic equilibrium equations yields

$$\frac{\mathrm{d}p}{p} = -\frac{g}{RT}\mathrm{d}h = -\frac{g_0}{RT}\mathrm{d}z \tag{2.61}$$

Suppose we assume $T =$ constant (isothermal atmosphere) and $g = g_0$:

$$\mathrm{d}p = -\rho g \mathrm{d}h = RT\mathrm{d}\rho \Rightarrow \frac{\mathrm{d}\rho}{\rho} = -\frac{g_0}{RT}\mathrm{d}h \tag{2.62}$$

Integrating the above equation yields

$$\ln\frac{\rho}{\rho_0} = -\frac{g_0}{RT}h \tag{2.63}$$

or, equivalently

$$\frac{\rho}{\rho_0} = e^{-\beta h} = e^{-\frac{h}{H_s}} \tag{2.64}$$

The above model is, for obvious reasons, known as the *exponential atmosphere*. ρ_0 is usually the density at sea level, but could in principle have any value to improve the fit of the model. $H_s = \frac{1}{\beta} = \frac{RT}{g_0}$ is called the scale height. For Earth, commonly used values are $\rho_0 = 1.225\,\mathrm{kg/m^3}$ and H_s in the range of 7,050 to 7,200 m. Given the lower value, this means that the (constant) temperature for this model is $T = 240\,\mathrm{K}$ (for a scale height of $H_s = 7{,}200\,\mathrm{m}$ this temperature would be $T = 246\,\mathrm{K}$). Furthermore, since the speed of sound is temperature dependent, this temperature value leads to a speed of sound of $a = 310.5\,\mathrm{m/s}$. Note that the scale height is the altitude range for which pressure and density change with a factor of e (≈ 2.718), also known as Euler's number.

Below is an example of an exponential atmosphere for the Earth (Example 2.2). This model is subsequently used in Chap. 5, where analytical expressions are derived for ballistic, gliding and skipping flight, and where some numerical examples are provided. However, an exponential model can in principle be made for any atmosphere, such that the mentioned analytical solutions can also be applied to missions on other planets. Section 2.3.3, Example 2.3 provides an exponential model for the Martian atmosphere, whereas in Sect. 2.3.4, Example 2.4, the more non-linear density profile of the Titan atmosphere is approximated by two profiles. The same approach is done for the Venusian atmosphere, see Sect. 2.3.5.

Example 2.2 (*Earth Exponential Atmosphere*) To obtain an exponential atmosphere model from a more complete model, the first step would be to obtain a table of the atmospheric density as a function of altitude. Once that is available one can fit an exponential model through these data and minimise the difference between the two models, by varying the two constants ρ_0 and H_s. For Earth entry study, it is common practise to use the United States Standard Atmosphere released in 1976, US76 (see also the upcoming Sect. 2.3.2.2). In this example the exponential atmosphere is compared with tabulated data obtained with that atmosphere model.

The exponential model will be evaluated with a scale height $H_s = 7{,}050\,\mathrm{m}$ and a density at sea level of $\rho_0 = 1.225\,\mathrm{kg/m^3}$. The corresponding temperature for this model is $T = 240\,\mathrm{K}$, as shown in Fig. 2.30, and compared with the temperature profile obtained with the US76 model. The atmospheric density for the two models is shown

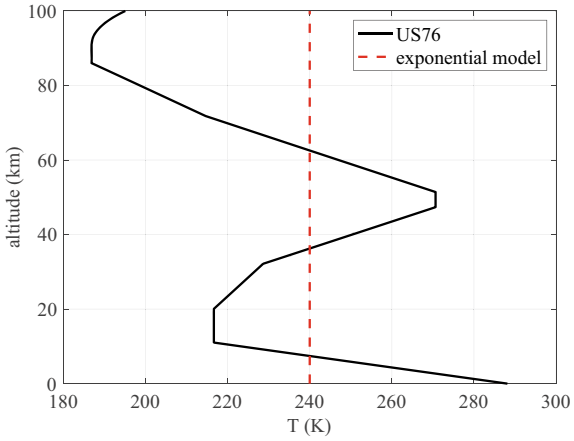


Fig. 2.30 Temperature profile according to the exponential atmosphere model ($H_s = 7{,}050\,\text{m}$, ρ_0 $1.225\,\text{kg/m}^3$) and the United States Standard Atmosphere (1976)

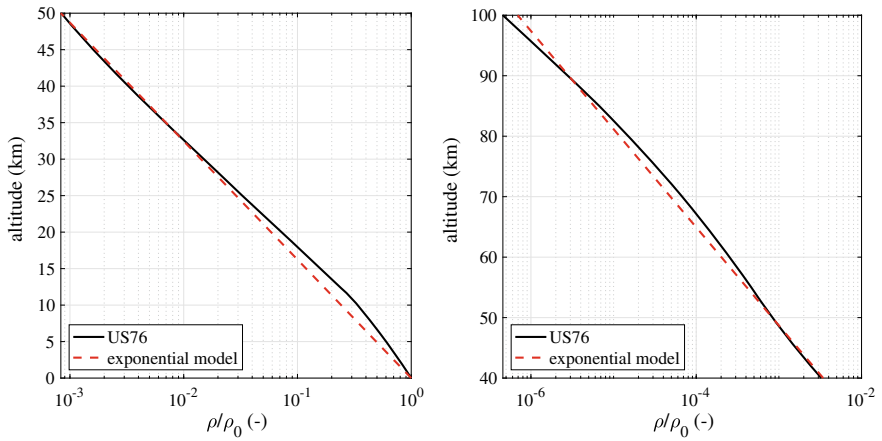


Fig. 2.31 Density profile according to the exponential atmosphere model ($H_s = 7{,}050\,\text{m}$, ρ_0 $1.225\,\text{kg/m}^3$) and the United States Standard Atmosphere (1976)

in Fig. 2.31. On a logarithmic scale, the density profile of the exponential atmosphere is of course linear. It is clear that the corresponding US76 profile is not so linear, and some discrepancies can be discerned. Depending on the application, one can either accept or reject that. In Chap. 5, where analytical solutions of the equations of motion are derived it can be accepted, but one should always be aware of the error that is made due to this assumption. ■

To conclude the discussion on exponential atmospheres, in Table 2.12, the values for scale height, and surface density, pressure and temperature for different gas giants are listed. These data will be sufficient to model the upper layers of the atmosphere and

Table 2.12 Scale heights for atmospheres of the gas giants (Data taken from https://nssdc.gsfc.nasa.gov/planetary/factsheet/index.html; last accessed April 12, 2020)

Planet	Scale height (km)	"Surface" density at 1 bar (kg/m^3)	"Surface" pressure (bar)	Avg. temperature at 1 bar (K)
Jupiter	27	0.16	$\gg$1000	165
Saturn	59.5	0.19	$\gg$1000	134
Uranus	27.7	0.42	$\gg$1000	76
Neptune	20.3	0.45	$\gg$1000	72

to perform a first assessment of, for instance, aero-gravity assists. For more detailed studies, the models should either be extended or sufficient uncertainty margins should be taken into account.

2.3.2 Earth

For the Earth many different atmosphere models have been developed over the years, as more data from atmospheric observations became available. Concerning the models there is a difference between a *standard* and a *reference* atmosphere. A standard atmosphere is a hypothetical vertical distribution of atmospheric properties (amongst others, temperature, pressure and density), which, by international agreement, is roughly representative of year-round, mid-latitude conditions. As these properties are average values, there is no temporal nor spatial (other than altitude) influence. This means that for different epochs and entry location, numerical simulations will always give the same result. Typical examples include the series of United States Standard Atmosphere with a first release in 1962, and updates in 1966 and 1976.

A reference atmosphere, on the other hand, includes latitudinal, seasonal, geomagnetic, and solar effects, to name a few, and does have a temporal and spatial influence. Typical examples are the Jacchia model for the upper layers of the atmosphere (90 to 2500 km) and the NASA Earth Global Reference Atmosphere Model (Earth-GRAM). The Earth-GRAM has been developed to have a model that provides complete global geographical variability, complete altitude coverage (surface to orbital altitudes), and complete seasonal and monthly variability of the thermodynamic variables and wind components. The development started in 1974, with updates in 1980, 1990, 1995, 1999, 2007, and the most recent update in 2010 (Leslie and Justus 2011). The Earth-GRAM2010 consists of a mixture of empirically-based models that represent different altitude ranges and the geographical and temporal variations within these altitude ranges. Figure 2.32 shows how GRAM divides the atmosphere into three regions from which mean values of atmospheric parameters are provided. If dispersions are required, one can use a perturbation model to compute variations about these means. Monthly mean values and standard deviations in the lower atmosphere come from the database of the National Centers for Environ-

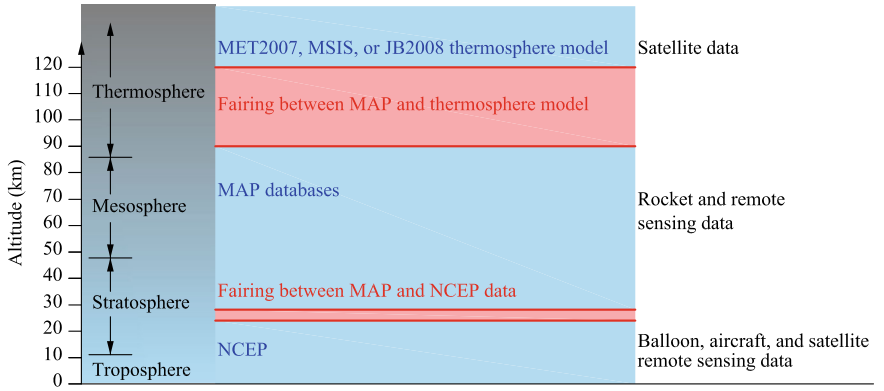


Fig. 2.32 Summary of the atmospheric regions in Earth-GRAM2010, sources for the model, and data on which the mean monthly values are based (Leslie and Justus 2011)

mental Prediction (NCEP). More details about the specific models, databases and use of the atmosphere model can be found in Leslie and Justus (2011) and the quoted references therein.

Current state-of-the-art for high-altitude atmosphere models is the so-called NLRMSISE-00 model, which has the added advantage of providing low-altitude temperature, pressure, and density profiles as well. It is the model of choice for studying orbital decay of (active and/or passive) satellites in low Earth orbits. Therefore, it can provide relatively accurate predictions for the initial conditions of space debris that will enter the Earth's atmosphere and either burn up in the atmosphere or impact on the surface.

Three different atmosphere models that are typically used in entry studies will be discussed. The first two belong together, the US Standard Atmospheres of 1962 and 1976, the latter having been developed from the former. The US Standard Atmosphere 1976 is commonly used for entry simulations. The third model discussed is the NLRMSISE-00 model that provides more accurate high-altitude data.

Wind is not always part of Earth atmosphere models, a well-known exception being the GRAM models, and should be included separately, when needed. A typical wind model may include predictions of the steady-state wind, horizontal and vertical wind gusts, and turbulence. It goes beyond the scope of this chapter to give an extensive treatment about the different aspects of wind, and suffice by giving a reference for one of the most commonly used steady-state wind models. This horizontal wind model, released in 2007 and known as HWM07, is an empirical model that describes a global, horizontal steady-state wind in the terrestrial atmosphere. It represents a trade-off between look-up tables and analytical models, and it integrates data coming from the previous versions (HWM87-HWM93) with new measurements sets, obtained with satellite-based observations. Details about the modelling and validation are provided by Drob et al. (2008). Figure 2.33 shows an example of the zonal and meridional wind velocities as a function of altitude.

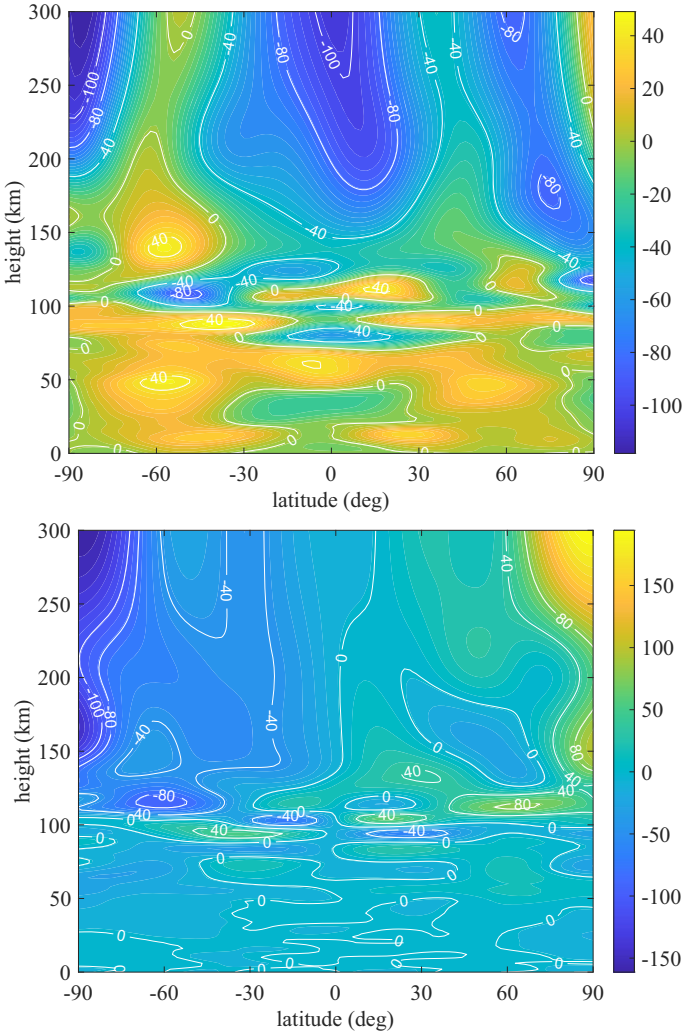


Fig. 2.33 Zonal (top) and meridional (bottom) wind maps computed with HWM07

2.3.2.1 The United States Standard Atmosphere 1962

The US Standard Atmosphere 1962 (US62) was developed to take into account increased knowledge and more accurate determinations of basic quantities, and to update the existing atmosphere models of that time. It is an idealised, steady-state representation of the Earth's atmosphere. For all practical purposes, US62 is in agreement with the ICAO standard over their common altitude range. This means that US62 is equal to the ICAO standard atmosphere to 20 km and the proposed ICAO

extension to 32 km. In the original publication (COESA 1962), tables and data are given for altitudes up to 700 km.

In this section, an analytic model of property variations with altitude to interpolate between measurements is presented. The tabulated measurements are taken from COESA (1962). The basis for the derivation is the hydrostatic equation, relating differential pressure $\mathrm{d}p$ to differential (geometric) altitude $\mathrm{d}h$, see Eq. (2.53).

Two additional equations to solve the dependency of ρ and g on h are needed. The first equation, assuming that the air is an ideal gas, is the ideal gas law, Eq. (2.52). By using the ideal gas law, two new variables, which are depending on altitude, were introduced: the kinetic temperature, T, and the molecular mass, M. For the sake of derivation, this dependency is combined in one variable, the molecular scale temperature T_M:

$$T_M = T\frac{M_0}{M} \tag{2.65}$$

with $M_0 =$ molecular mass of air at sea level $= 28.9644\,\mathrm{kg/kmol}$.

If we divide the total altitude range into a number of intervals, it appears that T_M varies at a linear rate (or is constant) in an altitude interval. So

$$T_M = T_{M_i} + L_{h_i}(h - h_i) \tag{2.66}$$

with the index i denoting the ith interval and T_{M_i} the molecular scale temperature at the lower boundary h_i of this interval. L_{h_i} is the thermal lapse rate and is constant per interval. If T_{M_i} is given as a function of h_i, L_{h_i} can be computed with

$$L_{h_i} = \frac{T_{M_{i+1}} - T_{M_i}}{h_{i+1} - h_i} \tag{2.67}$$

The second equation, to solve the dependency of g on h, assumes that the gravity field of the Earth is spherical, giving

$$g = \frac{\mu}{R^2} = \frac{g_0 R_e^2}{(R_e + h)^2} \tag{2.68}$$

g_0 in US62 is computed for $\delta = 45°$ and is taken as $g_0 = 9.80665\,\mathrm{m/s^2}$. Equation (2.68) will be approximated by a Taylor series. Retaining only the first two terms yields

$$g = g_0\left(\frac{R_e}{R_e + h}\right)^2 = g_0(1 - bh) \tag{2.69}$$

with $b = \frac{2}{R_e} = 3.139 \cdot 10^{-7}\ \mathrm{m^{-1}}$.

Based on the assumptions made, the following equations can be derived:

$$\boxed{L_{h_i} = 0}$$

$$p = p_i \exp\left[-\left\{\left(\frac{g_0(h-h_i)}{RT_{M_i}}\right)\left(1-\frac{b}{2}(h+h_i)\right)\right\}\right] \tag{2.70}$$

$$\rho = \rho_i \exp\left[-\left\{\left(\frac{g_0(h-h_i)}{RT_{M_i}}\right)\left(1-\frac{b}{2}(h+h_i)\right)\right\}\right] \tag{2.71}$$

$$T_M = T_{M_i} \tag{2.72}$$

$$T = T_M \frac{M_i}{M_0} \tag{2.73}$$

$\boxed{L_{h_i} \neq 0}$

$$p = p_i \left[\frac{L_{h_i}}{T_{M_i}}(h-h_i)+1\right]^{-\left\{\frac{g_0}{RL_{h_i}}\left[1+b\left(\frac{T_{M_i}}{L_{h_i}}-h_i\right)\right]\right\}} \cdot \exp\left[\frac{g_0 b}{RL_{h_i}}(h-h_i)\right] \tag{2.74}$$

$$\rho = \rho_i \left[\frac{L_{h_i}}{T_{M_i}}(h-h_i)+1\right]^{-\left\{\frac{g_0}{RL_{h_i}}\left[\frac{RL_{h_i}}{g_0}+1+b\left(\frac{T_{M_i}}{L_{h_i}}-h_i\right)\right]\right\}} \cdot \exp\left[\frac{g_0 b}{RL_{h_i}}(h-h_i)\right] \tag{2.75}$$

$$T_M = T_{M_i} + L_{h_i}(h-h_i) \tag{2.76}$$

$$M = M_i + \frac{M_{i+1}-M_i}{h_{i+1}-h_i}(h-h_i) \tag{2.77}$$

$$T = T_M \frac{M}{M_0} \tag{2.78}$$

In re-entry studies typically two similarity parameters are used, namely the Reynolds number Re and the Mach number M. They are defined as

$$Re = \frac{V L_{ref}}{\nu} \tag{2.79}$$

$$M = \frac{V}{a} \tag{2.80}$$

with

V = relative flow velocity (m/s)
L_{ref} = reference length (m)
ν = kinematic viscosity (m^2/s)
a = local speed of sound (m/s).

The kinematic viscosity is computed by dividing the dynamic viscosity, μ (in kg per m/s), by the atmospheric density:

$$\nu = \frac{\mu}{\rho} \tag{2.81}$$

For an ideal gas, μ is only dependent on temperature. Sutherland's formula gives accurate results for a large range of temperatures:

$$\mu = \mu_0 \left(\frac{T}{T_0}\right)^{1.5} \frac{T_0 + S}{T + S} = \beta \frac{T^{1.5}}{T + S} \tag{2.82}$$

with $\beta = \mu_0 \frac{T_0 + S}{T_0^{1.5}}$, and the index "0" denoting sea level. Sutherland's constant, S, and β are dependent on atmospheric composition and condition. For US62, the values read:

$\mu_0 = 1.7894 \cdot 10^{-5}$ kg/m/s
$T_0 = 288.15$ K
$S = 110.4$ K
$\beta = 1.458 \cdot 10^{-6}$ kg/m/s

The speed of sound is given by

$$a = \sqrt{\gamma R T} \tag{2.83}$$

with

$\gamma = \frac{c_p}{c_v}$ = the ratio of specific heats (–)
c_p = specific heat at constant pressure (J/kg K)
c_v = specific heat at constant volume (J/kg K)
$R = c_p - c_v$ = gas constant (J/kg K)
T = kinetic temperature (K)

In this model, an approximation will be used: γ and R are kept constant with the following numerical values, $\gamma = 1.4$ and $R = 287$ J/kg K.

2.3.2.2 The United States Standard Atmosphere 1976

The US Standard Atmosphere 1976 (US76) is a revised version of US62 (NOAA et al. 1976). For altitudes below 50 km it is identical with US62, and even for altitudes up to 86 km the same equations as described in the previous section can be used. The molecular mass is assumed to be constant up to 80 km with a value of $M = M_0 = 28.9644$ kg/kmole. However, US76 has a slightly different formulation, using the geopotential altitude z instead of the geometric altitude h. The relation between the two can be derived from Eq. (2.54):

$$g_0 \mathrm{d}z = g \mathrm{d}h \;\Rightarrow\; z = \int_0^h \frac{g}{g_0} \mathrm{d}h \;\approx\; \frac{R_0 h}{R_0 + h} \tag{2.84}$$

Here, the inverse squared field approximation for the gravity field was used, $g = g_0 \left(\frac{R_0}{R_0+h}\right)^2$, with $R_0 = 6356.766\,\text{km}$, the Earth's radius at the latitude, where the acceleration of gravity, g_0, is $9.80665\,\text{m/s}^2$.

For the sake of completeness, the revised equations are given.

$$T_M = T_{M_i} + L_{z_i}(z - z_i) \tag{2.85}$$

$$T = T_M \frac{M}{M_0} \tag{2.86}$$

and

$$p = p_i \exp\left[\frac{-g_0 M_0 (z - z_i)}{R^* T_{M_i}}\right] \tag{2.87}$$

for layers where $L_{z_i} = 0$ or

$$p = p_i \left[\frac{T_{M_i}}{T_{M_i} + L_{z_i}(z - z_i)}\right]^{K_i} \tag{2.88}$$

with $K_i = \frac{g_0 M_0}{R^* L_{z_i}}$ for layers where $L_{z_i} \neq 0$. The density ρ is computed with the ideal gas law, Eq. (2.52).

For $z \leq 80$ km, $M = M_0$, so $T = T_M$. From 80 to 86 km, a small correction for M is computed on basis of tabulated data. At altitudes from 86 to 120 km, the temperature-altitude profile is defined in terms of three successive functions. They are designed such, to represent the following conditions:

1. an isothermal layer from 86 to 91 km,
2. a layer in which $T = T(h)$ has the form of an ellipse from 91 to 110 km, and
3. a constant, positive-gradient layer from 110 to 120 km.

So, the temperature profiles for these layers are:

$\boxed{86\,\text{km} \leq h < 91\text{km}}$

$$T = 186.8673\,K = \text{constant} \tag{2.89}$$

$\boxed{91\,\text{km} \leq h < 110\,\text{km}}$

$$T = T_c + A\sqrt{1 - \left(\frac{h - h_8}{b}\right)^2} \tag{2.90}$$

where

$T_c = 263.1905\,\text{K}$
$A = -76.3232\,\text{K}$
$b = 19.9429\,\text{km}$

$h_8 = 91\,\text{km}$.

$$\boxed{110\,\text{km} \leq h \leq 120\text{km}}$$

$$T = T_9 + L_9(h - h_9) \tag{2.91}$$

where

$T_9 = 240\,\text{K}$
$L_9 = 12\,\text{K/km}$
$h_9 = 110\,\text{km}$.

The atmospheric pressure is computed with an alternative expression of the ideal gas law:

$$p = \frac{N R^* T}{N_A} \tag{2.92}$$

with N = total number density (m^{-3}), N_A = Avogadro's constant = $6.0225{\cdot}10^{26}$ kmol^{-1}.

The total number density is computed on basis of linear interpolation in a table taken from NOAA et al. (1976), because the direct computation would result in time consuming integration of complex formulas for each individual gas species. The atmospheric density, finally, is calculated with the ideal gas law in its normal form, where M is also computed on basis of linear interpolation.

The dynamic viscosity and the speed of sound are calculated with the same equations as for US62, Eqs. (2.82) and (2.83), although R is not considered to be constant in this case.

2.3.2.3 NRLMSISE-00

For simulations at higher altitudes that require an improved precision, e.g., the decay of satellites, or space-debris impact predictions, the US76 model no longer suffices, as it provides average data only. To include time-dependency in the model, the NRLMSISE-00 atmosphere model is a good candidate to use.[12] This is an MSIS-type atmosphere model, defining a number of atmospheric parameters as a function of altitude, latitude, longitude, and time (Picone et al. 2001). Similar to US76 it is an empirical model, based on past observations using orbit determination, accelerometer data, incoherent scatter radar data and other sources. However, main difference with the US76 model is that NRLMSISE-00 has a number of time-dependent inputs, such as solar and geomagnetic activity data. And, whereas the US76 provides global

[12] The name NRLMSISE-00 comes from the combination of several parts. NRL stands for the US Naval Research Laboratory that has developed the model, whereas MSIS is an abbreviation for Mass Spectrometer and Incoherent Scatter radar. This instrument was the primary data source for the development of earlier versions of the model. The "E" in the name indicates that the model extends from the ground through exosphere, and, finally, (20)00 is the year of release.

Table 2.13 Input and output parameters of the NRLMSISE-00 atmosphere model. Data based on Picone et al. (2001)

Input		
Longitude	τ	deg
Latitude	δ	deg
Altitude	h	m
Daily solar flux	$F_{10.7}$	W/m^2
81-day average daily flux	$\bar{F}_{10.7}$	W/m^2
Geomagnetic index	A_P	–
Day of year	doy	days
Seconds in day (UT)	sec	s
Local apparent solar time	lst	hours
Output		
Exospheric temperature	T_∞	K
Local temperature	T	K
Helium number density	n_{He}	cm^{-3}
Oxygen (O_2) number density	n_{O2}	cm^{-3}
Oxygen (O) number density	n_O	cm^{-3}
Nitrogen (N_2) number density	n_{N2}	cm^{-3}
Nitrogen (N) number density	n_N	cm^{-3}
Argon number density	n_{Ar}	cm^{-3}
Hydrogen number density	n_H	cm^{-3}
Oxygen (O^+) number density	n_{O+}	cm^{-3}
Total mass density	ρ	g/cm^3

average data independent from the surface location, there is also a dependence on longitude and latitude in NRLMSISE-00. Table 2.13 shows the model's inputs and outputs.

Although other thermospheric models exist (e.g., the Jacchia-series), in a comparative study between the MSIS and Jacchia models, based on decay prediction of multiple satellites, it was found that NRLMSISE-00 provides (slightly) more accurate density estimates in the majority of cases (Akins et al. 2003). Moreover, the fact that NRLMSISE-00 covers the whole (ground to space) altitude regime is considered an advantage, as the same atmosphere model can be used during the entire propagation, i.e., from orbit decay down to impact on Earth.

Figure 2.34 depicts the density and temperature as a function of altitude for both US76 and NRLMSISE-00 atmosphere models. The difference between high and low solar activity is more prominent at high altitudes (>150 km), and it may be clear that the US76 standard atmosphere provides an approximation based on average solar activity. Figure 2.35 shows the NRLMSISE-00 density output at two conditions in an alternative format, i.e., as a ratio to the US76 density. The large deviations at high

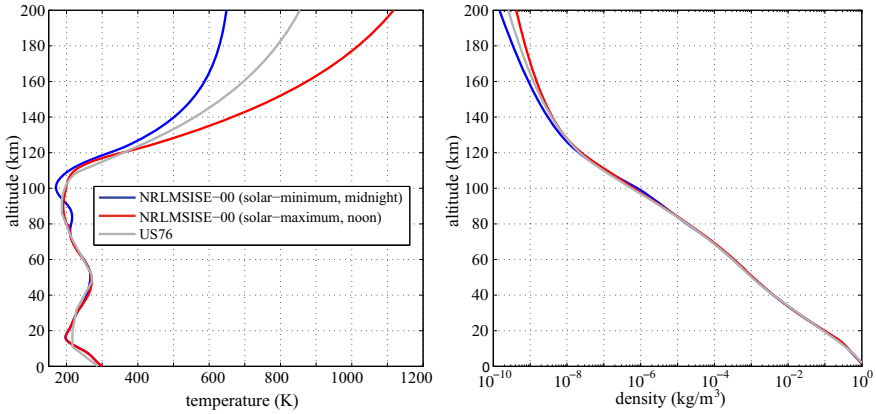


Fig. 2.34 Temperature and density as a function of altitude for the US76 Standard Atmosphere and NRLMSISE-00 model. For the latter, conditions for both solar maxima and minima are shown

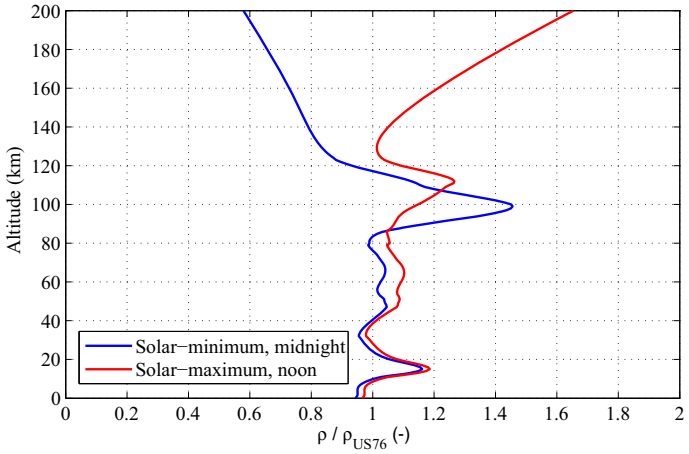


Fig. 2.35 Ratio of densities during periods of solar minimum and maximum activity. The US76 density is taken as (average) baseline

altitudes stand out, the density at daytime and high solar activity at 200 km being three times as high as the density at night for conditions of low solar activity. As an example of the importance to have accurate density data, even at high altitudes, the effect of selected atmosphere model on the orbital decay of a satellite is shown in Fig. 2.36. Choosing a static atmosphere model would lead to a severe offset in impact time and thus location.

Uncertainties

Although taking many atmospheric variations into account, NRLMSISE-00 may not be fully accurate in all of its outputs for certain applications. To study orbit decay, or

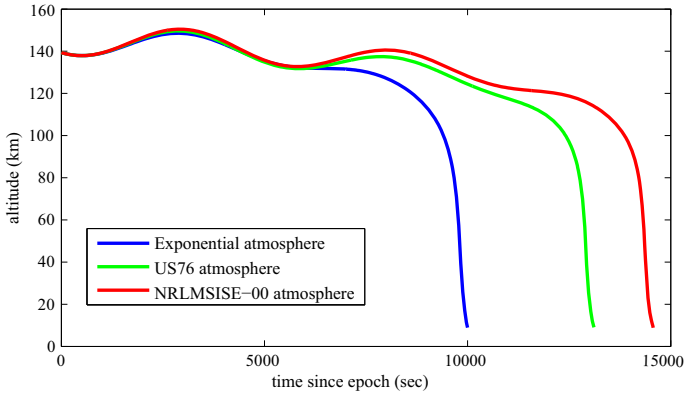


Fig. 2.36 Altitude as a function of time for a satellite's orbital decay (satellite ballistic coefficient, $K = 0.0088\,\mathrm{m}^2/\mathrm{kg}$). Selected atmosphere models are the exponential atmosphere, US76, and NRLMSISE-00

the statistical analysis of impact locations of decaying space debris, especially the uncertainties related to the total mass density (used in the determination of the drag term) are of importance.

Picone et al. (2001) analysed other data sources (accelerometer data, drag measurements, Space Shuttle data, etc.), and determined the offset with respect to density estimates given by NRLMSISE-00 across all geomagnetic-activity levels. The statistical results of these comparisons are tabulated in Table 2.14 for the 120 to 200 km altitude regime, expressing the residuals using a mean and standard deviation. A weighted average of both parameters is taken based on the number of points that were used in each comparison. As expected, the average offset is very small ($<1\%$), meaning that the model provides average values across several situations. The average standard deviation of 0.1176 ($\approx 12\%$) gives an idea of the inaccuracies of the NRLMSISE-00 model.

Space Weather Data

The NRLMSISE-00 model uses space weather data as an input. More specifically, indices are used for geomagnetic activity (A_P), the daily solar flux ($F_{10.7}$), and an 81-day average value for the solar flux ($\bar{F}_{10.7}$), the latter two at 10.7 cm wavelength. A_P gives a measure of the number of charged particles from solar flares interacting with the Earth's magnetic field and thus heating up the atmosphere. Daily or 3-hourly A_P-indices can be used, whereas the daily and long-term fluctuations in solar activity are covered by the $F_{10.7}$ and $\bar{F}_{10.7}$ values, respectively.

All necessary space-weather indices for NRLMSISE-00 are derived from worldwide measurements and grouped in the Celestrak space weather document, which is updated every three hours (Vallado and Kelso 2005). The document contains daily historic data (since 1 October 1957), as well as daily predictions, for the next six weeks, and monthly predictions for the following 90 months. For short-term predictions, e.g., the last few hours of orbital decay, one should initialise the simulations

Table 2.14 Comparison of NRLMSISE-00 density to drag, accelerometer and rocket database on total mass density across all geomagnetic activity levels (Picone et al. 2001)

Data set	Points	Weighted mean (–)	Standard deviation (–)
Barlier drag	3,285	0.01	0.22
SETA 79	30,611	0.01	0.10
SETA 82	48,445	−0.04	0.10
SETA 83	57,350	−0.07	0.09
SETA 84	56,816	−0.12	0.09
San Marco-5	2,935	0.35	0.13
AE-C MESA	54,353	0.11	0.17
AE-D MESA	28,499	0.01	0.16
AE-E MESA	43,798	0.05	0.12
Shuttle acceleration	52	0.07	0.23
Weighted average	–	−0.009	0.1176

with the latest information and keep it constant, especially when the daily solar activity fluctuations are moderate. For more long-term simulations in the higher atmospheric regions, it is advised to use an interpolation between subsequent A_P and $F_{10.7}$ values.

Molecular Composition

As will be described in the next chapter, Sect. 3.1, the determination of the Mach and Knudsen number requires the knowledge of molecular properties as molar mass, M, and the molecular collision diameter, σ. Although these quantities remain rather constant between 0 and 100 km altitude, gaseous diffusion of the atmospheric components causes a gradual change at higher altitudes. Based on the number densities of the separate constituents contained in the output of the NRLMSISE-00 atmosphere model, and the knowledge of the related molar masses and collision diameters of the individual components through basic chemistry, the average molecular weight and collision diameter can be determined using the following relations:

$$\bar{M} = \sum_{i=0}^{N} \frac{n_i}{n_{tot}} M_i \tag{2.93}$$

$$\bar{\sigma} = \sum_{i=0}^{N} \frac{n_i}{n_{tot}} \sigma_i \tag{2.94}$$

where subscript i denotes the considered element, n_i its number density and n_{tot} the cumulative number density. M_i and σ_i represent the molar mass and collision diameter of the molecule under consideration. These values are listed in Table 2.15 for the major atmospheric components. The resulting progress of average molecular

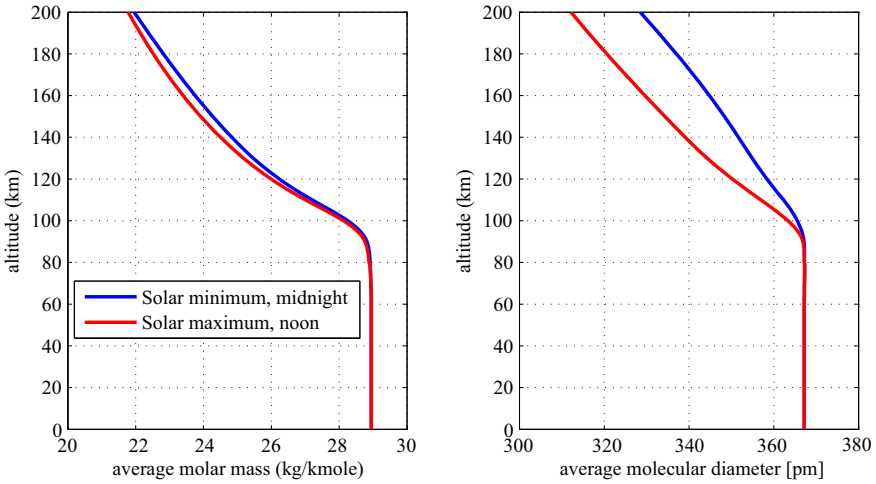


Fig. 2.37 Average molecular weight and collision diameter as a function of altitude. Derived from NRLMSISE-00 output for periods of high and low solar activity

weight and collision diameter as a function of altitude is depicted in Fig. 2.37, again between 0 and 200 km altitude and for cases of high and low solar activity.

Table 2.15 Molar masses for the major atmospheric components. Data obtained from NIST (2017), and Kaye and Laby (2005)

Element	M_i (kg/kmol)	σ_i (pm)
N_2	28.0	370
O_2	32.0	358
N	14.0	290
O	16.0	280
He	4.0	256
Ar	39.9	340
H	1.0	260

2.3.3 Mars

The atmosphere of Mars is very thin compared to the one of Earth. It is a mixture of 95% CO_2, 3% N_2, Ar, CO and H_2O. At the surface the mean pressure is 6 mbar, about the 0.6% of the corresponding value on Earth. This pressure is below the saturation pressure for liquid water at 215 K, which is the mean surface temperature, so that

water can be present either as vapour or ice. Over the equatorial regions formation of water-ice clouds is possible at about 10 km altitude, while at 50 km, where the temperature is about 150 K, CO_2-ice clouds take shape.

Winds are a frequent occurrence on Mars. In general, winds are caused by pressure gradients that, in turn, are caused by temperature gradients. The rotation of the planetary body and, in particular, the Coriolis force it generates, results in a varying direction of the wind velocity with respect to the surface. Because the atmosphere is so thin, it is very sensitive to solar heating. As a result, the day-night thermal excursion is large. Night temperature can be as low as 130 K, whereas it can rise to 300 K during the day. This causes thermal-tide winds, which are strong winds across the day-night line. In the Spring and Summer small tornadoes, called Dust Devils, can form in flat regions, with wind velocity up to 100 m/s. These "devils" raise dust in the atmosphere, and because the dust absorbs solar energy it gets locally heated. Local pressure gradients originate and, as a result, the Dust Devils eventually evolve into large-scale dust storms.

All things considered, the Martian atmosphere is very dynamic and difficult to model. However, due to the many missions flown to Mars and the large amount of observational data returned, many models for the Martian atmosphere have been developed and are continuously being updated when new data becomes available. One of the commonly-used engineering models is the Mars Global Reference Atmosphere Model (Mars-GRAM) developed for NASA (Justh 2010), with the latest version issued in 2010. Compared to the 2001 version, major changes concerned the user interface, and added adjustment factors to agree more with Mars Global Surveyor (MGS) Thermal Emission Spectrometer (TES) limb data from 0–60 km altitude. It also provided a better match with data at approximately 90 to 130 km altitude from the three aerobraking missions, MGS, Odyssee, and Mars Reconnaissance Orbiter (MRO).

From the European side, the Mars Climate Database (MCD) has been developed for ESA, and has currently reached version 5.3 (Millour et al. 2018). It has been derived numerically from the General Circulation Model (GCM) and validated using available observational data. The GCM includes water cycle, chemistry, and ionosphere models. The MCD provides mean values and statistics of the main atmospheric parameters (temperature, density, pressure, and winds), as well as atmospheric composition (including dust and water vapour and ice content). The database extends up to and including the thermosphere (about 350 km), and different so-called scenarios can be selected, i.e., a nominal, warm (high dust) or cold (low dust) atmosphere, or one with a dust storm. As an example of its output, in Fig. 2.38 the water-ice column calculated by the MCD is compared with a mosaic of pictures obtained with the Mars Orbiter Camera on Mars Global Surveyor, displaying water-ice clouds. This figure also shows some of the validation that has been done while developing the climate database.

Below, we will begin with a simple evaluation of an exponential atmosphere model (Example 2.3), to compare with exponential models derived for other atmospheres, and to be able to apply the analytical expressions for planar motion to Mars missions (Chap. 5). Then a comparison between three Mars climate databases is made, i.e.,

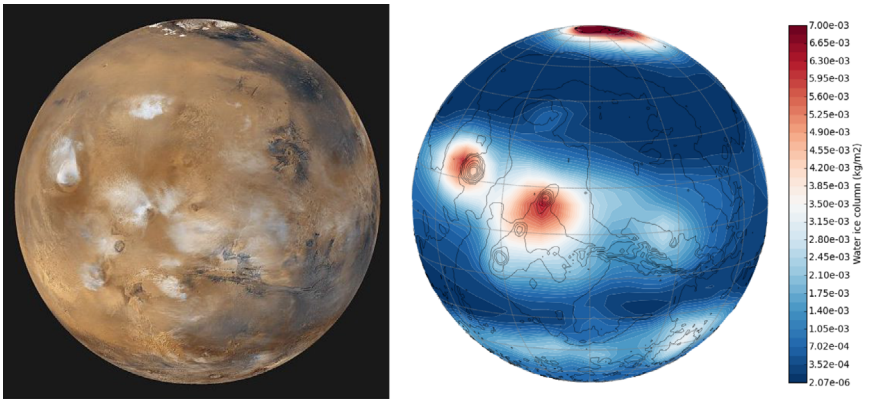
ESA's Mars Climate Database Versions 3.1 and 4.1. and Mars-GRAM2001 (which is publicly available, whereas the 2010 version is licensed), based on an older study by Mooij et al. (2006). These databases have originated around the same time and will make for a fair comparison. Also, since the main point is to show how the selection of a particular database can affect the outcome of a mission analysis, these three will serve this purpose. To support this claim, in Fig. 2.39 different temperature profiles are shown for different scenarios. Despite large differences for the upper layers of the atmosphere, these profiles are more or less identical for the altitude range relevant for Mars entry and descent missions.

Example 2.3 (*Mars Exponential Atmosphere*) For Mars Seiff and Kirk (1977) published an atmosphere model based on in situ Viking-1 measurements. These data can be fitted by an exponential atmosphere model. In Fig. 2.40 the exponential atmosphere model fitted through these data is shown. Also, the error between the two models has been plotted. The scale height and reference density are given by $H_s = 8.8\,\text{km}$ and $\rho_0 = 0.0168\,\text{kg/m}^3$.

It is clear that the Martian atmosphere is not altogether exponential; the error that will be made by using the exponential atmosphere is below 100 km maximum 50%. In between 120 and 180 km, however, the error increases to almost 200%. One has to decide on a per-application basis what is the altitude region of interest and the maximum acceptable error, and let the curve fit follow from those requirements. ■

2.3.3.1 ESA Mars Climate Database Version 3.1 (EMCD 3.1)

The ESA Mars Climate Database Version 3.1 is a database of statistics (Lewis et al. 1999), which describe the climate and surface environment of Mars for 12 seasons,



(a) Mars Global Surveyor mosaic displaying water ice clouds (Image: JPL).

(b) MCD calculated water-ice column. Image: http://www-mars.lmd.jussieu.fr, date accessed April 10, 2020.

Fig. 2.38 Comparison between MGS-measurement and MCD-calculated data

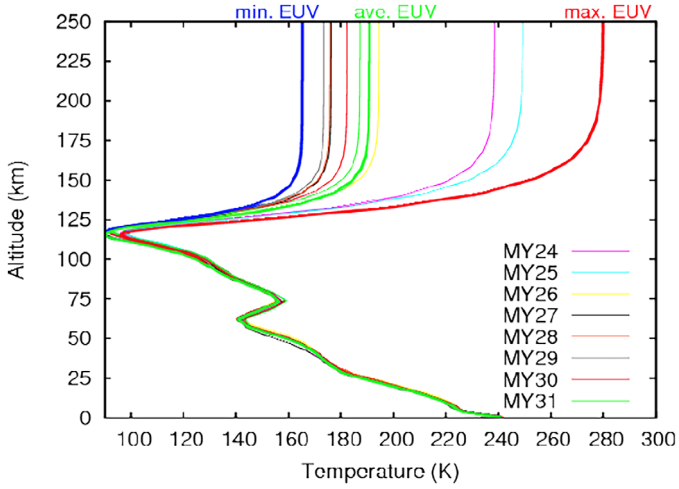


Fig. 2.39 MCDv5.3 temperature profiles using a minimum, average and maximum Extreme Ultra Violet (EUV) input from the Sun, obtained at same location and time of year when using the Mars Year 24–32 scenarios (Millour et al. 2018)

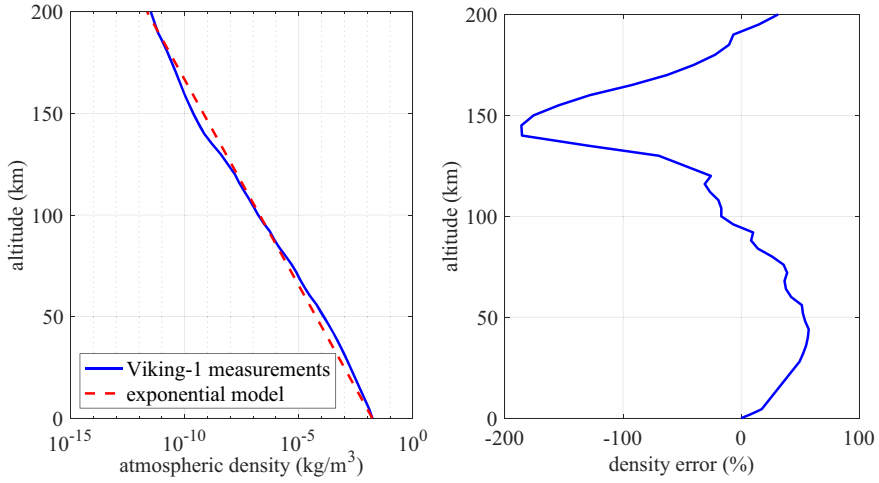


Fig. 2.40 Exponential atmosphere model fitted through Viking-1 measurements. Data source: Seiff and Kirk (1977)

12 times of day at all locations in an altitude range 0–120 km. It has been constructed directly on the basis of output from multi-annual integrations of two general circulation models developed jointly at Laboratoire de Météorologie Dynamique (LMD) du Centre National de la Recherche Scientifique, France, and the University of Oxford, United Kingdom, with support from the European Space Agency (ESA)

and Centre National d'Etudes Spatiales (CNES). The models have been developed and validated to reproduce the main features of the meteorology of Mars, as observed by past spacecraft missions. As well as the more standard statistical measures for mission design studies, the Mars Climate Database includes a novel representation of large-scale variability, using empirical eigenfunctions derived from an analysis of the full simulations, and small-scale variability using parameterisations of processes such as gravity-wave propagation. It provides different scenarios, including a standard Martian year, labeled Mars Global Surveyor (MGS) scenario, dust-storm scenarios, and a somewhat warmer scenario than the MGS scenario, called Viking scenario.

2.3.3.2 ESA Mars Climate Database Version 4.1 (EMCD 4.1)

The Mars Climate Database Version 4.1 is a database of atmospheric statistics as well, compiled from state-of-the art General Circulation Model simulations of the Martian atmosphere. The models used to compile the statistics have been extensively validated using available observational data and represent the current best knowledge of the state of the Martian atmosphere given the observations and the physical laws, which govern the atmospheric circulation and surface conditions on the planet (Forget et al. 2005).

The scientific content of version 4.1 is very similar to that of version 3.1, the major difference being an extension of the model to an altitude of 220 km, covering the thermosphere. EMCD 4.1 also includes different scenarios, equivalent to those in EMCD 3.1. Moreover, a surface-pressure correction has been added to match the calculated surface pressures with Viking measurements.

2.3.3.3 Mars-GRAM2001

The Mars Global Reference Atmospheric Model (Mars-GRAM) is an engineering-oriented model of the atmosphere of Mars (Justus and Johnson 2001). Recent documented applications of Mars-GRAM include aero-braking operations of Mars Global Surveyor (Lyons et al. 1999), trajectory design and entry dispersion studies of Mars Pathfinder (Braun et al. 1995; Spencer and Braun 1996), and aerothermodynamic and entry dynamics studies for Mars Polar Lander (Queen et al. 1999).

Earlier Mars-GRAM versions were based on ad-hoc parameterisations to data observed by the Mariner and Viking missions. The previous and current (Mars-GRAM2001) version are based on input data tables of output from the NASA Ames Mars General Circulation Model (MGCM) and the University of Arizona Mars Thermospheric General Circulation Model (MTGCM). Fully global in scope, and based on first-principles physics (e.g., atmospheric thermodynamics and equations of atmospheric motion), both MGCM and MTGCM have been somewhat "tuned" to insure that they represent Mariner and Viking results at the locations and times for which these data are available.

Table 2.16 Potential landing sites (Epoch 01 Jul 2013 13:00:00)

Landing	τ_0	δ_0
Site	[°]	[°]
CASE A	−14.8	28.6
CASE B	−75.2	−38.3
CASE C	−93.0	8.0

2.3.3.4 Assessment of Databases

The basic evaluation of all three databases will be carried out by calculating the atmospheric properties for three different locations, which are possible landing locations for a potential Mars mission. This mission will be discussed in more detail in Sect. 5.3.5. For each evaluation the epoch and a point defined by the longitude and latitude are selected (see Table 2.16), and the main variables p (pressure), T (temperature) and ρ (density) will be evaluated along an altitude tube. This tube starts at 120 km altitude (usually the entry interface) and will end at 0 km on the mean Mars surface. In the case of EMCD 3.1 we have used Scenario #1 (MGS) and a similar one for EMCD 4.1, designated Martian Year 24 (MY24). The MGS (Mars Global Surveyor) scenario represents a moderate realistic Mars Atmosphere without big (dust) storms. The MY24 scenario has been designed to mimic Mars as observed by Mars Global Surveyor in between 1999 to June 2001, a Martian year thought to be typical. The dust fields were derived from MGS Thermal Emission Spectrometer (TES) observations using a data-assimilation technique.

The MY24 scenario is provided with three solar EUV conditions: solar minimum, solar average and solar maximum. For this evaluation Scenario #2 is selected, i.e., the solar average. The MY24 scenario of EMCD 4.1 is very closely related to the MGS scenario of EMCD 3.1, with the essential difference that the aerosol optical thickness as measured by TES during that year has been assimilated into the model to give the most realistic temperature prediction. Since Mars-GRAM2001 does not have a direct scenario selection, the related individual settings have been set such as to come as close to the EMCD 3.1 and 4.1 scenarios as possible. No detailed analysis of changing individual settings has been performed. As such, the database is applied using common engineering judgement.

With the introduction of EMCD 4.1, there is a possibility to correct the pressure on the basis of the Viking surface measurements. This correction can reach rather high values, as is shown in Fig. 2.41. It seems that the magnitude of the correction exactly follows the relief of the Mars surface. Especially at high (e.g., Mons Olympus) and low (e.g., Valles Marineris) the correction factor can reach high (low) values. The way the correction is taken into account is that both the pressure and density are scaled with this factor for all altitudes (a so-called atmospheric tube). However, this does not seem to be the best approach considering the large correction factors for surface areas with large gradients. A limited altitude range for this correction seems

to be a better approach. One can use the original correction for the complete tube, especially when for the considered reference trajectories the correction factor is in the order of a few percent only and has no large impact on the results.

In Fig. 2.42, the atmospheric pressure, density and temperature—evaluated with respect to the MOLA areoid—are plotted as the relative difference from the EMCD 4.1 values. The (preliminary) comparison of EMCD 3.1 and 4.1 with Mars-GRAM2001 is somewhat misleading, because the surface pressure of the EMCD may not have been properly tuned to the Viking values, whereas it has been tuned for GRAM2001. As mentioned before, this tuning has been applied to the pressure data and should bring the pressure and density results together. A second cause for the discrepancies may be differences in MOLA height calculations that were noticed to vary from database to database. The real and important difference is the colder temperature of the atmosphere in Mars-GRAM2001, which seems to be a specific feature of that database. This feature is not present in either EMCD 3.1, which has been generated by the Oxford GCM, or in EMCD 4.1, which was generated with the LMD GCM.

It should be stressed, that irrespective of the surface-pressure issue, large differences in density arise, because of the uncertainty on the temperature distribution along the vertical, $T(z)$. Density at an altitude z is given in the hydrostatic approximation by

$$\rho = \frac{M p_s}{R T_s} \int_0^z e^{-\frac{gM}{RT(z)}} \mathrm{d}z \tag{2.95}$$

where p_s and T_s are surface pressure and temperature, respectively, M is the molar mass of the Martian atmosphere, g is the acceleration of gravity, and R is the ideal gas constant. It is hoped that in the future, temperature differences between models will

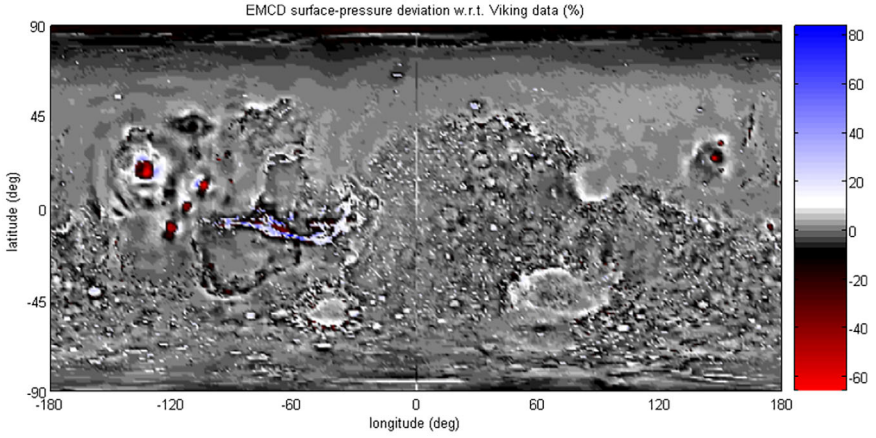


Fig. 2.41 Surface-pressure correction for EMCD 4.1

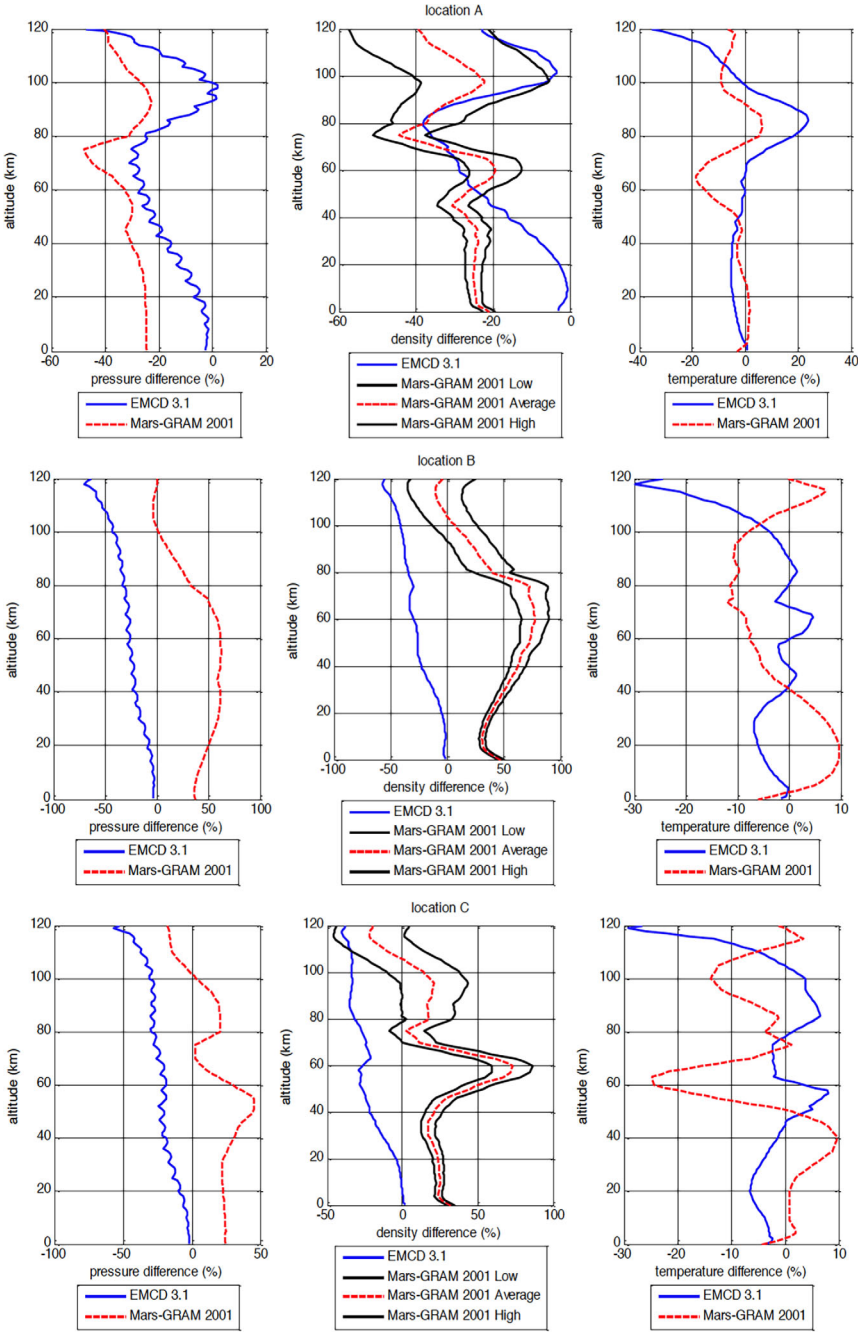


Fig. 2.42 Comparison of p, ρ and T with respect to EMCD 4.1 values, for the three landing locations

gradually become less important as more and more observations become available to constrain results, but in the meantime this source of uncertainty should be kept in mind and translated into appropriate design margins.

The last comparison that we do is that of the wind profile for a single location (Case A). In Fig. 2.43 the north-south and east-west wind components have been plotted for the three databases. Significant variations in wind vector are observed for this location, which means that the influence on, for instance, a footprint analysis can be significant. Special care should be taken in terms of the settings of a Monte-Carlo analysis to obtain comparable results.

In Chap. 5, Sect. 5.3.5, the above discussed atmosphere models will be used to assess the impact of mission performance of a ballistic Mars entry vehicle.

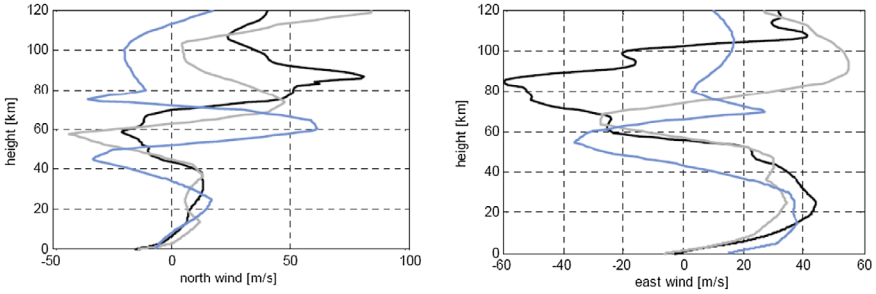


Fig. 2.43 Wind-vector components for an altitude tube above landing Location A

2.3.4 Titan

For the design and trajectory analysis of the Huygens mission, the so-called Lellouch-Hunten engineering model was used. This atmosphere model was derived from Voyager-1 radio-occultation refractivity profiles of Titan in the 0–200 km altitude range. The major atmospheric constituents were assumed to be N_2 and CH_4 (Lellouch and Hunten 1987). With launch of Cassini/Huygens on October 15, 1997, mission analysis did not stop, and also the understanding of Titan's atmosphere improved. Yelle et al. (1997) developed new engineering models based on the old ones, using improved insight in the thermal structure of the upper atmosphere and in its chemical composition. This came from re-analysing the Voyager Radio Science Subsystem (RSS), Infrared Spectrometer (IRIS) and Ultraviolet Spectrometer (UVS) data, and constructing a model that fitted all three datasets. The range of uncertainty on the atmospheric profiles of temperature, density, and pressure was, as before, taken into account in the form of a minimum, nominal, and maximum density as a function

of altitude, with corresponding values for temperature and pressure. Composition of the atmosphere was defined to be 95% N_2, 3% CH_4 and 2% Ar.

With the descent through the Titan atmosphere the Huygens Atmospheric Structure Instrument (HASI) provided a much more accurate dataset, albeit it was for a specific time and location only, and could therefore not be used to construct a global atmosphere model. Also the additional measurements of Cassini have improved understanding of Titan's atmosphere and will, without doubt, contribute to improved atmosphere models.

The current state-of-the-art engineering-level atmospheric model for Titan is the Titan Global Reference Atmosphere Model (Titan-GRAM), which has been developed for use in NASA's systems analysis studies of aerocapture and entry, descent and landing applications in potential missions to Titan (Justus et al. 2004). The basis for development is similar to those for Earth and Mars, but the main input comes from the Yelle atmosphere models.

From *Voyager* data it was observed that there is a substantial wind environment, which gave rise to the development of a Global Circulation Model (GCM). The three-dimensional hydrostatic GCM discussed by Tokano et al. (1999) is based on primitive equations of hydrodynamics, and can predict both zonal and meridional winds on Titan. The core of the model consists of the zonal and meridional momentum equations, including zonal, meridional and vertical advection, a Coriolis, centrifugal and pressure-gradient term, as well as vertical turbulent diffusion, and surface friction. Temporal aspects in the form of Saturn's and Titan's position with respect to the Sun play an important role as well.

For an epoch in 2035, with a Kronocentric (= Saturn-centred) orbital longitude of $L_s = 306°$, the model has been evaluated for an altitude of 10 km.[13] The resulting wind field was used for a conceptual study of the Titan Mare Explorer (Lorenz et al. 2012) and to simulate long-duration balloon trajectories (Garg and Mooij 2016). Here, we restrict ourselves to showing the zonal and meridional wind fields, Figs. 2.44 and 2.45, and relate these to the motion of an aerobot.

When a balloon is deployed near the equator, it stays within ±30° latitude as there are no strong winds near the equator that can push the balloon towards the polar region. Balloons will keep circumnavigating the equatorial region, till the time there is a wind gust that can carry it towards a higher latitudinal region. Also, near the equatorial region zonal winds (east to west) are stronger than meridional winds (north to south), which is why the motion in the direction of east to west predominates. When deployed near the polar region, the wind velocity in zonal direction is high and it is impossible for the balloon to drift towards the equator. However, at 100° longitude, the speed of meridional winds is higher, and that is why patterns are formed in the balloon movement.

We will conclude the discussion on Titan's atmosphere, with a third example of constructing an exponential atmosphere.

Example 2.4 (*Titan Exponential Atmosphere*) In this third example we will look at a way to locally improve the accuracy of the curve fit for the exponential atmosphere

[13] The data have been produced by Ashima Research Centre, Pasadena, and used with permission.

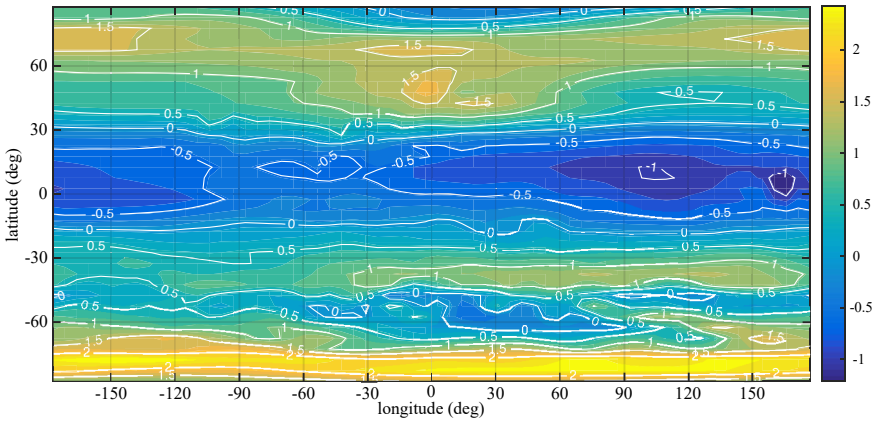


Fig. 2.44 Zonal winds (m/s) on Titan (10 km altitude), year 2035 ($L_s = 306°$), based on external data[14]

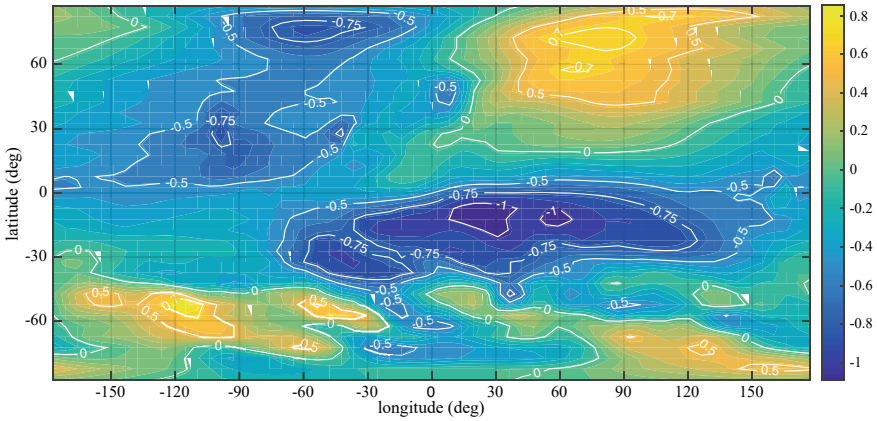


Fig. 2.45 Meridional winds (m/s) on Titan (10 km altitude), year 2035 ($L_s = 306°$), based on external data[14]

model. In Fig. 2.46 a density profile for the Titan atmosphere has been plotted. This profile has been derived by Yelle et al. (1997) and is known as the nominal Yelle profile. It is clear that due to the gradient change around $h = 140$ km, curve fitting this profile with a single function would lead to unacceptably large errors at either low or high altitudes (in a log-plot, an exponential function is a straight line). Therefore, one could consider two exponential curve fits that will intersect at $h = 140$ km. Since the two models will not have the same gradients at this point, this introduces a discontinuity in the density evaluation. Such a discontinuity shows as a small spike in vehicle acceleration, which can, for instance, lead to problems when designing guidance and control systems. Therefore, some form of bridging function needs to

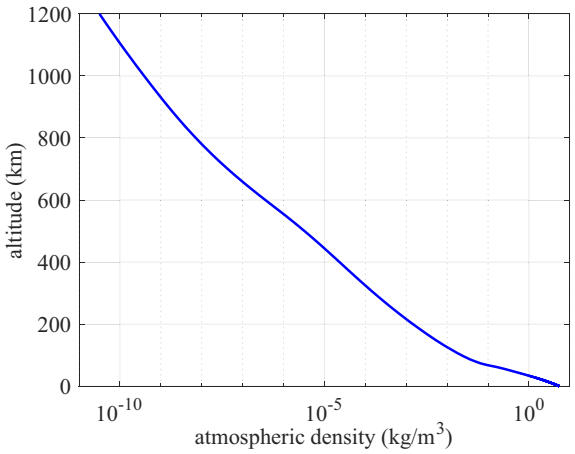


Fig. 2.46 Titan atmosphere model according to Yelle. Data source: Yelle et al. (1997)

be defined to get a smooth transition from one model to the other. One can think of, for instance, linearly interpolating between the two values for ρ_0 and H_s, or using a polynomial that can guarantee zero-order and first-order continuity.

Both parts of the Titan atmosphere above and below 140 km are now approximated by an exponential model. The scale height and reference density are given by $H_s = 21.5\,\text{km}$ and $\rho_0 = 3.2623\,\text{kg/m}^3$ for the part $h < 140$ km, and $H_s = 50\,\text{km}$ and $\rho_0 = 0.0053\,\text{kg/m}^3$ for $h > 140$ km (note that the latter model is in this case given by $\rho = \rho_0 e^{-\frac{h-140}{H_s}}$, with h in km). The curve fits and associated errors are shown in Fig. 2.47. Due to local non-linearities the error still amounts to about 50%, but this is a far smaller error than the one that would have resulted from a single curve fit. ■

2.3.5 Venus

Venus has a very dense and hot atmosphere, i.e., the temperature at the surface is 740 K, whereas the pressure is 93 bar (Basilevsky and Head 2003). It is composed primarily of carbon dioxide (96.5%) with nitrogen as second (3.5%), and some minor traces of sulphur dioxide, argon, water vapour, and carbon monoxide. Despite the harsh conditions on the surface, pressure and temperature at about 50 to 65 km above the surface are nearly the same as that of the Earth. This makes the Venusian upper atmosphere the most Earth-like area in the Solar System, even more so than the surface of Mars. For this reason NASA even proposed Venus as a potential target for permanent colonisation by means cities designed to float at about 50 km altitude in the atmosphere (Landis 2003).

Also for Venus a Global Reference Atmosphere Model has been developed (Venus-GRAM), which provides mean density, temperature, pressure and wind com-

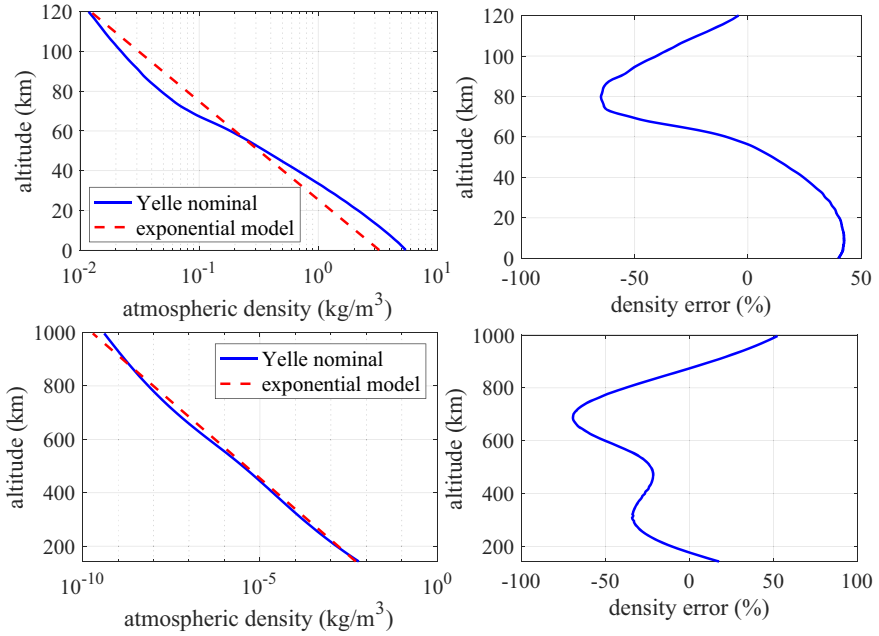


Fig. 2.47 Exponential atmosphere model fitted through the nominal Yelle atmosphere model

ponents at any height from 0 to 1000 km. Included in the model are random perturbations about the mean, which enables the user to do mission planning and system

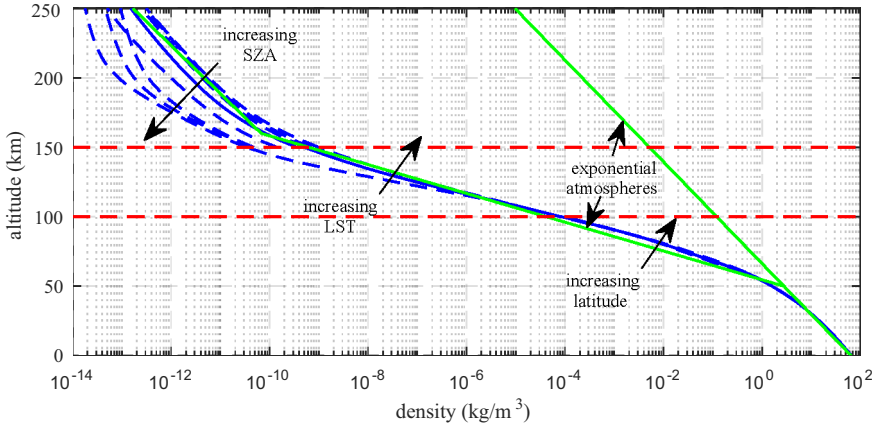


Fig. 2.48 Original and averaged profiles of the Venus Global Reference Atmosphere Model (blue lines, with two exponential approximations (green lines). Marked in red is the altitude range for aerocapture missions (see Chap. 11)

analysis. Venus-GRAM is a combination of three other atmosphere profiles for different altitude ranges, based on the Venus International Reference Atmosphere (VIRA) profiles for those ranges. The profile for low altitudes, between the surface and 100 km altitude, is altitude and latitude dependent. The profile for the mid altitudes, between 100 and 150 km, is dependent on altitude and local solar time. The high-altitude profile, between 150 and 250 km, is a function of altitude and solar zenith angle. Of course, whenever new measurement data become available, Venus-GRAM may be updated (Justh and Dwyer Cianciolo 2017).

For engineering purposes it is often justified to simplify the atmosphere model, and only keep an altitude-dependent profile by averaging the low profile over all latitudes, the middle profile over all solar-times, and the high profile over the zenith-angles. The latitude, local-solar-time, or solar-zenith-angle dependent profiles, together with the average profile, and two exponential atmosphere profiles are presented in Fig. 2.48. The first of these exponential-atmosphere profiles is based on a single value for the surface density and scale-height. These values are listed in Table 2.17. However, it is clear that for heights above about 50 km, this profile and the data obtained from Venus-GRAM start to diverge significantly by multiple orders of magnitude. This can be avoided by building a custom exponential atmosphere by approximating the averaged profile with three segments with constant scale height, similarly to what we did for Titan in Example 2.4. This gives a much closer agreement with the averaged profile.

The segments can be fitted using he standard formulation for a part of the standard atmosphere with consistent scale-height, starting at a particular altitude, for which the equation is:

$$\rho = \rho_0 e^{-\frac{h-h_0}{H_s}} = \rho_0 e^{\frac{h_0}{H_s}} e^{-\frac{h}{H_s}} = \rho_0^* e^{-\frac{h}{H_s}} \tag{2.96}$$

where ρ_0 is the starting density for that layer of the atmosphere, h_0 the starting altitude of that layer of the atmosphere, h the altitude with respect to the surface, and H_s the scale height. ρ_0^* is the density at the surface if one would use the profile described by the scale-height for all altitudes. If the exponential atmosphere is used for a specific altitude range only, as indicated in the figure, it is preferred to use a single segment only. As explained in Example 2.4, the transition from one segment to the other may give a sudden change in the slope of the density profile that should be avoided.

From Fig. 2.48 it can also be seen that for the lower part of the atmosphere, there seems to be little latitude variation. Only halfway through the middle altitude profile and above can a "significant" dependency of density be observed. Whilst the deviation seems large, it should be noted that at this point, the density is already approximately a hundred-millionth of the surface density of Venus, or a millionth of the Earth's surface density. This is furthermore illustrated in Fig. 2.49. In addition to the error between the average profile and the dependent profiles, the error between the average profile and the custom-fitted exponential atmosphere profile is shown.

Table 2.17 Venus exponential atmosphere model parameters

Model	Scale-height (H_s) (m)	ρ_0 (kg/m^3)	h_0 (m)	ρ_0^* (kg/m^3)
Standard exp. atmosphere	15,900	64.79	0	64.79
Low-altitude exp. atmosphere	15,900	64.79	0	64.79
Mid-altitude exp. atmosphere	4,500	2.7913	50,000	186765.3
High-altitude exp. atmosphere	15,000	$6.7594 \cdot 10^{-11}$	160,000	$2.8986 \cdot 10^{-6}$

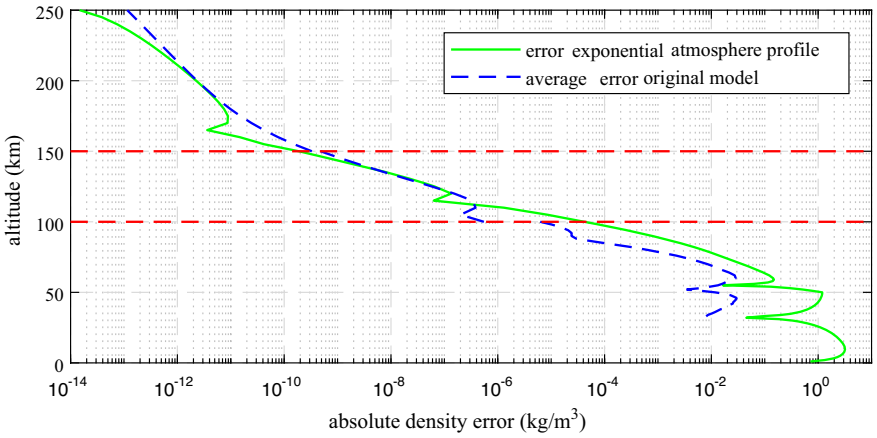


Fig. 2.49 Absolute density error between the averaged Venus atmosphere profile and the custom fit exponential atmosphere model and the average error between the averaged Venus atmosphere profile and the original model. Marked in red is the altitude range for aerocapture missions (see Chap. 11)

Problems

Judge whether the following statements are TRUE or FALSE (*Problems* 2.1 to 2.12).

2.1. In an exponential atmosphere model, the density follows an exponential function whereas the pressure change is linear.
2.2. An accurate shape model of an asteroid is sufficient to provide the user with an accurate gravity field using a polyhedra model.
2.3. A spherical-harmonics model can be used for any location on and above the Moon.
2.4. To accurately calculate the atmospheric density, not only a three-dimensional position should be taken into account, but also the epoch for which one is evaluating.

2.5. The difference between geocentric and geodetic latitude is independent from the flattening of the central body.
2.6. Since the maximum height variation of the Moon's surface is less than 20 km its gravity field can be considered "regular".
2.7. The hydrostatic equation gives the relation between pressure and density.
2.8. The "surface" of a gas planet is not properly defined.
2.9. In general, the steady-state wind is proportional to the altitude above the surface.
2.10. A spherical planet can have a very irregular gravity field, requiring high degree and order of the spherical-harmonics model to properly account for these irregularities.
2.11. The strong flattening of Saturn is caused by its relatively large planetary rotation.
2.12. In an exponential atmosphere model, the speed of sound decreases with increasing altitude.

Exponential model of the atmosphere (*Problems* 2.13 to 2.17).

2.13. What is the physical interpretation of the scale height? Confirm this with a numerical example.
2.14. Given the data in Table 2.18, fit an exponential atmosphere model for the density.
2.15. For an Earth exponential atmosphere model, the (constant) temperature is assumed to be 220 K.

(a) What does this mean for the speed of sound?
(b) What changes if the temperature increases by 50 K?
(c) What is the consequence when an exponential atmosphere model is used for an entry starting at $h_E = 100$ km down to landing?

2.16. Suppose a planetary atmospheric density can be modelled with two separate exponential models. What is (are) the condition(s) for a continuous density profile at the junction?
2.17. Given two atmosphere layers, one from $h = 0$ to 20 km, with $\rho_{0,1} = 1.4\,\text{kg/m}^3$ and $H_{s,1} = 9{,}000$ m, and the second one from 20 km to 40 km, with $\rho_{0,2} = \rho_1$ at 20 km (for the sake of continuity in ρ) and $H_{s,2} = 25{,}000$ m. Provide a bridging function for a smooth transition between the two layers.

Gravity-field modelling, GOCE's orbit around Earth (*Problems* 2.18 to 2.23).

The ESA satellite GOCE (altitude 250 km, inclination 96.5°, circular orbit) observed the gravity field of Earth. One of the elements of interest is the term $J_{3,1}$. The gravity potential of Earth can be modelled as a spherical-harmonics expansion, of which a particular formulation is given by the following equation:

Table 2.18 Temperature and density values

h (m)	ρ (kg/m^3)
0	1.5262
100	1.3472
200	1.3520
300	1.2733
400	1.2217
500	1.2085
600	1.2828
700	1.3331
800	1.2077
900	1.2429
1,000	1.2159
1,500	1.0814
2,000	1.0096
2,500	0.8279
3,000	0.7477
3,500	0.6970
4,000	0.6491
4,500	0.5439
5,000	0.5130
5,000	0.4816
6,000	0.3970
7,000	0.3516
8,000	0.2789
9,000	0.2344
10,000	0.1835
11,000	0.1457
12,000	0.1199
13,000	0.1072
14,000	0.0788
15,000	0.0724
16,000	0.0559
17,000	0.0434
18,000	0.0352
19,000	0.0306
20,000	0.0235

$$U = -\frac{\mu}{r}\left[1 - \sum_{n=2}^{\infty} J_n \left(\frac{R_e}{r}\right)^n P_n(\sin\delta)\right] - \frac{\mu}{r}\sum_{n=2}^{\infty}\sum_{m=1}^{n} J_{n,m}\left(\frac{R_e}{r}\right)^n P_{n,m}(\sin\delta)\cos[m(\lambda - \lambda_{n,m})] \tag{2.97}$$

with R_e the reference radius, and μ the gravitational parameter. Furthermore, $P_n(\sin\delta)$ and $P_{n,m}(\sin\delta)$ represent the Legendre polynomials and functions, respectively:

$$P_n(x) = \frac{1}{(-2)^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(1-x^2)^n \tag{2.98}$$

$$P_{n,m}(x) = (1-x^2)^{m/2}\frac{\mathrm{d}^m P_n(x)}{\mathrm{d}x^m} \tag{2.99}$$

Data that can be used: $\mu = 398600.4415\,\mathrm{km}^3/\mathrm{s}^2$, $R_e = 6378.137\,\mathrm{km}$, $J_{3,1} = -1.72{\cdot}10^{-6}$, and $\lambda_{3,1} = -1.0°$.

2.18. What do the parameters J_n, $J_{n,m}$, and $\lambda_{n,m}$ represent?
2.19. What is the definition of the parameters r, δ, and λ?
2.20. Give the general expression to derive the radial acceleration (in $\mathrm{m/s}^2$) from the potential formulation for the gravity field.
2.21. Derive the general equation for the radial acceleration due to the term $J_{3,1}$ for an arbitrary satellite. Do not fill in values for individual parameters, yet.
2.22. What is the numerical expression for the radial acceleration due to $J_{3,1}$ as experienced at GOCE's altitude (expressed in $\mathrm{m/s}^2$, still for arbitrary latitude and longitude)?
2.23. What are the values for δ for which this radial acceleration changes sign?

Gravity-field modelling, the effect of J_2 on entry flight (*Problems* 2.24 to 2.26).

For an Earth-entry analysis, two options for gravity-field modelling are used: a central field and a central field with the J_2 effect included. The current set of problems deals with the differences resulting from the use of either model. Data to be used: $R_e = 6{,}378.137\,\mathrm{km}$, $\mu = 398{,}600.4415\,\mathrm{km}^3/\mathrm{s}^2$, and $J_2 = -1.08263529{\cdot}10^{-3}$. For the sake of calculation, the radius of the Earth can be considered constant.

2.24. For an equatorial entry, how much does the (central-field) gravitational acceleration change from entry altitude ($h_E = 120\,\mathrm{km}$) down to 25 km altitude?
2.25. What is the error in acceleration when the J_2-effect is included? What if the latitude is 40° or 60°?
2.26. For a gliding entry due north, starting from $\delta = 40°$ it takes 20 min to reach $\delta = 60°$ at $h = 25\,\mathrm{km}$. In case J_2 is ignored, discuss the effect on (radial) position and velocity. State your assumptions, if any.

References

Akins KA, Healy LM, Coffey SL, Picone JM (2003) Comparison of MSIS and Jacchia atmospheric density models for orbit determination and propagation, AAS 03-165. In: 13th AAS/AIAA space flight mechanics meeting, Ponce, Puerto Rico, 9–13 Feb 2003

Araki H, Tazawa S, Noda H, Ishihara Y, Goossens S, Sasaki S, Kawano N, Kamiya I, Otake H, Oberst J, Shum C (2009) Lunar global shape and polar topography derived from Kaguya-LALT laser altimetry. Science 897–900

Basilevsky AT, Head JW (2003) The surface of Venus. Rep. Prog. Phys. 66:1699–1734

Braun RD, Powell RW, Engelund WC, Gnoffo PA, Weilmuenster KJ, Mitcheltree RA (1995) Mars Pathfinder six-degree-of-freedom entry analysis. J Spacecr Rockets 32:6

Drob DP et al (2008) An empirical model of the Earth's horizontal wind fields: HWM07. J Geophys Res-Space Phys 113(A12304):1–18

Forget F, Dassa K, Wanherdrick Y, Lewis SR, Collins M, Bingham SJ (2005) "Mars Climate Database v4.0 User Manual", ESTEC contract 11369/95/NL, Jan 2005

Garg K, Mooij E (2016) "Autonomous navigation of a balloon over Saturn's moon Titan", AIAA-2016-0887. Guidance, navigation, and control conference (SciTech 2016), San Diego, 4–8 Jan 2016

Hirt C, Kuhn M, Featherstone WE, Goettl F (2012) Topographic/isostatic evaluation of new-generation GOCE gravity field models. J Geophys Res—Solid Earth

Jorda L, Lamy PL, Gaskell RW, Kaasalainen M, Groussin O, Besse S, Faury G (2012) Asteroid (2867) Šteins: shape, topography and global physical properties from OSIRIS observations. Icarus 221(2):1089–1100

Justh HL (2014) Mars global reference atmospheric model 2010 version: users guide. NASA/TM-2014-217499, National Aeronautics and Space Administration, Feb 2014

Justh HL, Dwyer Cianciolo AM (2017) Venus global reference atmospheric model status and planned updates. Venus modelling Workshop, Cleveland, OH, May, pp 9–11

Justus CG, Johnson DL (2001) "Mars global reference atmospheric model 2001 version (Mars-GRAM 2001): users guide", NASA/TM-2001-210961, National Aeronautics and Space Administration, Apr 2001

Justus CG, Duva11 A, Keller VW (2004) Engineering-Level model atmospheres for Titan and Mars. In: Proceedings international workshop 'planetary probe atmospheric entry and descent trajectory analysis and science', Lisbon, Portugal, 6–9 Oct 2003. Published in ESA-SP-544, Feb 2004

Kaye G, Laby T (2019) Tables of physical and chemical constants. https://web.archive.org/web/20190506031327/, http://www.kayelaby.npl.co.uk/ (Date last Access: 1 Aug 2019)

Landis GA (2003) Colonization of Venus. In: Conference on human space exploration, space technology and applications. International Forum, Albuquerque, NM, 2–6 Feb 2003

Lellouch E, Hunten DM (1987) Titan atmosphere engineering model. Report ESLAB 87/199, ESLAB

Leslie FW, Justus CG (2011) The NASA Marshall space flight center Earth global reference atmospheric model-2010 version, NASA/TM-2011-216467. National Aeronautics and Space Administration

Lewis SR, Collins M, Read PL, Forget F, Hourdin F, Fournier R, Hourdin C, Talagrand O, Huot J-P (1999) A climate database for Mars. J Geophys Res-Planets 104(E10):24177–24194

Li Z, Zhu Q, Gold C (2004) Digital Terrain modelling, principles and methodology. CRC Press

Lorenz RD, Tokano T, Newman CE (2012) Winds and tides of Ligeia Mare, with application to the drift of the proposed time TiME (Titan Mare Explorer) capsule. Planetary Space Sci 60(1):72–85

Lyons DT, Beerer JG, Esposito P, Johnston MD, Willcockson WH (1999) Mars Global Surveyor: aerobraking mission overview. J Spacecraft Rockets 36(3):307–313

Mayer-Gürr T, Eicker A, Kurtenbach E, Ilk KH (2010) ITG-GRACE: global static and temporal gravity field models from GRACE data. In: Flechtner F et al (eds), System Earth via geodetic-geophysical space techniques. Advanced technologies in Earth sciences, pp 159–168. Springer

Melman JCP, Mooij E, Noomen R (2009) Autonomous navigation at asteroids, IAC-09-A3.I.10. In: 60th international astronautical congress, Daejeon, Republic of Korea, 12–16 Oct 2009
Millour E et al. (2018) The Mars climate database (Version 5.3), Scientific Workshop: "From Mars Express to ExoMars", ESAC Madrid, Spain, 27–28 Feb 2018
Montebruck O, Gill E (2000) Satellite orbits. Methods, and applications. Springer, Models
Mooij E, Huot J-P, Ortega G (2006) Entry trajectory simulation using ESA Mars climate database version 4.1, AIAA-06-6023, AIAA/AAS astrodynamics specialist conference and exhibit, Keystone, Colorado, 21–24 Aug 2006
NASA (2011) http://pds-geosciences.wustl.edu/default.htm. Last Accessed: 31 Oct 2011
National Institute of Standards and Technology (NIST), "Atomic Properties of the Elements" (2010). https://www.nist.gov/sites/default/files/documents/2017/02/02/periodic-table-2017-crop.pdf (Date last access: 1 Aug 2019)
NOAA, NASA and USAF (1976) U.S. Standard Atmosphere, 1976, Washington, DC
Ostro SJ, Hudson RS, Nolan MC, Margot JL, Scheeres DJ, Campbell DB, Magri C, Giorgini JD, Yeomans DK (2000) Radar observations of asteroid 216 Kleopatra. Science 288(5467):836–839
de Pater I, Littauer JJ (2010) Planetary Sciences, 2nd edn. Cambridge University Press
Picone JM, Hedin AE, Drob DP (2001) NRLMSISE-00 empirical model of the atmosphere: statistical comparisons and scientific issues. J Geophys Res 107, Nr. A12:1468
Queen EM, McNeil Cheatwood F, Powell RW, Braun RD, Edquist CT (1999) Mars Polar Lander aerothermodynamic and entry dispersion analysis. J Spacecraft Rockets 36(3), 421–428
Seiff A, Kirk D (1977) Structure of the atmosphere of mars in summer at mid-latitudes. J Geophys Res 82(28)
Smith D, Zuber MT, Frey HV, Garvin JB, Head JW, Muhleman DO, Pettengill B (2001) Mars orbiter laser altimeter: experiment summary after first year of global mapping of Mars. J Geophys Res 106(E10):23689–23722
Spencer DA, Braun RD (1996) Mars Pathfinder atmospheric entry: trajectory design and dispersion analysis. J Spacecraft Rockets 33(5)
Tokano T, Neubauer F, Laube M, McKay C (1999) Seasonal variation of Titan's atmospheric structure simulated by a general circulation model. Planetary Space Sci 47(3):493–520
United States Committee on Extension to the Standard Atmosphere (COESA) (1962) US Standard Atmosphere, 1962
Vallado DA (2001) Fundamentals of astrodynamics and applications. McGraw-Hill Inc, New York, NY
Vallado DA, Kelso TS (2005) Using EOP and space weather data for satellite operations, AAS 05-406. In: 15th AAS/AIAA astrodynamics specialist conference, Lake Tahoe, CA, 7–11 Aug 2005
Werner RA, Scheeres DJ (1997) Exterior gravitation of a polyhedron derived and compared with harmonic and mascon gravitation representations of asteroid 4769 Castalia. Celestial Mech Dyn Astron 65:313–344
Wieczorek MA (2007) Gravity and topography of the terrestrial planets. Treatise Geophys 10:165–206
Wieczorek MA (2011) http://www.ipgp.fr/~wieczord/SH/SH.html. Accessed 31 Oct 2011
Yelle RV, Strobell DF, Lellouch E, Gautier D (1997) Engineering models for Titan's Atmosphere, ESA SP-1177, pp 243–256
Yoder CF (1995) Astrometric and geodetic properties of Earth and the Solar System. Glob Earth Phys: A Handbook Phys Constants, AGU Ref Shelf 1:1–31

Chapter 3
Aerothermodynamics



After the discussion about the atmosphere and how to model the most relevant physical aspects, the moment has come to put a moving object into this atmosphere and find out how the two will interact with each other. After all, a lot has to happen between entering the atmosphere and landing on the surface of the planet below. To begin with, at entry the vehicle will have a large altitude and velocity, or in other words, potential and kinetic energy. This energy has to be "shed" before the vehicle can land at an acceptable speed. This could be achieved by using a propulsion system to brake, but this might be a very expensive way to do that. Alternatively, maybe in some way the atmosphere can be used. And indeed, the vehicle interaction with the atmosphere creates an aerodynamic force, which can be used to both decelerate the vehicle to an acceptable landing speed, and guide the vehicle towards a predefined landing location.

From a mission point-of-view this is ideal, of course, but from the side of the vehicle, the aerodynamic force determines the thermo-mechanical loads. These loads need to be sustained by the structural design of the vehicle, and therefore, a clear understanding of the vehicle's aerodynamic properties is required. This understanding does not come easy, as there is a three-way interaction between the environment (atmospheric properties), the mission (flow and velocity regimes), and the vehicle (through its shape and surface).

Studying aspects of aerodynamics dates back over 2,300 years, and (Anderson 2011) gives an interesting overview of its history. It began with Aristotle (350 B.C.), who described a model for a continuum and proposed that a body, which moves through this continuum, encounters a resistance. Archimedes came up with the notion that a fluid starts moving when a pressure difference acts on the fluid (250 B.C.). Jumping ahead, Leonardo da Vinci conceived the law of continuity, which says that in a flow tube the mass per unit time through a surface area remains constant, and that reducing the surface area results in a higher velocity. He observed various flow patterns and made sketches thereof (Fig. 3.1). He concluded that air resistance is directly proportional to the area of the body, and to reduce drag it would be

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_3

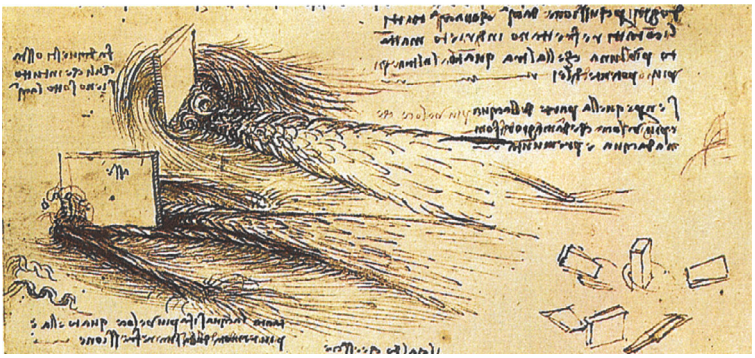


Fig. 3.1 Da Vinci's sketch of free-surface hydraulic turbulent flow behind obstacle (*Codex Leicester*)

necessary to streamline the body. Galileo added that aerodynamic resistance is also proportional to the density of the fluid. Edme Mariotte, in 1673, completed the picture by stating that it varied with the square of the velocity. In the end, it was Newton who theoretically verified that, and with his "Newtonian" mechanics developed the sine-squared law for the aerodynamic force.

Other fundamental contributions came from Swiss mathematician and physicist Daniel Bernoulli (1700–1782), who published his *Hydrodynamica* in 1738 with the initial statement of the pressure-velocity relationship (leading to Bernoulli's equation, formalised by Leonard Euler in 1752). In 1788, Joseph-Louis Lagrange (1736–1813), an Italian mathematician and astronomer, introduced the concepts of the velocity potential and stream function, and Pierre-Simon Laplace (1749–1827) published his famous Laplace equation in 1789. This equation became the fundamental equation for analysis of incompressible, inviscid, irrotational flow. He was also the first one to accurately calculate the speed of sound in air. All contributions of these scholars came together in 1840, when the Navier-Stokes equations were published, which combined all the equations necessary to characterise a real (viscous) fluid flow.

In 1903 the first flight of the Wright brothers was a fact, and one year later German engineer Ludwig Prandtl (1875–1953) introduced his concept for the boundary layer, that layer of the fluid closest to the boundary surface of, for instance, an aircraft fuselage or its wings. Four years after that, in 1908, he published his supersonic-shock and expansion-wave theory. At the onset of World War II the early applications of supersonic-flow theory were seen, and even though there were several claims as to supersonic flight during the war, the first acknowledged supersonic flight took place on October 14, 1947, when Chuck Yeager flew his experimental Bell X-1 research rocket plane. Commercial supersonic flight came with the Concorde, which operated from 1976 till 2003. In terms of hypersonic flight, the first man-made object to achieve that was the modified V-2 launched from White Sands Proving Grounds in 1949, and around 1955 significant research on hypersonic aerodynamics began.

With the introduction of more powerful computers, in 1962 the era of computational fluid dynamics started.

This chapter focuses on both aerodynamics and aerodynamic heating, which, of course, go hand-in-hand and are combined in the term aerothermodynamics. Starting with the fundamentals of aerodynamics, in Sect. 3.1 a qualitative overview of relevant flow effects is presented, and the key parameters used in the fields of aerothermodynamics are defined and discussed. The velocity regime of a typical entry mission shows the large Mach and Reynolds-number range involved, and the classification of Mach regimes leads to the conclusion that re-entry flight covers all the speed regimes: subsonic and transonic (Sect. 3.2), supersonic (Sect. 3.3), and hypersonic (Sect. 3.4). In the latter we will also introduce a few engineering approximations to calculate the pressure distribution (and thus the aerodynamic forces and moments) that can be used in conceptual-design phases. In the upper layers of the atmosphere the entry vehicle—but also satellites in low Earth orbit—will experience the atmosphere that is no longer a continuum. Particle dynamics and rarefied flow become important, which will be treated in Sect. 3.5. Section 3.6 continues this chapter and the focus there will be on aerodynamic heating aspects, and how one can estimate these with simplified models. The last Sect. 3.7 has been added as a kind of an appendix, with a summary of the Navier-Stokes equations. A detailed treatment is beyond the current scope, the reader is referred to literature for a detailed treatment.

In preparing this chapter, use has been made of many sources, most notably the books by Anderson (2011, 2019), Bertin (1994), and Hirschel (2015).

3.1 Basics

3.1.1 Flow-Field Parameters

In Sect. 2.3 the atmosphere was introduced as an ideal gas, for which the equation of state Eq. (2.52) holds, which relates pressure with density and temperature. In popular wording this equation is also known as *the ideal gas law*. For convenience, this law is restated below:

$$p = \rho RT = \frac{RT}{v} \tag{3.1}$$

with p being the pressure in the gas in N/m^2, ρ is the density in kg/m^3, T is the temperature in K, and R is a constant specific for the gas in J/(kg K). The alternative formulation is shown on the right-hand side, with v being the specific volume, i.e., the volume per unit of mass (and thus the reciprocal of ρ). In an ideal gas it is assumed that all intermolecular forces can be neglected, and the particles will only experience non-elastic collisions. The particles themselves are defined having a certain collision cross-section diameter d, from which expressions for the particle collision frequency and mean free-path, among others, can be derived. The mean free-path, λ, will

be important for the discussion of rarefied flow in Sect. 3.5, and an approximated expression is given by:

$$\lambda = \frac{1}{n\pi\bar{\sigma}^2} \tag{3.2}$$

Here, n is the average number of particles per unit volume and $\bar{\sigma}$ is the average particle diameter ($\bar{\sigma} = 3.69 \cdot 10^{-10}$ m or 3.69 Å), irrespective of atmospheric composition. The mean free path represents the mean distance that a particle will travel without experiencing a collision with other particles. Regan and Anandakrishnan (1993) provide a more accurate expression for λ that accounts for the motion of the (individual) particles, i.e.,

$$\lambda = \frac{m_p}{\sqrt{2}\pi\sigma_p^2\rho} \tag{3.3}$$

where m_p and σ_p are the particle's mass and collision diameter, respectively. The particle mass can be calculated using the molar mass and Avogadro's number ($N_A = 6.02214076 \cdot 10^{23}$):

$$m_p = \frac{M}{N_A} \tag{3.4}$$

In the case where the actual composition of the atmosphere is known, with M and σ_p specified for the different chemical components, the weighted average over all components can be used ($\bar{M}$ and $\bar{\sigma}$, following from Eqs. (2.93) and (2.94), respectively). The weighing is done using the elements' number densities, provided by the atmosphere model (see for more details Sect. 2.3). Up to an altitude of about 100 km, one can safely assume $\bar{M}$ (= 28.94 kg/kmol) and $\bar{\sigma}$ (= 3.69 Å) to be constant, but due to a change in the molecular composition of the atmosphere at higher altitudes, a gradual decrease of both quantities occurs, see also Fig. 2.37.

The first law of thermodynamics states that the change of internal energy (i.e., the energy of a closed system going from state 1 to state 2) is equal to the amount of heat Q supplied to the system, minus the amount of work W done by the system on its surroundings:

$$\mathcal{E}_2 - \mathcal{E}_1 = Q - W \tag{3.5}$$

For a constant-pressure process, for a volume change from $\mathcal{V}_1$ to $\mathcal{V}_2$, W is given by

$$W = p(\mathcal{V}_2 - \mathcal{V}_1) \tag{3.6}$$

Substituting Eq. (3.6) into Eq. (3.5) and regrouping yields:

$$(\mathcal{E}_2 + p\mathcal{V}_2) - (\mathcal{E}_1 + p\mathcal{V}_1) = Q \tag{3.7}$$

For the constant-pressure situation under consideration, the heat flow into the system is given by

$$Q = C_p(T_2 - T_1) \tag{3.8}$$

where C_p is a proportionality constant, aptly named the heat coefficient at constant pressure.

Now introduce a new parameter, namely the enthalpy $\mathcal{H}$, defined as the sum of the gas's internal energy, $\mathcal{E}$, and its ability to do work by means of pressure acting on the volume of gas, $p\mathcal{V}$:

$$\mathcal{H} \triangleq \mathcal{E} + p\mathcal{V} \tag{3.9}$$

such that for a thermally perfect gas

$$\mathcal{H}_2 - \mathcal{H}_1 = C_p(T_2 - T_1) \tag{3.10}$$

The specific enthalpy, h, is the enthalpy per unit of mass, and is obtained by dividing the left and right side of Eq. (3.9) by the gas's mass:

$$h \triangleq e + pv = e + \frac{p}{\rho} \tag{3.11}$$

such that Eq. (3.10) is rewritten as

$$\mathrm{d}h = c_p(T_2 - T_1) \tag{3.12}$$

where c_p is now the *specific* heat coefficient at constant pressure. More formally, c_p, as well as c_v, the specific heat at constant volume, are defined as follows:

$$c_p = \left(\frac{\partial h}{\partial T}\right)_p \qquad c_v = \left(\frac{\partial e}{\partial T}\right)_v \tag{3.13}$$

From the above definitions, c_p and c_v can be interpreted as the increase of energy and enthalpy when temperature increases. In contrast to c_v, c_p is also affected by how much the gas expands, i.e., how much work is done. This is due to the fact that to keep the pressure constant with increasing temperature, the (specific) volume should increase, see Eq. (3.1). It can therefore be concluded that $c_p > c_v$. For air, at $T = 300$ K, the values are $c_p = 1{,}005$ J/(kg K) and $c_v = 718$ J/(kg K).

At (low) temperatures (for air at 1 atmosphere up to about 800 K), the flow can be assumed to be calorically perfect. A calorically perfect gas is one where the specific heats are constant. Both h and e are a function of temperature only, and are expressed by Anderson (2011):

$$h = c_pT \qquad e = c_vT \tag{3.14}$$

The physical cause of this assumption is that the only degrees of freedom that the particles have are translational and rotational, so that the molecules may be considered "rigid rotors".

A related quantity to c_p and c_v is the *ratio* of specific heats γ, which is defined as:

$$\gamma \triangleq \frac{c_p}{c_v} \tag{3.15}$$

and is, for obvious reasons, a dimensionless parameter, with a value of $\gamma = 1.4$. Furthermore, the gas constant, R, relates to c_p and c_v as follows:

$$R = c_p - c_v \tag{3.16}$$

and using the definition of γ, R can be expressed alternatively as:

$$R = c_p \frac{\gamma - 1}{\gamma} \tag{3.17}$$

The commonly used value for R for the Earth atmosphere is $R = 287$ J/kg K, but its value, as well as those for c_p, c_v and γ, changes with composition (and temperature, if the gas cannot be considered ideal) at higher altitudes.

The ratio of specific heats is fundamental in calculating the speed of sound, a, which is the velocity with which a pressure disturbance propagates through the flow. The corresponding expression for a calorically perfect gas is:

$$a = \sqrt{\gamma RT} \tag{3.18}$$

Alternative expressions are obtained by using Eq. (3.1) (by substituting $T = \frac{p}{R\rho}$) and (3.14) (by substituting $T = \frac{h}{c_p}$):

$$a = \sqrt{\gamma RT} = \sqrt{\gamma \frac{p}{\rho}} = \sqrt{(\gamma - 1)h} \tag{3.19}$$

Yet another quantity that is of interest in the study of flow fields around bodies, is the entropy s. It is a measure of the unavailability of a system's energy to do work, or equivalently, a measure of the randomness of particles in a system. Entropy is a quantity that plays a key role in the behaviour of gases through the second law of thermodynamics, which states that (Anderson 2011):

$$\mathrm{d}s = \frac{\delta q}{T} + \mathrm{d}s_{irrev} \tag{3.20}$$

where δq is the actual amount of heat added, T is the temperature of the gas, and ds_{irrev} is the generation of entropy due to the irreversible, dissipative phenomena

of viscosity, thermal conductivity, and mass diffusion occurring within the system. These dissipative phenomena *always* increase the entropy, so $ds_{irrev} > 0$[1].

For reversible processes, one can write

$$\delta q - p\mathrm{d}v = \mathrm{d}e \tag{3.21}$$

which, together with Eq. (3.20) (and $ds_{irrev} = 0$), leads to

$$T\mathrm{d}s = \mathrm{d}e + p\mathrm{d}v = \mathrm{d}h - v\mathrm{d}p \tag{3.22}$$

where the differential form of the enthalpy definition, Eq. (3.11), has been used. With the help of Eq. (3.14) and using the ideal gas law, this can be written as

$$\mathrm{d}s = c_p \frac{\mathrm{d}T}{T} - R\frac{\mathrm{d}p}{p} \tag{3.23}$$

and after integration between two states 1 and 2, this yields

$$s_2 - s_1 = c_p \ln \frac{T_2}{T_1} - R \ln \frac{p_2}{p_1} \tag{3.24}$$

Processes for which $ds = 0$ are termed isentropic, and for such a process the so-called *isentropic relations* can be used to relate thermodynamic quantities at two points on a streamline:

$$\frac{p_2}{p_1} = \left(\frac{T_2}{T_1}\right)^{\frac{\gamma}{\gamma-1}} = \left(\frac{\rho_2}{\rho_1}\right)^{\gamma} \tag{3.25}$$

which are based on the assumption of a calorically perfect gas.

During the course of this chapter, most notably when discussing aerodynamic-heating aspects, the so-called stagnation point or stagnation area is an important concept. At this point the flow is assumed to come adiabatically to rest, and the related properties pressure, density, temperature, etc. are known as total (or stagnation) properties. When the flow comes to rest, this would result in an increase of p, ρ and T. A formulation often used for the total enthalpy is the following:

$$h_0 = h + \frac{1}{2}u^2 \tag{3.26}$$

[1] In the case a system for which the internal energy increases by de is considered, there can be infinite combinations of heat that can be added and work done on the system (also known as processes) to achieve this. The three most relevant processes for us are: (i) Adiabatic process, where no heat is added to or taken away from the system, (ii) Reversible process, where no dissipative phenomena occur, i.e., there are no effects of viscosity, thermal conductivity, and mass diffusion, and (iii) Isentropic process, which is both adiabatic and reversible.

For calorically perfect gasses, the total temperature can be obtained from the above equation, using Eqs. (3.14), (3.17) and (3.18).

$$h_0 = c_p T_0 = h + \frac{1}{2}u^2 = c_p T + \frac{1}{2}u^2 \tag{3.27}$$

or

$$c_p T_0 = c_p T + \frac{1}{2}M^2 \gamma R T = c_p T \left(1 + \frac{\gamma R}{2c_p} M^2\right) \tag{3.28}$$

and thus

$$\frac{T_0}{T} = 1 + \frac{\gamma - 1}{2} M^2 \tag{3.29}$$

Assuming that the stagnation process is also reversible (and thus isentropic), Eq. (3.25) can in that case be combined with Eq. (3.29) to determine the total pressure and density. Although only valid for an isentropic deceleration of a calorically perfect gas, it is a frequently used approach.

3.1.2 Similarity Parameters

Throughout this and following chapters, re-entry vehicles and missions will be introduced and compared. The flow fields over these vehicles may be different, but to remove dependencies on, for instance, vehicle shape and flight altitude, it is common practise to use so-called similarity parameters, such as the Mach and Reynolds number. In that way one can determine whether the flows are dynamically similar, which is the case if: (i) the streamline patterns are geometrically comparable, (ii) the distributions of V/V_∞, p/p_∞, T/T_∞, etc. throughout the flow field are identical when plotted against common non-dimensional coordinates, and (iii) the force coefficients are equal. Here, "∞" indicates the free-stream (or *infinity*) conditions; without subscripts, the parameters are related to local conditions.

In a simple manner of speaking, the similarity parameters should be the same for both flows. The most commonly used similarity parameters are defined and explained below.

The Mach number, M, is a measure for how fast the flow moves with respect to the speed of sound, or the speed at which information about flow disturbances is propagated:

$$M = \frac{V}{a} \tag{3.30}$$

with V being the flow velocity and a the speed of sound. Looking at ratios of different forms of energy in fluid flows, the Mach number can physically be interpreted as being proportional to the ratio of ordered and random (kinetic) energy:

$$\frac{\text{ordered energy}}{\text{random energy}} = \frac{\frac{1}{2}mV^2}{\frac{1}{2}m\bar{v}^2} = \frac{V^2}{3RT} = \frac{\gamma}{3}\frac{V^2}{a^2} = \frac{\gamma}{3}M^2 \tag{3.31}$$

where $\bar{v}$ is the mean molecular velocity, which can be approximated by $\bar{v}^2 = 3RT$ (Hankey 1988).

The Reynolds number, Re, is defined as

$$Re = \frac{\rho V L}{\mu} = \frac{V L}{\nu} \tag{3.32}$$

where L denotes a characteristic length of the geometry under consideration, e.g., the vehicle length or wing chord. It may also be replaced with a running variable x to indicate local conditions; in that case, the Reynolds number is indicated by Re_x. In Eq. (3.32), μ and ν are the dynamic and kinematic viscosity, respectively. Concerning physical interpretation, the Reynolds number indicates the relative strength of the viscous forces with respect to the inertial forces. A high Reynolds number generally indicates a small viscous contribution and a thin, laminar boundary layer.

$$\frac{\text{inertial force}}{\text{viscous force}} = \frac{\dot{m}V}{\tau A} = \frac{\rho A V^2}{\mu \frac{V}{L} A} = \frac{\rho V L}{\mu} = Re \tag{3.33}$$

with $p = mV$ being the momentum of the gas, and $\tau = \mu \frac{\partial u}{\partial z}$ being the shear stress. In Sect. 2.3.2.1, Sutherland's equation was introduced as an accurate approximation for μ:

$$\mu = \beta \frac{T^{1.5}}{T + S} \tag{3.34}$$

with $S = 110.4$ K and $\beta = 1.458 \cdot 10^{-6}$.

The use of the Reynolds number is also crucial when doing wind-tunnel experiments. Due to the scale of the wind-tunnel models—more often than not the model is a (much) smaller representation of the actual vehicle/wing/other element—it means that without compensating in another way, the Reynolds number goes down. This is, of course, because the characteristic length is in the numerator of Re. Also, the temperature of the wind-tunnel flow is commonly not the same as the actual flow temperature (see also Sect. 3.4.3, where we discuss the issue of real-gas effects), which has an effect on the viscosity. To compensate for this by increasing the wind-tunnel flow-velocity is not so straightforward, as this will affect the Mach number—which also suffers from a temperature effect, because of the speed-of-sound's dependency on temperature. So, in global terms it is already difficult (and for many flow situations, impossible) to match the real conditions, let alone for localised flow problems,

such as boundary-layer transition from laminar to turbulent flow, which is, amongst others, a function of Reynolds number (see Sect. 3.4.2). The above identified issues are but a few of the reasons that flight testing is necessary, because the actual flow conditions cannot be reproduced in ground-based facilities.

The Prandtl number, Pr, gives ratio of kinematic viscosity and thermal diffusivity (ratio of thermal conductivity and volumetric heat capacity):

$$Pr = \frac{\mu c_p}{k} \tag{3.35}$$

with k being the thermal conductivity in $\frac{\mathrm{W}}{\mathrm{mK}}$.

$$\frac{\text{viscous dissipation energy}}{\text{energy conducted}} = \frac{\tau A V}{qA} = \frac{\mu \frac{V^2}{L}}{\frac{kT}{L}} = \frac{\mu c_p}{k}\frac{V^2}{c_p T} = Pr\frac{M^2}{\gamma - 1} \tag{3.36}$$

In the above equation, $q = -k\frac{\partial T}{\partial z}$ is the heat transfer ($\mathrm{W/m^2}$) (in one dimension; the general form is given by Fourier's law, $\mathbf{q} = -k\nabla T$).

The Prandtl number is a measure of how well energy that is released through dissipation is conducted through the gas. It is a property of the flow, independent of the body under consideration, although high temperature effects do affect its magnitude due to variations in μ, c_p and k. For air, $Pr = 0.71$ is a good approximation at standard conditions, but in analytical approaches also $Pr \approx 1$ is often used. Correlations for high-temperature air are given by Hansen (1958). It should be noted that the results given there show that Pr never exceeds 1 and decreases to values of order of 0.01 for very high temperatures.

The Lewis number, defined as[2]

$$Le = \frac{\rho D c_p}{k} \tag{3.37}$$

relates the particle's diffused heat to its conducted heat (Hankey 1988):

$$\frac{\text{heat diffused}}{\text{heat conducted}} = \frac{\dot{m}_i h_0}{qA} = \frac{\rho D \frac{\partial C_i}{\partial z} A h_0}{k\frac{\partial T}{\partial z}A} = \frac{\rho D h_0}{kT} = \frac{\rho D c_p}{k}\left(\frac{h_0}{c_p T}\right) = Le\left(\frac{h_0}{c_p T}\right) \tag{3.38}$$

In Eq. (3.37), D is the mass diffusivity in $\mathrm{m^2/s}$ and is related to the mass flow of species i in a mixture as $\rho D\frac{\partial C_i}{\partial z}$, with $\frac{\partial C_i}{\partial z}$ being the gradient in concentration of species i. In case there are only two species involved, D is the mixture-averaged diffusion coefficient (as is it is usually assumed to be). Considering the temperature and density/pressure range of interest in this book, $0.5 \lessapprox Le \lessapprox 1.5$ (Hansen 1958).

[2] In textbooks on chemical engineering the Lewis number is defined as the inverse of the definition that is adhered to in aerothermodynamics literature.

The local heat-transfer coefficient can be indicated by one of several defined parameters, such as the Nusselt number, Nu, or the Stanton number, St (Anderson 2019). The Nusselt number defines the ratio of convective to conductive heat transfer at a boundary in a fluid:

$$Nu = \frac{q_w x}{k_e(T_{aw} - T_w)} \tag{3.39}$$

The Stanton number measures the ratio of heat convected to the body compared to the heat contained in the fluid (i.e., its thermal capacity):

$$St = \frac{C_h}{\rho V c_p} \tag{3.40}$$

where C_h is the heat transfer coefficient and c_p is the specific heat at constant pressure. This equation can be re-written for a body, where q_w is the heat transferred at the wall, h_{aw} and h_w are the adiabatic wall and wall enthalpies, respectively, and the subscript e represents the conditions at the edge of the boundary layer:

$$St = \frac{q_w}{\rho_e u_e(h_{aw} - h_w)} \tag{3.41}$$

In the above two expressions, q_w is the local heat-transfer rate (energy per second per unit area) at the wall, x is the distance along the wall measured from the leading edge, k_e is the thermal conductivity at the edge of the boundary layer, and T_{aw} and h_{aw} are the adiabatic-wall temperature and enthalpy, respectively. The Stanton number will be used mostly in Sect. 3.6; in case St is based on free-stream density and velocity it will be denoted by St_∞. Note that the Nusselt and Stanton numbers are related through $Nu = St \cdot Re \cdot Pr$.

For evaluating to what degree the flow can be considered a continuum, the Knudsen number, Kn, is used:

$$Kn = \frac{\bar{v}/L}{\bar{v}/\lambda} = \frac{\lambda}{L} \tag{3.42}$$

with $\bar{v}$ the average particle velocity, λ the mean free path of a particle in the flow, and L a characteristic body length. In simpler words, the Knudsen number gives the ratio between the number of particle collisions with the vehicle and the number of particle collisions with other particles. Knudsen numbers as low as 10^{-3} indicate continuum flow, and values above 10 generally mean free-molecular flow. This will be discussed in more detail in Sect. 3.1.5.

3.1.3 Conservation Laws

For the derivation of the normal and oblique shock equations (Sect. 3.3), we will use the basic conservation laws of mass, momentum, and energy. As the derivations will be done for two-dimensional flow, the general form of the conservation laws will be stated in 2D as well. Given a flow with density, ρ, and velocity, $\mathbf{v}$, and assuming a control surface, A, the general 2D-form of the conservation law of mass is given as a surface integral:

$$\iint_A \rho(\mathbf{v} \cdot \hat{\mathbf{n}}) \mathrm{d}A = 0 \tag{3.43}$$

where $\hat{\mathbf{n}}$ is the outward normal-vector on each of the sides of the control surface. The above equation basically means "what goes in must come out", i.e., no additional mass is added to (or extracted from) the control volume.

The general 2D-form of the Law for Conservation of Momentum is given by

$$\iint_A \rho \mathbf{v}(\mathbf{v} \cdot \hat{\mathbf{n}}) \mathrm{d}A + \iint_A p\hat{\mathbf{n}} \mathrm{d}A = 0 \tag{3.44}$$

with p being the pressure acting on the boundaries of the control surface.

Finally, the standard form for the Energy Equation is formulated as

$$\iint_A \rho(e + \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v})(\mathbf{v} \cdot \hat{\mathbf{n}}) \mathrm{d}A + \iint_A p(\mathbf{v} \cdot \hat{\mathbf{n}}) \mathrm{d}A = 0 \tag{3.45}$$

where e is the specific internal energy.

3.1.4 Forces and Moments

Any object in a moving medium will experience a force, being it the result of a pressure difference or friction between the object's exterior and the flow particles. Integrating over the complete surface area will result in the aerodynamic force, and if the resulting force components are not passing through the centre of mass, also an aerodynamic moment. A typical example of aerodynamic forces are the drag and lift force that one may recall from flight-performance studies. These forces can be made non-dimensional such that between different (aircraft) configurations there is a way to compare them. For example, the drag and lift coefficients, C_D and C_L, are thus formulated as:

$$C_D = \frac{D}{\bar{q}_\infty S_{ref}} \qquad C_L = \frac{L}{\bar{q}_\infty S_{ref}} \tag{3.46}$$

with D and L being the drag and lift force in N, S_{ref} a reference surface area in m^2, and the free-stream dynamic pressure in N/m^2, $\bar{q}_\infty$, defined as

$$\bar{q}_\infty = \frac{1}{2}\rho V_\infty^2 = \frac{1}{2}\gamma p M_\infty^2 \tag{3.47}$$

Here, ρ is the density of the medium (let us assume air for the sake of convenience), and V_∞ is the relative speed of the object (i.e., the vehicle) with respect to the free stream (or, equivalently for a calorically ideal gas, its Mach number, M_∞).

With the expressions for the coefficients at hand one can always re-calculate the corresponding forces acting on the vehicle, given its altitude (= density) and velocity, *if* the coefficients are constant, which unfortunately is not always the case. In a more general way, D and L are then written as

$$D = C_D(M_\infty, \alpha, \ldots)\bar{q}_\infty S_{ref} \qquad L = C_L(M_\infty, \alpha, \ldots)\bar{q}_\infty S_{ref} \tag{3.48}$$

As indicated above, C_D and C_L are a function of M_∞ and the vehicle's angle of attack, α, or in more general terms, its attitude with respect to the oncoming flow. The ellipsis in each equation suggests that there may be more dependencies, which could be, for instance, a Reynolds effect, or the deflection of a control surface (in other words, a change in vehicle shape). To complicate matters, one must realise that the (local) Mach number may vary at different locations of the vehicle.

Calculation of the aerodynamic coefficients is very much flow-regime dependent, but in any case the division between pressure (inviscid effect, normal to the vehicle surface) and friction (viscous effects, tangential to the surface) components holds. Accurate prediction of the two components for all velocity regimes is not always trivial, and even though various approximations exist for different regimes, these do not always match the required accuracy, especially not in transition phases between regimes, or when one is interested in very specific details about the flow around the body. In those cases one wants to resort to Computational Fluid Dynamics (CFD) simulations.[3]

However, during conceptual design of entry vehicles prediction of forces and moments with reasonable accuracy is often sufficient. The flight through the hypersonic velocity regime has the most impact on a re-entry vehicle. In terms of forces and moments, these are (mostly) determined by pressure forces alone, and once can refrain from the more difficult computation of the viscous effects. So, for the brief (mathematical) discussion in the remainder of this section we will ignore friction, and the aerodynamic forces and moments, $\mathbf{F}_a$ and $\mathbf{M}_a$, can be determined from the distribution of the pressure p over the closed vehicle surface as follows:

[3] The CFD methodology is an approach, where the flow properties around the vehicle is solved by discretising a volume around (part of) the vehicle into millions of cells, and solving the (non-linear) flow equations for each cell, taking the proper boundary conditions into account. The flow properties in the "surface" cells will lead to the surface pressure and friction components that can then be used to calculate aerodynamic force and moment coefficients, as well as the thermodynamic conditions at the vehicle wall.

$$\mathbf{F}_a = \oiint_S \left(-p\hat{\mathbf{n}}\right) \mathrm{d}S \tag{3.49}$$

$$\mathbf{M}_a = \oiint_S -p\left(\left(\mathbf{r} - \mathbf{r}_{ref}\right) \times \hat{\mathbf{n}}\right) \mathrm{d}S \tag{3.50}$$

where $\mathbf{r}_{ref}$ is a chosen reference point about which the aerodynamic moments are evaluated, $\mathbf{r}$ is a position on the vehicle surface and $\hat{\mathbf{n}}$ is the outward unit normal. These equations can be used to evaluate $\mathbf{F}_a$ and $\mathbf{M}_a$ in two ways, either by deriving analytical relations for the pressure distribution over the vehicle surface and (numerically) integrating the above, or by discretising the vehicle surface into a number of panels on each of which the pressure is determined analytically or numerically. In case of discretising the surface, the above integrals become summations over all vehicle panels that approximate the surface, so

$$\mathbf{F}_a = \sum_{i=1}^{n_p} -p_i A_i \hat{\mathbf{n}}_i \tag{3.51}$$

$$\mathbf{M}_a = \sum_{i=1}^{n_p} -p_i A_i \left(\mathbf{r}_i - \mathbf{r}_{ref}\right) \times \hat{\mathbf{n}}_i \tag{3.52}$$

where n_p is the number of panels. The vector $\mathbf{r}_i$ now denotes the centroid of the panel, p_i the (mean) pressure on the panel, A_i the panel area, and $\hat{\mathbf{n}}_i$ the panel unit normal. Non-dimensionalising the aerodynamic forces and moments as before, yields the following expressions for the aerodynamic force and moment coefficients (approximate, for a finite number of panels):

$$\begin{pmatrix} C_X \\ C_Y \\ C_Z \end{pmatrix} S_{ref} \approx -\sum_{i=0}^{n_p} C_{p_i} A_i \hat{\mathbf{n}}_i \tag{3.53}$$

$$\begin{pmatrix} C_l \\ C_m \\ C_n \end{pmatrix} S_{ref} \approx -\sum_{i=0}^{n_p} C_{p_i} A_i \left(\mathbf{r}_i - \mathbf{r}_{ref}\right) \times \hat{\mathbf{n}}_i \tag{3.54}$$

These coefficients commonly denote the non-dimensionalised forces and moments in a body-fixed reference frame (it is stressed that one should not confuse the roll- and yaw-moment coefficients, C_l and C_n, with the lift and normal force coefficients, C_L and C_N. In some texts considering two-dimensional motion only, the latter two are occasionally written with lower-case subscripts.). In the above expressions, C_p denotes the pressure coefficient, which is defined as:

$$C_p = \frac{p - p_\infty}{\bar{q}} = \frac{2}{\gamma M_\infty^2}\left(\frac{p}{p_\infty} - 1\right) \tag{3.55}$$

Note that the term involving p_∞ in the summation drops out, since integrating a constant value over a closed surface evaluates to zero.

Before moving on, a remark about the reference geometry, i.e., S_{ref}, c_{ref}, and b_{ref}, is in order. It may be more obvious which quantities to use for aircraft, e.g., the wing-surface area, wing chord, and wing chord. For winged entry vehicles this may also be clear as they resemble aircraft. For entry capsules, however, it may be less clear. Is the surface area the area of the flow-facing heat shield? Is the characteristic length the (maximum) diameter? The reader should realise that it *does not matter* which values one chooses, as long as one is consistent in normalising the aerodynamic coefficients, and *using the exact same values* to go back to forces and moments. Therefore, it is crucial that when aerodynamic data are gathered from literature, one should always confirm which were the reference values with which the coefficients were obtained. A consistent aerodynamic database should therefore include not just the aerodynamic coefficients, but *also* the reference geometry.

In subsequent chapters we will be using the aerodynamic forces and moments extensively. As we will see in Chap. 5, the (individual) lift and drag forces determine the thermo-mechanical loads on the re-entry vehicle, whereas the lift-to-drag *ratio* will drive the flight range it can cover: the higher this ratio, the better it is. Another important aerodynamic characteristic is the pitch moment. At the beginning of the entry, any winged vehicle will fly at large angle of attack, thus generating a large pitch moment. For the vehicle to remain stable and be easy to control, it should be *trimmable*, i.e., its body flaps and/or elevators should be able to provide a counter moment, despite the rapidly changing environment and flight regime. The moment-generating capabilities of the other control surfaces is a crucial input to the design of attitude controllers, as will be made clear in Chap. 7.

Now that the importance of knowing the aerodynamic characteristics of re-entry vehicles is established, and given a general formulation for the forces and moments as a function of pressure distribution, the step to flow-field description will be made. The flow field in combination with the vehicle's velocity regime will determine to a large extent how the pressure distribution around the vehicle can be determined.

3.1.5 Flow-Field Description

When the number of particles in (a layer of) the atmosphere is very high it would automatically lead to a very high number of interparticle collisions. The relaxation time for particles moving through such a flow field will be very short, and also the relaxation distance will be very small. If this relaxation distance, or, in the case of larger numbers of particles, the mean free path is small with respect to a characteristic length (of, for instance, the re-entry vehicle), then the macroscopic flow properties, such as pressure, temperature, and density, can be considered to change continuously. In that case the gas can be treated as a continuum. On the other hand, when the opposite is true, i.e., low number of particles, large mean free path and long relaxation times, the state of the individual particles and their interaction with other particles

and the boundaries must be considered. In that situation we talk about rarefied or free-molecular flow. And of course there is a regime in between, where the flow is neither continuum nor rarefied, which is therefore called the transition regime.

As mentioned in Sect. 3.1.2, the Knudsen number is typically used to indicate whether a flow is continuous or rarefied. Regan and Anandakrishnan (1993) define the Knudsen lines of demarcation to be 0.01 and under for continuum flow, 10 and higher for free-molecular flow, and the range in between for the transition regime. However, they also state that there are no sharp boundaries for when the flow characteristics change. For re-entry trajectories also another classification can be used, i.e., one based on a combination of Mach and Reynolds number: (i) free molecular flow occurs when $M/Re > 3$, (ii) transition takes place when $M/Re < 3$ and $M/\sqrt{Re} > 0.01$ (with $Re \gg 1$), and (iii) continuum flow happens when $M/\sqrt{Re} < 0.01$.

In Fig. 3.2 the three main flow regimes, i.e., continuum flow, free-molecular flow and the transition-flow regime in between are shown, in relation to several parameters. In the left figure, the Knudsen number (in relation to the mean free path length of $\lambda = 0.01$ and 10 m, respectively) constrains the transition regime between 86 and 131 km. This means that even for a re-entry from low Earth orbit, this regime should be taken into account. As we will see in Sect. 3.5, there is a special bridging function to be used, which means that for that regime we need to have knowledge about free-molecular and continuum regime. In the right figure, the criteria based on Mach and Reynolds number have been plotted. This alternative way of dividing the flow regimes, nicely links to, for instance, Fig. 3.4, which will be discussed shortly.

In terms of physical effects, the division is more or less the same, i.e., the continuum regime is discussed first, and via the viscous merged layer regime the transition regime will be covered, and through near-free molecular flow free molecular flow is reached. The individual aspects will now be briefly discussed.

Continuum regime

In this region, classical fluid mechanics of high Reynolds number apply, and the shock waves and boundary layer are treated as discontinuities. This regime is often subdivided into speed regimes, such as incompressible (subsonic), transonic, supersonic, and hypersonic, with boundaries between the speed regimes set by the Mach number. There is an example of a re-entry trajectory flying through the Mach and Reynolds regimes below.

Viscous-merged-layer regime

At high Mach numbers, the shock waves appear very close to the vehicle wall and interfere with the boundary layer. Thus, the viscous effects in the forming shock wave and boundary layer must be treated simultaneously. The presence of a boundary layer near the wall alters the boundary conditions for the shock wave, while at the same time the large pressure gradients due to the shock wave strongly alter the boundary layer flow.

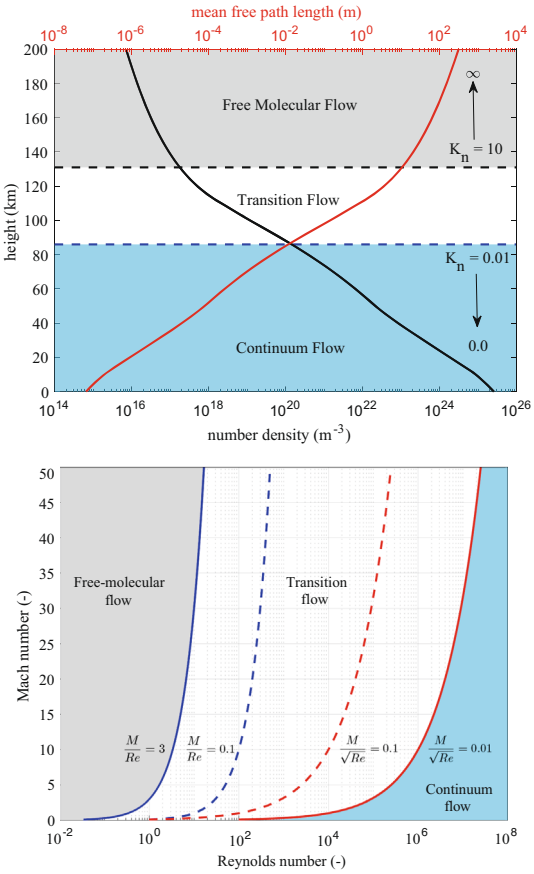


Fig. 3.2 Division in three main flow regimes, with lines of demarcation, defined by Knudsen number (top), or a Mach/Reynolds combination (bottom). Characteristic length is set at $L = 1$ m, such that $Kn = \lambda$

Transition regime

This regime represents a transition between the near-free molecular flow and the viscous-merged-layer regime. It is very difficult to treat analytically and only little is known about aerodynamic quantities such as lift, drag, or heat transfer. A typical engineering approach is to evaluate the properties in either adjoining regime and estimate these quantities by careful interpolation between the two.

Near free molecular flow

The gas molecule acquires the momentum of the boundary only after there have been several collisions. If on the average a molecule does not acquire the energy corresponding to the energy of the boundary after a single collision, one says that there is a lack of "accommodation". This regime is also called the slip-flow regime, because the fluid particles do not have a have a zero velocity relative to the boundary.

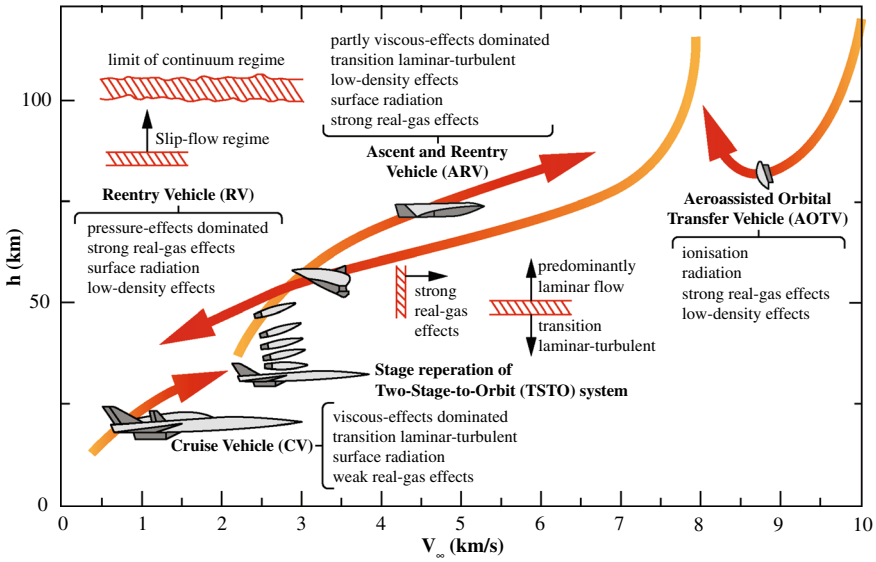


Fig. 3.3 Four categories of hypersonic vehicles and encountered aerodynamic phenomena. Image inspired by Hirschel (1991)

The no-slip condition for viscous fluids assumes that the fluid will acquire zero relative velocity.

Free molecular regime

The molecular mean free path is large compared to a characteristic body dimension, such as the nose radius. Molecules that collide with a moving boundary attain the state of that boundary after a single collision. This regime is encountered by satellites in very low Earth orbit, but also by spacecraft leaving the International Space Station to make an entry into the Earth's atmosphere.

In Fig. 3.3 four categories of hypersonic vehicles with their representative trajectories and encountered aerodynamic phenomena are shown: re-entry vehicles, ascent and re-entry vehicles (typically single-stage-to-orbit spaceplanes), cruise vehicles, and aero-assisted orbital transfer vehicles.

A re-entry vehicle coming from space experience (very) low-density free-molecular-flow effects above 100 km (for Earth). Despite the high velocity, dynamic pressure is still very low. Going through transition towards the hypersonic regime the vehicle encounters strong real-gas effects (above about Mach 12–15). The atmosphere cannot be considered an ideal gas at these high Mach numbers and altitude, and the initial phase of the hypersonic flight is dominated by strong heating effects, including surface radiation. In terms of aerodynamic forces, the pressure effects dominate the viscous effects.

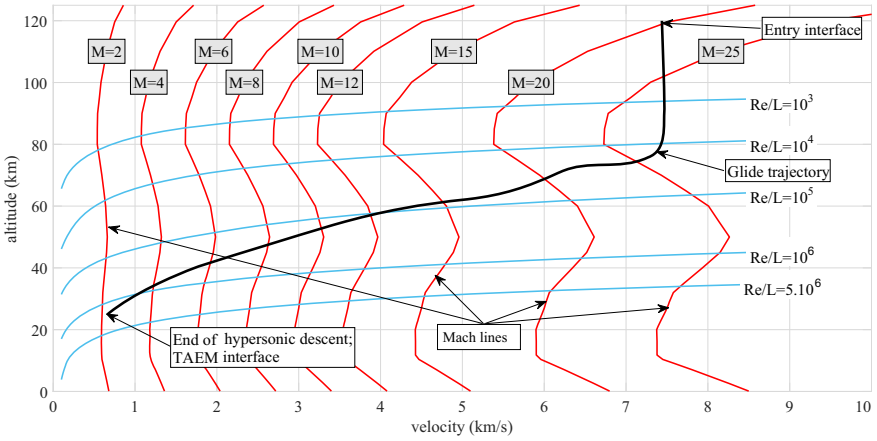


Fig. 3.4 The Mach and Reynolds regime of a typical gliding flight of a space-shuttle like vehicle (TAEM = Terminal Area Energy Management, see Chap. 8)

Ascent and re-entry vehicles will experience slightly different phenomena during the ascent flight, such as phases where the viscous effects dominate the pressure effects, and the transition from a laminar to turbulent boundary layer (typically below 50 km, at sufficiently high velocity). For the re-entry portion of its mission a vehicle will encounter the same aspects as mentioned before. Cruise vehicles are meant to fly at lower altitudes albeit with a reasonably high Mach number (between 5 and 10, although the higher limit may be very challenging for the system's thermal protection system). During its flight, viscous effects dominate, causing high aerodynamic heating, but also the surface radiation becomes high. The boundary layer trips from laminar to turbulent, and because of its lower Mach number the real-gas effects are weak at best.

Finally, the aero-assisted orbital transfer vehicles are vehicles that use the atmosphere to change their orbital parameters. Coming from space with high velocity, they remain in the upper layers of the atmosphere only, so they experience free-molecular flow and partially continuum flow. It is the high velocity, though, that induces very strong real-gas effects, such as ionisation and radiation of the plasma. In terms of forces, they are predominantly induced by low-density effects.

The re-entry flight of a Space-Shuttle-like winged re-entry vehicle is plotted in Fig. 3.4. A typical re-entry mission begins from low-Earth orbit and enters the noticeable atmosphere at 120 km with a relative velocity of about 7.4 km/s. As can be seen, the initial Mach number is about $M \approx 20$. Due to the temperature-dependency of the speed of sound, and the particular temperature profiles as a function of altitude (see Fig. 2.30), initially the Mach number increases to a value of about $M = 28$. Then, the gliding flight shows a gradual decrease in Mach number down to $M \approx 2.5$, when the hypersonic descent has ended and the final part of the vehicle's mission begins: the terminal area flight that brings the vehicle to the landing region.

In terms of Reynolds number, it begins at low values and at the beginning of the gliding flight it is about 10,000 m^{-1}. The Reynolds number only slowly increases, and the boundary layer can remain laminar for a very long time. Measurements on Space-Shuttle flights showed that even a small surface roughness due to gap fillers between the TPS tiles could trip the boundary layer. For instance, for STS-28 it occurred at $M = 18$, and for STS-55 this happened at $M = 14$, at Reynolds numbers of roughly between 80,000 and 100,000 m^{-1} (McGinley et al. 2006).

Obviously, the encountered Mach and Reynolds range depends on the entry velocity of the vehicles at a given altitude. For instance, the Apollo capsules returning from the Moon arrived from a hyperbolic orbit with a velocity of 11 km/s at 220 km altitude, and Mars sample return missions would have an even higher capture velocity, i.e., on the order of 12–15 km/s. However, in the end, all vehicles will pass through the atmosphere via the same velocity regimes:

(a) Hypersonic, with a Mach number greater than $M > 5$ (to be discussed in Sect. 3.4)
(b) Supersonic, $1 < M < 5$ (Sect. 3.3)
(c) Transonic, $0.8 < M < 1.2$ (Sect. 3.2)
(d) Subsonic, $M < 1$ (Sect. 3.2).

These velocity regimes will be discussed in the upcoming three sections, beginning with low-speed aerodynamics and moving up in Mach range, where the flow aspects become more complex. However, before doing so, some aspects of numerical and experimental techniques will be mentioned.

3.1.6 Numerical and Experimental Techniques

To determine the flow fields around objects, their aerodynamic characteristics, and, at the high speeds of hypersonic flight, the thermal loading on the object's surface, one can resort to dedicated flight experiments, either on existing vehicles or on test vehicles that are developed for this purpose alone. One can imagine that this is not a flexible approach: not only does it take a long time to design, build, and fly the vehicle, it is also very expensive. Fortunately, there are some alternatives that are typically used interchangeably: wind tunnels and numerical simulations (CFD). However, there are pros and cons to each of these techniques: wind tunnels cannot fully reproduce the operational conditions of re-entry vehicles, and the outcome of CFD codes can only be accepted when these codes are fully validated, and have the functionality to simulate said operational conditions. Validation of the codes can in part be done by comparing with wind-tunnel data, but given the earlier mentioned limitation, the better way to do this would be to do flight experiments. And so the circle is round again.

In this section we will briefly describe some aspects of both numerical and experimental techniques, albeit cursory. The interested reader is referred to the vast volume of literature on in particular CFD. Here, we mention the book by Anderson (1995),

introducing the basics of CFD, and the three-volume text by Hoffmann and Chiang (2000a, b, c), which is used in introductory, intermediate, and advanced courses in computational fluid dynamics and computational fluid turbulence. Cebeci et al. (2005) discuss the use of CFD from an engineering point of view, and have included a set of useful computer programs in their book, and, finally, Lewis et al. (2004) review the fundamentals of the finite-element method for heat and fluid flow. As a starting point for studying the use of wind tunnels, the book by Kaushik (2019) connects theoretical and experimental aerodynamics; even though that wind tunnels are covered in only a single chapter it gives a good overview. Chanetz et al. (2020) provide a more elaborate treatment of all types of wind tunnels, and also cover the measurement and visualisation techniques.

3.1.6.1 Computational Fluid Dynamics

Computational fluid dynamics is the science of predicting fluid flow, heat transfer, mass transfer, chemical reactions, and related phenomena, by solving the mathematical equations (e.g., the Navier-Stokes equations described in Sect. 3.7) that describe these processes using numerical techniques. One can use the outcome of this process for (rapid) analysis of conceptual designs, for detailed predictions of dimensioning load cases beyond conceptual-design studies, and to support testing and experimentation to reduce the total effort required in the laboratory.

Depending on the problem at hand, mathematical models and their numerical counterparts, representing one or more physical elements of the problem (e.g., turbulence, compressibility, chemistry, multiphase flow, etc.), should be available in the software. At all times, the fundamental laws of conservation of matter, momentum, and energy must be satisfied throughout the region of interest. Fluid properties are commonly modelled empirically. Whenever possible, simplifying assumptions can (and should) be made to reduce problem complexity and make the computations manageable. One may think about flows that are, for instance, steady-state, incompressible, inviscid, two-dimensional, etc. These assumptions can impact the accuracy of results, particularly for highly complex or unsteady flows.

To solve the physical problem, it has to be discretised, which results in approximations of the (algebraic) governing equations in the fluid region of interest. Typically, one would start with a detailed two-dimensional or three-dimensional model of the vehicle to be analysed. For the more complex shapes, such as a winged re-entry vehicle, such a model is made in computer aided design (CAD) software. An example is provided by Fig. 3.5, which shows a CAD model of the HORUS vehicle (see also Appendix B). Special care should be taken of the control surfaces, if analyses for a range of control-surface deflections are required.

The next step involves the definition of the computational domain around the object. This domain includes the volume where fluid flow will be simulated and analysed. It should extend sufficiently far to capture the important flow features. This domain is subsequently discretised into a large number of cells, the total of which is called the mesh. One can distinguish between the surface mesh, i.e., following

the contours of the vehicle (see also Fig. 3.6), and the volumetric mesh representing the flow-field domain (see Fig. 3.7). One can choose between two different mesh types, i.e., a structured mesh, where the grid lines are organised in a structured, regular pattern, and an unstructured mesh, where the grid elements are irregularly shaped and can adapt to complex geometries more flexibly. The mesh is commonly generated using specialised software tools. This involves dividing the computational domain into discrete elements (e.g., cells, elements, or nodes), and fine-tuning the mesh density near critical regions. This can be the surface of the object, where flow interactions are most significant. The fine-tuning is often achieved through mesh-refinement techniques.

The mesh should always be inspected to ensure that it has good-quality elements to prevent numerical instability and inaccuracies. This involves checking for skewed elements, aspect ratios, and other mesh-quality metrics. Once the quality of the mesh has been established, boundary conditions have to be assigned to it, specifying how the fluid interacts with the object's surface and the domain boundaries. These conditions include inflow, outflow, no-slip walls, and others.

Once the problem has been properly setup and an initial solution is provided, the set of algebraic equations are solved numerically (on a computer) for the flow field variables at each node or cell. The system of equations is solved simultaneously to provide solutions for each of the cells, and, once converged, the solution is post-processed to extract quantities of interest, such as the force and moment coefficients, heat transfer, separation, pressure loss, to name but a few. Convergence is reached when changes in the solution variables are below a (user-defined) tolerance. One can inspect the residuals to help monitor this trend, and to confirm that overall property conservation is achieved. It is noted that the accuracy of a converged solution depends on the mesh resolution and independence, as well as the problem setup. It goes without saying that the appropriateness and accuracy of the physical models play a crucial role in how close the solution is to "real life".

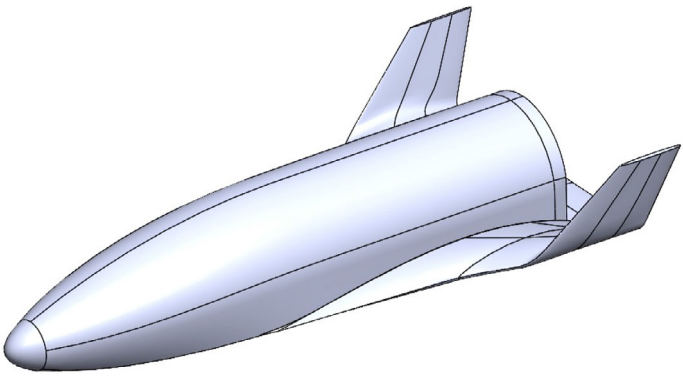


Fig. 3.5 CAD model of the HORUS re-entry vehicle

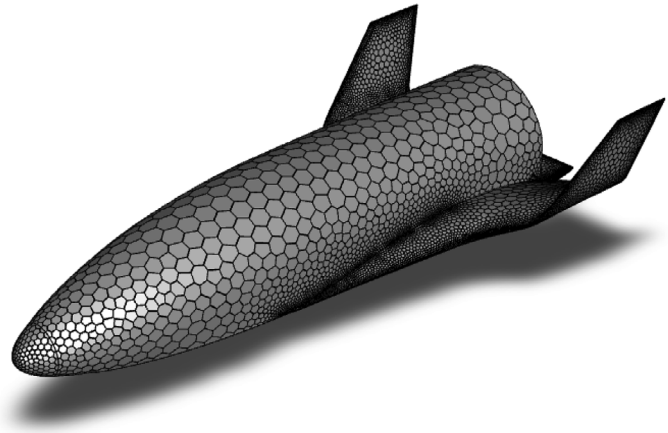


Fig. 3.6 Surface mesh of the HORUS re-entry vehicle in SolidWorks. Courtesy of Jeroen Reijne

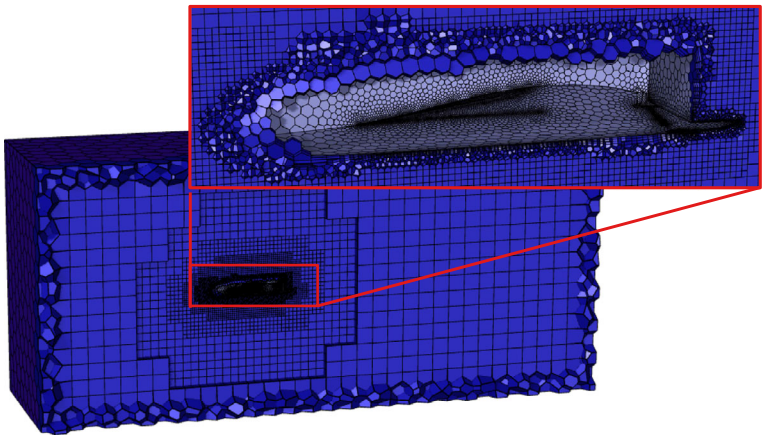


Fig. 3.7 Volume mesh of the HORUS re-entry vehicle in SolidWorks. Courtesy of Jeroen Reijne

The use of CFD has a few advantages. Once the software is available, the cost are relatively low compared to using physical experiments and tests, to get essential engineering data for design. This is specifically true when testing complex geometries or a wide range of conditions, as it reduces the need for building physical prototypes and conducting multiple wind-tunnel tests. CFD allows for rapid simulations and analysis of fluid-flow behaviour. This speed accelerates the design-iteration process and enables engineers to explore a larger design space in a shorter time. The quick turnaround means also that engineering data can be introduced early in the design process. By adjusting various parameters, such as geometry, boundary conditions, and operating conditions, optimisation of specific performance criteria can then be accommodated.

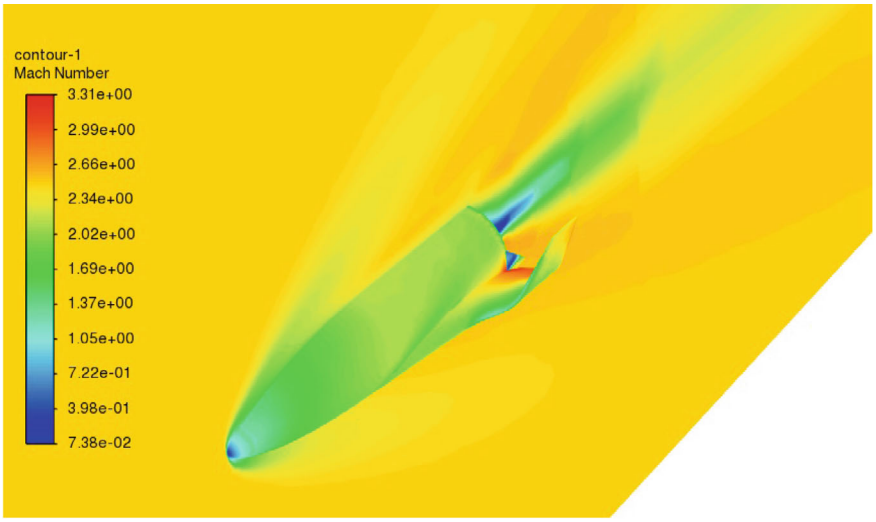


Fig. 3.8 Local Mach-number distribution for the HORUS re-entry vehicle at a free-stream condition of $M_\infty = 2.5$, obtained with SolidWorks. Courtesy of Jeroen Reijne

Some of the flow and heat-transfer processes may be challenging or even impossible to replicate in wind-tunnel tests (in particular at hypersonic conditions), and CFD can thus provide the insight that would otherwise be lacking. Alternatively, it can also be used to simulate ideal conditions, as it allows great control over the physical process. Doing so, gives the engineer the ability to isolate specific phenomena for study, such as heat-transfer processes with adiabatic, constant heat flux, or constant temperature boundaries. In contrast to experiments, where data can typically only be extracted at a limited number of locations in the system due to the installed sensors, with CFD a large number of locations in the region of interest can be analysed, which results in an extensive database of flow parameters for examination, including pressure distributions, velocity profiles, and turbulence patterns. An example of this is shown in Fig. 3.8, where the local Mach-number distribution for the HORUS re-entry vehicle at a free-stream condition of $M_\infty = 2.5$ is plotted.

However, CFD has its limitations and disadvantages as well. Using CFD software is not trivial, and its effective use requires specialised training and experience. Engineers must be skilled in setting up simulations, selecting appropriate models, and interpreting results correctly. It requires a solid understanding of fluid dynamics, numerical methods, and software tools; misinterpretation of results due to inadequate knowledge can easily lead to inaccurate conclusions. Generating a suitable computational mesh that accurately represents the geometry and captures the flow physics is a critical step, and requires a lot of expertise, although the quality of automatic mesh generators has improved significantly over time. Poor mesh quality or inappropriate mesh resolution can lead to inaccurate results.

A major issue is the verification and validation of CFD simulations against experimental data, a crucial step to ensure accuracy. This process can be time-consuming and may require additional physical testing. The lack of data makes this process complicated and sometimes impossible. One should always keep this in mind, and not blindly accept the results of CFD simulations. Numerical errors are nowadays, with increasing word size, less of an issue, even though solving equations on a computer invariably introduces numerical errors. These include round-off errors, due to finite word size available on the computer, and truncation errors due to approximations in the numerical models. The latter can be reduced to (almost) zero with mesh refinement. Of course, one should never forget that the accuracy of the CFD solution is only as good as the initial/boundary conditions provided to the numerical model. And, these models may have their inherent limitations, particularly turbulence models, in capturing certain flow behaviours accurately. Choosing the right model for a specific application is essential.

To conclude this brief discussion, a few words on the different levels of complexity of CFD simulations, ranging from simple to highly complex. The level of complexity depends on various factors, including the physics of the flow being studied, the accuracy required, the available computational resources, and the expertise of the CFD practitioner. Three main levels of complexity in CFD are:

1. Simplified or reduced-order models: These models are used for quick and approximate analyses. They involve simplifications and assumptions to reduce computational requirements and complexity. Commonly used for preliminary design studies and quick assessments. They are suitable for cases where qualitative insights are sufficient, but may lack accuracy for detailed predictions.
2. Reynolds-averaged Navier-Stokes (RANS) simulations: RANS simulations average out unsteady and turbulent flow effects, providing steady-state or time-averaged results. They require moderate level of complexity and computational resources, and are suitable for simulating many engineering flows, such as the flow around vehicles.
3. Large eddy simulation (LES) and direct numerical simulation (DNS): LES resolves large-scale turbulent structures and models the smaller scales. DNS directly solves all scales of turbulence, resulting in highly accurate, but computationally demanding simulations. High complexity and substantial computational resources are needed. The simulations are suitable for capturing detailed turbulence effects and complex flow phenomena, like turbulent mixing and combustion.

Within each level of complexity, there are further considerations:

(a) Geometry: The complexity of the geometry being simulated impacts the CFD setup. Simple geometries are easier to model accurately, while complex shapes might require more advanced techniques.
(b) Mesh: Mesh generation becomes more challenging for complex geometries. Structured meshes are simpler, but might not adapt well to complex shapes. Unstructured meshes allow for greater flexibility but require more computational resources.

(c) Physics Models: Different turbulence models and physical models are available at different complexity levels. Simple models like the laminar flow assumption, are used in simplified models, while more complex models are used in higher-fidelity simulations.
(d) Software and Resources: Complex simulations require specialised CFD software and powerful computational resources. High-fidelity simulations demand substantial memory and processing power.
(e) Validation: The more complex the simulation, the more challenging validation becomes. Validation against experimental data is crucial to ensure the accuracy of the simulations.

It is important to choose the appropriate level of complexity based on the specific goals of the simulation, the available resources, and the desired accuracy. Overcomplicating a simulation can lead to unnecessarily long computation times without providing significantly more accurate results, if the additional complexity is not justified by the physics of the problem.

On a final note, the outcome of CFD simulations can be trusted to a certain extent, but there are several factors that need to be considered to ensure the reliability and accuracy of the results. CFD simulations can provide valuable insights into fluid-flow behaviour, but their accuracy depends on various factors, including the quality of the simulation setup, the chosen models, the level of validation, and the expertise of the practitioner. CFD simulations have limitations in capturing certain flow behaviours, such as unsteady or complex flow phenomena. Users should always be aware of the limitations of CFD in reproducing certain physics. It is important to recognise that it is a tool that should be used alongside experimental data and engineering judgment, especially for critical applications.

3.1.6.2 Wind Tunnels

Wind tunnels are essential tools in the field of aerodynamics for studying the behaviour of objects, such as aircraft, spacecraft, vehicles, and other structures, in a controlled and simulated airflow environment. These tunnels help researchers and engineers understand how these objects interact with air, providing valuable insights into their performance, stability, and efficiency. Wind tunnels consist of a test section, where models or prototypes of the objects under study are placed. Air is forced through the tunnel at varying speeds to simulate the conditions the object would experience during real-world operation.[4] By measuring the forces, pressures, and aerodynamic characteristics of the object in the tunnel, engineers can refine designs,

[4] It is important to note that the flow situation in the wind tunnel is actually opposite to the real world, where the vehicle is moving and the atmosphere (without wind) is static. However, the wind-tunnel experiment is based on the principle of motion reciprocity according to which displacement of a body relative to gas or liquid can be replaced by gas or liquid flow around the body at rest. This is an example of *Galileo's Principle of Relativity*, which, in general terms, states that "It is impossible by mechanical means to say whether we are moving or staying at rest". This is clear from the example of being in one train that moves with the same speed and in the same direction

optimise performance, and ensure safety. One may also study how the air moves around the object by injecting smoke or dye into the tunnel and photographing its motion around the object.

Figure 3.9 shows the layout of a subsonic wind tunnel, with a closed measurement section. Several key components can be identified. First, there is a pre-chamber that supplies high-pressure gas, which is accelerated through a nozzle to achieve a specific velocity (V_∞) before entering the working section. The diffuser zone of the nozzle converts the gas's kinetic energy into a pressure potential. A (powerful) wind-tunnel drive unit, such as a fan or compressor, compensates for energy losses caused by factors, like friction and shock waves. For wind tunnels using compressed gas bottles, reciprocating compressors increase gas pressure.

These components are common in various wind-tunnel types, including subsonic and supersonic ones, although they can differ in design and size. The working section can be open or closed, and in open sections, an altitude chamber called an Eiffel chamber separates the gas jet from the atmosphere. The choice between open and closed sections depends on the ease of observation and recording of the measurements.

The working section typically has a cylindrical shape with various cross-sectional shapes like circular, rectangular, elliptic, or polygonal, depending on the object being tested (e.g., oval is most fitting for an aircraft and round for rockets). In addition to these basic components, wind tunnels may have specific elements like gas heaters, straightener vanes, lattices to regulate the gas flow, and return ducts in closed-circuit wind tunnels. These elements help control the gas flow, reduce turbulence, and optimise testing conditions. In closed-circuit wind tunnels, moisture-removal devices are used to prevent distortions in flow patterns caused by excessive humidity. Continuous-action wind tunnels incorporate gas-cooling heat exchangers to manage temperature increases due to friction and drive operation.

The wind tunnel must meet the conditions required to facilitate correlation of experimental data with benchmark test results, obtained under similar conditions on other setups, and, if available, flight data. That means that there are tight requirements on the operating conditions, as well as flow quality, such as velocity-field uniformity, the level of longitudinal pressure gradient, and the intensity of turbulence of the free stream. Disturbances may propagate far downstream distorting the stream. Therefore, the inner surfaces of the tunnel must be thoroughly finished, and have faired bend turns and changes in cross section. To prevent separation of the boundary layer, much attention should be given to shaping the regions with stream stagnation.

Depending on the size of the measurement section, one can either test the full-scale vehicle (for instance, in NASA's largest subsonic wind tunnel a Boeing 737 can be fitted, see below), or one has to resort to scaled-down models of aircraft or spacecraft. These models allow for cost-effective testing while still providing accurate insights into full-scale behaviour. Proper scaling ensures that the aerodynamic characteristics of the model accurately represent those of the full-scale vehicle. The models need to

---

as another train: a passenger will not be able to observe whether they are moving or at rest. So, for the wind-tunnel experiment, under the condition of a uniform rectilinear motion, the stream force and heat effect on a body are identical in both direct and reverse motion.

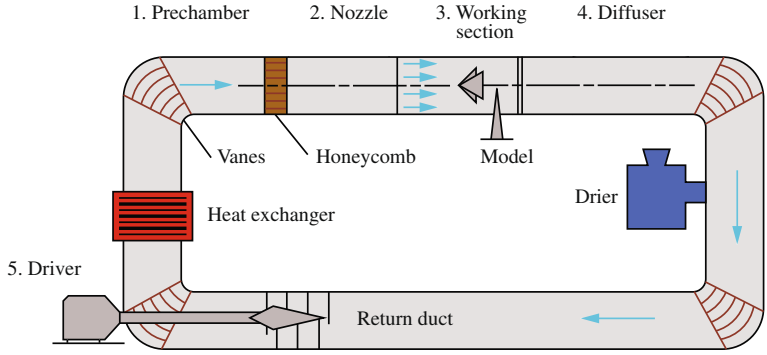


Fig. 3.9 Schematic overview of a (subsonic) windtunnel

be of high quality, as the smallest imperfection may scale to unacceptable proportions, when the model is scaled back to the original vehicle size. Figure 3.10 shows two examples of high-quality wind-tunnel models, one of the Blackbird, the YF-12 version, and one of the Space Launch System (SLS). The latter model was used for acoustic testing to identify vibrational problems during launch.

The model has to be carefully mounted in the test section, and precisely positioned. With the mounting system the model orientation and angle of attack can be controlled during the testing. Data-acquisition software records and analyses the data on aerodynamic forces, pressure distributions, temperature, and other relevant parameters, obtained from sensors and instruments. These may include pressure taps to measure pressure variations on the model's surface, a force balance to measure lift, drag, and other forces acting on the model, and anemometers and pitot-static probes to measure airflow velocity and pressure. Finally, a tunnel-control system is required to adjust and control the wind tunnel's airflow velocity, temperature, and other parameters. These systems ensure that the testing conditions are consistent and can be adjusted as needed.

It may be clear that for a wind-tunnel campaign adequate funding and resources to build, maintain, and operate the wind-tunnel facility are required. A further consideration of costs is associated with model fabrication, instrumentation, and data analysis. At different moments in the campaign, experts are required to plan, support, and execute the campaign.

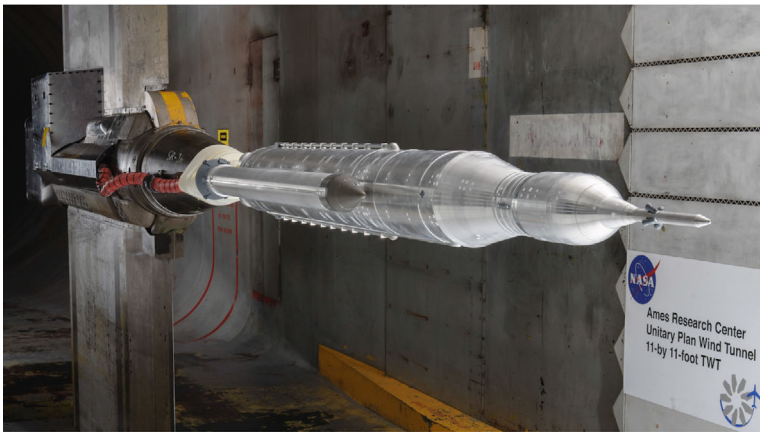
Wind tunnels are categorised based on the range of airspeeds they can generate, each suited for different types of applications. We conclude this section with a brief overview.

Subsonic

Subsonic wind tunnels simulate airflow at speeds below the speed of sound (Mach 1), typically used for testing ground vehicles, aircraft, and low-speed aerodynamics.



(a) Blackbird (YL-12)



(b) Space Launch System (SLS)

Fig. 3.10 Examples of wind-tunnel models in two of NASA's wind tunnels. Images: NASA

Subsonic tunnels are the most common type and come in various sizes, accommodating different scales of models for testing. Subsonic tunnels have either an open-loop or closed-loop configuration. In open-loop tunnels, air is drawn from the environment, compressed, and then directed through the test section. After passing through the test section, the air is usually released back into the atmosphere. These tunnels are suitable for testing large-scale models and full-sized aircraft components. In closed-loop tunnels, the air is recirculated through a closed circuit. This allows for more precise control over the airflow conditions and temperature. Closed-loop tunnels are often used for high-precision testing, such as evaluating small-scale models and intricate components.

Another categorisation can be based on the speed regime. Low-speed wind tunnels are designed to simulate the low-speed flight conditions experienced during takeoff, landing, and cruising of aircraft. They are crucial for studying lift, drag, and other aerodynamic characteristics at lower speeds. High-speed subsonic wind tunnels, on the other hand, are used to simulate near-cruising speeds and investigate transonic effects, which occur as an aircraft approaches the speed of sound (see also below). They help researchers understand issues like shock waves and wave drag.

Some subsonic wind tunnels are equipped with facilities to study aircraft noise and acoustics. This is crucial not only for developing quieter aircraft and understanding noise propagation patterns, but also to study the effect of noise-induced vibrations. A particular example of flow visualisation is provided in Fig. 3.11, which shows a (subsonic) wind-tunnel test to investigate what happens during launch, if the main engine of a Space Shuttle fails. The test was conducted in May 1988 in a wind tunnel at the John H. Glenn Research Center.

To conclude, an example of a large subsonic wind tunnel is shown in Fig. 3.12: NASA's 14×22 ft subsonic wind tunnel at NASA's Langley Research Center in Hampton, Virginia. It is known as such because of its test section with a height of 14.5 ft (4.42 m), a width of 21.75 ft (6.63 m), and a length of 50 ft (15.24 m). Despite its large dimensions, it is not even the largest one in the world. That honour is reserved for the tunnel at NASA's Ames Research Center.[5] This subsonic tunnel, which can test planes with wing spans of up to 100 ft (30.5 m), is over 1,400 ft (427 m) long and 180 ft (55 m) high! Its two test sections are 80 ft high and 120 ft wide (24.4 m × 36.6 m), and 40 ft × 80 ft (12.2 m × 24.4 m), respectively. Air is driven through these test sections by six 15-bladed fans, powered by six 22,500-horsepower motors. Each fan has a diameter equal to the height of a four-story building. In this wind tunnel, also the 16-m diameter parachute of Mars Science Laboratory was tested,[6] see also Fig. 9.126.

Transonic

[5] https://www.nasa.gov/content/behold-the-worlds-largest-wind-tunnel, date accessed: 12 September, 2023.

[6] https://www.nasa.gov/mission_pages/mars/images/msl0164-20090422.html date accessed: 12 September, 2023.

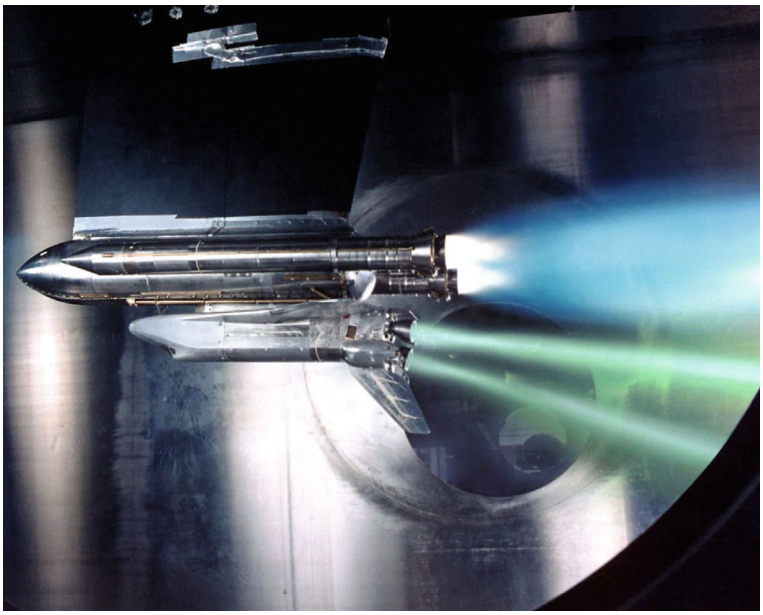


Fig. 3.11 Wind-tunnel test to simulate engine failure of the Space Shuttle during launch. Image: NASA



Fig. 3.12 Aerial images of the 14 × 22 ft subsonic wind tunnel at NASA's langley research center in hampton, Virginia. Original image: NASA

Transonic wind tunnels simulate airflow around the speed of sound, typically in the range of speeds around and just below the speed of sound (Mach 0.8 to Mach

1.2). This is a critical range as airflow behaviour changes significantly as an object approaches and goes beyond the speed of sound. Studying these effects is crucial for understanding the complex flow phenomena that occur in this speed regime. It encompasses the region where airflow around the aircraft can become supersonic in some areas and subsonic in others, leading to phenomena such as shock waves, wave drag, and the "transonic barrier". The shock waves form as airflow accelerates and decelerates around the aircraft, and can lead to increased drag and affect the aircraft's performance, stability, and control.

Transonic wind tunnels are often designed to accommodate larger-scale models compared to purely supersonic tunnels. This allows for more accurate representation of the real-world aerodynamic behaviour of aircraft. Accurate representation of the boundary layer is crucial in transonic testing, as it affects the interaction between the air and the aircraft's surface. Tunnel designs incorporate mechanisms to simulate realistic boundary layer behaviour. The tests provide valuable input for design modifications to reduce the effect of wave drag and enhance overall aircraft performance. They are also crucial for designing supersonic airfoils, which must maintain optimal performance both at subsonic and supersonic speeds. Finally, transonic wind tunnels help optimise the wing-sweep angle to ensure smooth airflow and minimal performance degradation; wings with swept-back configurations are commonly used in transonic and supersonic flight.

Transonic wind tunnels are different from other wind tunnels by the structure of the working section: their walls are perforated and slotted. The narrowest section in the wind tunnel will encounter a flow speed equal to the speed of sound. For that reason, in wind tunnels where this speed is reached, the effect of the model blocking up the cross section of the working section gives rise to local supersonic flows. This is called choking of the wind tunnel.

Just like their subsonic counterparts, transonic wind tunnels can be found all over the world: from aerospace research centres, universities, the larger aircraft manufacturers (e.g., Boeing and Airbus), government research facilities (e.g., NASA, ESA, JAXA, DLR), to military research centres.

Supersonic

Supersonic tunnels generate airflow speeds above the speed of sound (Mach 1). They are essential for studying the aerodynamics of supersonic flight and related phenomena, such as shock waves and sonic booms. Supersonic wind tunnels are used in the design and testing of supersonic and hypersonic aircraft.

The heart of a supersonic wind tunnel is the nozzle. It is designed with a convergent-divergent shape. The convergent section initially accelerates the airflow, and the divergent section further accelerates it to supersonic speeds. This acceleration is achieved by carefully shaping the nozzle to control the flow of air. To maintain continuous supersonic airflow, a compressor is used to compress and accelerate the air entering the wind tunnel. The compressor raises the pressure of the incoming air to match the pressure required for the specific test conditions.

As with the other tunnels, the test section is the part of the wind tunnel where the actual testing takes place, located downstream of the nozzle. The test section is designed to minimise disturbances and provide a stable supersonic flow. After passing through the test section, the supersonic airflow is typically slowed down and diffused to subsonic speeds in a diffuser section. The diffuser helps reduce turbulence and recover some of the kinetic energy from the airflow. The air is then exhausted.

A high-speed compressor is often used to supply and pressurise the air needed for the wind tunnel. The compressor raises the pressure of the incoming air to the levels required for supersonic flow. Compressors can be of various types, such as centrifugal or axial-flow compressors, and are chosen based on the facility's needs. In some supersonic wind tunnels, a driver gas is used to help accelerate and drive the airflow to supersonic speeds. Helium or nitrogen are commonly used as driver gases. These gases are introduced into the nozzle section of the wind tunnel and mixed with the test gas (usually air) to achieve the desired supersonic flow conditions. The driver gas can be stored in pressure vessels and injected into the system as needed.

Alternatively, a supersonic wind tunnel can be fully supplied by (compressed) air in a large pressure vessel. For instance, the supersonic wind tunnel at Delft University of Technology is based on such a design. As it is an older design, it is not equipped with a flexible-wall nozzle, but uses interchangeable sets of liners to obtain Mach numbers of either 1.5, 2, 2.5, or 3 in a 15 cm $\times$ 15 cm test section. It has a total running time on the order of 18 min before recharging the pressure vessel is required.[7]

Hypersonic

Hypersonic wind tunnels are used to simulate extremely high speeds, typically above Mach 5. These tunnels are crucial for understanding the aerothermodynamics of objects traveling at hypersonic speeds through the atmosphere. This type of tunnel simulates speeds that exceed Mach 5, which is over 6,000 km/hr or 1.7 km/s. These speeds are encountered during space travel, atmospheric re-entry, and high-speed missiles. At hypersonic speeds, vehicles experience extreme heat due to air compression and friction. Hypersonic wind tunnels are equipped to simulate these high-temperature conditions, allowing engineers to study thermal protection systems and materials that can withstand such heat. It should be mentioned, though, that testing TPS materials is better done in so-called arc-jet or plasma generators. For instance, the 1200 KW induction plasmatron at the Von karman Institute for Fluid Dynamics (Belgium) can provide heat fluxes up to 15 MW/m^2, albeit at supersonic flow conditions with a maximum run time close to 1 min.[8]

Hypersonic vehicles often generate complex shock-wave interactions at these speeds. Hypersonic wind tunnels allow researchers to understand how these shock

[7] https://www.tudelft.nl/lr/organisatie/afdelingen/flow-physics-and-technology/facilities/high-speed-wind-tunnels/st-15-supersonic-windtunnel, date accessed: 12 September, 2023.

[8] https://www.vki.ac.be/index.php/facilities-other-menu-148/plasma-facilities/50-research-and-consulting/facilities/plasma-facilities/71-1200-kw-induction-plasmatron, date accessed: 14 September, 2023.

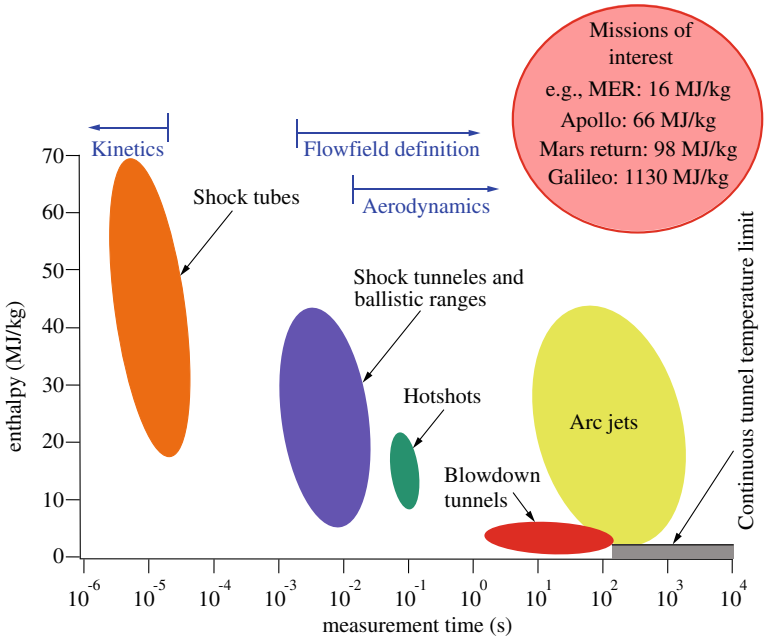


Fig. 3.13 Measurement range for several tunnels, in terms of enthalpy versus measurement time. Indicated are also the specific heat loads of several entry missions (Thermal & Fluids Analysis Workshop (TFAWS), Pasadena, CA, August 13–17, 2012. Data taken from the Aerothermodynamics course, Lecture #1: Stagnation Point Heating: https://tfaws.nasa.gov/TFAWS12/Proceedings/Aerothermodynamics%20Course.pdf, date accessed: 14 September, 2023)

waves interact with vehicle surfaces and affect vehicle stability. These tunnels reproduce the flow dynamics that occur during hypersonic flight, including the formation of shock waves, boundary-layer interactions, and rarefied gas effects that arise due to the low air density at high altitudes. They can be used to simulate the airflows and conditions faced by hypersonic propulsion systems, such as scramjets, which are air-breathing engines designed to operate at hypersonic speeds.

However, it is typically not feasible to reproduce the exact combination of flight conditions in terms of speed, Reynolds number, and enthalpy. Figure 3.13 gives a rough indication of the enthalpy versus measurement time (to be interpreted as actual duration of, for instance, an entry mission). It may be clear that due to the severe limitations in both heating environment and measurement time, wind-tunnel testing at high Mach number will not provide all the answers that a vehicle designer or an aerodynamicist may require. To acquire those, flight testing remains the only way. It should be noted, though, that for many tests the measurement time of around 100 ms is sufficient.

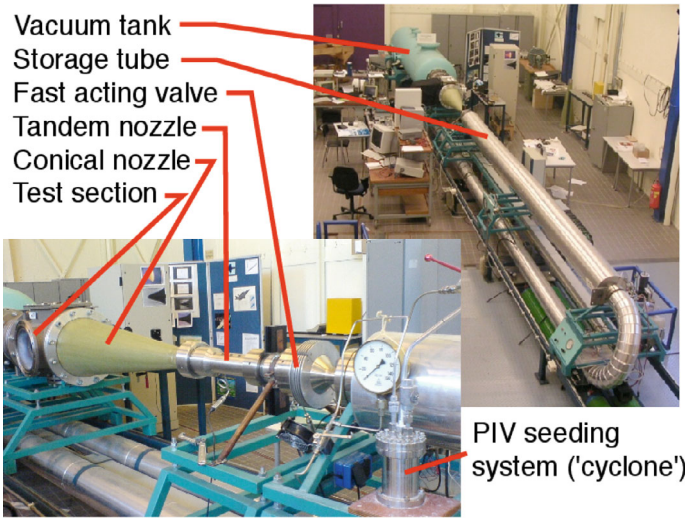


Fig. 3.14 The hypersonic test facillity Delt (HTFD) at Delft university of technology

Similar to other wind tunnels, hypersonic wind tunnels use scaled models to simulate full-scale behaviour. Proper scaling is essential to accurately replicate the aerodynamic behaviour of the actual vehicle. Since the measurement sections of hypersonic tunnels are usually relatively small, due to the large volumes of air involved, scaleability effects play a more important role than for other tunnels. In terms of layout and construction, hypersonic-tunnel design resembles that of supersonic tunnels. To prepare for the test, air is first compressed to high pressure by using a multistage compressor, and the high-pressure air is then passed through a dryer to remove the moisture content before it is stored in the large tanks. During operation, the compressed air is released through the tunnel, and axisymmetric convergent-divergent nozzles are generally used to expand the high-pressure and high-temperature air to the desired test-section Mach number. The nozzle throat is usually cooled by water to withstand the high temperature or high-enthalpy flow conditions.

As an example, Fig. 3.14 shows the hypersonic test facility Delft (HTFD) at Delft University of Technology, a relatively small facility based on the so-called Ludwieg-tube principle. Its test conditions cover a Mach-number range from 6 to 11, whereas the Reynolds number (referenced to the test-section diameter of 35 cm) may change from 3 to 8 million. Finally, the measurement time varies in between 0.07 and 0.13 s (70–130 ms). The figure shows the different main parts, to begin with a 29 m long storage tube, where dry air is stored under high pressure and temperature (6 m of the tube is heated). This tube is connected to the test section via a valve that can switch in about 1 ms, and a de Laval nozzle. This nozzle has a total opening angle of 150°, which provides a conical flow field to the test section. The last main component of the tunnel is the vacuum vessel, located directly after the test section, which acts as an expansion.

With a focus on developing hypersonic missiles, there is an interest in developing large, hypersonic wind tunnels, which, of course, can also be used for civilian applications. In this context, we mention the two large tunnels that have been built in China, i.e., the JF-12 and JF-22.[9] The former can reach speeds of 2.5–11.5 km/s (Mach 33). Its test section is 80 centimetres wide, twice as large as the largest high-speed wind tunnel that had been built to that time. The JF-22 underwent qualification testing in the first half of 2023; it has a total length of 167 m, and can reproduce flight conditions at altitudes of 40 to 90 km, with a maximum speed of up to 10 km/s (Mach 30). It is said that the flow will reach a temperature up to 10,000 °C, sufficient to break atoms apart.

3.2 Low-Speed Aerodynamics

Re-entry vehicles are typically designed for hypersonic flow, mainly to keep the aerodynamic heating below the constraint values. Entry capsules will spend the final portion of their mission under a (cluster of) parachute(s). The aerodynamics of the parachute dominate the flight, although local aerodynamic effects around the capsule may lead to oscillations of the capsule under the parachute. This particular field of aerodynamics, i.e., parachute aerodynamics and parachute-payload interaction, is beyond the scope of this introductory discussion on aerodynamics. For winged vehicles, the effect of low-speed aerodynamics is a bit more pressing, especially since the aerodynamic design is not optimised for this flight phase.

Even though we do not underestimate the importance of low-speed aerodynamics, it is not the main focus of this chapter, and therefore the discussion will remain brief. It is split into two parts, i.e., one on subsonic aerodynamics, where some more general statements about subsonic flow will be made, and one on transonic aerodynamics, which forms the bridge between subsonic and supersonic flow. For more detailed discussions, the interested reader is referred to one of the many text books on aerodynamics, e.g., those by Anderson (2011), Houghton and Carpenter (2003), Kaushik (2019), and Vos and Farokhi (2015).

3.2.1 Subsonic Flow

Subsonic aerodynamics describes flows for which the free-stream Mach number, M_∞, is always (and everywhere) smaller than 1, i.e., the fluid motion is (much) lower than the speed of sound. Therefore, for any body both windward and leeward surfaces are important for the calculation of the pressure distribution and thus the

[9] https://frontierindia.com/china-opens-hypersonic-capable-wind-tunnel-for-insane-speeds-of-33-mach/?expand_article=1 and https://newatlas.com/aircraft/china-jf22-hypersonic-wind-tunnel/, date accessed: 11 September, 2023.

aerodynamic coefficients. For wings, lift can be determined from circulation around airfoil.

Within the subsonic Mach-number range both incompressible ($M_\infty < 0.3$) and compressible ($M_\infty > 0.3$) flow is possible. This distinction means that for incompressible flow the amount of change of density in the flow is small, i.e., a constant density can be assumed, and the effects on the solution can be ignored.

Another distinction that can be made is, whether the flow is inviscid or viscous. In an inviscid flow the effects of viscosity can be completely neglected, which will have no significant effect on the flow solution. Inviscid means that if the flow would be considered as a stack of layers, these layers all move at the same speed without any friction in between them. This commonly holds for flow away from the vehicle wall: close to the wall, though, there is a thin layer in which viscous effects force the flow velocity to zero at the wall, the so-called *no-slip* condition. This thin layer is called the boundary layer, and will be discussed in detail later (Sect. 3.3.5).

A third characteristic is whether a flow is irrotational or rotational. Irrotational flow means that if we would consider each particle to be a cube, it only deforms, but does not rotate. If the cubes do rotate then the flow possesses vorticity.

The special case when the flow is incompressible, inviscid and irrotational is called "potential flow", and the governing equations that describe the flow can be simplified, such that a range of quick and easy solutions is available. For all the points in a potential flow, the flow properties at any one point depend, in theory, on the flow properties at all other points in the flow. This is due to the fact that the propagation of information about disturbances travels faster than the flow velocity, so that in a steady flow, the conditions at all locations will be "known" to the flow everywhere.

Even though in principle only applicable to incompressible flow, it is sometimes also used for compressible flows, as well as both stationary and non-stationary flows. Its main application is found for the outer flow field for aerofoils, but when there are strong vorticity effects, the potential flow approximation is not applicable. It is beyond the scope of this chapter to treat potential flow theory in detail, but the reader is directed to the earlier referenced books for more details on this method.

It may be obvious with the many aircraft designs and developments that there is a large industrial base for subsonic calculation methods and advanced CFD tools. Two (out of many) engineering tools that allow for very fast calculations, compared to the CPU-heavy CFD codes, are PAN-AIR and APAS. PAN-AIR is a panel-based method based on potential flow that solves the so-called Prandtl-Glauert equation for subsonic and supersonic flow (Erickson 1990). The workings are based on the fact that there are fundamental analytic solutions to the Prandtl-Glauert equation, which are the source, doublet, and vorticity singularities.

In panel methods, the (approximated) surface of the vehicle (aircraft) is divided into small quadrilateral portions, the *panels*. For each panel, the surface distributions of the aforementioned singularities are superimposed, and the resulting distribution will automatically satisfy the Prandtl-Glauert equation.

APAS, or Aerodynamic Preliminary Analysis System, is a panel method that incorporates, besides the potential-flow method for subsonic and supersonic flow, an impact-type finite element method to calculate solutions at hypersonic conditions. It

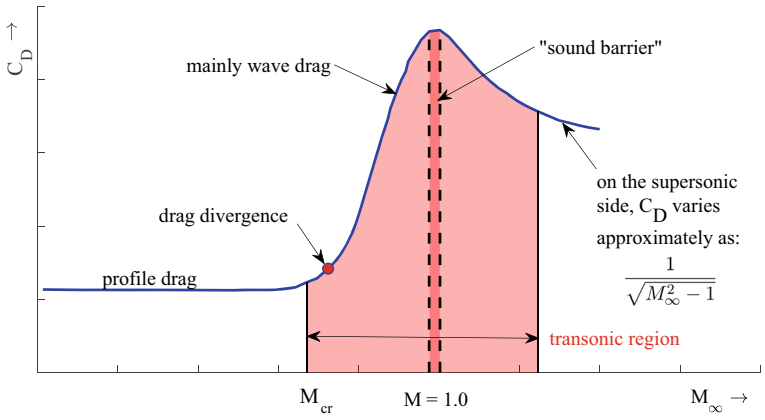


Fig. 3.15 The transonic-flow region: approaching supersonic-flow conditions ($M = 1$), the drag rapidly increases

is capable of handling three-dimensional configurations having multiple non-planar surfaces of arbitrary planform and bodies of non-circular contour. APAS can determine static, rotary, and control longitudinal- and lateral-directional characteristics (Bonner et al. 1981; Divan and Sova 1991).

3.2.2 *Transonic Flow*

As mentioned before, despite the flow region being called *subsonic*, with increasing Mach numbers (but still lower than $M = 1$), the flow around the vehicle will locally be larger and locally be smaller than one. The compressibility of the flow causes the formation of supersonic and subsonic pockets, leading to strong shifts in centre of pressure. This regime is called the *transonic flow regime*, and it is characterised by a sharp increase in drag (see Fig. 3.15). Whereas at low Mach numbers, the drag is mainly determined by the so-called parasitic drag, its magnitude driven by vehicle shape and wing profile. Viscous effects play an important role. At a certain critical Mach number, supersonic flow starts to form, and with increasing Mach numbers there is a strong increase in wave drag, due to the presence of shock waves. Contrary to the profile drag, wave drag is independent of viscous effects.

The drag peaks at the sonic Mach number and it led to the misplaced name *sound barrier*, as it was once believed that the "sound barrier" could not be "broken" due to this drag increase. Aircraft not designed to fly at supersonic speed were simply lacking the propulsive power to overcome the large drag.

In Fig. 3.16, the flow around an airfoil for increasing Mach numbers is shown. Up to a free-stream Mach number of $M_\infty \approx 0.7$ to 0.8, compressibility effects have only minor effects on the flow pattern and thus the drag. Even though the flow is subsonic

everywhere (a), locally the Mach number may be larger than M_∞, as the flow velocity increases when it follows the curvature of the airfoil. It is thus not hard to imagine that at the so-called critical Mach number a sonic point appears somewhere on the airfoil surface (b), usually near the point of maximum thickness.

With a further increase of M_∞, larger and larger regions of supersonic flow appear on the airfoil surface (c and d), and to return to subsonic conditions the supersonic flow passes through a shock wave.[10] The loss in velocity results in an increase in temperature and extracts energy from the flow, causing an increase in wave drag. Due to the interaction with the boundary layer, a shock-induced (boundary-layer) separation occurs immediately behind the shock. This results in a large increase in drag, and the Mach number for which this happens is called the drag-divergence Mach number. If an airplane has insufficient thrust to further accelerate, this will be its limiting Mach number.

When M_∞ approaches 1, there will be large supersonic regions and the shocks become very strong (e and f). After exceeding Mach 1 a bow shock appears around the airfoil nose, with a small subsonic pocket behind the shock, as well as a subsonic region behind the trailing edge shock wave. The supersonic flow begins to realign itself parallel to the body surface and stabilise, and the shock-induced boundary-layer separation is reduced. As a result, the drag reduces significantly.

In general, transonic flow is highly non-linear and unsteady, and the shock waves on the body surface may jump back and forth along the surface. The combination of supersonic flow, shock waves, and large subsonic pockets makes it very difficult to model transonic flow, as the effects of both of these flow types should be sufficiently captured by the methods that are used. There is, therefore, limited applicability of engineering tools to establish the aerodynamic coefficients, and the use of CFD tools is required.

To conclude, Fig. 3.17 beautifully captures the transonic flow around an airfoil. This schlieren image shows a nearly vertical shock wave, followed by boundary-layer separation that adversely affects lift, drag, and other flight parameters.

3.3 Supersonic Flow

With the description of transonic flow in the previous section, it was necessary to already introduce supersonic flow, with the formation of shock waves and their interaction with the boundary layer. In this section, these concepts will be explored in more detail. Since most aspects of supersonic flow occur in hypersonic flow as well, albeit with an effect of the increasing Mach number, the discussion will be quite elaborate. After a brief introduction, Sect. 3.3.2 begins by describing the formation of a stationary normal shock in one-dimensional flow. Section 3.3.3 zooms in on the

[10] In the next section, the fundamental aspects of supersonic flow will be discussed in detail. Most notably, the formation of normal and oblique shock waves, expansion waves, and boundary layers will pass review.

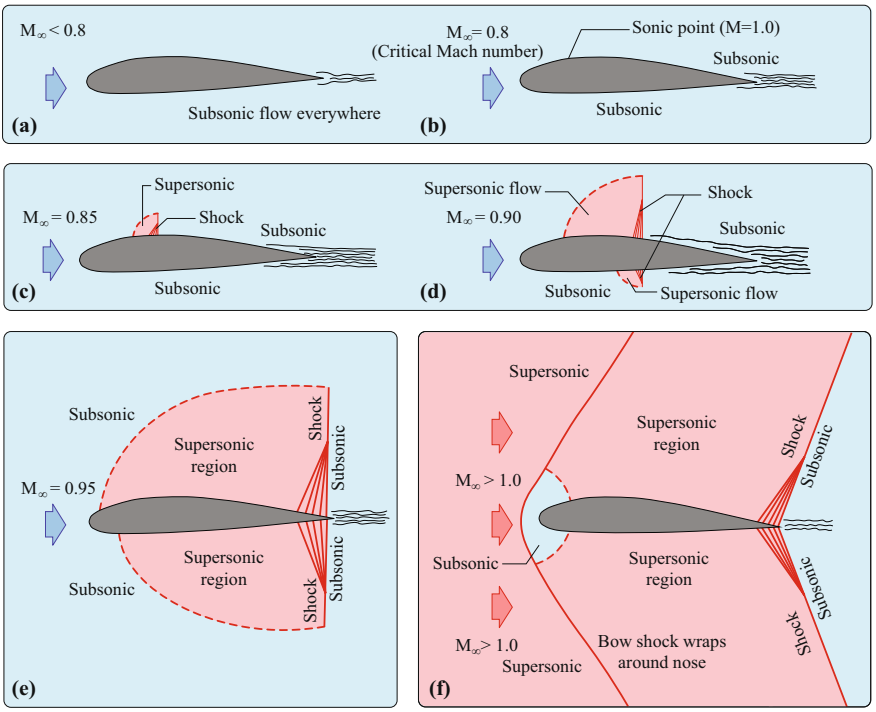


Fig. 3.16 Transonic flow over an airfoil, and the formation of shock waves and boundary-layer separation

concept of oblique shock waves that occur when the flow is deflected by a (concave) curvature of, for instance, the vehicle wall. In case this curvature is convex, it results in Prandtl-Meyer expansion waves instead, which are detailed in Sect. 3.3.4. Finally, a discussion about boundary layers in Sect. 3.3.5 concludes the treatment of supersonic flow.

Using the flow phenomena shock waves, expansion waves, and boundary layers, as well as the interaction between any of these, the flow field of re-entry vehicles can be described, both in supersonic and hypersonic flow. This complete description will be provided in Sect. 3.4.

3.3.1 Introduction

It was in 1897 when Swedish engineer Gustav de Laval experimented with exhaust nozzles and came to the conclusion that to reach maximum efficiency the ratio of the minimum area to the inlet and outlet areas was critical. Assuming constant density, with the principle of mass continuity and Bernoulli's law, the flow in a pipe

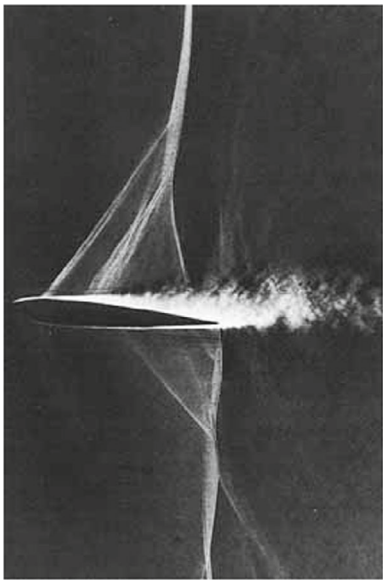


Fig. 3.17 Schlieren photograph of transonic flow over an airfoil. Image: NASA (Baals and Corliss 1981)

with a narrowing throat will reach maximum velocity and minimum pressure at the throat. After expansion towards the exhaust nozzle, velocity will drop, but pressure will rise again and could for a certain throat geometry reach its maximum. De Laval also found that when narrowing down the throat too much, the exit pressure dropped dramatically, whereas the exit velocity rose significantly. This was opposite to what Bernoulli's law predicted, only because de Laval had unknowingly generated supersonic flow!

During World War II, Spitfire pilots, amongst others, found that, when they were in a dive to escape enemy aircraft, their aircraft became exceedingly difficult to control. Reaching speeds close to the speed of sound they encountered a strong nose-down pitch and a loss of pitch-control authority, which often resulted in a crash into the ground. The nose-down pitching moment was a result of localised supersonic flow and air compressibility.[11]

As mentioned earlier, sound waves are caused by infinitesimally small pressure disturbances, and travel through a medium at the speed of sound. As long as the vehicle has a velocity well below sonic speed, it will not "feel" these disturbances (see Fig. 3.18). In the (compressible) flow, Mach waves will occur, which are pressure waves traveling with the speed of sound (which in term depends on gas properties and temperature), caused by a slight change of pressure added to a compressible

[11] At subsonic speeds, the aerodynamic centre (a.c.), i.e., the point where the pitch moment does not vary with the lift force (with changing angle of attack), is usually located at 25% of the wing-chord length. However, at supersonic speeds the a.c. moves backward to 50%, and since the location of the a.c. is then behind the centre of mass, this gives rise to a strong nose-down pitch moment. This phenomenon is known as "Mach tuck", and the diving tendency is also known as "tuck under".

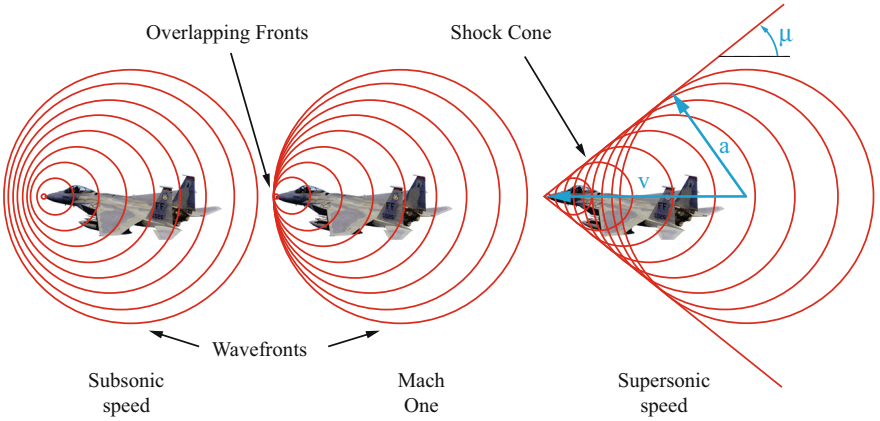


Fig. 3.18 Forming of shock wave

flow. For flight velocities well under the sonic velocity, the Mach waves travel out uniformly similar to waves on the surface of a pond. Mach waves are typically weak, but in supersonic flow, when there are sufficient waves present at any location, i.e., at vehicle velocities close to and above sonic speed, they can combine to become a shock wave. Shock waves are characteristic of supersonic flows, that is when the fluid velocity (u) is greater than its speed of sound, a. Flow across a shock gives rise to abrupt changes in fluid properties, is irreversible and cannot be approximated as isentropic.

A Mach wave propagates across the flow at an angle μ, appropriately called the Mach angle. It is defined by

$$\mu = \arcsin\left(\frac{1}{M_\infty}\right) \tag{3.56}$$

where M_∞ is the (free-stream) Mach number in front of the wave front. For example, when $M_\infty = 2$, the corresponding angle of the Mach cone is $\mu = 30°$. Note that with increasing Mach number, the shock cone moves closer to the vehicle wall.

Supersonic flight poses substantial technical challenges, as the aerodynamics of supersonic flight are dramatically different from those of subsonic flight. Most notably, aerodynamic drag rises sharply as the aircraft passes the transonic regime, demanding much greater engine power and more streamlined airframes. Supersonic vehicle designs entail slender bodies with a sharp nose, and the wings and tail surfaces are very sharp. The main reason for this is to minimise the wave drag, which can amount to 50% of the total drag during supersonic flight. With this kind of configuration, the shock waves will remain relatively weak. An example of a supersonic aircraft is, besides the military jets F-15, F16, F-18, etc., the Concorde (Fig. 3.19), a British-French turbojet-powered passenger aircraft that operated from 1976 to 2003. Its maximum speed was just above Mach 2.



Fig. 3.19 The Concorde aircraft, exhibited at the Technik Museum Sinsheim (Images: author)



Fig. 3.20 The SR-71 *Blackbird* (Images: NASA)

Another example of supersonic design is provided in Fig. 3.20, where the SR-71 *Blackbird* is shown. Developed as a long-range, strategic reconnaissance aircraft that was operated by the United States Air Force, it had a sustained Mach 3+ cruise capability. Its official maximum speed was Mach 3.3, whereas a sustained high-altitude flight was flown at 25,929 m. It was operated by USAF from 1964 to 1998, whereas NASA, which used the aircraft as a high-speed, high-altitude test-flight environment, retired the *Blackbird* one year later. Looking at the pictures, one can recognise the delta-like planform, and the sharp leading edges. The inlet spikes of the turbo-jet engines are sharp-pointed cones, whereas in the frontal view the very thin surface profiles are clear.

3.3.2 Normal Shock

Shockwaves can occur as normal shocks and as oblique shocks, and both will be discussed here and the next section, respectively. A normal shock wave can develop,

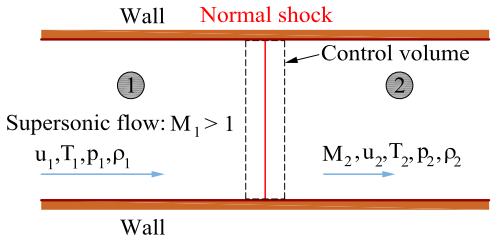


Fig. 3.21 Stationary normal shock in a duct flow

for example, in the intakes to the engines in some supersonic aircraft, in the exhaust of rocket engines, or in long-distance gas pipe-lines. Consider a steady flow, parallel to a wall or, alternatively, in a duct (note that the wall of the duct can be treated as a boundary streamline, parallel to the wall of a vehicle). Somewhere downstream is a stationary normal shock wave, as shown in Fig. 3.21. This shock wave is very thin, on the order of $\approx 10^{-4}$ mm, and in that short distance a large change in pressure occurs, causing an enormous energy dissipation.

Before the shock, the density of the flow that moves with velocity u_1, is ρ_1. Temperature and pressure are equal to T_1 and p_1. Around the shock, a control volume has been placed such that the three conservation laws for mass, momentum, and energy can be applied to derive the shock equations. A two-dimensional flow will be considered.

The Continuity Equation, Eq. (3.43), with $A_1 = A_2$ is

$$\rho_1 u_1 A_1 = \rho_2 u_2 A_2 \tag{3.57}$$

or

$$\rho_1 u_1 = \rho_2 u_2 \tag{3.58}$$

In the Momentum Equation, Eq. (3.44), the only force is a pressure force, pA, so:

$$\rho_1 u_1^2 A_1 + p_1 A_1 = \rho_2 u_2^2 A_2 + p_2 A_2 \tag{3.59}$$

Dividing out A_1 and A_2 it follows that

$$p_1 + \rho_1 u_1^2 = p_2 + \rho_2 u_2^2 \tag{3.60}$$

Finally, with the Energy Equation, Eq. (3.45)

$$\rho_1 u_1 (e_1 + \frac{1}{2} u_1^2) + p_1 u_1 = \rho_2 u_2 (e_2 + \frac{1}{2} u_2^2) + p_2 u_2 \tag{3.61}$$

Dividing both sides by $\rho_1 u_1$ (and using Eq. (3.58)) this results in

$$e_1 + \frac{1}{2}u_1^2 + \frac{p_1}{\rho_1} = e_2 + \frac{1}{2}u_2^2 + \frac{p_2}{\rho_2} \tag{3.62}$$

Finally, substituting the expression for enthalpy $h = e_1 + \frac{p_1}{\rho_1}$ this yields:

$$h_1 + \frac{1}{2}u_1^2 = h_2 + \frac{1}{2}u_2^2 \tag{3.63}$$

The above equation states that the total or stagnation enthalpy, i.e., $h_0 = h + \frac{1}{2}u^2$ is constant, before and after the shock.

Assuming that we know the flow properties u_1, p_1, and ρ_1 before the shock, we can derive the corresponding properties u_2, p_2, and ρ_2 after the shock (note that through Eq. (3.1) T_1 and T_2 are derived parameters). First, we are interested in the Mach number after the shock, M_2, so to find an expression for that the momentum Eq. (3.60) is divided by the continuity Eq. (3.58):

$$\frac{p_1}{\rho_1 u_1} + u_1 = \frac{p_2}{\rho_2 u_2} + u_2 \tag{3.64}$$

With the ideal gas law, $p_1 = \rho_1 R T_1$,

$$\frac{p_1}{\rho_1 u_1} = \frac{RT_1}{u_1} = \frac{\gamma RT_1}{\gamma u_1} = \frac{a_1^2}{\gamma u_1} \tag{3.65}$$

which gives

$$u_1 - u_2 = \frac{1}{\gamma}\left(\frac{a_2^2}{u_2} - \frac{a_1^2}{u_1}\right) \tag{3.66}$$

The second equation to be rewritten is the energy Eq. (3.63). For the enthalpy we substitute $h = c_p T$, with c_p being constant ($c_{p,1} = c_{p,2}$):

$$c_p T_1 + \frac{1}{2}u_1^2 = c_p T_2 + \frac{1}{2}u_2^2 = \frac{\gamma RT_1}{\gamma - 1} + \frac{1}{2}u_1^2 = \frac{\gamma RT_2}{\gamma - 1} + \frac{1}{2}u_2^2 \tag{3.67}$$

where $c_p = \frac{\gamma R}{\gamma - 1}$, which easily follows from Eq. (3.17). With the expression for the speed of sound substituted it follows that

$$\frac{a_1^2}{\gamma - 1} + \frac{1}{2}u_1^2 = \frac{a_2^2}{\gamma - 1} + \frac{1}{2}u_2^2 \tag{3.68}$$

and since both left-hand and right-hand side are equal to the stagnation enthalpy, it equates to a constant.

This notion can be used to set up an expression for a known flow condition, i.e., $M = 1$, for which holds $u = a^*$, with a^* being the speed of sound for the sonic condition. After a rather lengthy derivation—see the Intermezzo below—a^* follows

from the so-called *Prandtl condition*, i.e., $(a^*)^2 = u_1 u_2$. Combined with Eqs. (3.66) and (3.68) this leads to a relation between M_1 and M_2 (with the derivation again to be found in the Intermezzo):

$$M_2^2 = \frac{M_1^2 + \frac{2}{\gamma-1}}{\frac{2\gamma}{\gamma-1} M_1^2 - 1} \tag{3.69}$$

The momentum Eq. (3.60) is now written as function of Mach number rather than velocity, with the definition of the speed of sound, as well the ideal gas law:

$$u^2 = M^2 \gamma RT = M^2 \gamma \frac{p}{\rho} \quad \Rightarrow \quad \frac{p_1}{\rho_1}\left(1 + \gamma M_1^2\right) = p_2 + \rho_1 u_2^2 \tag{3.70}$$

which rewrites to

$$\frac{p_2}{p_1} = \frac{1 + \gamma M_1^2}{1 + \gamma M_2^2} \tag{3.71}$$

With the expression for M_2 substituted into the above equation, after some manipulation this results in

$$\frac{p_2}{p_1} = \frac{2\gamma}{\gamma + 1} M_1^2 - \frac{\gamma - 1}{\gamma + 1} \tag{3.72}$$

With the help of $u = M\sqrt{\gamma RT} = M\sqrt{\gamma R \frac{p}{\rho}}$ the continuity equation is reformulated as

$$\frac{\rho_2}{\rho_1} = \frac{u_1}{u_2} = \frac{M_1}{M_2}\sqrt{\frac{p_1}{p_2}\frac{\rho_2}{\rho_1}} \tag{3.73}$$

so

$$\sqrt{\frac{\rho_2}{\rho_1}} = \frac{M_1}{M_2}\sqrt{\frac{p_1}{p_2}} \quad \Rightarrow \quad \frac{\rho_2}{\rho_1} = \frac{M_1^2}{M_2^2}\frac{p_1}{p_2} \tag{3.74}$$

which yields, after substituting Eqs. (3.69) and (3.72):

$$\frac{\rho_2}{\rho_1} = \frac{(\gamma + 1) M_1^2}{2 + (\gamma - 1) M_1^2} \tag{3.75}$$

The last ratio to be found follows easily by dividing the ideal gas law after the shock by the corresponding values before the shock, and substituting Eqs. (3.72) and (3.75):

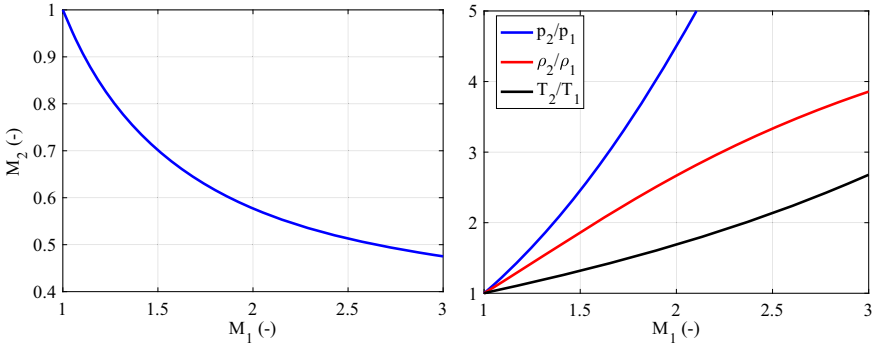


Fig. 3.22 Ratios of flow parameters before and after a normal shock, following the *Rankine–Hugoniot relations*

$$\frac{p_2}{p_1} = \frac{\rho_2}{\rho_1}\frac{T_2}{T_1} \quad \Rightarrow \quad \frac{T_2}{T_1} = \frac{p_2}{p_1}\frac{\rho_1}{\rho_2} \tag{3.76}$$

such that

$$\frac{T_2}{T_1} = \left[1 + \frac{2\gamma}{\gamma+1}(M_1^2 - 1)\right]\frac{2 + (\gamma-1)M_1^2}{(\gamma+1)M_1^2} \tag{3.77}$$

Note that because $h = c_p T$ it automatically follows that $\frac{h_2}{h_1} = \frac{T_2}{T_1}$.

Equations (3.69), (3.72), (3.75) and (3.77) are known as the *Rankine–Hugoniot relations*[12] and link the conditions after the normal shock to those before. In Fig. 3.22 these relations have been plotted as a function of M_1. If $M_1 = 1$, then also $M_2 = 1$ and no shock wave exists. In case $M_1 > 1$, then $M_2 < 1$, which means that the flow after a normal shock is *always* subsonic. Finally, it is noted that when $M_1 < 1$ the equations give $M_2 > 1$, such that somehow a subsonic flow has been converted to a supersonic flow. This is clearly impossible, because it results in a positive production of entropy, and this is a violation of the second law of thermodynamics. In other words, the *Rankine–Hugoniot relations* are valid for $M_1 > 1$ only.

*** Intermezzo 3.1 (Mach relation for a normal shock)**
In this Intermezzo, we will derive the *Prandtl condition* as well as the relation between M_2 and M_1, Eq. (3.69). Starting point for the derivation is Eq. (3.68):

$$\frac{a_1^2}{\gamma-1} + \frac{1}{2}u_1^2 = \frac{a_2^2}{\gamma-1} + \frac{1}{2}u_2^2 = \text{constant} = c \tag{3.78}$$

[12] William John Macquorn Rankine (1820–1872) was a Scottish mechanical engineer, mathematician and physicist, and Pierre Henri Hugoniot (1851–1887) was a French inventor, mathematician, and physicist.

As the above expression is constant for all conditions, we can apply the equation to a known condition, i.e., the sonic point for which $M = 1$. The speed of sound for this point will be indicated by a^*, such that $u = a^*$ and thus:

$$\frac{(a^*)^2}{\gamma - 1} + \frac{(a^*)^2}{2} = \frac{2(a^*)^2}{2(\gamma - 1)} + \frac{(\gamma - 1)(a^*)^2}{2(\gamma + 1)} = \frac{(\gamma + 1)(a^*)^2}{2(\gamma - 1)} = c \tag{3.79}$$

The found constant can now be plugged into Eq. (3.78), on both the left-hand and right-hand part, leading to two expressions for a_1 and a_2. For a_1 we get

$$\frac{a_1^2}{\gamma - 1} + \frac{1}{2}u_1^2 = \frac{(\gamma + 1)(a^*)^2}{2(\gamma - 1)} \quad \Rightarrow \quad a_1^2 = \frac{\gamma + 1}{2}(a^*)^2 - \frac{\gamma - 1}{2}u_1^2 \tag{3.80}$$

and similarly for a_2:

$$a_2^2 = \frac{\gamma + 1}{2}(a^*)^2 - \frac{\gamma - 1}{2}u_2^2 \tag{3.81}$$

Recall now the mass-momentum Eq. (3.66):

$$\frac{a_1^2}{\gamma u_1} + u_1 = \frac{a_2^2}{u_2} + u_2 \tag{3.82}$$

The expressions for a_1^2 and a_2^2 found above are substituted, resulting in

$$\frac{\gamma + 1}{2\gamma u_1}(a^*)^2 - \frac{\gamma - 1}{2\gamma u_1}u_1 + u_1 = \frac{\gamma + 1}{2\gamma u_2}(a^*)^2 - \frac{\gamma - 1}{2\gamma u_2}u_2 + u_2 \tag{3.83}$$

which simplifies to

$$\frac{\gamma + 1}{2\gamma u_1}(a^*)^2 + \left(1 - \frac{\gamma - 1}{2\gamma}\right)u_1 = \frac{\gamma + 1}{2\gamma u_2}(a^*)^2 + \left(1 - \frac{\gamma - 1}{2\gamma}\right)u_2 \tag{3.84}$$

$$\frac{\gamma + 1}{2\gamma u_1}(a^*)^2 - \frac{\gamma + 1}{2\gamma u_2}(a^*)^2 = \left(1 - \frac{\gamma - 1}{2\gamma}\right)(u_2 - u_1) \tag{3.85}$$

$$\frac{\gamma + 1}{2\gamma u_1 u_2}(a^*)^2(u_2 - u_1) = \left(1 - \frac{\gamma - 1}{2\gamma}\right)(u_2 - u_1) \tag{3.86}$$

With the $u_2 - u_1$ term cancelling out, and $1 - \frac{\gamma - 1}{2\gamma} = \frac{\gamma + 1}{2\gamma}$ we get

$$\frac{\gamma + 1}{2\gamma u_1 u_2}(a^*)^2 = \frac{\gamma + 1}{2\gamma} \quad \Rightarrow \quad \frac{(a^*)^2}{u_1 u_2} = 1 \tag{3.87}$$

and we have thus reached the *Prandtl relation*:

$$(a^*)^2 = u_1 u_2 \tag{3.88}$$

The next step will be to find a relation between M_1 and M_2. Recall Eq. (3.80), which can be written in a non-specific form:

$$\frac{(\gamma+1)}{2(\gamma-1)}(a^*)^2 = \frac{a^2}{\gamma-1} + \frac{1}{2}u^2 \tag{3.89}$$

Dividing by u^2 and multiplying by $2(\gamma-1)$ yields

$$(\gamma+1)\frac{(a^*)^2}{u^2} = 2\frac{a^2}{u^2} + (\gamma-1) \tag{3.90}$$

Substituting the definition of the Mach number

$$(\gamma+1)\frac{(a^*)^2}{u^2} = \frac{2}{M^2} + (\gamma-1)\frac{M^2}{M^2} = \frac{2+(\gamma-1)M^2}{M^2} \tag{3.91}$$

and dividing by $\gamma+1$

$$\left(\frac{a^*}{u}\right)^2 = \frac{2+(\gamma-1)M^2}{(\gamma+1)M^2} \tag{3.92}$$

With the *Prandtl relation* rewritten, i.e., $\frac{a^*}{u_1} = \frac{u_2}{a^*}$, and the above expression substituted this results in

$$\frac{a^*}{u_1} = \sqrt{\frac{2+(\gamma-1)M_1^2}{(\gamma+1)M_1^2}} = \sqrt{\frac{(\gamma+1)M_2^2}{2+(\gamma-1)M_2^2}} = \frac{u_2}{a^*} \tag{3.93}$$

Squaring the middle two terms and rearranging gives

$$\left[2+(\gamma-1)M_1^2\right]\left[2+(\gamma-1)M_2^2\right] = \left[(\gamma+1)M_1^2\right]\left[(\gamma+1)M_2^2\right] \tag{3.94}$$

$$2\left[2+(\gamma-1)M_1^2\right] + (\gamma-1)\left[2+(\gamma-1)M_1^2\right]M_2^2 = \left[(\gamma+1)^2 M_1^2\right]M_2^2 \tag{3.95}$$

$$4 + 2(\gamma-1)M_1^2 + 2(\gamma-1)M_2^2 + \left[(\gamma-1)^2 M_1^2\right]M_2^2 = \left[(\gamma+1)^2 M_1^2\right]M_2^2 \tag{3.96}$$

Moving all M_2 terms to the right-hand side gives

$$4 + 2(\gamma-1)M_1^2 = \left\{ \overbrace{\left[(\gamma+1)^2 - (\gamma-1)^2\right]}^{4\gamma} - 2(\gamma-1) \right\} M_2^2 \tag{3.97}$$

and thus

$$4 + 2(\gamma - 1)M_1^2 = [4\gamma - 2(\gamma - 1)]\, M_2^2 \tag{3.98}$$

Finally, dividing both sides by 4 and solving for M_2 yields the relation between M_2 and M_1, Eq. (3.69):

$$M_2^2 = \frac{M_1^2 + \frac{2}{\gamma-1}}{\frac{2\gamma}{\gamma-1} M_1^2 - 1} \tag{3.99}$$

■

3.3.3 *Oblique Shock*

An oblique shock wave, unlike a normal shock, is inclined with respect to the incident upstream flow direction. It will occur when a supersonic flow encounters a corner that effectively turns the flow into itself and compresses (Fig. 3.23). When the curvature of the wall (i.e., the obstruction) is gradual, Mach waves will occur first. By itself, a Mach wave is weak, but as discussed before, when in one location they are combined they turn into a shock wave.

To derive a relation between the (finite) wedge angle (or deflection angle), θ, and the inclination of the oblique shock wave, the shock angle, β, we make use of the basic conservation laws of mass, momentum, and energy, as before while deriving the normal-shock equations. Additionally, this exercise will lead to the relations between the flow properties ahead of and behind the oblique shock, as a function of the upstream Mach number, M_1, and θ. Finally, we will introduce the so-called $\theta - \beta - M$-diagram that provides a visual aid for the relationship between these three parameters.

Contrary to the normal-shock equations that were derived for a one-dimensional flow, now a two-dimensional flow is considered. The derivation is facilitated by decomposing the velocity vector, $\mathbf{u}$, into a component normal to the shock, $\mathbf{u_n}$, and one tangential to the shock, $\mathbf{u_t}$. To distinguish between upstream and downstream

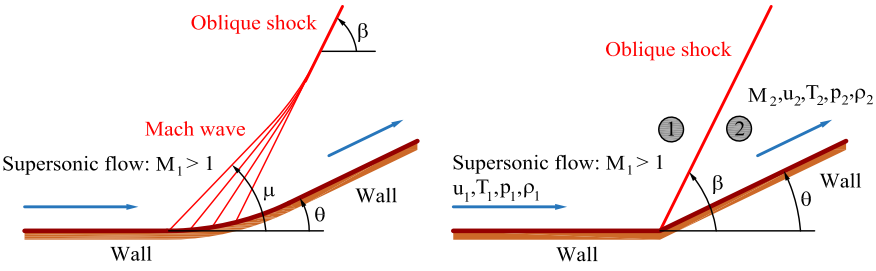


Fig. 3.23 Occurrence of Mach waves, combining into an oblique shock

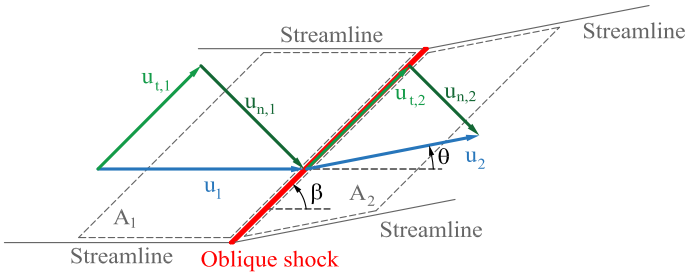


Fig. 3.24 Schematic representation of flow over an oblique shock

conditions, subscripts 1 (upstream) and 2 (downstream) are introduced. The related parameters are shown in Fig. 3.24.

To apply the conservation laws, we need to define a control volume before and after the shock. In the two-dimensional case it is assumed that the flow properties do not change in the third dimension, and we will consider the corresponding width to be 1 m, effectively reducing the volume to a surface area. Now, because the shock is assumed to be infinitely thin, the two control surfaces have the same area, A (even though the shape of the surface behind the shock is twisted). The top and bottom sides of the control surface are parallel to the respective streamlines, which means that there is no perpendicular component of velocity and thus no ingoing or outgoing flow through these sides.

Conservation of Mass

With the general form of the conservation law of mass given by Eq. (3.43), it follows that

$$\iint_{A_1} \rho_1(\mathbf{u_1} \cdot \hat{\mathbf{n}}) + \iint_{A_2} \rho_2(\mathbf{u_2} \cdot \hat{\mathbf{n}}) = 0 \tag{3.100}$$

with $\hat{\mathbf{n}}$ being the outward normal-vector on each of the sides of the control surface.

As stated, we only need to concern ourselves with the incoming flow through the left side of control surface A_1, and the outgoing flow from the right side of A_2, i.e., the flow parallel to the streamlines. To obtain the normal velocity, we need to evaluate $\mathbf{u_1} \cdot \hat{\mathbf{n}}$ and $\mathbf{u_2} \cdot \hat{\mathbf{n}}$, yielding $-u_{n,1}$ and $+u_{n,2}$. The minus sign in the first term is introduced, because the (outward) normal on A_1 is in opposite direction of $\mathbf{u_1}$. The final result is, with $A_1 = A_2$:

$$\rho_1 u_{n,1} A_1 = \rho_2 u_{n,2} A_2 \tag{3.101}$$

or

$$\rho_1 u_{n,1} = \rho_2 u_{n,2} \tag{3.102}$$

Conservation of Momentum

Given the general formulation, Eq. (3.44), the momentum equation has to be solved for both normal and tangential direction. As before, $\mathbf{v} \mathrm{d}A = 0$ for the sides of the control surface parallel to the streamlines. To begin with the tangential direction,

$$\rho_1 u_{t,1} u_{n,1} A_1 = \rho_2 u_{t,2} u_{n,2} A_2 \quad \Rightarrow \quad \rho_1 u_{t,1} u_{n,1} = \rho_2 u_{t,2} u_{n,2} \tag{3.103}$$

The pressure term $p\hat{\mathbf{n}}\mathrm{d}A$ is zero, as by definition the pressure acts in normal direction on the control surface. Substituting Eq. (3.102), this gives an important result:

$$u_{t,1} = u_{t,2} \tag{3.104}$$

or, in words, the tangential velocity component before and after the shock is the same. For the normal direction we get:

$$\rho_1 u_{n,1} \underbrace{(\mathbf{u_1} \cdot \hat{\mathbf{n}}_{\mathbf{A_1}})}_{-u_{n,1}} A_1 + \underbrace{p_1 \hat{\mathbf{n}}_{\mathbf{A_1}}}_{-p_1} A_1 + \rho_2 u_{n,2} \underbrace{(\mathbf{u_2} \cdot \hat{\mathbf{n}}_{\mathbf{A_2}})}_{u_{n,2}} A_2 + \underbrace{p_2 \hat{\mathbf{n}}_{\mathbf{A_2}}}_{p_2} A_2 = 0 \tag{3.105}$$

or

$$\rho_1 u_{n,1}^2 A_1 + p_1 A_1 = \rho_2 u_{n,2}^2 A_2 + p_2 A_2 \tag{3.106}$$

With Eq. (3.102) substituted, and again using $A_1 = A_2$:

$$p_1 + \rho_1 u_{n,1}^2 = p_2 + \rho_1 u_{n,2}^2 \tag{3.107}$$

Conservation of Energy

Starting point is the energy equation, Eq. (3.45), and evaluating this equation in a similar manner as before, yields:

$$-\rho_1 u_{n,1}(e_1 + \frac{1}{2}u_1^2) + \rho_2 u_{n,2}(e_2 + \frac{1}{2}u_2^2) - p_1 u_{n,1} + p_2 u_{n,2} = 0 \tag{3.108}$$

or

$$-\rho_1 u_{n,1}(e_1 + \frac{1}{2}u_1^2 + \frac{p_1}{\rho_1}) + \rho_2 u_{n,2}(e_2 + \frac{1}{2}u_2^2 + \frac{p_2}{\rho_2}) = 0 \tag{3.109}$$

With Eq. (3.102) substituted this simplifies to

$$e_1 + \frac{1}{2}u_1^2 + \frac{p_1}{\rho_1} = e_2 + \frac{1}{2}u_2^2 + \frac{p_2}{\rho_2} \tag{3.110}$$

Furthermore, the norm of the velocity is the vector sum of the normal and tangential components, or $u_1^2 = u_{n,1}^2 + u_{t,1}^2$ and $u_2^2 = u_{n,2}^2 + u_{t,2}^2$, and this substituted into Eq. (3.110) results in

$$e_1 + \frac{1}{2}u_{n,1}^2 + \frac{1}{2}u_{t,1}^2 + \frac{p_1}{\rho_1} = e_2 + \frac{1}{2}u_{n,2}^2 + \frac{1}{2}u_{t,2}^2 + \frac{p_2}{\rho_2} \tag{3.111}$$

and with $u_{t,1} = u_{t,2}$ found in Eq. (3.104), the two corresponding terms on the left-hand and right-hand side cancel out. Substituting Eq. (3.11)—the definition of enthalpy—the Conservation of Energy is given by:

$$h_1 + \frac{1}{2}u_{n,1}^2 = h_2 + \frac{1}{2}u_{n,2}^2 \tag{3.112}$$

Summarised, the main results are given by Eqs. (3.102), (3.104), (3.107), and (3.112). Note that apart from Eq. (3.104), these equations also hold for a normal shock. For this reason, the relations for the flow properties before and after the normal shock also hold for the oblique-shock case, with the difference that for the velocity components now the normal-velocity components are used. For the sake of completeness, the corresponding expressions are stated below.

$$M_{n,2}^2 = \frac{1 + \frac{\gamma-1}{2}M_{n,1}^2}{\gamma M_{n,1}^2 - (\gamma-1)/2} \tag{3.113}$$

$$M_{t,2}^2 = M_{t,1}^1 \tag{3.114}$$

$$\frac{p_2}{p_1} = 1 + \frac{2\gamma}{\gamma+1}\left(M_{n,1}^2 - 1\right) \tag{3.115}$$

$$\frac{\rho_2}{\rho_1} = \frac{(\gamma+1)M_{n,1}^2}{2 + (\gamma-1)M_{n,1}^2} \tag{3.116}$$

$$\frac{T_2}{T_1} = \left[1 + \frac{2\gamma}{\gamma+1}(M_1^2 - 1)\right]\frac{2 + (\gamma-1)M_1^2}{(\gamma+1)M_1^2} \tag{3.117}$$

It can be seen from these relations that pressure, density and temperature are increased over an oblique shock. This is due to the decrease in (normal) velocity, and the conservation of mass and momentum. Also, the rise in these quantities increases with increasing values of $M_{n,1}$, so with an increase in either free-stream Mach number or shock angle. The Mach number after the shock can be determined, using the geometry shown in Fig. 3.24:

$$M_2 = \frac{M_{n,2}}{\sin(\beta - \theta)} \tag{3.118}$$

With $M_{n,2}$ given by Eq. (3.113) and $M_{n,1}^2 = M_1^2 \sin^2 \beta$ this results in

$$M_2 = \frac{1}{\sin(\beta - \theta)} \sqrt{\frac{1 + \frac{\gamma-1}{2} M_1^2 \sin^2 \beta}{\gamma M_1^2 \sin^2 \beta - \frac{\gamma-1}{2}}} \tag{3.119}$$

After applying the three conservation laws it has been shown that a shock wave can indeed exist. However, it has not indicated whether the shock is compressive (i.e., $p_2/p_1 > 1$) or expansive (i.e., $p_2/p_1 < 1$). To understand what is happening, use the second law of thermodynamics and determine the entropy change across the shock wave, the starting point is Eq. (3.24):

$$s_2 - s_1 = c_p \ln \frac{T_2}{T_1} - R \ln \frac{p_2}{p_1} \tag{3.120}$$

and substituting Eqs. (3.115) and (3.117) we obtain:

$$\begin{aligned} s_2 - s_1 = c_p \ln \left\{ \left[1 + \frac{2\gamma}{\gamma + 1} (M_{n,1}^2 - 1) \right] \frac{2 + (\gamma - 1) M_{n,1}^2}{(\gamma + 1) M_{n,1}^2} \right\} - \ldots \\ \ldots - R \ln \left[1 + \frac{2\gamma}{\gamma + 1} (M_{n,1}^2 - 1) \right] \end{aligned} \tag{3.121}$$

After substituting relevant parameters, this confirms that $s_2 > s_1$, and thus the flow after the shock is compressive. Because the shock wave is very thin, the gradients of velocity and temperature in the shock are very high. As a result, the effects of viscosity and heat conduction are important within the shock leading to the entropy increase. Since the flow across a shock is adiabatic, the stagnation temperature does not change across a shock wave. However, entropy increases across a shock and as a result the stagnation pressure always decreases.

After crossing a shock wave that is not perpendicular to the incoming flow, the flow will be deflected over an angle θ with respect to its initial direction. With the result from before, $u_{t,2} = u_{t,1}$, the velocity triangles before and after the shock can be related by (see Fig. 3.24):

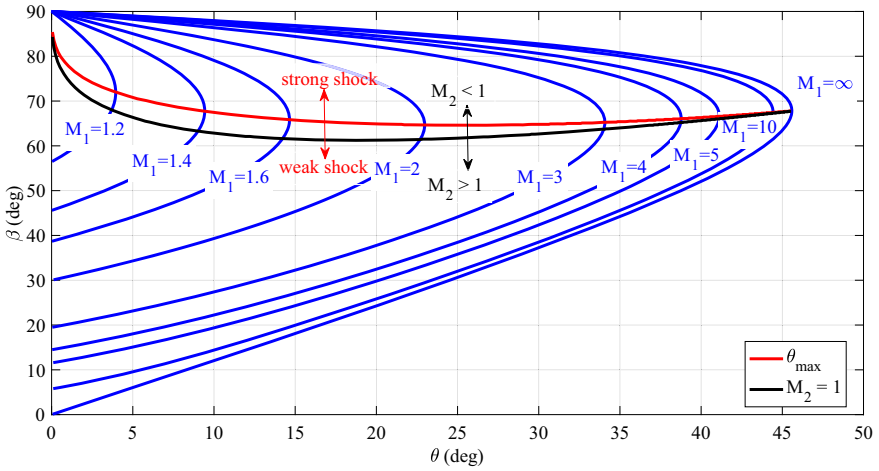


Fig. 3.25 The $\theta - \beta - M$ plot for varying free-stream Mach numbers, M_1

$$\frac{\tan(\beta-\theta)}{\tan\beta} = \frac{u_{n,2}}{u_{n,1}} = \frac{\rho_1}{\rho_2} = \frac{(\gamma+1)M_1^2\sin^2\beta}{2+(\gamma-1)M_1^2\sin^2\beta} \tag{3.122}$$

and solving for θ yields the $\theta - \beta - M$-equation:

$$\tan\theta = 2\cot\beta\frac{M_1^2\sin^2\beta - 1}{M_1^2(\gamma+\cos 2\beta)+2} \tag{3.123}$$

The above relation between θ, β, and M is visualised in Fig. 3.25. Some important features van be discerned. There is a maximum turning angle $\theta_{\max}$ for any given upstream Mach number M_1. If the wall angle is larger than this maximum value ($\theta > \theta_{\max}$), no oblique shock can be formed and instead a detached shock forms ahead of the concave corner. For $\theta < \theta_{\max}$ it is easy to see that two solutions for β exist. The smaller value leads to a so-called weak shock, whereas the larger β gives rise to a strong shock. The required pressure rise determines if a weak or a strong shock will occur, and this pressure rise is determined by flow conditions. In practise, a weak shock is most likely to occur in supersonic flow, where the velocity after the shock remains supersonic; in case of a strong shock the flow after the shock would be subsonic ($M_2 < 1$). Note that the limit case of the strong shock ($\theta \to 0°$ and thus $\beta \to 90°$), to be found in the upper-left corner of Fig. 3.25, the shock is a normal shock.

The two solutions to Eq. (3.123) are more explicitly shown in Fig. 3.26. For $\theta = 16°$, and $M_1 = 2$, the two solutions for the shock angle are $\beta_1 = 46.7°$ and $\beta_2 = 78.9°$. On the left of the figure, both the weak and strong shock solutions have been visualised. Also the detached-shock case where $\theta > \theta_{\max}$ is shown. The detached shock is a very common phenomenon for hypersonic vehicles with blunt noses, in

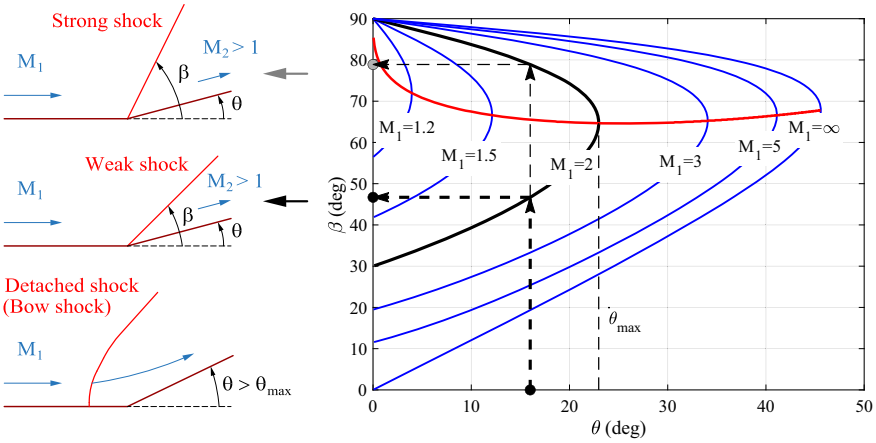


Fig. 3.26 Oblique-shock example for a wall deflection of $\theta = 16°$ at free stream $M_1 = 2$

which case it is a detached bow shock. This will be considered in more detail in Sect. 3.4.

3.3.4 Expansion Wave

In the previous section we analysed the turn of a supersonic flow towards the forming oblique shock wave, a so-called compression deflection. However, when the change in deflection is away from the flow and is thus an expansion deflection (i.e., around a *convex* corner), no shock wave will occur. In case the deflection is gradual (Fig. 3.27, on the left-hand side), there are infinite Mach waves appearing, which diverge with distance from the surface, and forming an expansion fan. Alternative names are rarefaction fan or Prandtl-Meyer fan, and the Mach waves are called Prandtl-Meyer expansion waves.[13] During this expansion, the Mach number downstream of the expansion increases ($M_2 > M_1$), whereas pressure, density, and temperature decrease. The current section will focus on expansion waves.

The right schematic in Fig. 3.27 shows the expansion waves in a simplified manner, as if they emanate from the vertex and the flow will be isentropic. The angle with the horizontal of the leading Mach wave is indicated by the Mach angle, μ. To differentiate from the geometry for the oblique shock wave, the deflection angle is indicated by θ^*, which is defined positively downwards, such that $\theta^* = -\theta$.

[13] The theory of the Prandtl-Meyer expansion waves was described by mathematician Theodor Meyer (1882–1972) in his PhD dissertation in 1908, along with his thesis advisor Ludwig Prandtl. The latter had already discussed the problem a year before.

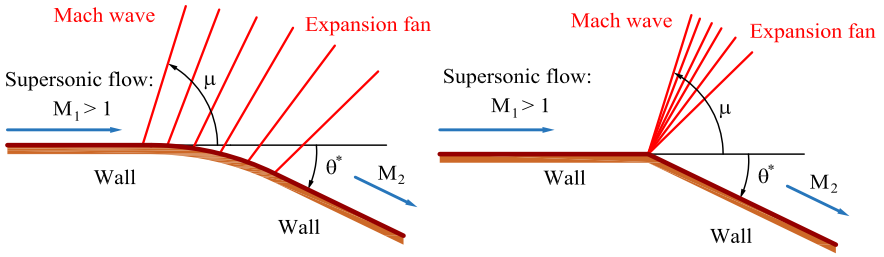


Fig. 3.27 Occurrence of an expansion fan due to a large wall deflection

To obtain a relation between the deflection angle and the change in Mach number, we will set up a difference equation relating the two, see also the book by Anderson (2003) for more details and applications.

In the Intermezzo below, a relation between a change in total flow speed, $\mathrm{d}V$, and a deflection-angle increment, $\mathrm{d}\theta^*$, is derived:

$$\frac{\mathrm{d}V}{V} = \frac{\mathrm{d}\theta^*}{\sqrt{M^2-1}} \quad \Rightarrow \quad \int_{\theta_1^*}^{\theta_2^*} \mathrm{d}\theta^* = \int_{M_1}^{M_2} \sqrt{M^2-1}\frac{\mathrm{d}V}{V} \tag{3.124}$$

To relate $\mathrm{d}V$ with $\mathrm{d}M$, we can derive from the definition of Mach number:

$$V = Ma \quad \Rightarrow \quad \ln V = \ln M + \ln a \tag{3.125}$$

such that

$$\frac{\mathrm{d}V}{V} = \frac{\mathrm{d}M}{M} + \frac{\mathrm{d}a}{a} \tag{3.126}$$

From Eq. (3.29), the adiabatic energy equation can be rewritten as

$$\frac{T_0}{T} = 1 + \frac{\gamma-1}{2}M^2 = \left(\frac{a_0}{a}\right)^2 \tag{3.127}$$

and solving for a yields

$$a = a_0 \frac{1}{\sqrt{1+\frac{\gamma-1}{2}M^2}} \tag{3.128}$$

To obtain the relation between $\mathrm{d}a$ and $\mathrm{d}M$ we differentiate the above equation, such that

$$\mathrm{d}a = -a_0 \frac{1}{(1+\frac{\gamma-1}{2}M^2)^{1.5}} \frac{\gamma-1}{2} M\,\mathrm{d}M \tag{3.129}$$

and thus, realising that a_0 is constant, in combination with Eq. (3.128) ($a_0 = \cdots$):

$$\frac{\mathrm{d}a}{a} = -\frac{\frac{\gamma-1}{2}M}{1+\frac{\gamma-1}{2}M^2}\mathrm{d}M \tag{3.130}$$

Equation (3.130) substituted into Eq. (3.126) results in

$$\frac{\mathrm{d}V}{V} = \frac{1}{1+\frac{\gamma-1}{2}M^2}\frac{\mathrm{d}M}{M} \tag{3.131}$$

With this expression at hand we can solve Eq. (3.124):

$$\int_{\theta_1^*}^{\theta_2^*} \mathrm{d}\theta^* = \theta_2^* - 0 = \int_{M_1}^{M_2} \frac{\sqrt{M^2-1}}{1+\frac{\gamma-1}{2}M^2}\frac{\mathrm{d}M}{M} \tag{3.132}$$

In the above equation, the integral

$$\nu(M) = \int_{M_1}^{M_2} \frac{\sqrt{M^2-1}}{1+\frac{\gamma-1}{2}M^2}\frac{\mathrm{d}M}{M} \tag{3.133}$$

is called the *Prandtl-Meyer function*. The solution to this integral is

$$\nu(M) = \sqrt{\frac{\gamma+1}{\gamma-1}}\arctan\left[\sqrt{\frac{\gamma-1}{\gamma+1}(M^2-1)}\right] - \arctan\left[\sqrt{M^2-1}\right] \tag{3.134}$$

The constant of integration is chosen here as zero, such that $\nu(M) = 0$ when $M = 1$. With the help of above solution, we rewrite Eq. (3.132) to obtain the basis for the Prandtl-Meyer expansion:

$$\theta_2^* = \nu(M_2) - \nu(M_1) \tag{3.135}$$

Summarised, the steps taken to do the relevant calculations are the following (Anderson 2003), given a supersonic flow with M_1 approaching a convex corner oriented at θ_2^*:

1. For the known M_1, evaluate $\nu(M_1)$ using Eq. (3.134), or obtain it from an $M-\nu$ table or graphical aid (see Fig. 3.28), when available;
2. Calculate $\nu(M_2)$ from Eq. (3.135), using the known θ_2^* and the evaluated $\nu(M_1)$;
3. Given $\nu(M_2)$, obtain M_2 from the $M-\nu$ table/figure, or use an iterative (= numerical) process to solve Eq. (3.134);

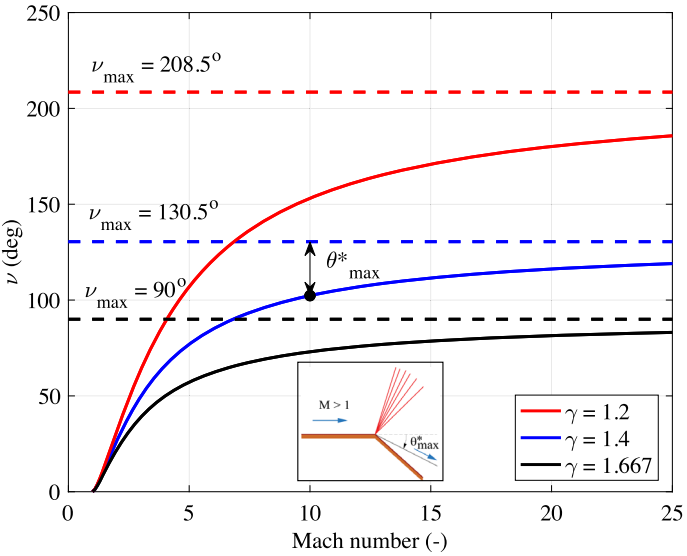


Fig. 3.28 The Prandtl-Meyer function, with its limit value and maximum deflection angle

4. As the process is isentropic (such that T_0 and p_0 are constant through the wave), the temperature and pressure after the expansion wave can be obtained with:

$$\frac{T_1}{T_2} = \frac{1 + \frac{\gamma-1}{2} M_2^2}{1 + \frac{\gamma-1}{2} M_1^2} \qquad \frac{p_1}{p_2} = \left(\frac{1 + \frac{\gamma-1}{2} M_2^2}{1 + \frac{\gamma-1}{2} M_1^2} \right)^{\frac{\gamma}{\gamma-1}} \tag{3.136}$$

Last, but not least, we devote a few words to the Prandtl-Meyer function. It has been plotted in Fig. 3.28. As can be seen, the function approaches a limit value when the Mach number goes to infinity. This limit value is easily found, realising that for $M \to \infty$: $\sqrt{..} \to M$, and $\arctan(M) \to 90°$. Thus

$$\nu_{\max} = \left(\sqrt{\frac{\gamma+1}{\gamma-1}} - 1 \right) 90° \tag{3.137}$$

and is solely dependent on γ. For $\gamma = 1.4$, $\nu_{\max} = 130.5°$ (some other values are: $\gamma = 1.667 \to \nu_{\max} = 90°$, $\gamma = 1.3 \to \nu_{\max} = 159.2°$, and $\gamma = 1.2 \to \nu_{\max} = 208.5°$). So, as γ decreases (especially at higher temperatures), the maximum turn angle, $\delta_{\max} = \nu_{\max} - \nu(M)$, increases. The maximum turn angle implies that, for a given upstream Mach Number, M_1, there is a maximum angle that the flow can negotiate. For example, for $M_1 = 10$ and $\nu(M_1) = 102.3°$, the maximum negotiable deflection is $\delta_{\max} = 28.2°$, at which, in theory, $M_2 \to \infty$. In practice, however, it means that the flow separates from the wall before it reaches this deflection.

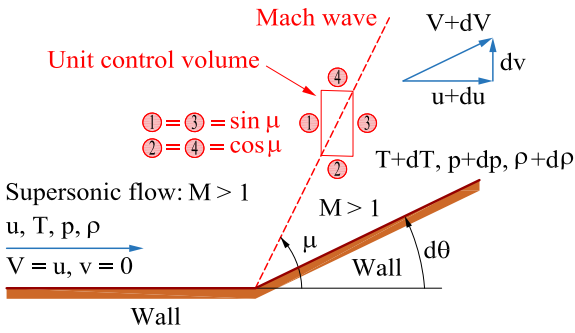


Fig. 3.29 Supersonic flow experiencing a small deflection, dθ

*** Intermezzo 3.2 (Two-dimensional supersonic flows for small deflections)**
In this Intermezzo, we will derive the relation between a small deflection angle, dθ, and the resulting change in Mach number, dM. Consider the supersonic-flow condition in Fig. 3.29. Due to a small deflection, dθ, in the wall a Mach wave with Mach angle μ occurs. This Mach wave gives rise to a small change in fluid properties p, ρ, and T. The two velocity components, u and v, experience small changes du and dv. A normalised control volume is placed around the Mach wave and serves as the basis for the derivation of a relation between dθ and dM, the change in Mach number. At all times, the flow remains supersonic due to the small deflection only.

Applying the principle of continuity before and after the Mach wave yields:

$$\rho u \sin\mu = [(u + \mathrm{d}u)\sin\mu - \mathrm{d}v\cos\mu]\,(\rho + \mathrm{d}\rho) \tag{3.138}$$

where the corresponding changes in u, v, and ρ are small. Neglecting higher-order terms, expanding the above equation results in:

$$u\sin\mu\mathrm{d}\rho + \rho\sin\mu\mathrm{d}u - \rho\cos\mu\mathrm{d}v = 0 \tag{3.139}$$

The momentum parallel to the upstream wall can be formulated as

$$-\sin\mu\mathrm{d}p = (\rho + \mathrm{d}\rho)\left[(u + \mathrm{d}u)^2\sin\mu - (u + \mathrm{d}u)\cos\mu\mathrm{d}v\right] - \rho u^2\sin\mu \tag{3.140}$$

Retaining only the nominal and first-order terms yields:

$$-\sin\mu\mathrm{d}p = (u^2\mathrm{d}\rho + 2\rho u\mathrm{d}u)\sin\mu - \rho u\cos\mu\mathrm{d}v \tag{3.141}$$

The momentum perpendicular to the upstream wall is given by

$$\cos\mu\mathrm{d}p = (\rho + \mathrm{d}\rho)\,[(u + \mathrm{d}u)\sin\mu\mathrm{d}v - (\rho + \mathrm{d}\rho)\mathrm{d}v\mathrm{d}v] - \rho u^2\sin\mu \tag{3.142}$$

which reduces to

$$\cos\mu \mathrm{d}p = \rho u \sin\mu \mathrm{d}v \tag{3.143}$$

Combining Eqs. (3.139), (3.141), and (3.143), and realising that $\mathrm{d}v = u\mathrm{d}\theta$ (to first order) eliminates $\mathrm{d}p$ and $\mathrm{d}\rho$:

$$\frac{\mathrm{d}u}{u} = \frac{\mathrm{d}V}{V} = \tan\mu \mathrm{d}\theta = \frac{\mathrm{d}\theta}{\sqrt{M^2 - 1}} \tag{3.144}$$

where the definition of the Mach angle, $\sin\mu = \frac{1}{M}$, Eq. (3.56), has been used. ■

3.3.5 *Boundary Layer*

In several supersonic flow situations the effect of viscosity can be neglected in the larger part of the flow, except in a small portion of the flow close to any solid surface, e.g., the vehicle wall. This small portion of the flow is called the boundary layer, and viscous effects dominate here. A boundary can either be laminar or turbulent, or can at some point transition from laminar to turbulent. In Fig. 3.30 this transition is shown; the laminar layer is a very thin layer, whereas the turbulent boundary layer is much thicker with many vortices. At all times a laminar boundary layer is preferred during normal flight, as as a turbulent boundary layer can potentially cause flow separation with accompanying loss of lift and high-pressure drag. Moreover, because laminar flow is a smooth flow with smaller viscous effects it gives lower skin-friction drag. Skin-friction drag increases the convective heating on the wall, so a turbulent boundary layer will put a larger demand on the thermal-protection system during hypersonic flight. It is noted, though, that transition is hard to predict, and it may always occur in hypersonic flight. Therefore, to have a design margin in the thermal protection system, often it is designed for a "turbulent" heat load. Finally, since the transition location is very uncertain, while doing wind-tunnel experiments and/or CFD computations, one often trips the transition on purpose, such that experimental and numerical data can be compared.

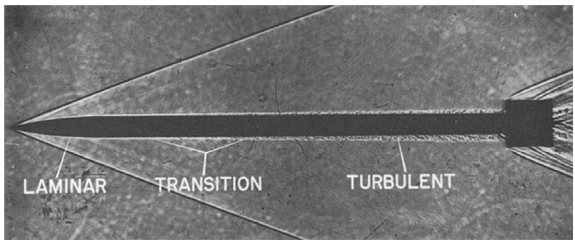


Fig. 3.30 Schlieren image of boundary-layer transition at Mach 3, one of the many wind-tunnel tests done in the 1950 and 1960s (Image: NASA)

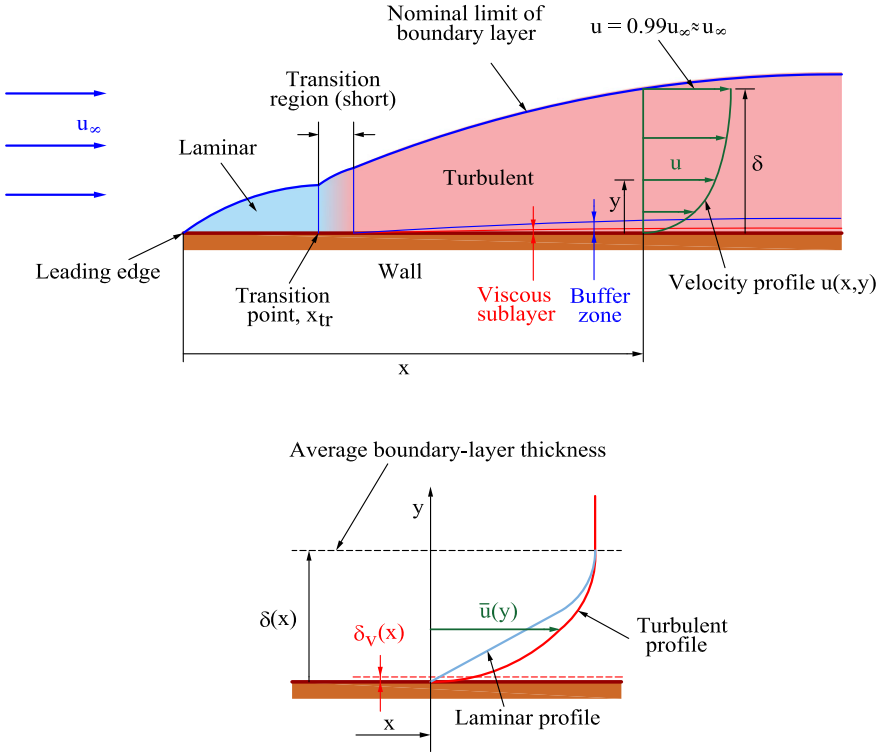


Fig. 3.31 Schematic representation of a boundary layer over a flat wall

Typically the potential-flow theory as mentioned before provides the velocity just outside the boundary layer and the pressure at the surface. In this section, for general understanding we will provide a simplified analysis of the boundary layer over a flat plate. The complete boundary-layer equations are difficult to solve for curved surfaces, and too complicated for the current introductory text. The reader is referred to literature for more details, as well as the formal derivation of the general case based on the Navier-Stokes Equations (White 2005; Schetz 2010; Schlichting and Gersten 2017).

Let us consider the schematic overview of a (two-dimensional) boundary layer, shown in Fig. 3.31. A flat plate is immersed in a flow with free-stream velocity u_∞. A laminar boundary layer originates at the leading edge, with running coordinate x in the direction of u_∞, and y perpendicular to the wall, pointing upwards. The point where transition occurs is indicated by x_{tr}. In the laminar boundary layer, the flow is regular with only slowly diverging streamlines. After transition, the flow directly at the wall (the viscous sub-layer) follows the wall, but in the turbulent region, the flow is chaotic with many vortices. The layer in between the two is called the buffer layer. The relative velocity is zero at the wall, i.e., the no-slip condition,[14]

[14] The no-slip condition can be deduced from the physical interpretation of viscosity, which is the transfer of momentum between particles. As the wall is "stationary", the momentum of the flow here must have a boundary condition of zero, since the coefficient of viscosity is finite.

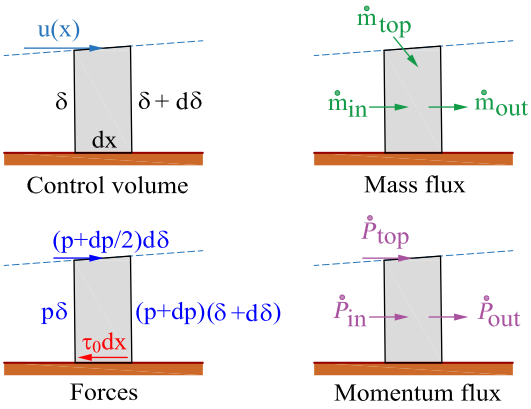


Fig. 3.32 Control volume in a boundary layer over a flat wall

and reaches u_∞ at the edge of the boundary layer. The boundary layer ends when the local-flow velocity is 99% of the free-stream velocity. The velocity profiles of laminar and turbulent boundary layers are different, with the turbulent profile being a bit "bulkier" than the laminar one.

To proceed, we define a small control volume in the boundary layer, with thickness dx and the height being the boundary-layer thickness, δ (Fig. 3.32). Considering a two-dimensional case, this assumes a "depth" of the boundary layer (in z-direction) of 1 m (and being consistent with SI units for the related variables). Forces acting on the volume are due to the flow pressure, p, and shear stress at the wall, τ_0, due to the flow's viscosity. Successively, we will look at mass continuity, and set up the momentum and force balance. Because the streamlines at the top and bottom of the control volume are not parallel, there is a small inflow of mass through the top. The continuity equation provides this mass flux, $\dot{m}_{top}$:

$$\dot{m}_{top} = \dot{m}_{out} - \dot{m}_{in} = \frac{\partial}{\partial x}\left(\int_0^\delta \rho u \mathrm{d}y\right)\mathrm{d}x \tag{3.145}$$

The sum of forces in x-direction follows from Newton's second law, i.e., force being equal to change in momentum, $\mathcal{P}$ (Fig. 3.32):

$$\sum F_x = \dot{\mathcal{P}}_{out} - \dot{\mathcal{P}}_{in} - \dot{\mathcal{P}}_{top} \tag{3.146}$$

where we have used $\mathcal{P} = mu$ for the definition of linear momentum.[15] Expanding this equation, and neglecting products of small terms,[16] it leads to

[15] Commonly, the symbol p is used for momentum, but to distinguish from pressure p, we use $\mathcal{P}$ instead.

[16] Small terms are not only dpdδ, but also pdδ, because δ itself is small, which makes dδ of second order.

$$\tau_0 \mathrm{d}x - \delta \mathrm{d}p = \frac{\partial}{\partial x}\left(\int_0^\delta \rho \bar{u}^2 \mathrm{d}y\right)\mathrm{d}x - u(x)\left(\int_0^\delta \rho \bar{u}\mathrm{d}y\right)\mathrm{d}x \tag{3.147}$$

with the realisation that $\dot{p}_{top} = u(x)\dot{m}_{top}$. Dividing Eq. (3.147) by $-\mathrm{d}x$ yields the so-called *von Karman integral equation*:

$$\tau_0 + \delta\frac{\mathrm{d}p}{\mathrm{d}x} = \rho u(x)\frac{\mathrm{d}}{\mathrm{d}x}\int_0^\delta \bar{u}\mathrm{d}y - \rho\frac{\mathrm{d}}{\mathrm{d}x}\int_0^\delta \bar{u}^2\mathrm{d}y \tag{3.148}$$

where we have changed from ∂x to $\mathrm{d}x$, because after integration the remaining functional is a function of x only (also $\delta = \delta(x)$). Note that we have assumed that ρ is constant over the boundary layer (incompressible flow).

For a flow over a flat plate with zero pressure gradient, i.e., $\frac{\mathrm{d}p}{\mathrm{d}x} = 0$, this means that $u(x) = \text{constant} = u_\infty$, such that Eq. (3.148) reduces to

$$\tau_0 = \rho\frac{\mathrm{d}}{\mathrm{d}x}\int_0^\delta \bar{u}(u_\infty - \bar{u})\mathrm{d}y \tag{3.149}$$

So if we know or assume a velocity profile $\bar{u}(x, y)$, then Eq. (3.149) together with $\tau_0 = \mu\left.\frac{\partial \bar{u}}{\partial y}\right|_{y=0}$, will give us $\delta(x)$ and $\tau(x)$. Note that because the boundary layer is thin, it is possible to approximate the velocity profile with a parabolic or cubic profile for a laminar layer, and a power-law profile for a turbulent layer. We will follow this up shortly.

Two additional lengths are commonly used in relation to boundary layers, which are the displacement thickness, δ_d, and the momentum thickness, θ, defined by

$$\delta_d = \frac{1}{u(x)}\int_0^\delta (u(x) - \bar{u})\mathrm{d}y \quad \Rightarrow \quad \delta_d = \frac{1}{u_\infty}\int_0^\delta (u_\infty - \bar{u})\mathrm{d}y \tag{3.150}$$

$$\theta = \frac{1}{u^2(x)}\int_0^\delta \bar{u}(u(x) - \bar{u})\mathrm{d}y \quad \Rightarrow \quad \delta_d = \frac{1}{\bar{u}_\infty^2}\int_0^\delta \bar{u}(u_\infty - \bar{u})\mathrm{d}y \tag{3.151}$$

The displacement thickness is the distance the streamline outside the boundary layer is displaced, because of the slower moving fluid inside the boundary layer (so it would also correspond with the decrease in mass flow, when compared to the inviscid case). The momentum thickness represents the loss of momentum that flows through the boundary layer as compared to the inviscid case.

To solve integral equation (3.149), we need to have the following boundary conditions for the flat plate: (i) $\bar{u} = 0$ at $y = 0$, (ii) $\bar{u} = u_\infty$ at $y = \delta$, and $\frac{\partial \bar{u}}{\partial y} = 0$ at $y = \delta$. If, for a laminar boundary layer, we assume a parabolic velocity profile, i.e.,

$$\frac{\bar{u}}{u_\infty} = c_0 + c_1 y + c_2 y^2 \tag{3.152}$$

then by substituting the boundary conditions we find $c_0 = 0$, $c_1 = \frac{2}{\delta}$ and $c_2 = -\frac{1}{\delta^2}$, yielding the laminar-flow velocity profile:

$$\frac{\bar{u}}{u_\infty} = \frac{2y}{\delta} - \frac{y^2}{\delta^2} \tag{3.153}$$

Substituting this into Eq. (3.149) results in

$$\tau_0 = \frac{\mathrm{d}}{\mathrm{d}x} \int_0^\delta \rho u_\infty^2 \left(\frac{2y}{\delta} - \frac{y^2}{\delta^2}\right)\left(1 - \frac{2y}{\delta} + \frac{y^2}{\delta^2}\right) \mathrm{d}y = \frac{2}{15} \rho u_\infty^2 \frac{\mathrm{d}\delta}{\mathrm{d}x} \tag{3.154}$$

Equivalently, the shear stress at the wall can also be written as

$$\tau_0 = \mu \left.\frac{\partial u}{\partial y}\right|_{y=0} = \mu u_\infty \frac{2}{\delta} \tag{3.155}$$

so equating Eqs. (3.154) and (3.155) gives

$$\delta \mathrm{d}\delta = 15 \frac{\nu}{u_\infty} \mathrm{d}x \tag{3.156}$$

With $\delta = 0$ for $x = 0$ (the leading edge), the solution for the *boundary-layer thickness* $\delta(x)$ is found to be:

$$\delta(x) = 5.48 \sqrt{\frac{\nu x}{u_\infty}} \tag{3.157}$$

Looking at the gradient of $\delta(x)$, and using the definition of Reynolds number, we get

$$\frac{\mathrm{d}\delta}{\mathrm{d}x} \sim \sqrt{\frac{\nu}{u_\infty x}} = \sqrt{\frac{1}{Re_x}} \tag{3.158}$$

where Re_x is the Reynolds number based on u_∞ and the distance measured along the surface from the point where the boundary layer began, x. So, in case $Re_L >> 1$ and $x < L$ it leads to conclude that $\frac{\mathrm{d}\delta}{\mathrm{d}x}$ is very small, except very close to $x = 0$. This

means that the velocity vector in the boundary layer is almost parallel with the wall and the velocity component perpendicular to it is very small. For convenience, we will rewrite Eq. (3.157):

$$\delta(x) = \left(\frac{5.48}{\sqrt{Re_x}}\right) x \tag{3.159}$$

The *wall shear stress* is found by substituting $\delta(x)$ back into Eq. (3.155):

$$\tau_0(x) = 0.365 \rho u_\infty^2 \sqrt{\frac{\nu}{u_\infty x}} = \left(\frac{0.365}{\sqrt{Re_x}}\right) \rho u_\infty^2 \tag{3.160}$$

Other derived parameters of interest are the *local skin-friction coefficient*, c_f:

$$c_f(x) = \frac{\tau_0}{\frac{1}{2}\rho u_\infty^2} = \frac{0.730}{\sqrt{Re_x}} \tag{3.161}$$

and the dimensionless drag-force coefficient, C_f:

$$C_f(x) = \frac{F_D}{\frac{1}{2}\rho u_\infty^2 L} = \frac{\int_0^L \tau_0 \mathrm{d}x}{\frac{1}{2}\rho u_\infty^2 L} = 1.46\sqrt{\frac{\nu}{u_\infty L}} = \frac{1.46}{\sqrt{Re_L}} \tag{3.162}$$

It is stressed that the above equations are approximations that depend on the assumed velocity profile and the assumption of incompressible flow. Therefore, for this simple case the derivation is quite easily done. However, for more accurate solutions (and for more complex surfaces!), the Navier-Stokes equations need to be solved.[17] Doing so for the incompressible laminar flow over the flat plate discussed above, this would give as coefficient values for $\delta(x)$, τ_0, c_f, and C_f: 5.0, 0.332, 0.664 and 1.33, respectively (White 2005). This means that the assumption of a parabolic velocity profile for a laminar boundary-layer flow over a flat plate gives an error of about 10%.

For turbulent boundary layers, often a power-law velocity profile is assumed, with a varying exponent depending on the Reynolds-number range (Potter and Wiggert 2008):

$$\frac{\bar{u}}{u_\infty} = \left(\frac{y}{\delta}\right)^{\frac{1}{n}} \qquad n = \begin{cases} 7 & Re_x \leq 10^7 \\ 8 & 10^7 < Re_x \leq 10^8 \\ 9 & 10^8 < Re_x \leq 10^9 \end{cases} \tag{3.163}$$

As an example, the profile with $n = 7$ is substituted into Eq. (3.149), resulting in

[17] Although a complete treatment is beyond the scope of this text, Sect. 3.7 gives a summary of the Navier-Stokes equations.

$$\tau_0 = \frac{7}{72}\rho u_\infty^2 \frac{\mathrm{d}\delta}{\mathrm{d}x} \tag{3.164}$$

but this solution gives $\tau_0 = \mu \frac{\partial u}{\partial y} = \infty$ at the wall, so it cannot be used for the complete boundary. One way to solve this is to combine an alternative (analytical) solution for τ_0 with Eq. (3.164), and extract an expression for $\frac{\mathrm{d}\delta}{\mathrm{d}x}$. One such solution is the Blasius solution for the local skin-friction coefficient (Anderson 2011):

$$c_f = \frac{\tau_0}{\frac{1}{2}\rho u_\infty^2} = 0.046 \left(\frac{\nu}{u_\infty \delta}\right)^{1/4} \quad \Rightarrow \quad \tau_0 = 0.023 \rho u_\infty^2 \left(\frac{\nu}{u_\infty \delta}\right)^{1/4} \tag{3.165}$$

such that

$$\delta^{1/4} \mathrm{d}\delta = 0.237 \left(\frac{\nu}{u_\infty}\right)^{1/4} \mathrm{d}x \tag{3.166}$$

In case we can assume that the laminar portion of the boundary layer is short, we can integrate the above equation from the leading edge onwards (0 to x), which yields:

$$\delta = 0.378x \left(\frac{\nu}{u_\infty x}\right)^{1/5} = \left(\frac{0.378}{Re_x^{1/5}}\right) x \tag{3.167}$$

and, consequently,

$$c_f = \frac{0.059}{Re_x^{1/5}} \qquad C_f = \frac{0.073}{Re_L^{1/5}} \tag{3.168}$$

The displacement and momentum thicknesses can be assessed using the same power-law velocity profile:

$$\delta_d = \left(\frac{0.048}{Re_x^{1/5}}\right) x \qquad \theta = \left(\frac{0.037}{Re_x^{1/5}}\right) x \tag{3.169}$$

Other quantities of interest in the study of turbulent boundary layers are the shear velocity, u_τ, and the thickness of the highly fluctuating viscous wall layer, δ_v, indicated in Fig. 3.31. For the sake of completeness, they are provided here without further derivation:

$$u_\tau = \sqrt{\frac{\tau_0}{\rho}} \qquad \delta_v = \frac{5\nu}{u_\tau} \tag{3.170}$$

In this discussion it is worthwhile to mention the *energy thickness* as well, which can be defined as the distance, measured perpendicular to the boundary of a solid body, by which the boundary should be displaced to compensate for the reduction in kinetic energy of the flowing fluid because of boundary-layer formation. It can also

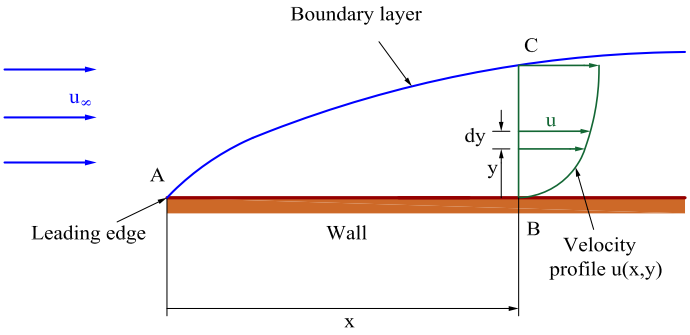


Fig. 3.33 Development of a boundary layer until separation

be seen as the thickness of flow (transverse distance measured perpendicular to the boundary of the solid surface) moving at the free-stream velocity, u_∞, and having kinetic energy equal to the deficiency of kinetic energy due to the formation of the boundary layer.

To look at this in more detail, consider the flow situation shown in Fig. 3.33. A flat surface is exposed to a flow with free-stream velocity, u_∞, parallel to the wall. From the leading edge downwards, a boundary layer has formed. At a downstream location, x, a velocity profile defines the distribution over the height of the boundary layer, indicated by the two points B and C. At a height y, an elemental strip with thickness dy and velocity $u = u(y)$, has been defined. The strip is thin such that u is constant over dy. Assume that this strip has a width b, which can be considered equal to the width of the wall, but is not relevant for the final result. Thus, the area of the elemental strip is $dS = bdy$, and the mass flow of the fluid through this surface is

$$\dot{m}_y = \rho\, u\, dS = \rho\, u\, b\, dy \tag{3.171}$$

The kinetic energy per second due to this mass flow is

$$\dot{\mathcal{E}}_{k,dy} = \frac{1}{2}\dot{m}_y u^2 \tag{3.172}$$

or

$$\dot{\mathcal{E}}_{k,dy} = \frac{1}{2}\rho\, u\, b\, dy\, u^2 = \frac{1}{2}\rho\, b\, dy\, u^3 \tag{3.173}$$

The kinetic energy of $\dot{m}$, if it would be moving with u_∞ (effectively meaning that the effect of the boundary layer is absent), would be

$$\dot{\mathcal{E}}_{k,u_\infty} = \frac{1}{2}\rho\, u\, b\, dy u_\infty^2 \tag{3.174}$$

By subtracting Eq. (3.173) from (3.174), the *energy loss* per second is then found to be

$$\Delta\dot{\mathcal{E}}_{k,y} = \frac{1}{2}\rho\, u\, b\, dy u_\infty^2 - \frac{1}{2}\rho\, b\, dy u^3 = \frac{1}{2}\rho\, b\, dy u(u_\infty^2 - u^2) \tag{3.175}$$

and, integrating over the full height of the boundary layer (i.e., $BC = \delta$):

$$\Delta\dot{\mathcal{E}}_k = \int_0^\delta \frac{1}{2}\rho b u(u_\infty^2 - u^2)\mathrm{d}y \tag{3.176}$$

which results in

$$\Delta\dot{\mathcal{E}}_k = \frac{1}{2}\rho b \int_0^\delta u(u_\infty^2 - u^2)\mathrm{d}y \tag{3.177}$$

Suppose that the distance BC is increased by an amount δ^{**}, where the fluid is flowing with velocity u_∞. The corresponding kinetic energy per second of this fluid flowing through the area $b\,\delta^{**}$ is then:

$$\dot{\mathcal{E}}_{k,\delta^{**}} = \frac{1}{2}\rho\, b\delta^{**}u_\infty^3 \tag{3.178}$$

Equating this with $\Delta\dot{\mathcal{E}}_k$, Eq. (3.177), yields

$$\frac{1}{2}\rho\, b\delta^{**}u_\infty^3 = \frac{1}{2}\rho b \int_0^\delta u(u_\infty^2 - u^2)\mathrm{d}y \tag{3.179}$$

so

$$\delta^{**}u_\infty^3 = \int_0^\delta u(u_\infty^2 - u^2)\mathrm{d}y \tag{3.180}$$

and thus

$$\delta^{**} = \frac{1}{u_\infty^3}\int_0^\delta u(u_\infty^2 - u^2)\mathrm{d}y \tag{3.181}$$

$$\delta^{**} = \int_0^\delta \frac{u\left(u_\infty^2 - u^2\right)}{u_\infty^3}\mathrm{d}y \tag{3.182}$$

and, finally

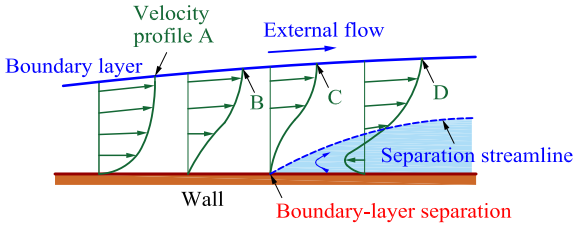


Fig. 3.34 Development of a boundary layer until separation

$$\delta^{**} = \int_0^{\delta} \frac{u}{u_\infty}\left(1 - \frac{u^2}{u_\infty^2}\right) \mathrm{d}y \tag{3.183}$$

In the above expressions, δ^{**} is the *energy thickness*. Note that δ^{**} is, just like the boundary-layer and momentum thickness, also a function of the down-stream location, x.

When a laminar boundary layer faces an instability of the flow it may lead to transition from laminar to turbulent flow. Instability first occurs when the velocity profile develops an inflection point (see Fig. 3.34, velocity profile C). Alternatively, the laminar boundary layer may separate from the surface, at a point where the velocity gradient, $\frac{\mathrm{d}u}{\mathrm{d}y}$, the surface shear stress, and the vorticity all reach zero (White 2005). The flow becomes unsteady, but could reattach to the wall and transition into an attached turbulent boundary layer.

In the next section, when discussing hypersonic flows, we will revisit the boundary layer and relate the discussion to the case at hand. And, since the boundary layer is the driving region for heating the vehicle, in Sect. 3.6 the boundary-layer expressions are used to establish expressions for the thermal loading.

3.4 Hypersonic Flow

In terms of vehicle performance, the supersonic phase of a re-entry mission is of less importance. This phase has little impact on the landing footprint—it started at about 100 km from the landing site for the Space Shuttle—and also the major thermal loads have already long since happened. However, the terminal-area targeting with all related manoeuvres typically begins during the supersonic phase, as will be discussed in more detail in Chap. 8. The (winged) vehicle will experience the "Mach Tuck", though, but using a flight control system reduces this effect to an insignificant one for the pilot. One exception, when the supersonic phase is more relevant, is supersonic parachute deployment (used for Mars descent, see also Chap. 9). However, supersonic parachute aerodynamics, no matter how important, is beyond the scope of the current discussion.

In hypersonic flow, the basic (supersonic) phenomena as discussed in the previous Sect. 3.3 occur here as well: normal and oblique shocks, expansion waves, and there is, of course, a boundary layer close to the vehicle wall. However, Mach number and Reynolds number are much higher in this flow regime, so one can expect an influence of that on these phenomena. Moreover, due to the excessive heat that is generated due to friction with the vehicle, the atmosphere behaves differently, and cannot always be treated as an ideal gas anymore. With this in mind, we move on to the world of hypersonic aerodynamics. In Sect. 3.4.1 we will discuss the main differences in flow properties compared to supersonic flow. Section 3.4.2 presents some more details on hypersonic boundary layers, whereas in Sect. 3.4.3 several flow complications are introduced, most notably the viscous, vorticity, and shock interactions, and the real-gas effects. Finally, Sect. 3.4.4 summarises several engineering methods to calculate the pressure distribution for different parts of the entry vehicle, which allows for force and moment calculation.

3.4.1 Flow Properties

The particular supersonic-vehicle shape, i.e., a slender body with a sharp nose, and thin wing and tail profiles with sharp leading edges, is the best configuration to reduce drag, due to the attached shocks. This comes at the expense of an increased heating, although this effect remains small and leads to a vehicle-surface temperature of a few hundreds °C. For higher Mach numbers, the heating may become excessive and different vehicle shapes have to be considered. For this reason we find a strong influence of nose bluntness on hypersonic vehicle design. This lowers the heating, but significantly increases the drag. However, minimum drag is not necessarily a requirement for an entry vehicle, as its main function is to shed entry velocity to reach an acceptable landing speed.

The Rankine-Hughoniot relations, Eqs. (3.115)–(3.117), define the relation of flow properties temperature, pressure, and density before and after an oblique shock-wave, as a function of free-stream Mach number, M_1, and (wall) deflection angle, θ. For hypersonic flow with large Mach number (and theoretically approaching ∞), the limit case for these properties is given by

$$\frac{p_2}{p_1} \approx \frac{2\gamma}{\gamma+1} M_1^2 \sin^2 \beta \tag{3.184}$$

$$\frac{\rho_2}{\rho_1} \approx \frac{\gamma+1}{\gamma-1} \tag{3.185}$$

$$\frac{T_2}{T_1} \approx \frac{2\gamma(\gamma-1)}{(\gamma+1)^2} M_1^2 \sin^2 \beta. \tag{3.186}$$

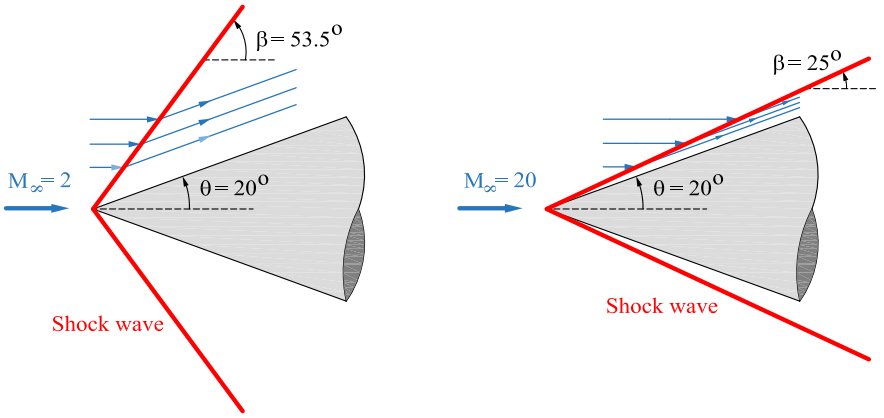


Fig. 3.35 The shock-cone geometry for two different Mach numbers

The relation between M, θ and the shock angle, β, was shown in Figs. 3.25 and 3.26. Whereas at low Mach numbers the shock wave appears to be "far away" from the wall, the encapsulating shock cone becomes smaller and smaller with increasing Mach number. In Fig. 3.35 the effect of Mach number on shock angle is shown. For $M_\infty = 2$, β is about 53.5°, which leaves an angle of 23.5° between shock and vehicle wall. However, at $M_\infty = 20$, the shock angle has reduced to about 25°, which only leaves a very small distance. At even higher Mach numbers the shock will be almost touching.

The fraction of the total energy that appears as frictional heat in the body depends upon the characteristics of the flow around the body. The fraction of the total heat load that is transferred to the atmosphere is directly proportional to the strength of the shock wave, which means that the stronger the shock wave, the smaller the frictional component imposed on the body. Because a blunt nose produces the strongest shock waves, an entry vehicle is given a blunt shape rather than the slender, streamlined configuration that has been the classic solution to aerodynamic problems at lower speeds (Fig. 3.36).

Figure 3.26 has already highlighted the detached-shock case where $\theta > \theta_{\max}$. A blunt vehicle shape automatically falls into this category, and a detached bow shock is a very common phenomenon for hypersonic vehicles. However, despite the fact that the total energy entering the vehicle compared to the amount transferred to the atmosphere is small, the remaining amount is a quantity not to be underestimated. The aerodynamic heating effects on the vehicle are very strong, and the vehicle's structure is loaded to its maximum capacity. The aerodynamic heating drives the vehicle and mission design: the vehicle shape determines the lift and drag characteristics that limit the flyable trajectories, and the resulting thermo-mechanical loading has to be sustained by the vehicle's structure and thermal-protection system.

The flow analysis is not easy, though, and to understand the complete picture extensive (and complicated) CFD analysis is required. Moreover, validation of the

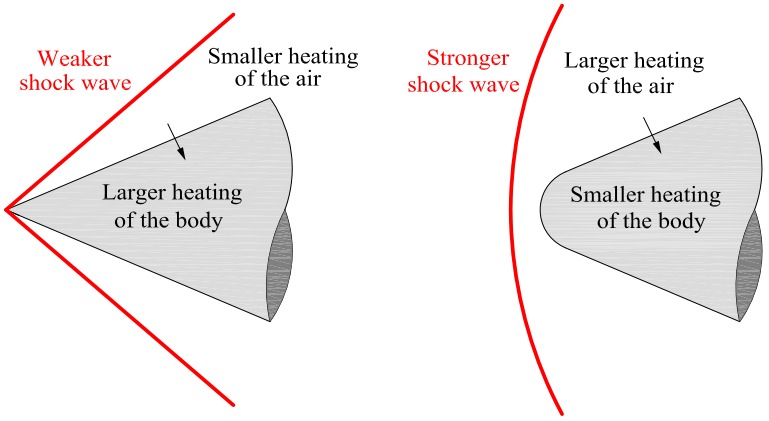


Fig. 3.36 Comparison of flow behaviour of sharp and blunt vehicles in supersonic and hypersonic flow

calculated results is not straightforward, due to the limited availability of experimental data. Real-gas effects play an important role, because of the large amount of energy that the flow absorbs. Effects, such as dissociation, vibrational excitation, ionisation, and shock-layer radiation occur, and also the chemical properties of the flow are no longer constant, but become at the very least temperature and time dependent. The boundary layer can develop into a thick layer that can transition from laminar to turbulent, but when (or if) this happens is difficult to predict.

With a thin shock layer close to the vehicle a strong interaction between the viscous and inviscid flow originates. Therefore, the viscous boundary layer and inviscid region behind the flow field can no longer be treated separately. This gets even more complicated due to the interactions between shock waves, and shock wave and boundary layer, which can cause local peaks in the heating. And last, but not least, a thick entropy layer behind the curved shock wave will emerge, which further complicates the analysis.

The development of the hypersonic boundary layer will be treated in more detail in the next section, whereas the other mentioned physical effects will be discussed in Sect. 3.4.3.

But before we go there, a few more words on the flow analysis. With all these (very non-linear) effects, is conceptual flow analysis possible at all? Or do we have to rely on existing aerodynamic databases that have been compiled for other missions?

In the hypersonic regime, application of limit cases can greatly simplify the analysis, since $M = 10$ or $M = \infty$ makes little difference in many equations. This *Mach-number independence principle* of inviscid flow facilitates the analysis of, for instance, pressure distribution, lift, pressure and wave drag, shock-wave shapes, and Mach-wave patterns, even though real-gas effects (slightly) break this independence. This means that with so-called local-inclination methods for inviscid approximation conceptually the aerodynamic characteristics can be determined. These methods are

very fast (in the order of seconds) compared to CFD, and the accuracy is reasonably good. Especially when we realise that viscous forces are generally quite small compared to the pressure forces.

However, the heating effects on the surface, the occurring shock and expansion-wave pattern, and the aforementioned interactions between shock waves and boundary layer are much more difficult (not to say impossible) to obtain with conceptual analyses. Therefore, the main focus of CFD is on aerodynamic-heating analysis, and not the exact determination of lift and drag curves. For this reason, in Sect. 3.4.4 several local-inclination methods will be discussed to calculate aerodynamic force and moment coefficients, with examples showing the results for the Apollo capsule and the Space Shuttle.

3.4.2 Boundary Layer

In Sect. 3.3.5 the development of an incompressible boundary was discussed, to create understanding of boundary-layer dynamics and to get insight in the most important parameters. With increasing Mach number, the flow can no longer be considered incompressible, and the simplified analysis no longer holds. Understanding of a compressible boundary layer is important for setting up an aerodynamic-heating analysis that is covered in Sect. 3.6. To avoid an overlapping discussion, the detailed analysis of the hypersonic boundary layer is postponed until later, and here we will summarise the main aspects.

The thickness of the boundary layer, δ, which is defined as the distance from the wall where the flow reaches 99% of the free-stream velocity, grows as follows for hypersonic laminar flows:

$$\frac{\delta}{x} \sim \frac{M_e^2}{\sqrt{Re_x}} \tag{3.187}$$

where the e subscript denotes that it has been evaluated at the "edge" of the boundary layer. Hirschel and Weiland (2009) give a relation for the thickness of a compressible boundary layer on a flat plate:

$$\delta = C \frac{x^{1-n}}{Re_e^n} \left(\frac{\rho_e \mu^*}{\rho^* \mu_e} \right) \tag{3.188}$$

with the * denoting quantities evaluated at a reference temperature (to be discussed later, in Sect. 3.6). For laminar flow, $C = 5$ and $n = 0.5$, while for turbulent flow, $C = 0.37$ and $n = 0.2$.

Parameters that can be used to quantify how much the shape that the inviscid flow "sees", is affected by the boundary layer—this can be significant for high Mach numbers—are the momentum thickness, θ, and the displacement thickness, δ^*. The

momentum thickness is physically related to the skin friction and sometimes used in turbulence predictions. It is defined as:

$$\theta = \int_0^\infty \frac{\rho u}{\rho_e u_e}\left(1 - \frac{u}{u_e}\right) \mathrm{d}y \tag{3.189}$$

with the y-direction taken perpendicular to the wall again. Physically, it represents the loss of momentum that flows through the boundary layer as compared to the inviscid case. Similarly, δ^* is defined as the decrease in mass flow, when compared to the inviscid case, and is defined as:

$$\delta^* = \int_0^\infty \left(1 - \frac{\rho u}{\rho_e u_e}\right) \mathrm{d}y \tag{3.190}$$

Because the boundary layer encloses the body, its behaviour is crucial in determining the flow's effects on the vehicle. Since $\frac{\partial p}{\partial y}$ is assumed to be zero over the boundary layer, pressure forces may be computed using the pressure at the edge of the boundary layer. Skin friction and heating, however, must be computed from the interface between the fluid and the wall as follows:

$$\tau_w = \left(\mu \frac{\partial u}{\partial y}\right)_w \qquad q_w = \left(k \frac{\partial T}{\partial y}\right)_w \tag{3.191}$$

which can interpreted as the viscous transport of momentum and energy at the wall. For the heat transfer, an additional term may need to be added to account for recombination at the wall and the resultant release of heat. Such matters will be dealt with in Sect. 3.6.

Besides the mechanical no-slip condition at the wall, there is a thermal no-slip condition as well, $T_w = T|_{y=0}$, such that the temperature of the wall equals the temperature of the gas at the wall (at $y = 0$). In general, the wall temperature will be quite different from the free-stream temperature. This results in a thermal boundary layer near the wall, and in many ways its description follows that of the velocity (momentum) boundary layer. It is a layer with strong temperature gradients, and for a no-slip condition reduces from the wall temperature to a value of 99% of the free-stream temperature, see Fig. 3.37. The corresponding thickness is indicated by δ_T.[18] For laminar flow over a flat plate at zero incidence, the thermal boundary-layer thickness is given by Schlichting and Gersten (2017):

$$\frac{\delta_T}{\delta} \approx Pr^{-\frac{1}{3}} \tag{3.192}$$

[18] Note that the maximum temperature in the thermal boundary layer occurs at the wall, which is the equilibrium hot-wall situation. This is not necessarily always the case, but could happen slightly above the wall, after which the temperature decreases to the wall value. See the extensive treatment of White (2005) on this subject matter.

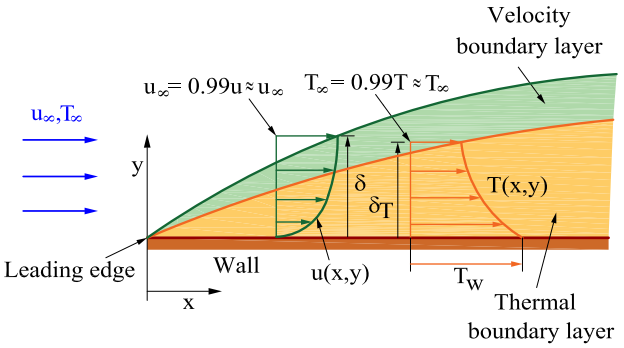


Fig. 3.37 Velocity (momentum) and (hot-wall) thermal boundary layers

whereas White (2005) gives a slightly different value:

$$\frac{\delta_T}{\delta} \approx Pr^{-\frac{2}{5}} \tag{3.193}$$

Alternatively, for turbulent flow over a flat plate δ_T is not determined by thermal diffusion, but instead, the driving force that determines δ_T originates from random fluctuations in the outer layers of the fluid's boundary layer. So the thickness for turbulent flow does not depend on the Prandtl number, but rather on the Reynolds number. Hence, the turbulent thermal boundary-layer thickness can be approximated by

$$\delta_T = \left(\frac{0.378}{Re_x^{1/5}}\right) x \tag{3.194}$$

which is the same as the turbulent velocity boundary layer thickness for incompressible flow, Eq. (3.167).

As mentioned, for boundary layers, and viscous flow in general, in general two types of flow exist, laminar and turbulent. Certain "intermediate" types, such as intermittent or transitioning flow, will not be considered here. Laminar flow can be seen as a succession of layers of constant velocity, starting at the wall (with zero velocity) and up to the external flow, where it reaches the free-stream value. In laminar flow, the particles mainly stay in their own layer, so that there is no macroscopic mass exchange between the layers. When the flow is disturbed, however, this laminar behaviour may break down and turbulent flow arises. In turbulent flow, the particles behave chaotically, not following a path that can be clearly pre-defined, as shown in Fig. 3.38.

(Semi-)empirical models which can be used to predict mean turbulent-flow quantities are available, such as the Baldwin-Lomax model (Anderson 2019). These models often involve the addition of "turbulent" or *eddy* viscosity and conduction coefficients to those that are already present in the equation. Although the occurrence of turbulence does not cause an increase in the physical viscosity coefficient, its effect may

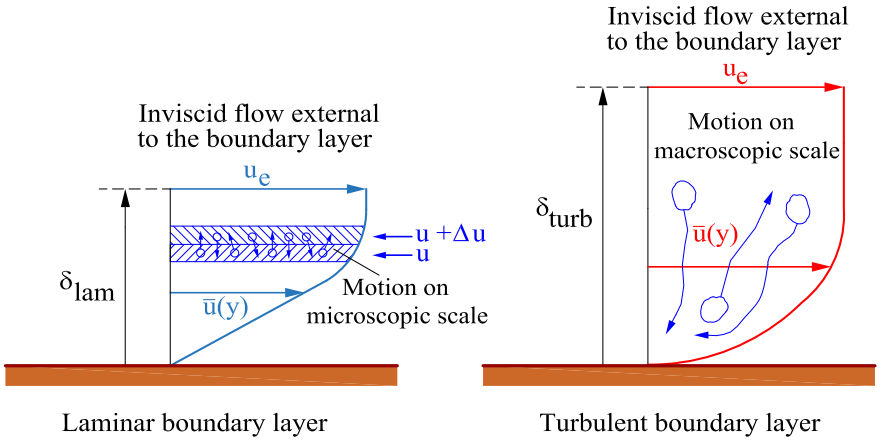


Fig. 3.38 Comparison of velocity profile and particle motion in laminar and turbulent boundary layers

be modelled as such. Such models, which relate the effects of turbulence to the governing equations, are called turbulence-closure models.

Aside from determining flow quantities involving turbulence, one of the major challenges is to predict where and how approximately the flow will transition from laminar to turbulent flow. This is important to know, since the skin friction and heating increase in turbulent flow when compared to laminar flow. This will lead to higher thermal loads on the vehicle, and may affect the design of the thermal-protection system.

Flow transition in the hypersonic regime is discussed by many, e.g., by Bertin (1994), White (2005), and Anderson (2019). A more extensive discussion, which also specifically deals hypersonic flow, is provided by Arnal and Delery (2004). Although a full discussion of this topic is by far too extensive to present here, some general comments that can be used in the design process are in order. Unfortunately, no single theory or correlation exists as yet to predict the onset of turbulence in hypersonic flow. Some of the correlations that we present here can be used as "guidelines", though. These correlations are usually based on some local Reynolds number. One of these correlations for hypersonic flow is the following, which was used in the design of the Space Shuttle (Anderson 2019):

$$\frac{Re_{\theta_{tr}}}{M_e} = 100 \tag{3.195}$$

where $Re_{\theta_{tr}}$ is the Reynolds number based on the momentum-loss thickness θ as characteristic length. Regrettably, no suggestions on its region and degree of applicability are given. A seemingly less conservative estimate is given by Cruz and Ware (1992), which is used for the analysis of the HL-20 lifting body. It distinguishes between axisymmetric and flat-plate regions:

$$\frac{Re_{\theta_{tr}}}{M_e} = \begin{cases} 150 & \text{axisymmetric body} \\ 350 & \text{flat plates} \end{cases} \tag{3.196}$$

Only flat-plate and axisymmetric-body estimates are given in this paper, since the inviscid analysis of the entire configuration is done using either tangent-wedge or tangent-cone methods (see the upcoming Sect. 3.4.4). Also now, no more information on the estimates is offered, apart from the stated numbers.

Other quantities that affect transition are, among others, the pressure gradient, wall temperature, and wall roughness. The effect of wall roughness will be clear, at least qualitatively, in that a rougher surface will perturb the flow and thus initiate a more rapid transition to turbulent flow. The effect of wall temperature can be understood through the fact that cooling the wall stabilises the boundary layer and delays transition. This is possibly associated with the smaller total amount of thermal energy that the boundary layer possesses in such a case. Pressure gradients, which will occur over a curved surface, will have an influence on the stability of the flow. A convex surface (e.g., an airfoil), typically has a stabilising effect, while a concave surface destabilises the flow. One well-known exception to this is the so-called "blunt-nose paradox", where a strongly decelerated flow is "tripped" to turbulence very early on a blunt nose, at a very low Reynolds number. This effect can, of course, have very unfavourable effects, resulting in a flow that is turbulent over a large portion of the vehicle. To understand the consequences of turbulent flow, it is advised not to restrict oneself to an analysis of a laminar-transition-turbulent flow, using some (or possibly multiple) transition criteria. Also, as a worst case, a fully turbulent flow should be simulated and analysed.

An additional boundary condition, which is not unique to viscous boundary-layer flows, but is instead a result of more general physical considerations, is the following

$$\mathbf{V_w} \cdot \mathbf{n} = \mathbf{0} \tag{3.197}$$

which states that the velocity normal to the body at the wall equals zero, called the flow tangency condition. This condition may not be satisfied in case of mass injection into the flow at the wall, for instance, when ablative or transpiration cooling is used (see Chap. 10). In addition to specifying the velocity at the wall, the temperature should also have a boundary condition. The choice of this condition depends on the model that is being used and will be discussed later (Sect. 3.6).

3.4.3 Flow Complications

In Fig. 3.3 a number of aerodynamic phenomena was shown for different categories of vehicles, re-entry vehicles being one of those. For a typical re-entry vehicle, some of the effects have been highlighted in Fig. 3.39. A characteristic for vehicles with a blunt nose is a detached bow shock that is at a stand-off distance δ_{so} from the

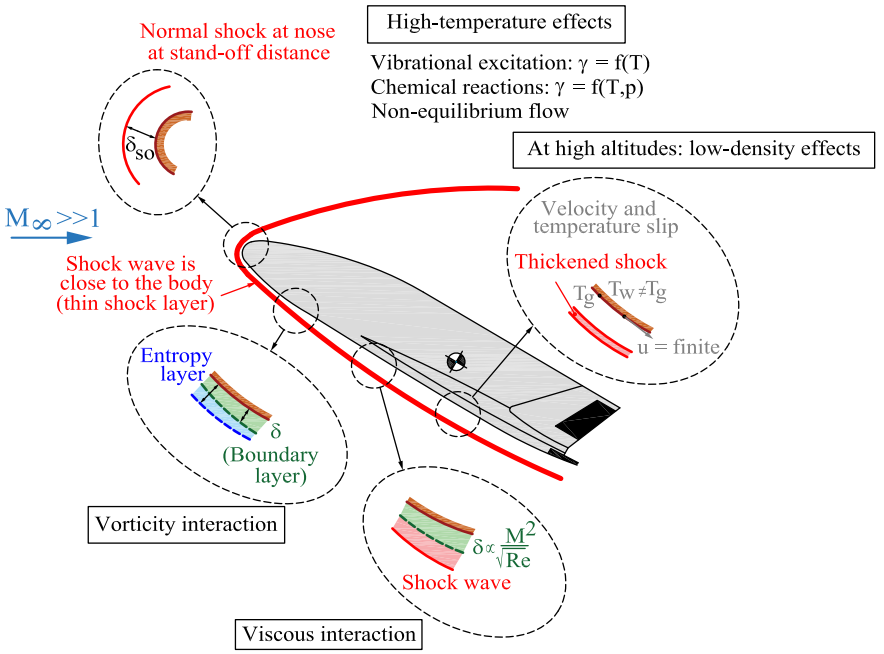


Fig. 3.39 The complete picture of hypersonic flow

nose. The (initially thin) shock wave is close to the vehicle, which means that an interaction between the shock layer and the boundary layer occurs, the so-called viscous interaction. Another interaction is that between the boundary layer and the entropy layer, an effect that becomes stronger with increasing Mach number. This interaction is called the vorticity interaction. Depending on the (complexity of the) vehicle shape, there may be a whole series of shock and expansion waves that also interact, complicating matters even more.

When the Mach number is high, the gas around the vehicle can no longer be considered ideal, as high-temperature effects start playing a role. Real-gas effects, such as vibrational excitation, ionisation, dissociation, and recombination, will alter the properties of the gas, and become temperature dependent. Finally, at high altitude where the atmospheric density is low, an effect that is called velocity and temperature slip occurs. Normally, in a boundary layer the flow will be zero at the wall due to friction, and also the temperature of the gas will be equal to the wall temperature. Velocity and temperature slip indicate that this will no longer be the case.

In the following, we will briefly discuss some of the aforementioned effects, notably the viscous, vorticity, and shock interactions, as well as the real-gas effects. A summary of the effects and how these complicate the design of entry vehicles will conclude the discussion about hypersonic flow.

3.4.3.1 Viscous Interaction

In hypersonic flow at high altitude (high M and low Re), the boundary layer becomes thick and its influence more pronounced. Because the shock wave is close to the body, the shock layer and the boundary layer can reach similar sizes, and the shock layer can become (nearly) fully viscous. This complicates the flow-field analysis, as the use of inviscid methods is no longer justified. An additional effect is that the boundary layer can affect the inviscid flow field far from the body, a phenomenon called viscous interaction.

For a (viscous) flow over a flat plate, the free-stream inviscid streamlines are displaced upward, creating a shock wave at the leading edge. Here, the pressure is increased and only downstream the pressure over the flat plate converges to the inviscid-flow value. The increased pressure (and thus increased density) compresses the boundary layer a bit, which in turn causes increased velocity and temperature gradients at the wall. As we know from Eq. (3.191), the result is that the skin friction and heat transfer are increased compared to their values that would have occurred if the pressure had just been the free-stream value.

The inviscid flow is most strongly affected near the leading edge (Fig. 3.40), where the rate of growth of the boundary-layer displacement thickness, $\mathrm{d}\delta^*/\mathrm{d}x$, is large, and leads to a so-called *strong interaction*. In principle this means that the incoming free stream "sees" an effective body with rapidly growing thickness. The phenomenon of *weak interaction* occurs further downstream, where $\mathrm{d}\delta^*/\mathrm{d}x$ is quite small. It means that the inviscid flow is not strongly affected and, *vice versa*, the feedback to the boundary layer is negligible and can be ignored.

The strength of the viscous interactions can be characterised by the parameter χ (Anderson 2019):

$$\chi = \frac{M_\infty^3}{\sqrt{Re}}\sqrt{C} \qquad C = \frac{\rho_w \mu_w}{\rho_e \mu_e} \tag{3.198}$$

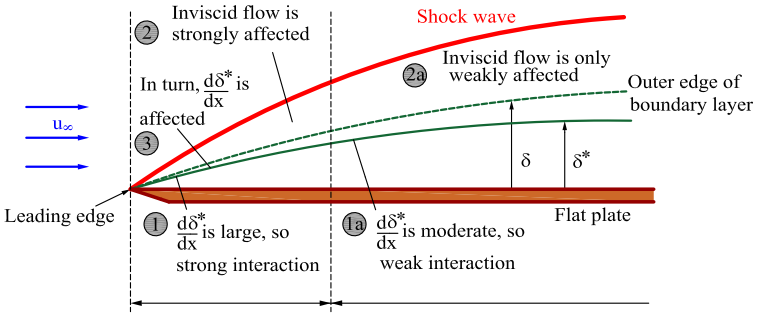


Fig. 3.40 Viscous interaction on a flat plate

with, as usual, the subscript w indicating wall conditions, and e representing the conditions at the edge of the boundary layer. The strong interaction occurs, approximately, for $\chi > 3$, and the weak interaction for $\chi < 3$. As mentioned, the effect of the viscous interaction will increase the pressure at the edge of the boundary layer due to the deflection of the flow. This increase can be quantified by the following equations for the interaction effect over a cold-wall flat plate:

$$\left(\frac{p}{p_\infty}\right)_{strong} = 1 + 0.15\chi \qquad \left(\frac{p}{p_\infty}\right)_{weak} = 1 + 0.078\chi \tag{3.199}$$

Since the viscous interaction works on both sides of the body, the direct effect on lift is minor. Both the skin friction and wave drag are increased, though, which means that viscous interaction reduces the lift-to-drag ratio, L/D, of the body.

3.4.3.2 Vorticity Interaction

To discuss the development of the entropy layer, consider the flow situation around the blunt nose of a re-entry vehicle (Fig. 3.41). In front of the nose, at a stand-off distance, δ_{so}, a detached shock wave has formed. Anderson (2019) reports a semi-empirical method for determining the shape of the shock around sphere-cone and circular cylinder-wedge configurations. This method shows good agreement with numerical inviscid results for moderate hypersonic Mach numbers. It assumes a hyperbolic shock shape (or presumably a hyperboloid in the sphere-cone case) and experimental data are used to correlate shock shape to the free-stream Mach number. In addition, correlations for the shock stand-off distance can be found in literature, and these could be used in, for instance, a preliminary heating analysis. Two correlations, given by Johnson et al. (2007) and Olivier (2000), respectively, are:

$$\frac{\delta_{so}}{R_N} = \left(\frac{\frac{\rho_2}{\rho_1} - 1}{\frac{\rho_2}{\rho_1} - \sqrt{2\frac{\rho_2}{\rho_1} - 1}} - 1\right)^{-1} \tag{3.200}$$

$$\frac{\delta_{so}}{R_N} = 0.82\frac{\rho_1}{\rho_2} \tag{3.201}$$

The former of these is based on wind-tunnel data obtained for radiative-heating analysis, while the latter is based on a correlation for numerical data. The two correlations give somewhat different results, with the latter consistently producing higher values for δ_{so}. This difference is possibly caused by dissociation of air that occurred during the wind-tunnel experiments. It is further noted by Anderson (2019) that for very high Mach numbers, when the density ratio over the shock becomes very large, the second of the correlations given here, with a coefficient of 1 instead of 0.82, becomes the one of choice. For more details, the reader is referred to the quoted references.

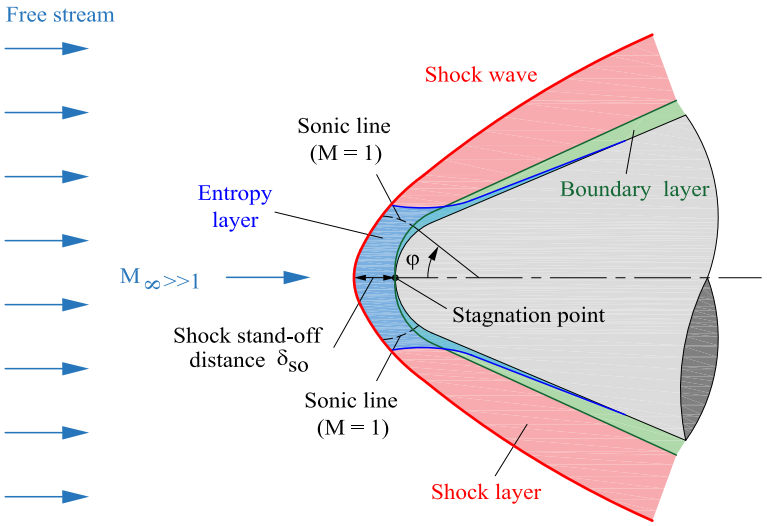


Fig. 3.41 Flow field around a blunt body, with swallowing entropy layer and interaction with the boundary layer

At zero angle of attack, a streamline aligned with the capsule's centre line will pass through a normal shock, and any flow particle will come to a stand still in the stagnation point. When we define the stagnation pressure, we assume an isentropic deceleration to the zero velocity state. For an isentropic process, Eq. (3.25) gives us

$$\frac{p_2}{p_1} = \left(\frac{T_2}{T_1}\right)^{\frac{\gamma}{\gamma-1}} \tag{3.202}$$

Since the velocity in the stagnation point is zero, pressure and temperature, p_2 and T_2, are equal to the total pressure and temperature (e.g., the total temperature, $T_0 = T + \frac{u^2}{2c_p} = T$), such that

$$\frac{p_0}{p_1} = \left(\frac{T_0}{T_1}\right)^{\frac{\gamma}{\gamma-1}} \tag{3.203}$$

Therefore, the stagnation state is defined by the total state, p_0 and T_0; in addition, the stagnation entropy is defined by its static value. From Eq. (3.23)

$$\mathrm{d}s = c_p \frac{\mathrm{d}T_0}{T_0} - R\frac{\mathrm{d}p_0}{p_0} \tag{3.204}$$

which integrates to

$$\Delta s = c_p \ln\left(\frac{T_{02}}{T_{01}}\right) - R \ln\left(\frac{p_{02}}{p_{01}}\right) \tag{3.205}$$

The constants of integration are zero, in the case of integrating total values.

For a process with $Q = W = 0$, the stagnation enthalpy, and hence the stagnation temperature, is constant, and with $\frac{T_{02}}{T_{01}} = 1$, the stagnation pressure is related directly to the entropy as,

$$\Delta s = -R \ln \left(\frac{p_{02}}{p_{01}} \right) \tag{3.206}$$

This relation, as well as the total-pressure drop over the normal shock, has been plotted in Fig. 3.42. The significant drop in total pressure, as well as the large entropy increase at higher Mach number reduces the heat in-flux in the stagnation point. This is not the whole story, though. Because the flow near the nose passes through a nearly normal shock, it will experience a much larger change in entropy, compared to flow passing through the much shallower shock angle further from the body centre line. As a consequence, the entropy of the inviscid flow behind the shock is not uniform, but instead shows large variations in its gradient. This affects several aspects of the flow field, one of which follows from Crocco's theorem, which reads (Hirschel 2015):

$$\mathbf{V} \times (\nabla \times \mathbf{V}) = -\mathbf{T} \nabla \mathbf{s} \tag{3.207}$$

This theorem shows that entropy gradients give rise to rotation (or *vorticity*) in the flow. Since the largest entropy gradients occur in the flow that passes through the strongly curved part of the shock, the streamlines closest to the body are most affected by it. This means that the region of strong entropy gradient, called the *entropy layer*,

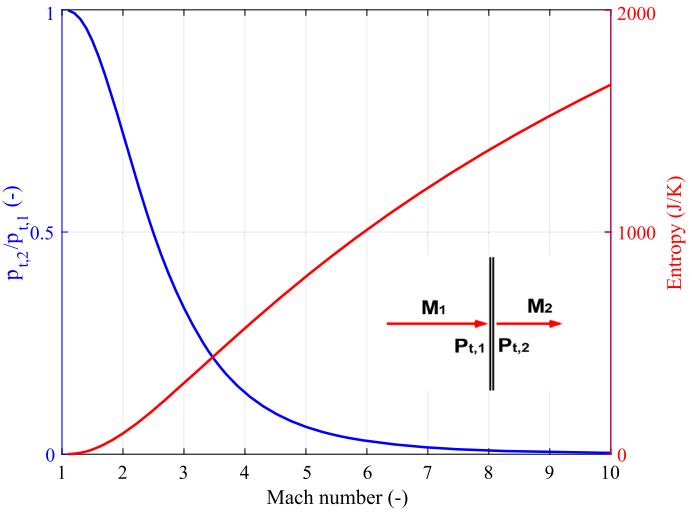


Fig. 3.42 The ratio of total pressure over a normal shock as a function of Mach number, as well as the entropy increase

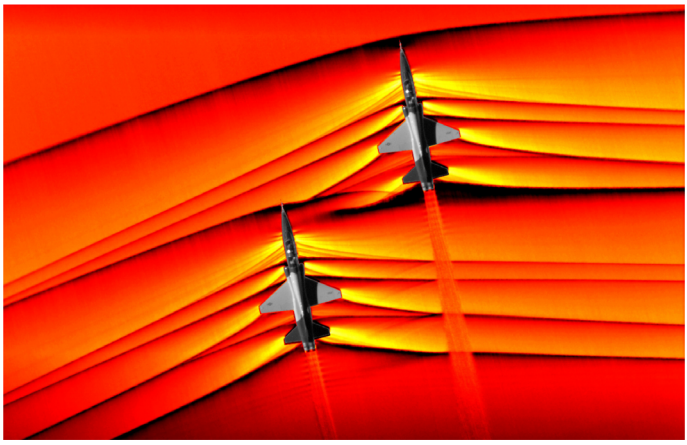


Fig. 3.43 This schlieren image shows the complex shock-wave pattern of two supersonic jets. Created with dedicated image-processing software (Image: NASA)

will interact with the boundary layer (often called *vorticity interaction*). The direct effect is that it compresses the boundary layer and decreases its thickness (Rasmussen 1994), such that the skin friction and heat transfer increase. This is a direct result of a thinner (thermal) boundary layer, since this will logically increase the velocity and temperature gradients at the wall: either quantity has to reach its free-stream value over a shorter distance.

3.4.3.3 Shock Interactions

A more complex vehicle shape, such as that of the Space Shuttle, can lead to complicated shock patterns. Shocks may impinge on the vehicle wall, leading to shock-wave-boundary-layer interaction, the effects of which can be destructive. An example of the complex shock patterns is shown in Fig. 3.43, where the first air-to-air image of the interaction of shockwaves from two supersonic T-38 aircraft flying in formation is shown. Originally in monochrome, this colour-enhanced schlieren picture shows the complex shock-wave patterns originating from the aircraft. This series of tests should provide insight into how shocks interact with aircraft plumes, as well as with each other. During this test, the two T-38s flew about 9 m from one another and at a 3-m difference in height.[19]

In the late 1960s, an X-15 rocket plane carried an experimental ramjet attached to a pylon, and during a Mach 6.7 test flight the shock wave originating at the nose spike of the ramjet impinged with the pylon, causing severe damage (Watts 1968). Figure 3.44 shows the pylon with the damage, obviously of a melt-back nature. The skins of

[19] https://www.nasa.gov/centers/armstrong/features/supersonic-shockwave-interaction.html. Accessed: August 6, 2019.

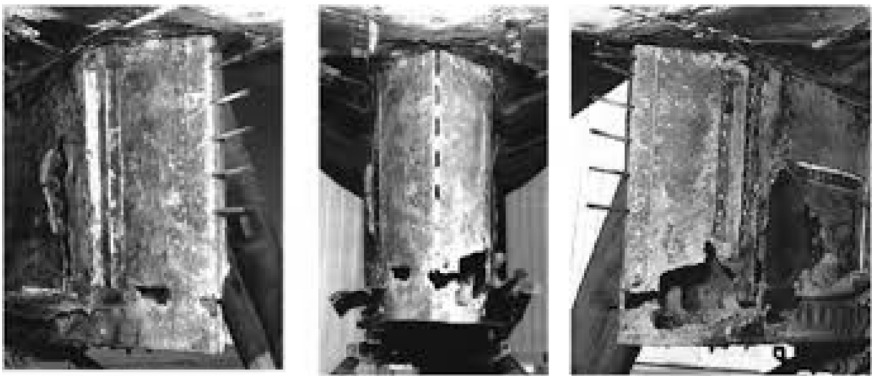


Fig. 3.44 Damage to the pylon carrying the experimental ram jet due to shock impingement (Image: NASA)

the pylon were made of Inconel-X, with a maximum operational temperature of about 815 °C (1500 F) and a melt point well beyond 1370 °C (2500 F), but far less than 1650 °C (3000 F). The material was simply overheated by shockwave impingement, and the structural damage led to dropping the ramjet.

Due to the high Mach numbers associated with hypersonic flow, the shock waves will generally lie relatively close to the vehicle body. This makes it more probable that shocks near the body interact, which can have serious consequences. Predicting the exact interaction pattern between shocks is a very complicated problem and well beyond the scope of this book to analyse. Some general considerations and a discussion of the various interaction types is in order, though. Lind (1997) distinguishes six types of shock-shock interactions between oblique shocks and bow shocks. He analysed the interactions for an X-43 like vehicle, where the shock originating at the nose impinges on the scramjet inlet cowl. Such interactions also occurred on the Space Shuttle. There, the shock from the first part of the wing impinged the bow shock ahead of the rear part of the wing, due to the double delta nature of the wing planform (Anderson 2019).

As can be seen in Fig. 3.45, the type of interaction depends on the angle with which the impinging shock hits the bow shock. Both Lind (1997) and Anderson (2019) identified the Type-IV interaction to be an especially important type, because such an interaction results in a supersonic jet that is shot in the direction of the body. This can cause extremely high heat fluxes, which the vehicle will most likely not be able to withstand.

Besides interacting with other shocks, a shock wave may also impinge directly on the vehicle body, in which case it will interact with the boundary layer. This effect is known as Shock-Wave Boundary-Layer Interaction (SWBLI). This interaction may follow from, for instance, the Type-III and Type-V shock/shock interactions, and are discussed in detail by Arnal and Delery (2004). If the analysis of shock-shock interaction is already not trivial, the occurrence and effects of SWBLIs are

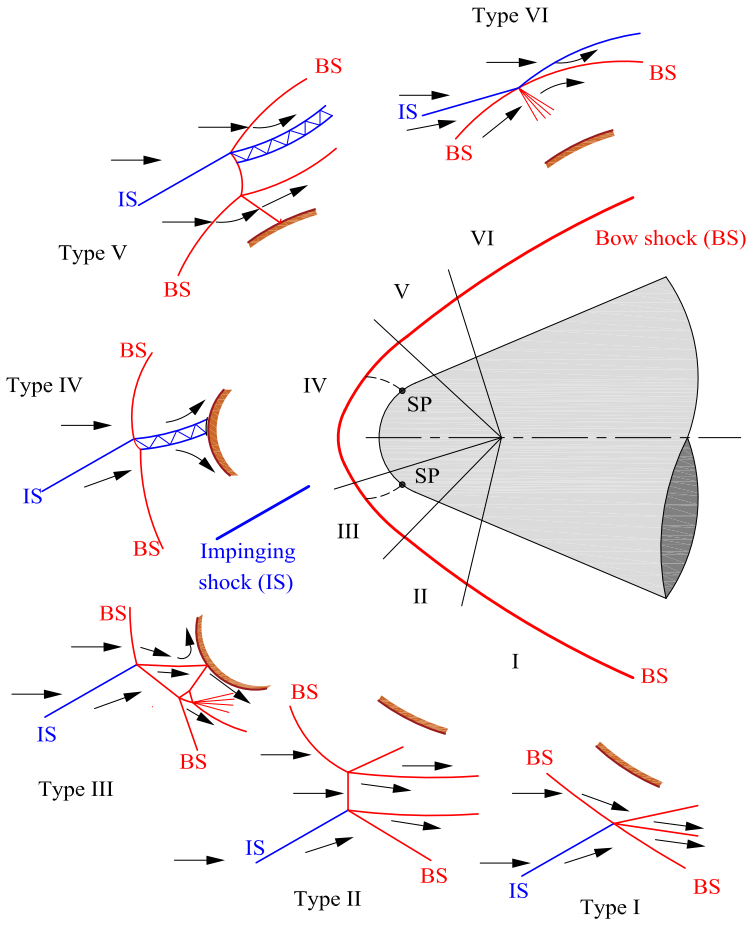


Fig. 3.45 Various types of shock/shock interaction patterns: SP = sonic point, IS = impinging shock, and BS = bow shock

even more complicated, and very much dependent on either the laminar or turbulent nature of the boundary layer. Moreover, the various models that are available often yield unsatisfactory results when transition occurs.

All in all, the effects of interactions are very difficult to assess, although some general statements can be made. Due to the impingement of the shock, the boundary layer may locally separate from the wall and later reattach, and this can cause very high peaks in both pressure and heat transfer. The strength of a shock that is required to cause boundary-layer separation increases with decreasing Reynolds number for laminar boundary layers. However, for turbulent boundary layers this is not true for the complete range of Reynolds numbers, amplifying the difference in reaction to shock impingement.

3.4.3.4 Real-Gas Effects

A jump over a normal shock, especially at higher Mach numbers, results in a very large temperature increase behind the shock in the nose area and beyond. These high temperatures cause chemical changes in the gaseous environment through which the vehicle flies, as the constituents begin to dissociate and the products of the dissociation react with one another. The gas can no longer be treated as a calorically perfect gas, which means that the ideal gas law and relations that have been derived under that assumption, are no longer valid. This obviously complicates computational matters significantly.

At a pressure of 1 atm., the dissociation of oxygen begins around 2,000 K (see Table 3.1) and is virtually completed around 4,000 K. At that temperature (coincidentally) nitrogen begins to dissociate and is fully dissociated around 9,000 K. The composition of air at 1 atm. as a function of T is shown in Fig. 3.46. At higher temperatures, the newly formed oxygen and nitrogen atoms ionise, further complicating the analysis of the flow, because electromagnetic effects may need to be included. It should be noted that the decrease in concentration of O around 5,000 K is not due to an actual reduction of the number of oxygen atoms, but is a result of the increase of the total number of separate particles due to the dissociation of N_2.

A positive aspect of the high temperature effects is that the specific heat of the gas increases, because exciting the vibrational degrees of freedom and dissociating the molecules requires quite some energy, energy that is no longer available to increase the temperature. Figure 3.47 provides the normal-shock temperature ratio for a calorically perfect and equilibrium reacting air. The plot indeed confirms that the gas temperature decreases significantly. This is obviously beneficial, as the gas temperature that would be reached if the gas were to stay calorically perfect, rises very quickly with increasing free-stream velocity. Even though both c_p and c_v are rising due to the high-temperature effects, it results, in fact, in a decrease of their ratio, γ, which gets closer to one. Any reduction in γ will show as a change in the limiting rations Eqs. (3.184) through (3.186). For instance, for air with $\gamma = 1.4$ the limiting density ratio at high M is six, but for high-velocity equilibrium air, this ratio can reach 15–20 for an entry from LEO (Anderson 2019).

A change in γ, or in broader sense, the real-gas effects, will also have its impact on the pressure distribution and the location of the centre of pressure. A change in either one, but even more so their combination, can have a significant impact

Table 3.1 High-temperature effects on air

Temperature (K)	Chemical change
800	Molecular vibration
2,000	Oxygen molecules (O_2) dissociate
4,000	Nitrogen molecules (N_2) dissociate, nitric oxide (NO) forms
9,000	Oxygen and nitrogen atoms ionise

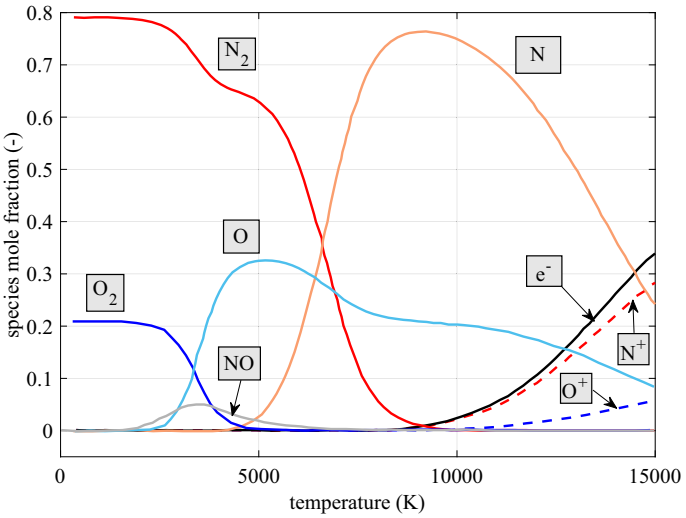


Fig. 3.46 Dissociation and ionisation of air at one atmosphere. Data taken from Barbante (2004)

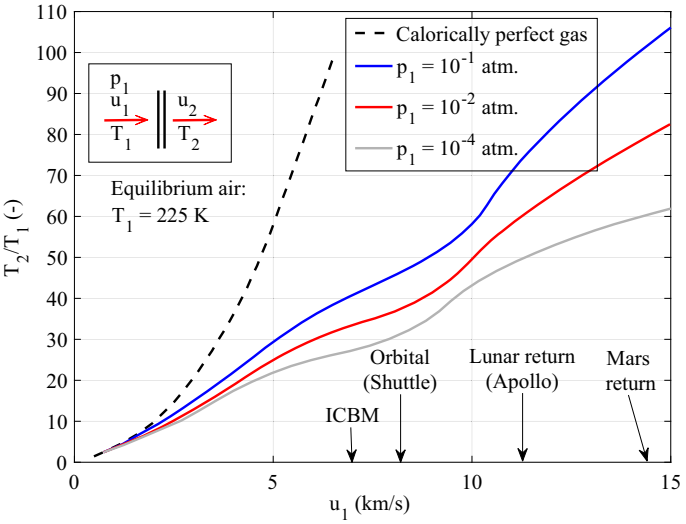
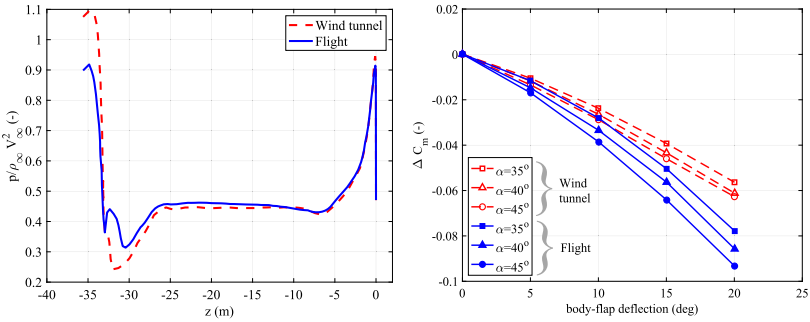


Fig. 3.47 Comparison of equilibrium and calorically perfect temperature ratios over a normal shock (Data taken from Anderson (2003))

on the pitch moment. Figure 3.48a shows the pressure distribution in longitudinal direction for the Space Shuttle, obtained from wind-tunnel measurements and in-flight measurements/estimates. Especially in the nose region, there are substantial differences that would lead to a change in pitch moment. In Fig. 3.48b, the body-flap pitch-moment increments are shown for three different angles of attack. With



(a) Pressure distribution (b) Body-flap pitch-moment increment

Fig. 3.48 Difference in Space-Shuttle aerodynamic characteristics between wind-tunnel and flight data (Data taken from Brauckmann et al. (1994))

increasing body-flap deflection the discrepancies become larger, and are increasingly diverging with angle of attack. The gravity of these discrepancies becomes apparent when one realises that the combination of errors in vehicle pitch moment and body-flap efficiency can potentially lead to a trimmability problem.

And this is exactly what happened during the first Space Shuttle flight, STS-1, flown by *Columbia*. During this flight, a significant anomaly in the modelled and actual required body-flap deflection was found (Maus et al. 1983). Whereas pre-flight predictions indicated that the Shuttle was well-trimmed at high Mach numbers, during flight it appeared that the body flap had to be deflected close to its maximum limit (Fig. 3.49). At lower Mach numbers, the discrepancy became a bit smaller, but was still significant nonetheless. Post-flight analysis confirmed a measurable influence of real-gas effects, which, at the time, could not be predicted by numerical analyses and wind-tunnel tests.[20]

3.4.3.5 Design Complications

Until now we have seen that hypersonic flight is difficult, and fundamentally different from supersonic flight. We will highlight some of the important aspects of hypersonic flight, and relate them to some design challenges. These challenges are more significant for lifting/winged bodies than for capsules. Capsule shapes are inherently simpler, and will have a predictable shock pattern. This makes it easier to use analytical approximations, and to apply some corrections to at least partially compensate for the differences with full-blown numerical design tools.

[20] Figure 3.65 shows a more detailed breakdown of the differences between pre-flight predictions, wind-tunnel tests, and flight data of the Space Shuttle's pitch-moment coefficient. An additional conclusion derived from these results is that also wind-tunnel data should be considered with care, and may not always be used as validation data.

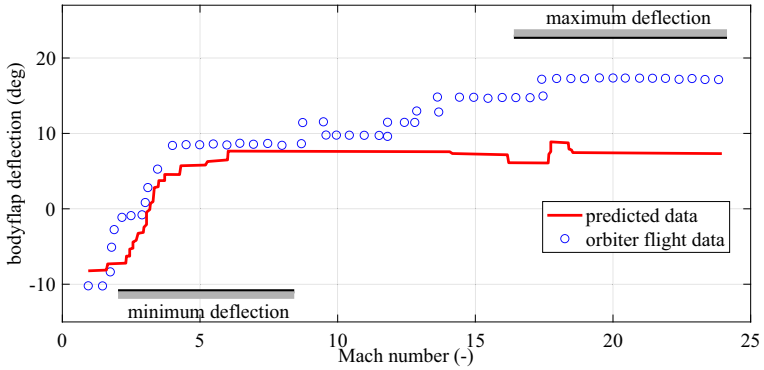
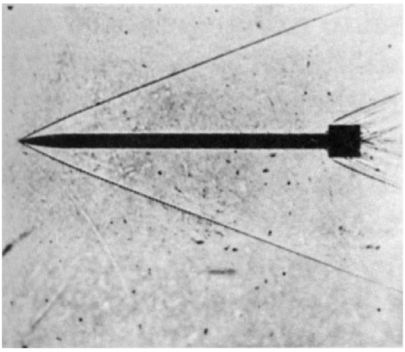


Fig. 3.49 Space Shuttle pitch moment underprediction and trim problem (Date taken from Maus et al. (1983))

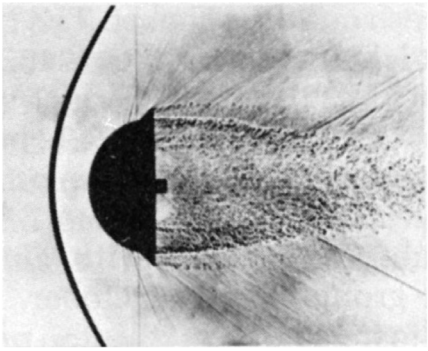
In Fig. 3.50 the development progress of the Mercury programme is shown. Extensive wind-tunnel tests supported the programme, and the acquired schlieren images show the impact of the vehicle nose on the shock pattern. The initial concept resembled a classical missile with a sharp nose; the shockwave is therefore attached to the nose (Fig. 3.50a). Engineers soon realised that this would not work for hypersonic flight, as the nose would burn away. The next concept was a blunt-body concept, and it is immediately clear that the shock is a detached normal shock (Fig. 3.50b). The wide base created a large, turbulent wake that resulted in a large wave drag. An updated configuration had a more flattened heat shield, which created a very thick bow shock (Fig. 3.50c); the smaller base reduced the wake effect significantly, although the short length created strong expansion and aft-body shock waves. The final, chosen configuration was the combination of a blunt, flattened but still rounded front shield, with an inverted conical aft body (Fig. 3.50d). The expansion waves are much weaker now, and also the turbulent wake flow has reduced.

Close to a blunt leading edge, detached shock becomes highly curved. Since flow near the nose passes through a nearly normal shock, it will experience a much greater change in entropy compared to flow passing through the much shallower shock angle further from the body centre line. Thus, strong entropy gradients exist near the leading edge generating an "entropy layer" that flows downstream along the body surface.

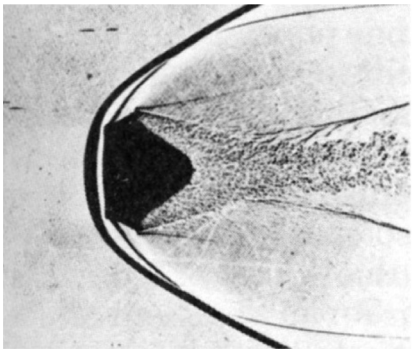
As the Mach number increases, the shock angle becomes smaller, and distance between surface and shock wave decreases with increasing speed. For a hypersonic body, the shock distance from the surface can become very small over a large portion of the body. The resulting flow field between the surface and the shock is often referred to as a shock layer. This thin layer can produce many complications in vehicle design, e.g., the shock layer may merge with the boundary layer at low Reynolds numbers to form a fully viscous shock layer. In the limit case of the Mach number approaching infinity (or at least becoming very large), the shock layer forms an infinitely thin and infinitely dense sheet, or, essentially, a flat plate.



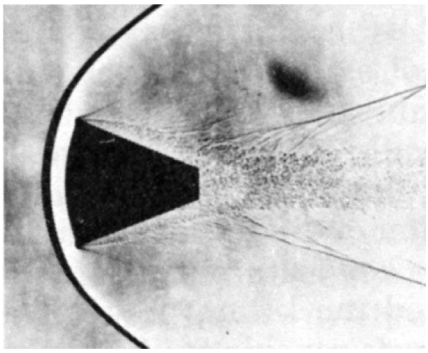
(a) Initial concept



(b) Blunt-body concept 1953



(c) Missile nose-cones 1953-1957



(d) Manned-capsule concept 1957

Fig. 3.50 Wind-tunnel research to support the mercury programme (Images: NASA)

An increasing Mach number does not only affect the shock angle, also the boundary layer can grow rapidly, which results in very high drag. If the boundary layer becomes thick enough, it may even affect the inviscid flow field far from the body, a phenomenon called viscous interaction. This viscous interaction can have a great influence on the surface-pressure distribution and skin friction on the body, thereby affecting the vehicle's lift, drag, stability, and heating characteristics.

In Fig. 3.39 some of the primary differences with supersonic flow were shown, which may create complications in hypersonic flow. Expanding our view, in Fig. 3.51 the flow situation around a potential entry capsule is shown. Due to the very large (and blunt) aeroshell, the bow shock is very well detached. In the shock layer many phenomena can be seen. A growing laminar boundary layer, which at some point transitions to turbulent. Corner flow around the heat shield, leading to flow separation and successive reattachment, but the flow impinges the aft body and creates a local hot spot. The heat shield itself is an ablator, and the ablation material mixes with the boundary layer. Due to the strong normal shock, the gas behind the shock heats up and becomes very energetic, such that all kinds of real-gas effects

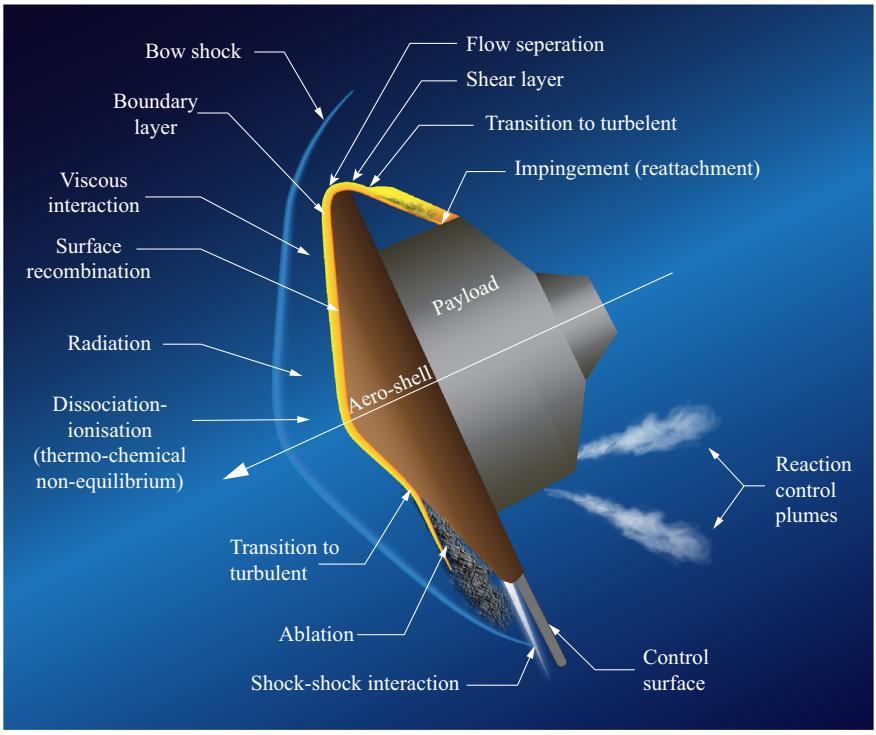


Fig. 3.51 Detailed flow field around an entry capsule

occur: ionisation, dissociation, and recombination, and the flow will never be in thermo-chemical equilibrium. As a result, the aerodynamic-heating effects are very strong.

A strong coupling between viscous and inviscid flow exists due to the thick boundary layer and thin shock layer, and the viscous boundary layer has merged with the shock layer. Strong entropy gradients occur, and the rise in entropy creates vortices that interact with the boundary layer. Finally, a potential trim surface has been added to the edge of the heat shield, to counter the induced pitch moment when a non-zero angle of attack has to be flown. The occurring expansion and shock waves interact with the bow shock, leading to complicated flow patterns. Last but not least, the capsule's primary control system uses reaction-control thrusters, of which the plumes interact with the wake flow.

The described flow field is complicated to solve, when taking all phenomena into account. However, for aerodynamic-heating analysis this is usually required. Hypersonic vehicles create so much heat and such high temperatures due to shock-wave compression and boundary-layer frictional heating, that they cause chemical changes to occur in the fluid through which they fly. These gas-chemistry effects are

important factors in heat dissipation on hypersonic vehicles, and, therefore, must be extensively analysed.

When one is "only" interested in certain flow phenomena or just the pressure distribution, simplified methods can be used; e.g., the local inclination methods for inviscid approximation. One can then also make use of the Mach-number independence of several aspects of inviscid flow, for instance, C_p, C_L $C_{D,wave}$, shockwave shapes, and Mach-wave patterns, such that only limited calculations have to be performed to obtain a complete picture. It is noted that even though the windward side of the re-entry vehicle has a dominating effect on the aerodynamic loads, the leeward side cannot be ignored, as it still has a noticeable contribution. Therefore, windward and leeward calculation methods will be discussed in the next section.

3.4.4 Local Inclination Methods

3.4.4.1 Introduction

During the conceptual design phase of a re-entry vehicle, the vehicle shape is one of the variables taken into account. Changing the shape will, of course, have an effect on the aerodynamic characteristics, which in turn affects the thermo-mechanical flight loads for a given trajectory, or change the trajectory when the (maximum) loads are taken as path constraints. In case the shape variation is part of an optimisation process, such as described by Dirkx and Mooij (2017), calculating the aerodynamic properties by solving the Navier-Stokes equations is not an option. Due to a continuous variation of the shape and the many evaluations involved,[21] fast (engineering) methods are preferred. Once the design has converged, validation runs with more complex CFD codes can be used.

The main goal of the engineering codes is to provide the aerodynamic coefficients, whereas the main objective of Navier-Stokes solvers would be detailed flow and/or aerodynamic heating analysis. So-called local inclination methods for aerodynamic analysis have proven sufficiently accurate for use at conceptual design stages, especially when low-speed and high-speed calculation models can be combined. An example of such software is the Aerodynamic Preliminary Analysis System (APAS), already mentioned in Sect. 3.2.

Cruz and Wilhite (1989) compared the calculated high-velocity results with several other inviscid and viscous techniques, as well as Shuttle-derived wind-tunnel data, and found that APAS agrees well with the other calculations and data. Pamadi and Brauckmann (1999) used APAS for engineering-level analysis and to fill the gaps in the wind-tunnel test data that were measured for the X-34 reusable launch vehicle. The realised aerodynamic database covered a Mach range up to $M = 8$, for

[21] Such a continuous variation of the shapes works best if the vehicle surface can be parametrised, either by using simple geometric shapes (cylinders, cones, etc.) or by defining cross-sectional areas with Hermite polynomials. A very efficient approach to optimise capsule and winged-vehicle shapes has been implemented by Dirkx and Mooij (2017), and served as the basis for this section.

Table 3.2 Qualitative assessment of the ability of APAS to predict longitudinal and lateral stability and control derivatives categorised by speed range, adapted from Maughmer et al. (1993)

M_∞	C_L	C_{L_α}	$C_{L_{\delta_e}}$	C_D	$C_{D_{\delta_e}}$	C_m	C_{m_α}	$C_{m_{\delta_e}}$	$C_{S_{\delta_a}}$	$C_{n_{\delta_a}}$	$C_{l_{\delta_a}}$	$C_{S_{\delta_r}}$	$C_{n_{\delta_r}}$	$C_{l_{\delta_r}}$
Subsonic $M_\infty < 0.8$	○	○	○	◑	◑	◑	◑	○	○[c]	○[c]	○	◑	◑[b]	◑
Transonic $0.8 < M_\infty < 1.2$	○	○	○	◑	◑	◑	◑	◑	○	○	○	◑	◑	◑
Supersonic $1.2 < M_\infty < 4.5$	○	○	○	○	○	◑	◑	◑	○	●	◑	○	○	●
Low hyperrsonic[a] $4.5 < M_\infty < 8.0$	○	○	○[c]	○	○	◑[b]	◑[b,c]	○[b,c]	◑[b]	○[b]	○[b]	○[c]	○[c]	○[c]
High hypersonic[a] $8.0 < M_\infty < 20.0$	○	○	○[c]	◑	○	◑	◑	○	○	○	○[c]	○	○	○
	○	Trends predicted well and magnitudes are within 40% over a significant portion of the operating range												
	◑	Results should be regarded cautiously: either magnitudes near zero angle of attack or angle-of-attack trends												
	●	are predicted reasonably well Not useful												

[a] Lateral/directional results are generally better without shielding
[b] Some problems in predicting results for X-15
[c] Some problems in predicting results for Space Shuttle Orbiter

angles of attack, sideslip, and control-surface deflections that were anticipated in the complete flight envelope.

Maughmer et al. (1993) did a qualitative assessment of the ability of APAS to predict longitudinal and lateral stability and control derivatives for, amongst others, the X-15 and the Space Shuttle, see Table 3.2. They compared the predicted control forces and moments with previously published wind-tunnel and flight-test data, and found all results for the longitudinal stability and control derivatives were acceptable for the conceptual design stage. This also holds for most lateral/directional control derivatives, except for the yaw-moment derivative to aileron deflection and the roll-moment to rudder defection at supersonic Mach numbers.

In summary, local inclination methods provide a powerful tool to characterise (most) aerodynamic properties, which is why we discuss a few of them here. Different methods exist for compression and expansion surfaces, and even per category there are variations, depending on, for instance, the curvature of the surface. In Table 3.3 the most frequently used methods have been listed, as well as when to use which, see also the papers by Maughmer et al. (1993), Moore and Williams (1989), and Gomez (1990). For an automated design process, selecting the right method per vehicle part should preferably be done in an automated manner, as has been developed by Dirkx and Mooij (2011).

Table 3.3 Selection of applicable methods per vehicle part type (Dirkx and Mooij 2011)

	Low hypersonic compression	High hypersonic compression
Blunt	Modified newtonian	Modified newtonian
Low inclination "round"	Tangent cone	Modified newtonian
Low inclination "flat"	Tangent wedge	Modified newtonian
	Low hypersonic expansion	High hypersonic expansion
Blunt	ACM empirical	High mach base pressure
Low inclination "round"	Prandtl-meyer expansion	Prandtl-meyer expansion
Low inclination "flat"	Prandtl-meyer expansion	Prandtl-meyer expansion

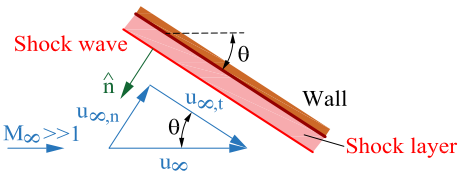


Fig. 3.52 Velocity components in relation to vehicle wall, relevant for Newtonian-flow method

3.4.4.2 Methods

Roughly speaking, a vehicle immersed in a flow can be divided into a windward and a leeward side. Windward sides will be facing the flow and thus be in compression, whereas leeward sides are either shielded from the flow due to a particular vehicle orientation (also known as *shadowed* or *in shadow*), or experience expansion. Engineering methods to calculate the pressure distribution are, therefore, also divided into compression and expansion methods. In the following, we will examine a few of the most commonly used methods, following the applicability of the methods given in Table 3.3. Although they are based on fundamental principles discussed earlier in this chapter, the methods themselves will not be derived here. The interested reader is referred to, for instance, Anderson (2019), who gives an excellent introduction. More details on their formulation and implementation can be found in the texts by, e.g., Gentry et al. (1973), Bonner et al. (1981), and Cremers (1996).

As the name already suggests, local-inclination methods only depend on the pressure perpendicular to the surface, as friction due to the tangential component is not considered. Therefore, only the angle by which the surface is (locally) inclined with respect to the free-stream flow is needed for the calculation. Given an outward surface normal $\hat{\mathbf{n}}$, the geometry mesh is related to the pressure coefficient by (Fig. 3.52):

$$\sin\theta = -\frac{\mathbf{u}_\infty \cdot \hat{\mathbf{n}}}{u_\infty} \tag{3.208}$$

The simplest local inclination method for the estimation of the pressure distribution on a hypersonic body—which, after integration, provides forces and moments for a given location of the centre of mass—is the Newtonian method. This method

assumes that after hitting a surface, the flow velocity loses its component normal to the surface while retaining all of its tangential motion. It was Newton who first described the effects of bodies in a fluid flow. Although it does not work well for low velocities, at hypersonic speeds it gives a reasonable estimation for the pressure force on a body, at least for relatively high inclinations.[22]

The assumptions of Newtonian flow, i.e., M going to infinity and $\gamma = 1$, lead to the following relation for the local pressure coefficient at a surface i:

$$C_{p,i} = 2 \sin^2 \theta_i \tag{3.209}$$

where θ_i is the local inclination of surface i with respect to the flow. As a consequence, the (maximum) Newtonian pressure coefficient at the stagnation point ($\theta_i = 90°$) equals 2.

Formulated as it is, the Newtonian method does not account for the pressure loss over the (thin) shockwave close to the surface. A modified version that does take this into account improves the accuracy of the prediction. To include the loss of total pressure, one uses the Rankine-Hughoniot relations, Eqs. (3.115)–(3.117), to determine the stagnation pressure coefficient behind a normal shock wave and consider an isentropic deceleration, Eq. (3.25). With the more accurate stagnation conditions, the stagnation pressure coefficient is found to be:

$$C_{p,s} = \frac{2}{\gamma M_\infty^2} \left[\left(\frac{(\gamma+1)^2 M_\infty^2}{4\gamma M_\infty^2 - 2(\gamma-1)} \right)^{\frac{\gamma}{\gamma-1}} \left(\frac{1-\gamma+2\gamma M_\infty^2}{\gamma+1} \right) - 1 \right] \tag{3.210}$$

The modified Newtonian method is then formulated as:

$$C_{p,i} = C_{p,s} \sin^2 \theta_i \tag{3.211}$$

which will yield results for the pressure coefficient slightly below the Newtonian values. For $\gamma = 1$ the limit case ($M_\infty \to \infty$ and $\theta_i = 90°$) would converge to $C_{p,i} = 2$; with $\gamma = 1.4$, on the other hand, the limit value is 1.84. Note that only the stagnation streamline passes perpendicularly through the bow shock (assuming an axisymmetric vehicle at zero angle of attack). Any other downstream point on the vehicle surface is impacted by a streamline that passes through a shock with $\beta \neq 90°$. As the loss in total pressure is less over such a shock wave compared to over a normal shock in front of the stagnation point, $C_{p,s}$ will be too low, when used in the modified Newtonian method. If one would know the geometry of the shock, a correction could be applied

[22] The reader should be aware that the theory behind this method is, in fact, wrong; particles do not lose all their normal momentum and γ will never go to 1. Fortunately, the physics neglected by this method happens to cancel out, which explains that the results are relatively accurate. When Prof. A. Busemann (1901–1986) tried to fit blunt, curved bodies, he took the curvature of the streamlines into account. Correcting the method for the presence of centrifugal forces, its results just got worse, even though it was physically more sound. Simply, because there was no longer a cancellation of effects!

and it would theoretically increase the fidelity of the results. However, this is usually not done, because the largely increased complexity of the method negates the benefits of having a fast and simple engineering method.

The Newtonian method works best when the inclination angle is large, but decreases in accuracy for small θ. For smaller inclination angles, a different method, such as the tangent-cone or tangent-wedge method, can be used. Instead of defining the pressure distribution for the vehicle, it will be modelled as one on an *equivalent* cone or wedge. For these simplified cases—and for a non-reacting, inviscid ideal gas—the (oblique) shock structure and pressure distribution can be calculated with the relations presented in Sect. 3.3.3.

Example 3.1 (*Newtonian Flow—Flat Plate*) The assumptions for Newtonian flow are that the mean free path of air particles is much larger than the dimensions of the spacecraft, such that there is no appreciable interaction of air molecules. Then, the interaction between the vehicle and the atmosphere can be modelled as independent perfectly elastic collisions, a situation that has been depicted in Fig. 3.53a. Here, a flat plate is exposed to an incoming particle with velocity $\hat{\mathbf{u}}_\infty$. It bounces at the plate under an angle α, and reflects on an outgoing path, with the same relative angle, α, with respect to the plate.

To determine the total effect of all incoming particles over a given span of time when the inclined flat plate is moving with velocity $\hat{\mathbf{u}}_\infty$, Fig. 3.53b shows a control volume. The sum of particles in this control volume is represented by the density, ρ. The mass flux is equal to the product of density, swept area, and velocity, so:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \rho S \sin\alpha u_\infty \tag{3.212}$$

where, for the sake of this example, a one-dimensional situation is considered.

Returning to the bouncing particle of Fig. 3.53a, it is obvious that its momentum perpendicular to the wall is reversed at impact. This so-called "bounce" momentum is transferred to the plate; one should realise that the momentum parallel to the plate is unchanged. From this momentum transfer a force originates (Newton's Second Law), which is easily derived. Figure 3.53c shows this situation; note that the flat plate can, of course, also be seen as the surface of a vehicle flying through the atmosphere.

The resulting force is equal to

$$F = \frac{\mathrm{d}m}{\mathrm{d}t}\Delta V \tag{3.213}$$

with ΔV the total change between incoming and outgoing velocity. Since the collisions are perfectly elastic, it holds that $\Delta V = 2u_\infty \sin\alpha$ (the velocity speed is simply inverted). Thus,

$$F = \frac{\mathrm{d}m}{\mathrm{d}t}\rho u_\infty S \sin\alpha(2u_\infty \sin\alpha) = 2\rho u_\infty^2 S \sin^2\alpha \tag{3.214}$$

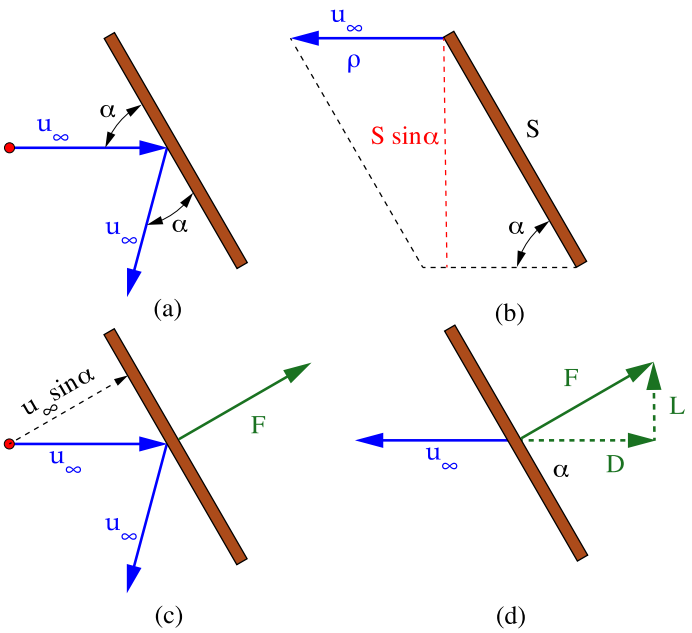


Fig. 3.53 Newtonian flow for a flat-plate application

The last step is to decompose F into a lift and drag force, as if the flat plate is a vehicle flying in the atmosphere, see Fig. 3.53d. This figures shows

$$D = F \sin\alpha \quad \text{and} \quad L = F \cos\alpha \tag{3.215}$$

so after substituting Eq. (3.214) this results in

$$D = 2\rho u_\infty^2 S \sin^3\alpha \quad \text{and} \quad L = 2\rho u_\infty^2 S \sin^2\alpha \cos\alpha \tag{3.216}$$

and for the drag and lift coefficients:

$$C_D = \frac{D}{\frac{1}{2}\rho u_\infty^2 S} = 4\sin^3\alpha \quad \text{and} \quad C_L = \frac{L}{\frac{1}{2}\rho u_\infty^2 S} = 4\sin^2\alpha\cos\alpha \tag{3.217}$$

Finally, the lift-to-drag ratio is

$$\frac{L}{D} = \frac{\cos\alpha}{\sin\alpha} \tag{3.218}$$

By inspecting the expressions it may be clear that C_D, C_L, and $\frac{L}{D}$ are only a function of the angle of attack, and independent from Mach number. In Fig. 3.54, Eqs. (3.217) and

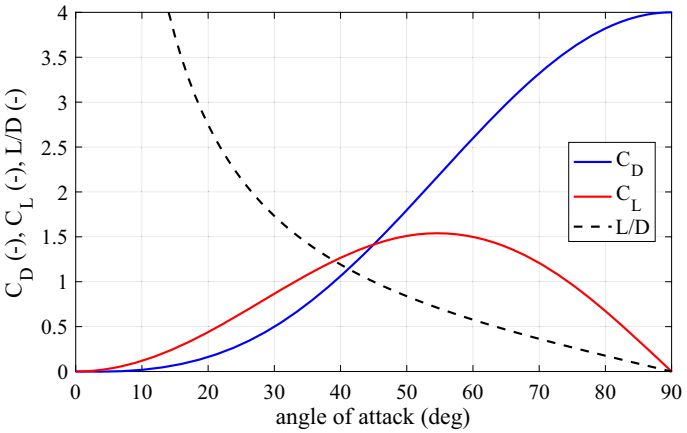


Fig. 3.54 Newtonian-flow coefficients as a function of angle of attack for a flat plate

(3.218) have been plotted, as a function of angle of attack. At zero angle of attack, both drag and lift are zero (no viscous force!), so L/D is undefined.

■

Example 3.2 (*Modified Newtonian Flow—Re-entry Capsule*) The drag coefficient of an arbitrary vehicle can be computed using the following equation, if viscous effects are ignored:

$$C_D = -\frac{1}{S_{ref}} \iint_S C_p \hat{\mathbf{u}}_\infty \cdot \hat{\mathbf{n}} \, \mathrm{d}S \tag{3.219}$$

where S is the wetted surface area, S_{ref} is the aerodynamic reference area, C_p is the pressure coefficient, $\hat{\mathbf{u}}_\infty$ the unit vector in the direction of the free-stream velocity, and $\hat{\mathbf{n}}$ the local outward-surface unit vector on the vehicle surface. In Fig. 3.55, the flow lay out of an Apollo-type re-entry capsule is shown.

Using modified Newtonian theory, the local pressure coefficient at any point on the surface of the vehicle can be computed as follows:

$$C_p = \begin{cases} C_{p,s} \sin^2 \alpha^* & \text{if } \gamma \leq \frac{\pi}{2} \\ C_p = 0 & \text{if } \alpha^* > \frac{\pi}{2} \end{cases} \tag{3.220}$$

where $C_{p,s}$ is the stagnation pressure coefficient (which can be treated treat as a free parameter in this example). The angle α^* is the angle between the local surface normal $\hat{\mathbf{n}}$ and $-\mathbf{u}_\infty$.

In this example, we will look for an expression to compute the drag coefficient on the entry capsule at zero angle of attack, which means that $\hat{\mathbf{u}}_\infty = -\hat{\mathbf{i}}$, with $\hat{\mathbf{i}}$ being the unit vector in $x-$direction.

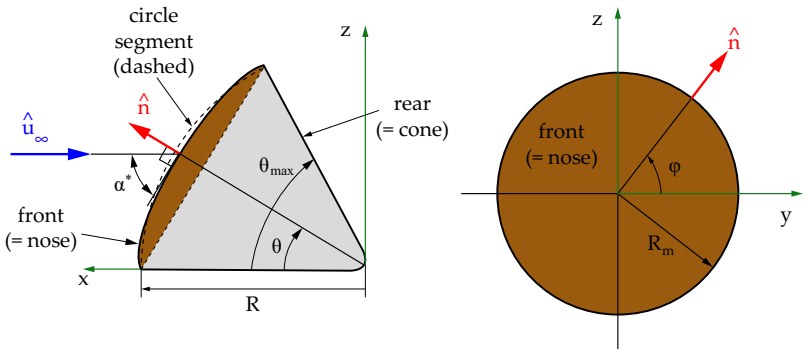


Fig. 3.55 Flow layout for an Apollo-type entry capsule: side view (left) and front view (right). The two angles θ and ϕ define the location of a point on the capsule; $\hat{\mathbf{n}}$ is perpendicular to the three-dimensional surface

For the part of the capsule facing the flow, the shape is best parameterised in spherical coordinates, i.e., R, θ, ϕ, with θ the angle from the $x-$axis and ϕ the circumferential angle. In this particular configuration, the cross-sectional area is assumed to be a circle segment (rather than the flattened configuration of the Apollo heat shield). The capsule shape is rotationally symmetric about the x-axis, with the aerodynamic reference area defined as:

$$S_{ref} = \pi R_m^2 \approx \pi R^2 \tag{3.221}$$

The first step in the derivation is to express the inner product $\hat{\mathbf{u}}_\infty \cdot \hat{\mathbf{n}}$ in terms of spherical coordinates.

$$\hat{\mathbf{u}}_\infty \cdot \hat{\mathbf{n}} = -\hat{\mathbf{i}}_\infty \cdot \hat{\mathbf{n}} = -\cos\theta \tag{3.222}$$

since $\hat{\mathbf{i}}_\infty \cdot \hat{\mathbf{n}}$ is the projection of $\hat{\mathbf{n}}$ on the x-axis.

The second step is to express the local pressure coefficient C_p on the capsule surface, from Eq. (3.220), in terms of the spherical coordinates. Looking at the figure, we may find that $\alpha^* + \theta = \frac{pi}{2}$, so $\alpha^* = \frac{pi}{2} - \theta$. This means that $\sin\alpha^* = \sin(\frac{pi}{2} - \theta) = \cos\theta$. In combination with Eq. (3.220):

$$C_p = C_{p,s} \sin^2\left(\frac{\pi}{2} - \alpha^*\right) = C_{p,s} \cos^2\theta \tag{3.223}$$

In the third step, we will set up the surface integral for C_D, thereby making use of Eqs. (3.222) and (3.223). To solve is the following:

$$C_D = -\frac{1}{S_{ref}} \left(\iint_{front} C_p \hat{\mathbf{u}}_\infty \cdot \hat{\mathbf{n}} \mathrm{d}S + \iint_{back} C_p \hat{\mathbf{u}}_\infty \cdot \hat{\mathbf{n}} \mathrm{d}S \right) \tag{3.224}$$

where, in spherical coordinates, the surface differential of the *front* (windward side) of the capsule is given by:

$$\mathrm{d}S = R^2 \sin\theta \mathrm{d}\theta \mathrm{d}\phi \tag{3.225}$$

Substituting the appropriate equations, realising that at the back $\alpha^* > \frac{\pi}{2}$, so $C_p = 0$:

$$C_D = -\frac{1}{S_{ref}} \left(-\iint\limits_{front} R^2 C_{p,s} \cos^3\theta \sin\theta \mathrm{d}\theta \mathrm{d}\phi + \iint\limits_{back} 0 \hat{\mathbf{u}}_\infty \cdot \hat{\mathbf{n}} \mathrm{d}S \right) \tag{3.226}$$

or

$$C_D = -\frac{1}{S_{ref}} \left(-\iint\limits_{front} R^2 C_{p,s} \cos^3\theta \sin\theta \mathrm{d}\theta \mathrm{d}\phi \right) \tag{3.227}$$

The fourth and final step is to solve the above integral. Taking out the constant R and $C_{p,s}$, the equation to solve is:

$$C_D = \frac{R^2 C_{p,s}}{S_{ref}} \int_0^{\theta_{\max}} \int_0^{2\pi} \cos^3\theta \sin\theta \mathrm{d}\phi \mathrm{d}\theta \tag{3.228}$$

Solving the integral for ϕ yields

$$C_D = \frac{2\pi R^2 C_{p,s}}{S_{ref}} \int_0^{\theta_{\max}} \cos^3\theta \sin\theta \mathrm{d}\theta \tag{3.229}$$

The solution to the integral follows from one of (many) standard integrals, here written as an indefinite integral:

$$\int \sin ax \cos^n ax \mathrm{d}x = -\frac{1}{a(n+1)} \cos^{n+1} ax + C \tag{3.230}$$

with $a = 1$ and $n = 3$ for the case at hand. Thus, $\int \sin x \cos^3 x \mathrm{d}x = -\frac{1}{4} \cos^4 x + C$.

Using this result, the solution to Eq. (3.229) is then

$$C_D = \frac{2\pi R^2 C_{p,s}}{S_{ref}} \left[-\frac{1}{4} \cos^4\theta \right]_0^{\theta_{\max}} \tag{3.231}$$

Substituting the boundaries, as well as Eq. (3.221) for S_{ref}, results in

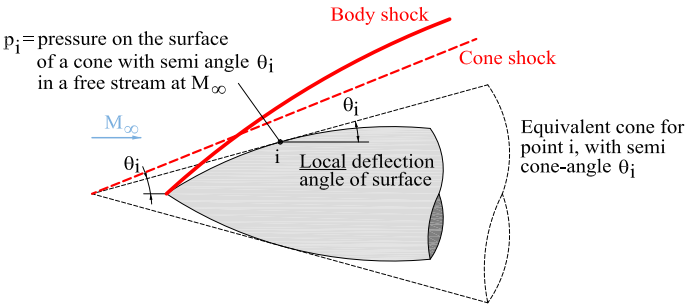


Fig. 3.56 Schematic representation of the tangent-cone method

$$C_D = \frac{C_{p,s}}{2}\left(1 - \cos^4 \theta_{\max}\right) \tag{3.232}$$

In the above answer, $C_{p,s}$ and $\theta_{\max}$ have been kept as free parameters. It is noted that this solution only holds for zero angle of attack. For other angles, the solution process is more complicated, albeit with the same steps.

■

The tangent-cone method is shown schematically in Fig. 3.56, where the geometry is locally approximated by an "equivalent cone" with the same semi-cone angle, θ_i, as the local surface deflection angle. This deflection angle cannot be too large, as the shock waves are assumed to be attached (or in practice, a very small stand-off distance). This means that the shock angle of the body should be larger than the one of the cone (or wedge), i.e., $\beta_{body} > \beta_{c/w,i}$. If this condition holds (and commonly true for slender bodies), the local pressure coefficient is given by

$$C_{p,i_c} = \frac{4}{\gamma+1}\frac{1}{M_\infty^2}(K_{c,i}^2 - 1) + \frac{2}{M_\infty^2}(K_{\beta,i} - K_{c,i})^2 \frac{\gamma+1}{\gamma - 1 + \frac{2}{K_{c,i}^2}} \tag{3.233}$$

with

$$K_{c,i} = M_\infty \theta_i \tag{3.234}$$

$$K_{\beta,i} = M_\infty \beta_{body} = \frac{\gamma+1}{\gamma+3}K_{c,i} + \sqrt{\left(\frac{\gamma+1}{\gamma+3}\right)^2 K_{c,i}^2 + \frac{2}{\gamma+3}} \tag{3.235}$$

For the tangent-wedge method, a similar approach is taken, albeit with an equivalent shape of a wedge, instead of a cone. This is more applicable for "flat" geometries. The expression for C_p is

$$C_{p,i_w} = \theta_i^2 \left[\frac{\gamma+1}{2} + \sqrt{\left(\frac{\gamma+1}{2}\right)^2 + \frac{4}{K_{w,i}^2}}\right] \tag{3.236}$$

with $K_{w,i} = M_\infty \theta_i$, and under the assumption of small wedge angles.

Recently, real-gas effects were included as an adjustment to the tangent-wedge method by Whitmore (2007), making the technique more suitable for calculations up to Mach 20. Note: the limitation of attached shocks does not seem to make the tangent-cone/wedge method a logical choice for re-entry vehicles. However, it has been used with quite good results for entry vehicles, see, amongst others, the papers by Cruz and Wilhite (1989), and Cruz and Ware (1992).

The tangent-wedge, tangent-cone, and Newtonian methods could also be classified in terms what is assumed to cause the flow deflection. In the tangent wedge method, the only deflection is assumed to be caused by the shock wave. For the pure Newtonian methods, the body performs all the flow deflection. For the tangent-cone method, the flow is first deflected by the shock wave and subsequently deflected by the presence of the body.

As mentioned before, impact methods as the ones discussed above can only be used for surfaces directly exposed to the external flow. To determine the pressure coefficient at the leeward side of the vehicle, an alternative method must be used. The simplest method is by simply assuming that the pressure coefficient is zero, which follows directly from the Newtonian-flow method:

$$C_{p,b} = 0 \tag{3.237}$$

at any leeward surface or shielded panel. Inspecting Eq. (3.55) shows that this follows from $p \to p_\infty$, an often used hypersonic approximation. An alternative expression may be obtained by only assuming that $p/p_\infty \ll 1$, so that the leeward pressure coefficient becomes:

$$C_{p,b} = -\frac{2}{\gamma M_\infty^2} \tag{3.238}$$

The physical situation that typically occurs on a hypersonic after-body is influenced by viscous effects, resulting in flow circulation (the so-called *wake*). This can have a significant impact on the pressure characteristics there, the so-called *base pressure*. A rough modification of the above relation can be used to account for such effects, and is given by:

$$C_{p,b} = -\frac{1}{M_\infty^2} \tag{3.239}$$

A further modification of this equation is the so-called ACM empirical method,[23] from Krieger (1981). This method takes into account the fact that the flow expands along the low absolute inclination-angle expansion region to a more or less constant low pressure coefficient for sufficiently low inclination angle. The pressure coefficient is determined as follows using this method:

$$C_{p,b} = \max\left(\frac{\theta}{16M_\infty^2}, \frac{-1}{M_\infty^2}\right) \tag{3.240}$$

[23] ACM = Aerodynamic Configured Missile Development Program.

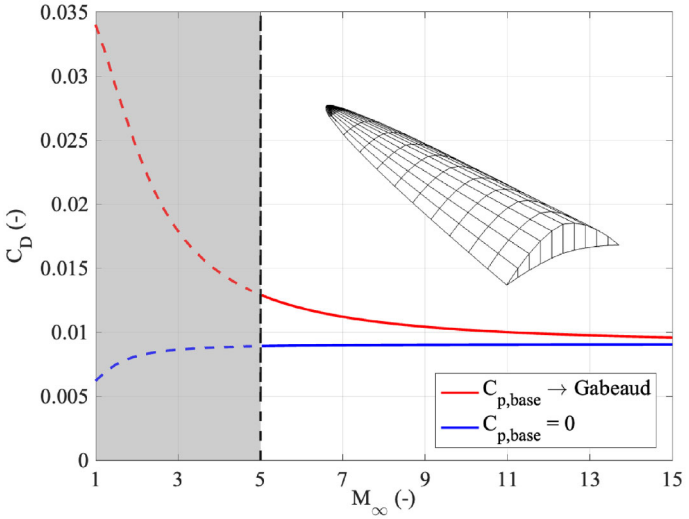


Fig. 3.57 The influence of the Gabeaud base pressure on the modified Newtonian drag coefficient for a waverider configuration

with θ in degrees.

For high Mach numbers it is advised to use an alternate expression for the base pressure (Gabeaud 1950), which may replace any of the other high-Mach-number base-pressure correlations:

$$C_{p,b} = \frac{2}{\gamma M_\infty^2}\left\{\left(\frac{2}{\gamma+1}\right)^{1.4}\left(\frac{1}{M_\infty}\right)^{2.8}\left[\frac{2\gamma M_\infty^2-(\gamma-1)}{\gamma+1}\right]-1\right\} \tag{3.241}$$

Whichever base-pressure method is chosen, it is clear that in the limit $M_\infty \to \infty$ for all base-pressure expressions (3.238)–(3.241), $C_{p,b}$ becomes zero. This means that for large Mach numbers, the contribution of the correction term becomes negligible and the local-inclination methods become more accurate. The result of adding the Gabeaud base pressure to the pressure distribution obtained with the modified Newtonian method is shown in Fig. 3.57. In this figure, the drag coefficient for a waverider configuration is shown, as a function of free-stream Mach number. Without base-pressure correction, the drag coefficient only increases as the Mach number increases from $M_\infty = 1$ onward, but with the Gabeaud base pressure at lower Mach numbers the drag increases substantially, as it should. It is noted that the modified Newtonian method should in principle not be applied at Mach numbers lower than $M_\infty = 5$, hence the grey area in the figure.

The last method that we will discuss is the shock-expansion method, which, according to Anderson (2019), has more theoretical justification than the tangent-cone and tangent-wedge methods, but at the cost of higher complexity and larger

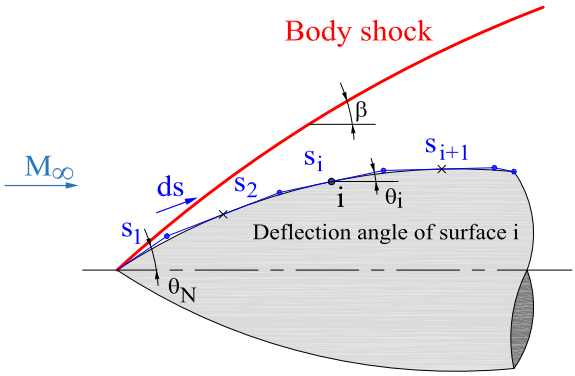


Fig. 3.58 Shock-expansion method along the surface contour of a slender body

computational load. Applicable to slender bodies, i.e., low "round" and "flat" inclined surfaces, the shock-expansion method can be used on the windward side of the configuration until the body slope relative to the free stream becomes the maximum wedge angle for an attached shock wave, as well as on the vehicle's leeward side.

Consider the two-dimensional hypersonic flow around a slender body with a sharp nose and an attached shock (Fig. 3.58). The semi nose angle is indicated by θ_N. We consider the flow to be a flow over a wedge, with deflection angle θ_N. The pressure, p_N and Mach number, M_N, immediately behind the shock at the nose, as well as the shock angle, β_N, can be determined using the oblique-shock theory from Sect. 3.3.3, more specifically Eqs. (3.115), (3.119), and (3.123). Using this as a starting point, we assume a local Prandtl-Meyer flow along the wall by taking small steps, ds, along the contour. For an arbitrary point i at the contour, the Mach number, M_i can be calculated with Eqs. (3.134)–(3.135):

$$\Delta\theta^* = \theta_N - \theta_i = \sqrt{\frac{\gamma+1}{\gamma-1}}\left\{\arctan\left[\sqrt{\frac{\gamma-1}{\gamma+1}(M_N^2-1)}\right] - \arctan\left[\sqrt{\frac{\gamma-1}{\gamma+1}(M_i^2-1)}\right]\right\} + \\ - \left\{\arctan\left[\sqrt{M_N^2-1}\right] - \arctan\left[\sqrt{M_i^2-1}\right]\right\} \tag{3.242}$$

The above equation has to be solved iteratively, or, equivalently, M_i could be obtained from the Prandtl-Meyer function table or figure ($M - \nu$), if available.

Because the Prandtl-Meyer expansion is isentropic, the pressure follows from Eq. (3.136):

$$\frac{p_i}{p_N} = \left(\frac{1+\frac{\gamma-1}{2}M_N^2}{1+\frac{\gamma-1}{2}M_i^2}\right)^{\frac{\gamma}{\gamma-1}} \tag{3.243}$$

and thus, with Eq. (3.55), the pressure coefficient for the panel with surface area s_i reads:

$$C_{p,i} = \frac{2}{\gamma M_\infty^2}\left(\frac{p_i}{p_\infty} - 1\right) \tag{3.244}$$

The approach for conic-shaped bodies is in principle the same, if the initial conditions right behind the shock, i.e., M_N and p_N, are determined for a conic flow with semi-cone angle θ_N. This can be done with the exact Taylor-Maccoll theory (Anderson 2019). Also now it is assumed that the local flow is two-dimensional, which implies that divergence of the streamlines in a plane perpendicular to the surface is more important than the divergence in a plane tangential to the surface. This is commonly true for bodies of revolution at zero angle of attack.

It is clear from Eqs. (3.243)–(3.244) that for high hypersonic velocities the contribution from the leeward pressure distribution to the aerodynamic coefficients is quite small, as the pressure coefficient approaches zero for infinite Mach number. This is also the case for the Prandtl-Meyer expansion method, where the minimum pressure coefficient that can be attained is the vacuum pressure. For lower hypersonic/supersonic velocities, however, its influence becomes more noticeable, as we have already seen in Fig. 3.57.

Although the above discussed local-inclination methods are obviously highly simplified, acceptable results for conceptual studies can often be obtained from them (as will be shown in Examples 3.3 and 3.4). Due to their combination of simplicity and acceptable fidelity, this class of methods finds wide use in preliminary design and analysis.

An example of a pressure distribution around the Space Shuttle is shown in Fig. 3.59 for a flight condition at $M = 20$, at three angles of attack, i.e., $\alpha = 40°$, 30°, and 20°. For gliding entry, as we will discuss at length in Chap. 5, Sect. 5.4, the vehicle will enter the atmosphere at large angle of attack, around $\alpha = 40°$. It will keep α large to keep the thermal loads within limits, and only later during the gliding descent it will lower α to maximise its downrange and crossrange capability. The figure clearly shows that the expansion surfaces add little to the total aerodynamic coefficients, as the pressure coefficient becomes ≈ 0.

Each of the above methods is quite easy to use, and runs very fast on a personal computer; the user has to be aware of their limitations, though. A more accurate, yet still approximate, method to calculate three-dimensional inviscid hypersonic flow fields was developed by Riley and DeJarnette (1990), to ascertain the flow characteristics over blunt-nose bodies during hypersonic speeds. This technique leverages an explicit expression proposed by Maslen[24] to describe pressure across the shock layer, combined with the assumption of approximately streamlined surfaces within

[24] Maslen (1964) presents a simple inverse method to compute the entire viscous flow field in the shock layer encapsulating a sufficiently smooth axisymmetric body, either blunt or slender, moving at hypersonic speed. Key assumption is that the shock layer is thin, such that the streamlines are essentially parallel to the shock wave. Furthermore, the shape of the shock wave is assumed and the body that supports this shock is calculated, hence the classification of an *inverse* method. Because of this assumption, all quantities just behind the shock follow from oblique-shock theory. With known and assumed stream-line functions just behind the shock wave and inside the flow field, entropy is known, and enthalpy and density follow, as well as the velocity. Once this procedure is followed for multiple steps, the complete flow field can be generated and the body shape that supports this.

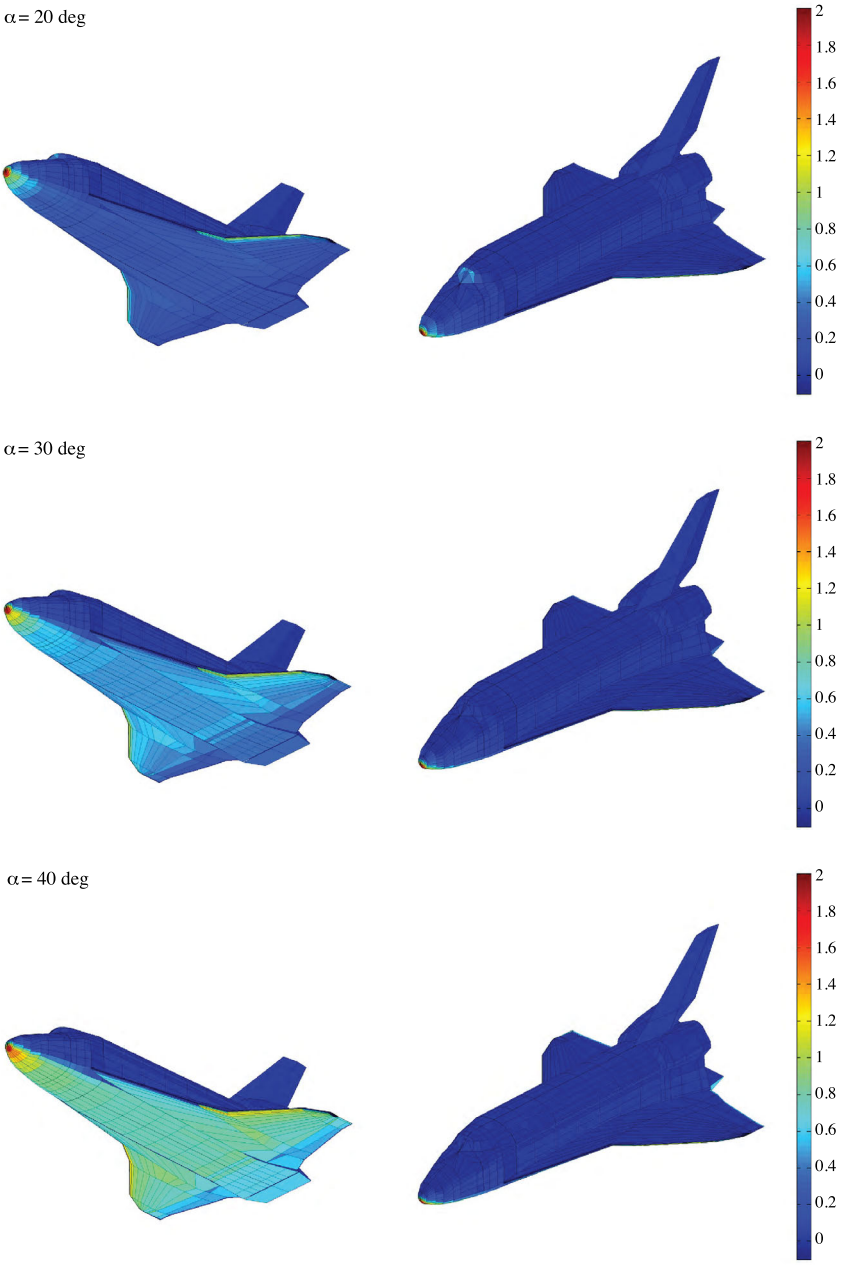


Fig. 3.59 Space-Shuttle pressure distribution for $M = 20$ at $\alpha = 20°$ (top), $\alpha = 30°$ (middle), $\alpha = 40°$ (bottom), with zero sideslip angle (Dirkx and Mooij 2017). Analysis-method selection was done according to Table 3.3

the shock layer. By doing so, the complexity of the three-dimensional flow is simplified. The application of this method encompasses the analysis of spherically blunted cones, paraboloids, and elliptical cones in various angles of attack, assuming a perfect gas behaviour.

The results of this technique showed a good agreement with more precise numerical solutions and experimental findings, particularly demonstrating excellent agreement adjacent to the central windward symmetry plane. As the run time is still relatively low—the authors claim a factor five reduction compared to an Euler code—this method may also be a good candidate for use in the conceptual design phase. As the method is quite a bit more complex than the local inclination methods discussed above, the reader is referred to the cited report for more technical details.

3.4.4.3 Shadowing

For fully convex bodies such as spheres, sharp and blunted cones, cylinders and capsules, every panel that has a positive inclination with respect to the flow is directly hit by the incoming gas particles. Thus, there are no compressive surface elements, which cause a flow interference for other panels, and for each panel the coefficients can be determined irrespective of the others. For concave bodies, however, this is not the case, as some panels will be shielded from the incoming flow by others.

But shielding may also occur due to extruding parts of the vehicle. Taking as an example the Delta-K rocket body (for which the aerodynamics will be calculated in Example 3.5), regions where shadowing occurs include the nozzle's external and internal surface, the region behind the pressure spheres, and the spherical bottom of the propellant tank (see Fig. 3.60). While studying orbit decay of a satellite, shadowing could be caused by extruding elements, such as its solar panels and antennas.

In their fundamental implementation, local-inclination methods as discussed before, as well as the panelised free-molecular flow method to be discussed in Sect. 3.5, use meshes with panels that are independent from each other. This means that any shadowing effect should be checked separately. The current section will discuss a possible approach to do this.

During the generation of aerodynamic databases for panellised surface models, at each angle of attack and sideslip combination, an analysis is done to determine which panels are shadowed by others. A panel is considered shadowed if its centroid is located behind another panel with respect to the incoming flow. This assumption leads to a simplified analysis when compared to other algorithms in which the size of the overlapped region is also determined, see, for instance, the shadow algorithm described by Gentry et al. (1973). The accompanied penalty in accuracy can be minimised by a decrease in panel-size for regions where shadowing occurs. If a

Besides the original paper, Anderson (2019) provides a comprehensive discussion about this "thin shock-layer theory" as well.

panel is identified to be shadowed, the aerodynamic coefficients for that surface element are set to zero.

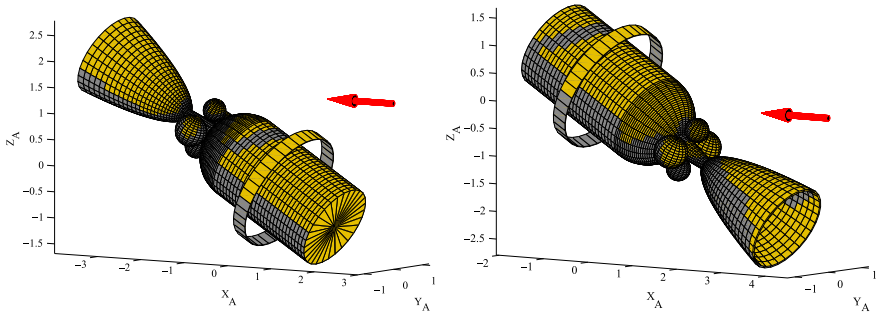
The following steps are taken during the analysis:

1. Remove any panels with negative inclination, because the effect of shadowing is only relevant for compression (positively inclined) panels.
2. All panel vertex and centroid coordinates are transformed to the aerodynamic reference frame and ordered according to their centroid x-coordinate, highest first.
3. Starting from the most upstream panel, it is checked whether the centroid of each panel is shadowed from the incoming flow by one of the previous panels. This is done by taking the yz-projection of the centroid and panel vertices, and applying a two-dimensional point-in-quadrilateral algorithm. If the centroid's projection is determined to lie within the quadrilateral described by the projection of the vertices of a more upstream panel, it is considered shadowed.

For the case of the modelled rocket body, a graphical representation of the result of the shadowing analysis is shown in Fig. 3.60. The orange areas are compression areas which are directly affected by the incoming flow. The panels, which are either shadowed or negatively inclined are indicated in grey. Especially note the shadow casted by the mini-skirt ring and spheres on the propellant tank and nozzle extension. Also notice the discrete way in which surface panels are either fully shadowed or not shadowed at all.

3.4.4.4 Examples

Two examples of aerodynamic-database calculations conclude this section, one for a representative entry capsule, i.e., the Apollo Command Module, and one for a winged



(a) $\alpha = 30°$ (b) $\alpha = 210°$

Fig. 3.60 Effect of shadowing on the Delta-K model. Orange panels are positively inclined and affected by the incoming flow. Grey regions are either negatively inclined or shadowed

re-entry vehicle, i.e., the Space Shuttle. The calculated data will be compared with wind-tunnel data. These examples are taken from Dirkx and Mooij (2011, 2017).

Example 3.3 (*Apollo*) The Apollo Command Module can be approximated by four matched analytical geometries (Fig. 3.61), a sphere segment, a torus segment, a conical frustum and a spherical segment. Although no unique set of parameters exists for defining this shape, the required number of parameters for defining it is five: (i) nose radius, R_N, (ii) side (or corner) radius, R_S, (iii) rear-cone half-angle, θ_c, (iv) mid radius R_m, and (v) rear conical-part length, L_c (alternatively the vehicle length could be defined, but this would complicate the constraint checking in the shape-generation algorithm (Dirkx and Mooij 2011).

Apollo has been defined with the following parameters: $R_N = 4.694$ m, $R_m = 1.956$ m, $R_S = 0.196$ m, $\theta_c = 33°$, $L_c = 2.662$ m, with derived parameters $S_{ref} = 39.441$ m^2, $c_{ref} = 3.9116$ m. The centre of mass location is taken as $\mathbf{r_{cm}} = (1.0367, 0, 0.1369)^{\mathrm{T}}$ m. The body frame has its origin in the apex, with the X-axis positive towards the heat shield, and the Z-axis positive pointing downwards.

Aerodynamic coefficients were then generated and compared to the data from wind-tunnel testing given by North American Aviation (1965). Figure 3.62a and b show the drag and lift coefficients of the Apollo capsule, which have been generated with the local-inclination methods compared to the wind-tunnel results. The results match the wind-tunnel data quite well. The most important difference between the two is the over-prediction at $M = 10$. At this Mach number the contribution of the expansion pressure coefficient is very small, as the vacuum pressure coefficient is very close to zero. As a result, the results are almost fully defined by the compression pressure distribution.

A comparison done by Dirkx and Mooij (2017) showed that of the methods used the Modified Newtonian method gives the lowest pressure coefficient at a given inclination angle. Since this method is used on all compression surfaces at $M = 10$ and still produces an axial force that is too high, this means that no there is no suitable local inclination method, which can produce sufficiently low results here. This is likely due to the fact that the streamlines pass through a shock wave at an angle just under 90°. As a result, the stagnation pressure coefficient will be lower and thus also reduce pressure coefficient according to Eq. (3.211).

The pitch moment coefficient has been plotted in Fig. 3.62c and shows a good agreement between the calculated data and the wind-tunnel measurements. For performance and stability analysis it is important that the trimmed α and corresponding L/D are well predicted. The presented results confirm that the value of α_{trim} is off by 0.1° for $M_\infty = 3$ and about 1° for $M_\infty = 10$. The resulting error in L/D_{trim} that combines errors in both force and moment coefficients, is $\sim 10\%$ in both cases. These discrepancies are well acceptable for a conceptual-design effort, but should of course be considered during a trajectory sensitivity-analysis.

■

Example 3.4 (*Space Shuttle*) In the second example, we evaluate the aerodynamic characteristics of the Space Shuttle. The geometry input file for the Shuttle is given in

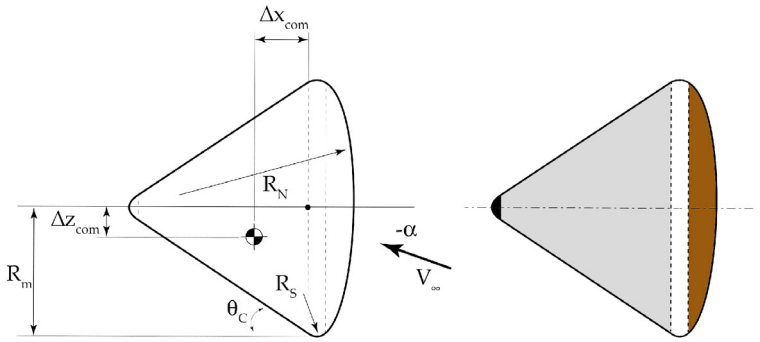


Fig. 3.61 Schematic representation of an entry capsule, with primary parameters and sub-division in fundamental shapes

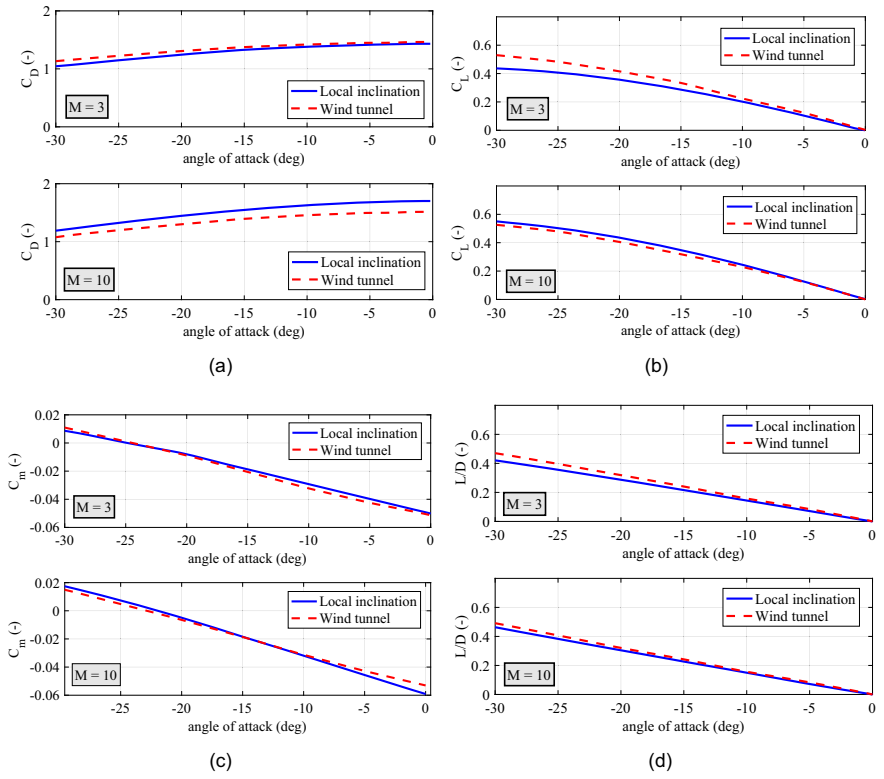


Fig. 3.62 Aerodynamic coefficients for the apollo command module

the Langley Wireframe Geometry (LaWGS) Format (Craidon 1985), see Fig. 3.63,

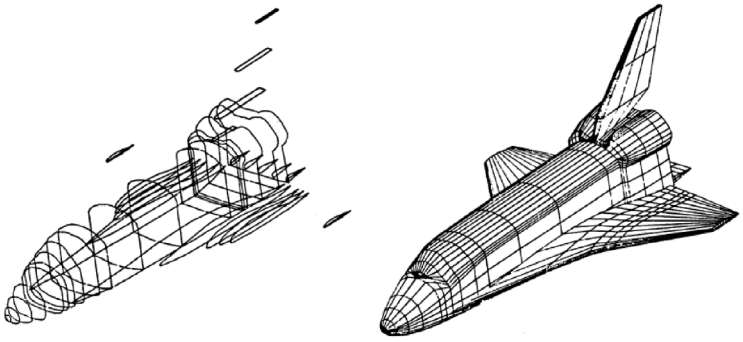


Fig. 3.63 Space Shuttle wireframe model from langley wireframe geometry standard, with left the cross-sectional areas and right the contour points connected by straight lines

and serves as the basis for creating the paneled surface mesh.[25] The aerodynamic coefficients have been generated for two Mach conditions ($M = 5$ and $M = 20$) with the methodology described earlier. The results are compared with coefficients obtained from wind-tunnel data, documented by Rockwell International (1980). Although these data are known to contain some discrepancies when compared to flight data, the data set is comprehensive and consistent. As such, it can be used to compare the calculated coefficients for other data points than those derived from flight data.

The analysis is performed using the default method selection criteria given in Table 3.3. When comparing the lift and drag coefficients from Fig. 3.64, it can be seen that this method predicts these coefficients quite well. In addition, for low Mach numbers where a non-Newtonian method is used as the compression method, the accuracy of the method selection yields results that are superior to a modified Newtonian approach. It is clear that shadowing has a marginal effect on the longitudinal force coefficients.

The moment coefficients differ strongly from the database values, though, as shown in Fig. 3.65b. However, there is a known discrepancy between the wind-tunnel data and flight data, consistently giving under-predicted values of the moment coefficients at high Mach numbers (Maus et al. 1983). The difference between the flight and wind-tunnel coefficients are shown in Fig. 3.65b. It is clear that there is a large offset between the pre-flight predictions in the Aerodynamic Design Data Book (ADDB) and the actual flight data. Since the generated data for this example show an over-prediction of the coefficients at $\alpha > 20°$, they likely approximate the actual coefficients more closely than can be seen from the figures here.

The effect of shadowing is greater for the moment coefficients than the longitudinal force coefficients. For high Mach numbers, it seems to cause the moment curve to better approximate the flight-measured curve. Shadowing has the most effect on

[25] The reference quantities of the coefficients are $S_{ref} = 249.9092$ m^2, $c_{ref} = 12.0579$ m, $\mathbf{r_{cm}} = (21.356, 0, 0.8224)$ m, with the Orbiter nose as origin, positive x-direction rearwards and positive z-direction upwards.

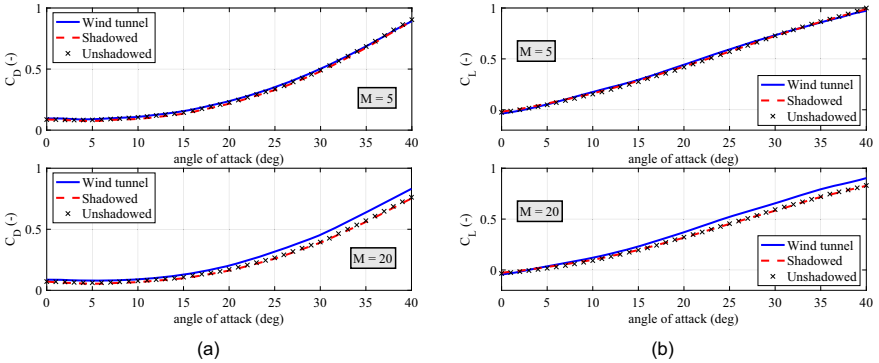


Fig. 3.64 Lift and drag coefficients produced for the Space Shuttle with and without shadowing compared to wind-tunnel data

the pitch moment at high angle of attack and low Mach numbers, which is adversely

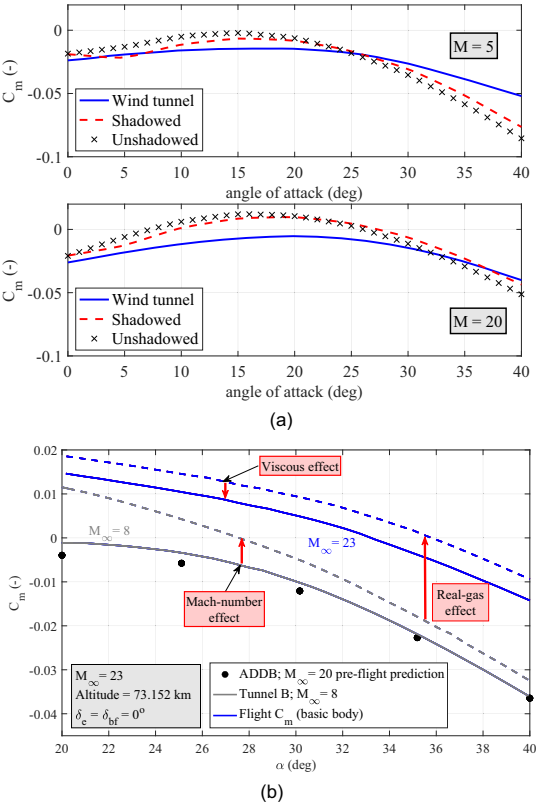


Fig. 3.65 Moment coefficients for the Space Shuttle **a** Local-inclination results with and without shadowing compared to wind-tunnel data **b** Wind-tunnel data compared to flight data (Data taken from Maus et al. (1983))

affected. However, gliding flight with winged vehicles will not fly at this combination of variables (see Sect. 5.4), so this error will have no effect on the trajectory results.

The stability derivatives of the lateral coefficients for small sideslip angles, which are important in the analysis of the vehicle stability during re-entry, can be seen in Fig. 3.66. In this case, shadowing has a more pronounced effect on the side force and yaw moment. In fact, due to the shadowing effect, the overall trend of the coefficients is followed quite well, contrary to the case without shadowing, where no such trend can be observed. The influence of shadowing is not quite as clearly positive for the roll-moment derivative (Fig. 3.66c), with as a result that the coefficients are a bit under-predicted rather than over-predicted in most cases. However, this will result in a more conservative estimate of the vehicle stability.

For the control increments, proper prediction of the pitch moment increments is quite often the most important, since these values determine the deflection of the control surfaces required to trim the vehicle. The resulting lift and drag increments can be $\sim$10% of the total lift and drag, and, therefore, should not be neglected during, for instance, trajectory optimisation. The control derivatives shown in Figs. 3.67, 3.68 and 3.69 indicate that the predictions for the elevon increments are reasonably good, with lift increments having the lowest accuracy, showing differences of up to about 0.01.

For positive deflection, the body-flap increments shown in Figs. 3.70 through 3.72 show larger differences. This could be caused by a number of factors. For low deflection angles, the body flap will be fully immersed in the fuselage boundary layer, reducing its effectiveness. For larger deflection angles, the compression shock that will occur on the ramp between the fuselage and body flap will cause an increase in the body-flap's surface pressure. This effect is not observed as strongly on the elevon, though, where a similar ramp effect is expected to occur. However, due to the shorter path over the wings, the boundary layer will be less pronounced on the elevons, reducing its effect on the control increments. Since a positive body-flap deflection is required for Space-Shuttle trim, these errors in the moment-increment prediction will result in an over-prediction of the required body-flap deflection. The consequences for Space-Shuttle trimmability has been discussed before, see, for instance, Fig. 3.49.

Concluding, it can be said that the longitudinal force coefficients show excellent agreement with wind-tunnel data. The pitch moment coefficients are reasonably well predicted, but show substantial discrepancies in some regions. To properly predict the side force and yaw moment coefficients, using shadowing is necessary, although this slightly decreases the accuracy of the yaw-moment coefficients. Elevon increments are better predicted than body flap increments. The body-flap moment and lift increments show the largest deviations, especially at high angle of attack and Mach numbers (Fig. 3.71).

■

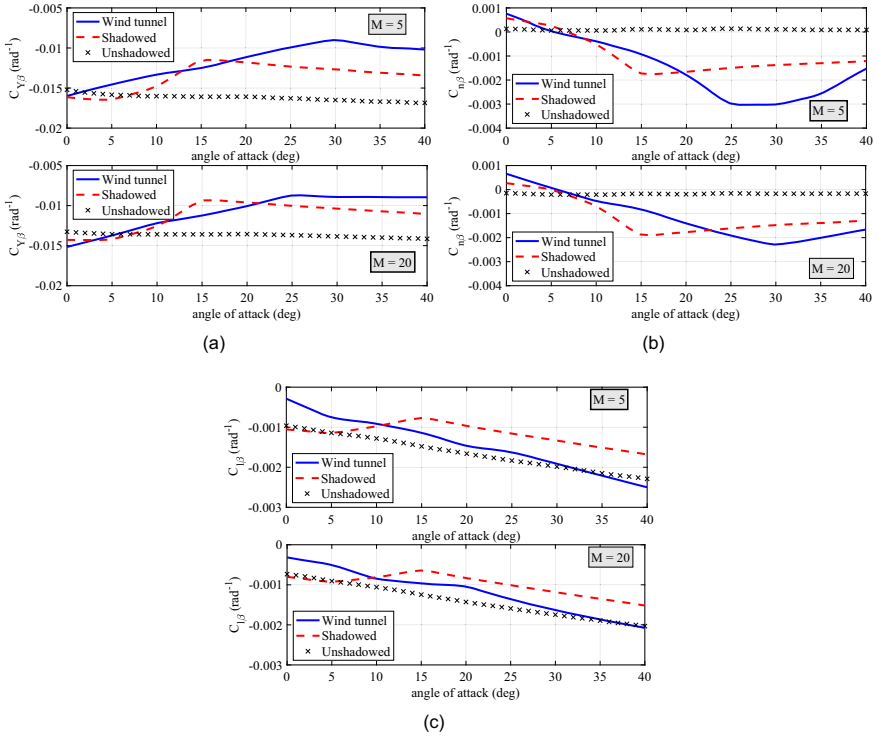


Fig. 3.66 Stability derivatives produced for the Space Shuttle with and without shadowing compared to wind-tunnel data

3.5 Free-Molecular Flow

In the rarefied or free-molecular regime, the assumption of collision-less flow can be made. The flow field close to the vehicle is dominated by particle-surface collisions whereas intermolecular collisions are neglected. The molecules can thus be assumed to hit the object surface undisturbed and with free-stream velocity. There are two approaches to model the molecule-vehicle interaction:

(i) Specular Reflection. The particle that hits the surface only transfers momentum in the direction perpendicular to the surface. The velocity component parallel to the surface is unchanged.
(ii) Diffuse Reflection. After collision, the particle can assume any magnitude and direction of the velocity, given by a Maxwell-Boltzmann distribution.

Experiments show that diffuse reflection is the method that represents reality the best, and is commonly assumed for general engineering surfaces as used on satellites, rocket bodies and entry vehicles (Regan and Anandakrishnan 1993). A characteristic of this regime is thus that friction drag is large compared to the pressure drag.

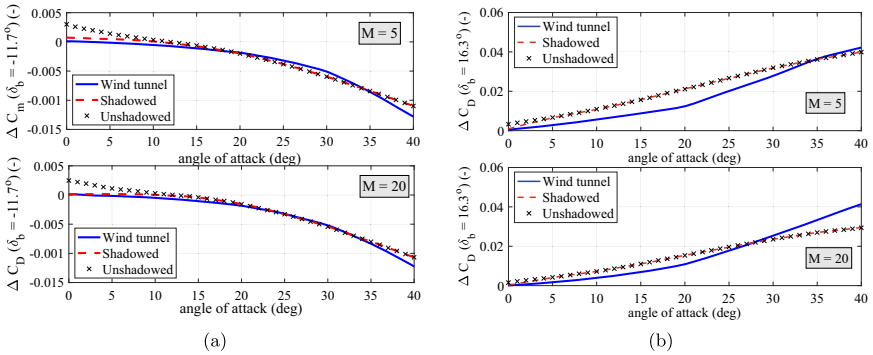


Fig. 3.67 Elevon moment increments produced for the Space Shuttle compared to wind-tunnel data

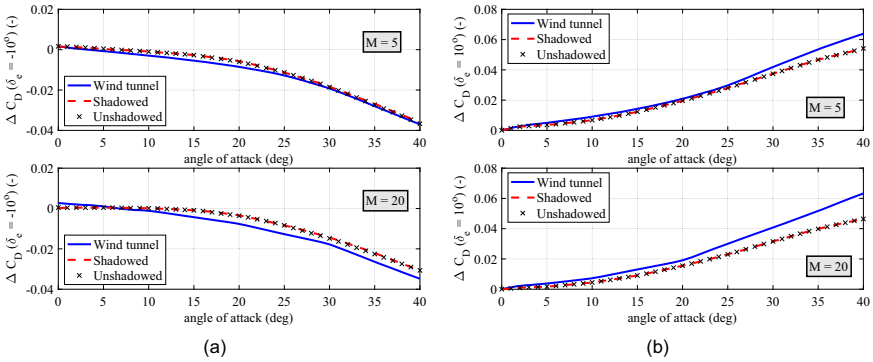


Fig. 3.68 Elevon drag increments produced for the Space Shuttle compared to wind-tunnel data

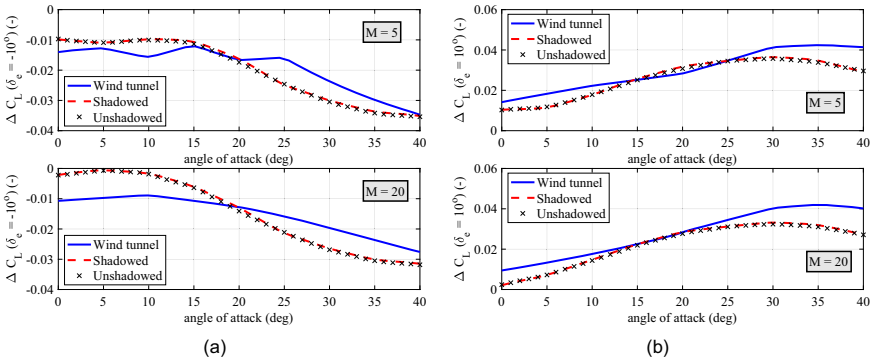


Fig. 3.69 Elevon lift increments produced for the Space Shuttle compared to wind-tunnel data

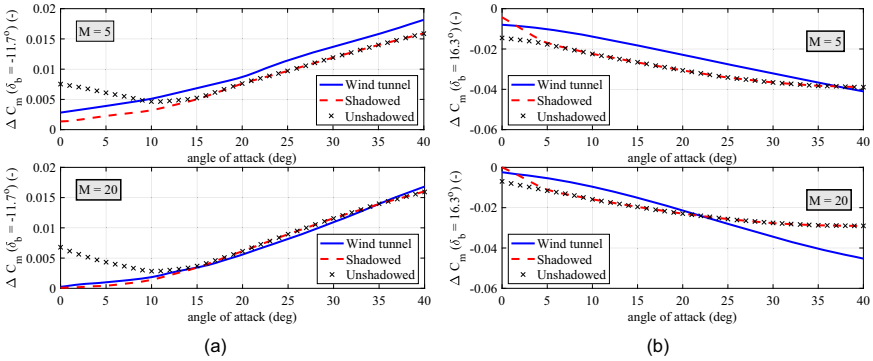


Fig. 3.70 Body-flap moment increments produced for the Space Shuttle compared to wind-tunnel data

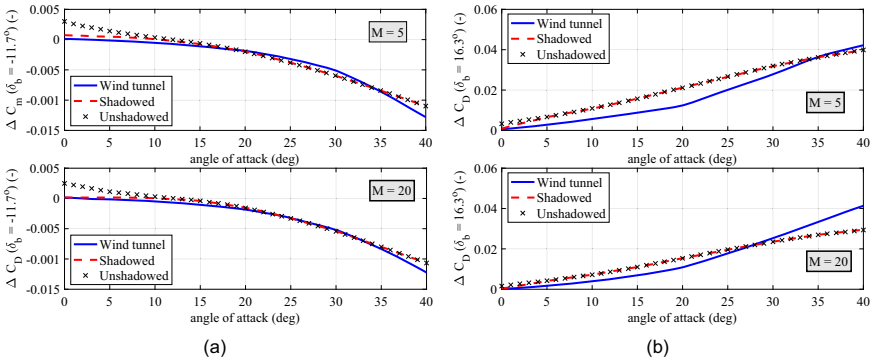


Fig. 3.71 Body-flap drag increments produced for the Space Shuttle compared to wind-tunnel data

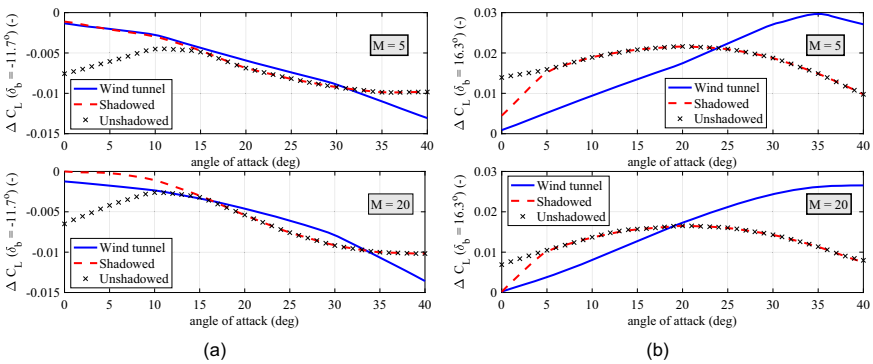


Fig. 3.72 Body-flap lift increments produced for the Space Shuttle compared to wind-tunnel data

To calculate the force acting on the vehicle, the Navier–Stokes equations do not apply anymore, since flow is no longer a continuum. Discrete particle models are typically used, driven by the Boltzmann equation, and for very high Knudsen numbers the collision-less Boltzmann equation (Moss 1992). Fortunately, Hart et al. (2014) provide a simple engineering method that can be used in conceptual design studies. In the free-molecular flow regime, under the assumption of the absence of inter-molecular collisions, the pressure and shear force on an elemental area inclined at angle θ_i to the flow are computed by:

$$C_{p,i} = \frac{1}{s^2}\left[\frac{2-\sigma_N}{\sqrt{\pi}} s\sin\theta_i + \frac{\sigma_N}{2}\sqrt{\frac{T_w}{T_\infty}}\right] e^{-s^2\sin^2\theta_i} + \frac{1}{s^2}\left[(2-\sigma_N)\left(\frac{1}{2}+s^2\sin^2\theta_i\right) + \frac{\sigma_N}{2}\sqrt{\frac{T_w}{T_\infty}}\sqrt{\pi} s\sin\theta_i\right][1+\text{erf}\,(s\sin\theta_i)] \tag{3.245}$$

$$C_{t,i} = \frac{\sigma_T\cos\theta_i}{s\sqrt{\pi}}\left[e^{-s^2\sin^2\theta_i} + \sqrt{\pi} s\sin\theta_i[1+\text{erf}\,(s\sin\theta_i)]\right] \tag{3.246}$$

where $\bar{q}_\infty$ represents the dynamic pressure, s the molecular speed ratio, $\frac{T_w}{T_\infty}$ the ratio between wall and free-stream temperature and σ_N and σ_T are the normal and tangential momentum accommodation coefficients, which describe the change in momentum due to the molecular-surface interaction in mentioned directions. For the applied diffuse reflection model, $\sigma_N \approx \sigma_T \approx 1$ (Regan and Anandakrishnan 1993). Finally, the molecular speed ratio, s, is related to the Mach number according to:

$$s = \frac{V_\infty}{\sqrt{2RT_\infty}} = M_\infty\sqrt{\frac{\gamma}{2}} \tag{3.247}$$

Note that erf () is the error function, defined as:

$$\text{erf}\,(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp\left(-\chi^2\right)\text{d}\chi \tag{3.248}$$

To obtain the forces (and moments) acting on the body, Eqs. (3.245) and (3.246) are integrated over the surface of the vehicle. As with the local-inclination methods, this boils down to discretising the vehicle surface into small panels, calculate the force felt by each of these panels, and then sum them to obtain the total. Normalising the forces gives us the value of C_D, C_S and C_L, the drag, side and lift force coefficients of the body, and, if need be, giving the location of the centre of mass, also the moment coefficients.

The results obtained with the method explained above may not always be accurate enough. A more reliable and accurate analysis can be done by using so-called Direct Simulation Monte Carlo (DSMC) methods, which are based on a probabilistic

simulation (of many particles) to solve the Boltzmann equation for finite Knudsen number fluid flows. An open-source software, freely available under GNU general public license, which implements the DSMC method is the Stochastic Parallel Rarefied-gas Time-accurate Analyzer (SPARTA) software, developed by the Sandia National Laboratories[26] (Gallis et al. 2014). Being a valuable tool that will be used in Chap. 11, in the Intermezzo below we will briefly outline this software.

*** Intermezzo 3.3 (SPARTA)**

SPARTA uses a DSMC method that "*models rarefied gases, using collision, chemistry, and boundary condition models*" (Plimpton and Gallis 2017). As described by Klothakis and Nikolos (2015), the DSMC method simulates the conditions in a rarefied gas by using kinetic theory of gases. In particular, the Boltzmann equation, which describes the change in particle distribution due to collisions, is used. Thus, the DSMC method simulates the motion of millions of particles, each representing a larger number of actual particles.

The solution obtained with a DSMC method is equivalent to the solution of the Boltzmann equation, when the number of particles tends to infinity and when the mesh and step sizes tend to zero (Bird 2013). The mesh (or grid) is the discretisation of the body being analysed. For SPARTA, for instance, the body is discretised by a triangulated surface mesh. Furthermore, sophisticated versions of the DSMC, just like SPARTA, manage to offer results that closely agree with experimental results, despite the non-ideal conditions (finite number of particles and non-zero mesh and step sizes). However, depending on the geometry of the body, the computation time and accuracy of the results will vary. For instance, a shape that generates large separation regions, will be more computationally challenging than the simulation for a blunt body (Klothakis and Nikolos 2015).

Theoretically, the DSMC method can be used for both the rarefied and continuum regimes, but for the continuum regime the computation times would be exponentially higher, due to the very large number of intermolecular interactions. For practical purposes one would use such a software only for the rarefied regime.

■

The continuum-flow region does not discretely change into the free-molecular region, but there is a rather large transitional regime. There is no general theoretical formulation for the pressure and shear coefficients in this regime, as its character is a mixture of both free-molecular and continuum flow. In principle, DSMC methods should be used to estimate the coefficients, but this may lead to excessive computation times.

An alternative and simpler method is to use a bridging function between the coefficients found by applying above theory and the ones obtained with continuum-flow models (Sect. 3.4.4). Such a bridging function is usually expressed as a function of Knudsen number and has the general form of:

[26] Sandia National Laboratories: SPARTA Direct Simulation Monte Carlo (DSMC) Simulator—https://sparta.sandia.gov/. Date of last access: August 19, 2019.

$$F(Kn) = \frac{C_{x,trans} - C_{x,cont}}{C_{x,fm} - C_{x,cont}} \tag{3.249}$$

where the subscript "fm" indicates the corresponding free-molecular coefficient, "cont" denotes the continuum variant, and "trans" is the transition coefficient to be calculated. The function F is empirically determined, by fitting experimental data and DSMC results. The results will, therefore, vary based on the overall shape of the body.

For the Space Shuttle, Blanchard and Buck (1986) derived the following function:

$$F(Kn) = \sin^2\left\{\frac{\pi}{8}\left[3 + \log_{10}(Kn)\right]\right\} \qquad 10^{-3} \leq Kn \leq 10^{1} \tag{3.250}$$

which can be used for both pressure and shear stress coefficients. Potter (1988), on the other hand, derived two separate bridging functions for C_L and C_D, and applied those to an aerocapture manoeuvre with heat shield. In this case, the bridging functions are not a function of Knudsen number, but rather depend on a combination of flow-field and vehicle parameters. Also Wilmoth et al. (1998) apply their bridging function directly to C_L and C_D, but in this case their solution is similar to Eq. (3.250).

Clearly, there exist many possibilities for the expression of the bridging function. Equation (3.250) has proven its worth in literature, and before moving to vehicle-shape dependent bridging functions, this may be a good starting point for an initial analysis.

Example 3.5 (*Delta-K Aerodynamics*) Ronse and Mooij (2014) have analysed the impact area of objects decaying from orbit, most notably Delta-K rocket stages. Unfortunately, there is no place to discuss the statistical impact study in more detail, but it involved simulations of the translational and rotational accelerations of a tumbling object due to aerodynamic forces and moments. To do so, one needs to have the related aerodynamic coefficients at different attitudes, flow regimes, and Mach numbers. Determining the aerodynamics of the Delta-K rocket stage is the topic of the current example.

To calculate the coefficients with the engineering methods discussed before, the object shape is first approximated by a combination of simple shapes, after which its surface is panelised by a quadrilateral meshing sequence. In case of the modelled Delta-K rocket body, the result is shown in Fig. 3.73. For every panel, the centroid and area are determined, which are subsequently used in the determination of aerodynamic coefficients with Eqs. (3.245) and (3.246). The panel's outward-normal direction is used by the shadowing algorithm, discussed in Sect. 3.4.4.3, to determine the panel's contribution to the overall coefficients.

As an example, the resulting pressure distribution at two angles of attack is shown in Fig. 3.74 for a free-molecular flow at $M = 20$. The shadowed panel segments of the object can be clearly distinguished behind the rocket body's mini-skirt and the inside of the nozzle. Since incoming molecules do not hit shadowed panels, the pressure coefficient of these is assumed zero.

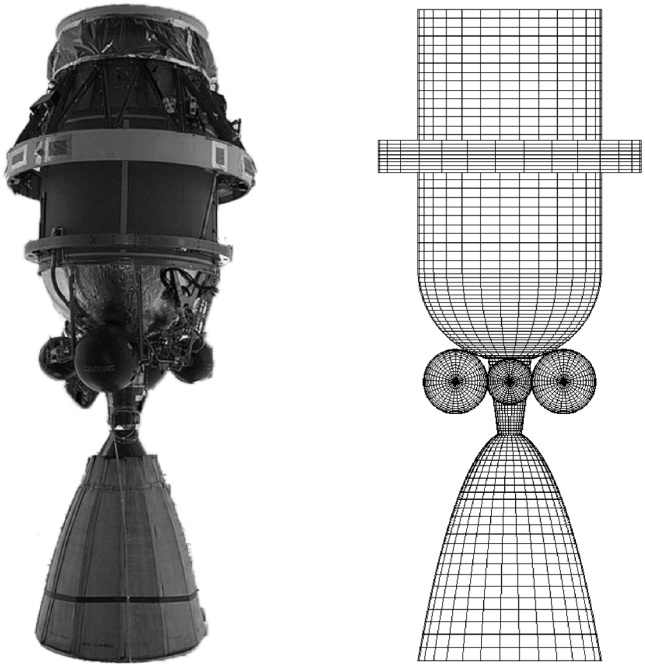


Fig. 3.73 Comparison of true Delta-K (Image: NASA) and modelled version consisting of simple geometric shapes

To realise the difference in aerodynamic coefficients, if either a free-molecular model or a continuum model is used, the pressure distribution can also be calculated with the modified-Newtonian method for the hypersonic regime (Sect. 3.4.4). An object in a decaying orbit will encounter arbitrary attitudes, so a full range of angle of attack and angle of sideslip is required. In terms of Mach number, in Fig. 3.75 a typical decay trajectory of the Delta-K is shown, starting at 200 km altitude. Based on the division by Knudsen number, the different flow regimes have been coloured. It can be seen that the Mach number during the flight in the free-molecular regime stays between 10 and 15. However, when further descending, the changing velocity and atmospheric conditions lead to a larger Mach variation in the transitional and hypersonic regimes. Concluding, the ranges for α, β, and M are $-180° \leq \alpha < 180°$, $-90° \leq \beta \leq 90°$, and $10 \leq M \leq 30$.

A selection of the obtained data is plotted in Fig. 3.76. A few observations are highlighted. The free-molecular drag coefficient is approximately twice as large as the hypersonic version at similar conditions. Knowing that the object passes through the free-molecular and transitional regime during its decay, it is thus of importance to take into account both flow regimes when simulating end-of-decay trajectories. Considering the flow field as continuum during the whole trajectory, would lead to

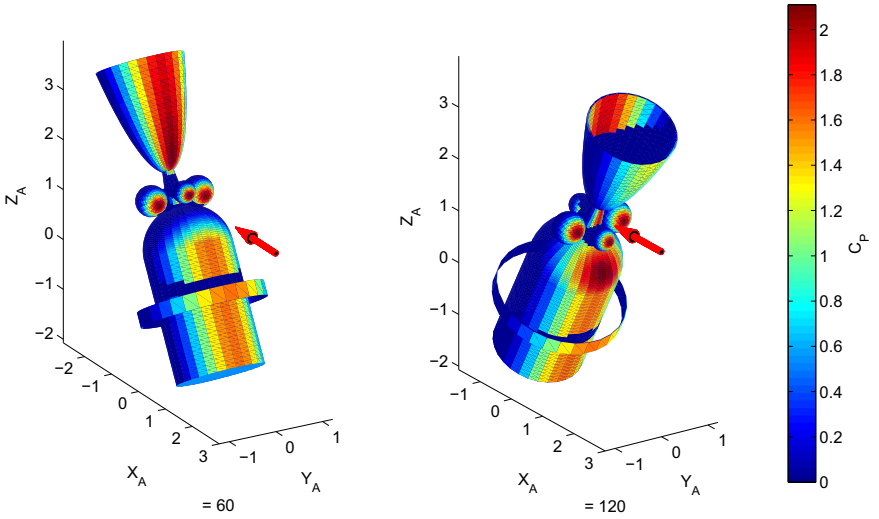


Fig. 3.74 Pressure distribution over modelled Delta-K rocket body in the free-molecular flow regime at two angle-of-attack conditions ($M = 20$)

an underestimation of drag at altitudes larger than ∼85 km. On the other hand, the assumption of purely free-molecular flow would result in an overestimation of drag at altitudes lower than ∼150 km.

The lift coefficient is much more pronounced in the hypersonic regime, and the effect of lift in the free-molecular regime is very small. If simplifying assumptions are required, the aerodynamic effects in free-molecular flow may be limited to the consideration of drag only.

The effect of the difference in flow regime is much larger than that of a changing Mach number. Especially for the hypersonic coefficients, acquired using modified Newtonian theory, the coefficients are quasi Mach-independent, having a mere 2% difference between the $M = 5$ and $M = 30$ case. For the free-molecular coefficients, the Mach-dependency is more pronounced, with a ∼17% decrease in C_D when increasing M from 5 to 30.

The drag fluctuates a lot with changing attitude. For free-molecule flow at $\beta = 0°$ and $M = 18$, the lowest drag coefficient, $C_D = 2.3$, is observed at $\alpha = 180°$, while the maximum drag coefficient, $C_D = 8.5$, occurs at $\alpha \pm 78°$. For the hypersonic case, these values amount to $C_D = 0.7$ for $\alpha = 180°$ and $C_D = 4.4$ for $\alpha = \pm 89°$, respectively.

From the $C_m - \alpha$ graph, it can be concluded that the modelled vehicle is aerodynamically stable when having a nose-forward attitude. This is characterised by a negative moment coefficient for positive α and vice versa. Obviously, this all depends on the centre-of-mass location, which was determined to lie in front of the vehicle's centre of pressure. Also, a limited region of stability can be seen for a nozzle-forward attitude ($\alpha = 180°$), but only for free-molecular flow conditions. In all fairness, the

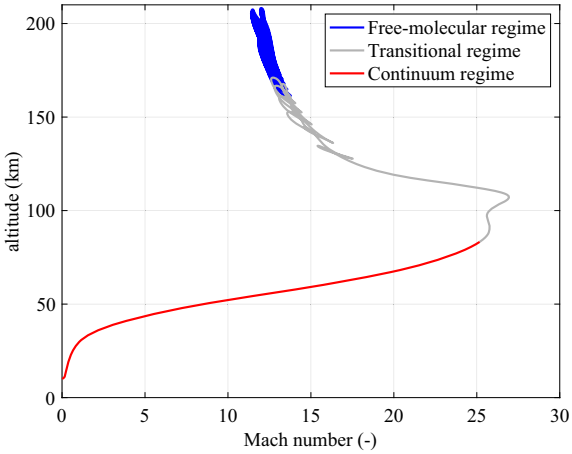


Fig. 3.75 Evolution of Delta-K's Mach number in a typical decay trajectory

determination of the centre-of-pressure location is not trivial, as we also saw in Example 3.4.

It is noted that because the Delta-K is quasi axial-symmetric about its longitudinal axis, the influence of the sideslip angle, β, on the side-force coefficient, C_S, and yaw moment coefficient, C_n, is very similar to the depicted dependency of the lift-force coefficient, C_L, and pitch moment coefficient, C_m on the angle of attack, α.

■

In the above example we have seen the importance of using the right method in the right flow regime, and if the mission covers both regimes then also a bridging function is required. In the example below, the selected bridging function, Eqs. (3.249)–(3.250), is verified by comparing the outcome with data from literature. This verification was done to support the impact analysis done by Ronse and Mooij (2014).

Example 3.6 (*Verification of Bridging Function*) The verification of the bridging function is done for two configurations, i.e., a blunted conical shape with 9° semi-cone angle and the Apollo Command Module. For the blunted cone, Padilla and Boyd (2006) have acquired wind-tunnel data and performed DSMC computations for the rarefied and transient regimes. Moss et al. (2006) studied the aerodynamics of the Apollo capsule using the 3D DSMC code, *DS3V*, in the free-molecule and transitional regime and used a CFD approach to get the numerical solution to the Navier-Stokes equations at hypersonic continuum conditions.

Based on the provided geometrical data in the quoted references, the two shapes were modelled and a panel mesh was generated, see Fig. 3.77. Our calculations are done using the panelised modified Newtonian and free-molecular flow methods, as discussed above, and include the Space-Shuttle bridging function. The results for both configurations are discussed separately.

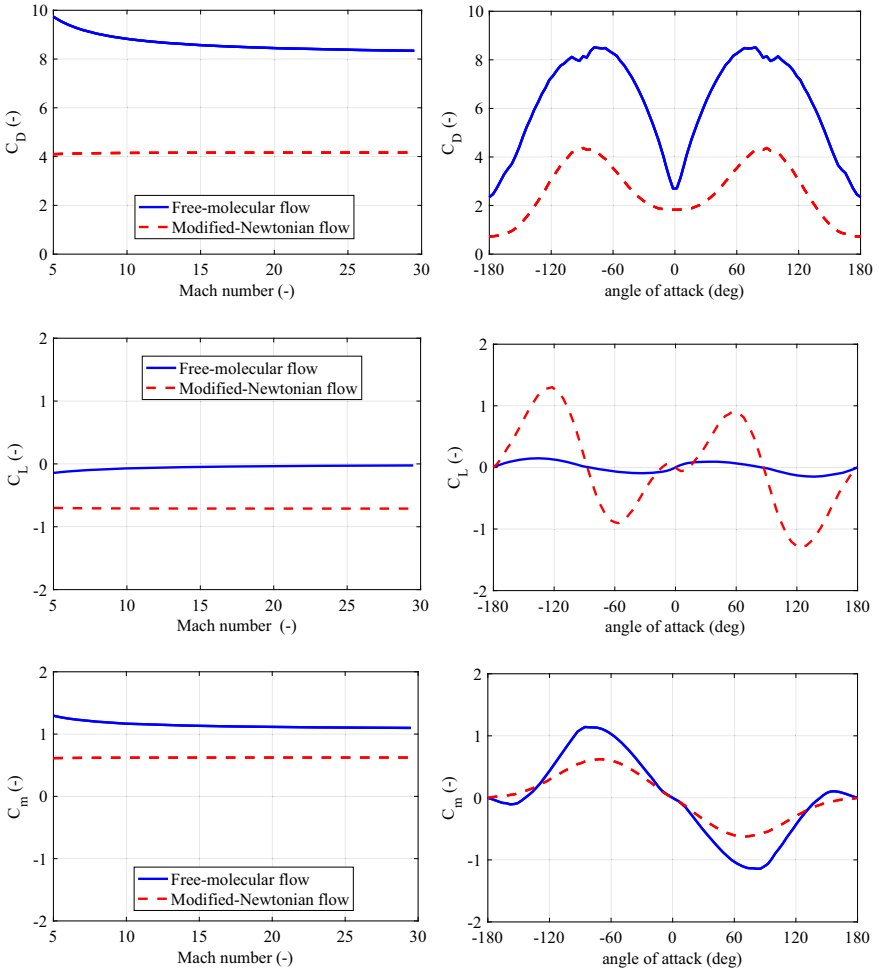


Fig. 3.76 Overview of some Delta-K aerodynamic characteristics, calculated using determined using panelised engineering methods. Left figures: $\alpha = -71°$, $\beta = 0°$; right figures: $M = 18$, $\beta = 0°$

Blunted cone

The L/D and C_m of the fully-revolved free-molecular cases were compared to the exact integral solution and panelised method, and are depicted in the top two graphs of Fig. 3.78. Moreover, C_D, C_L and C_m for all three flow regimes are compared to the panelised (free-molecular and modified Newtonian methods), DSMC and experimental data (transitional flow) for low angles of attack (0° to 30°). This comparison is shown in the lower four graphs of Fig. 3.78.

Overall, the free-molecular and modified Newtonian results are in good agreement with the published data. The small offsets can be attributed to minor modelling

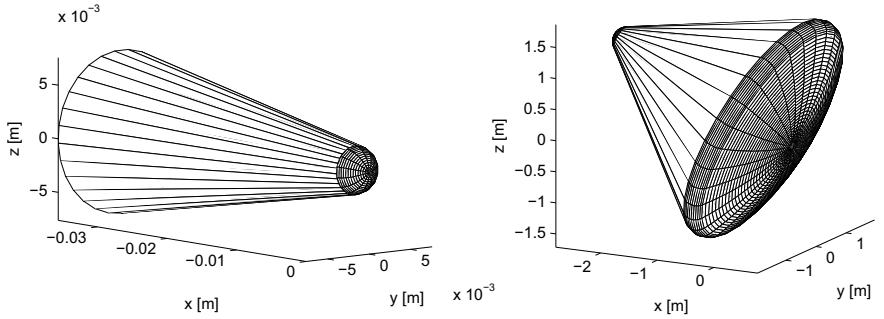


Fig. 3.77 3D meshed models of the blunted cone (left) and the Apollo (right)

differences of the blunted cone (mesh resolution, dimensions, etc.) and a different computation of the pressures at leeward surfaces. An interesting observation is found in the bridging functions for determining the coefficients in the transitional regime. The approximation of the wind-tunnel measurements is similar, and for C_D even better compared to the DSMC estimates. Among the bridging functions, the Shuttle-type performs best for estimating C_D, C_L and C_m. It is therefore advised to use bridging function Eqs. (3.249)–(3.250) for the simulation of decay and re-entry trajectories.

Apollo

For the Apollo capsule, two scenarios are analysed and compared. The first studies the evolution of C_D and C_L as a function of Kn (and thus altitude), at a constant angle of attack of $-25°$. These results are shown in Fig. 3.79. The high Knudsen numbers at the right of the figure ($Kn > 10$) correspond with free-molecular flow, whereas the far left ($Kn < 10^{-3}$) of the figure represents continuum flow. For each data point given by Moss et al. (2006), a corresponding computation was done with the (bridged) panel methods. Note that the validation data were computed by two methods: for altitudes above 100 km, DSMC software was used, while for the lower altitudes, CFD solutions have been used. As seen, the panelised C_D has comparable values to the validation data across the full regime. The bridged C_L and L/D values have a slightly larger offset from Moss' solutions. They are still reasonably accurate, especially when one considers the large difference in computational complexity between the techniques.

For the second scenario, the aerodynamic coefficients of the Apollo were evaluated at different angles of attack. Therefore, a (transitional) Kn of 0.081 is used (equivalent to $h = 105$ km), again with $V_\infty = 9.6$ km/s. The results are shown in Fig. 3.80. In the left plot, the coefficients resulting from a free molecular and hypersonic analysis are also depicted. The bridged results compare well with the DSMC data points. Also for C_m there is good consistency between both methods for the major part of the curve. Only around $\alpha \approx 0°$ there is a larger offset between both approaches.

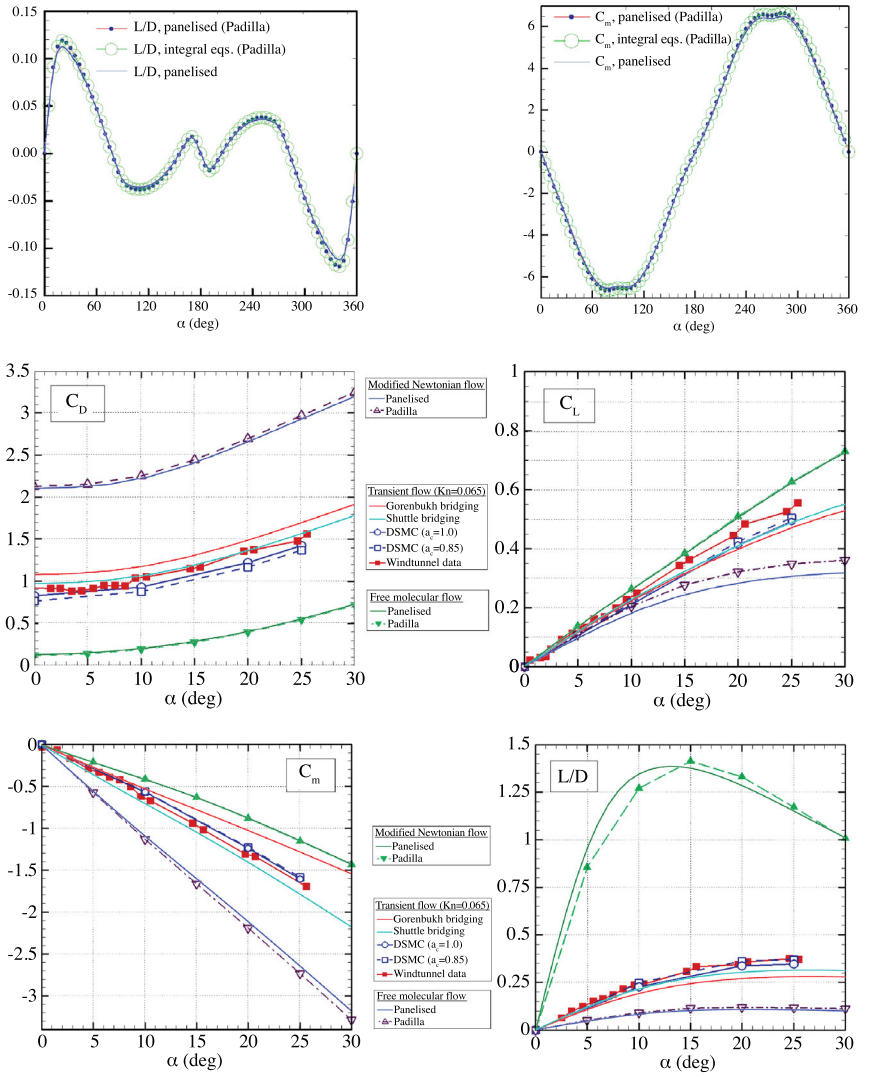


Fig. 3.78 Comparison of aerodynamic parameters in different flow regimes to data from Padilla and Boyd (2006)

■

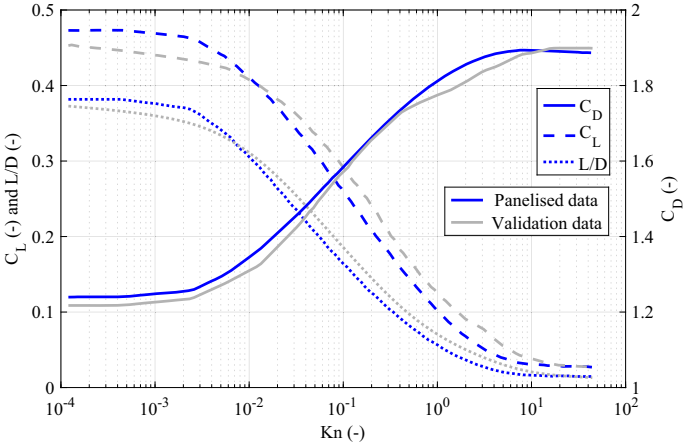


Fig. 3.79 Comparison of Apollo's lift and drag terms between the (bridged) panel method and DSMC/CFD techniques ($\alpha = -25°$ and $V_\infty = 9.6$ km/s). Validation data from Moss et al. (2006)

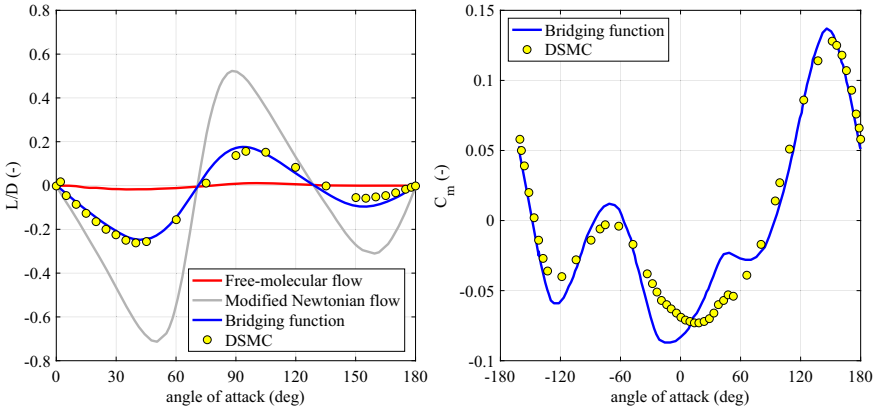


Fig. 3.80 Comparison of Apollo's lift-to-drag ratio and pitch-moment coefficient between the (bridged) panel method and DSMC/CFD techniques ($Kn = 0.081$ and $V_\infty = 9.6$ km/s). Validation data from Moss et al. (2006)

3.6 Aerodynamic Heating

Let us consider one kg of "something" at an orbital altitude of about 200 km. At this altitude, the circular orbital velocity is 7,784 m/s, which means that its specific kinetic energy is $\mathcal{E}_k = \frac{1}{2}V^2 \approx 30$ MJ/kg. Together with its specific potential energy, using a flat-Earth approximation, $\mathcal{E}_p = gh \approx 2$ MJ/kg, it means that the total energy is about 32 MJ/(kg of something) in low Earth orbit. Assuming now that this "something" needs to be going back "home", towards Earth's surface. After landing, it would be

at zero altitude with no velocity, so between leaving orbit and landing, 32 MJ/kg has to be dissipated by friction with the atmosphere.

If we now somehow manage to relate the energy absorption to the object's thermal protection system, then we can say something about the object's temperature increase. Suppose we have a fictitious "heat shield" of pure graphite (carbon), which is a high-temperature material with the highest heat-absorption coefficient: $c_p = 709$ J/kg K. Given the 32 MJ/kg, and assuming that all thermal energy is absorbed by the heat shield, this means that the orbital energy would cause a temperature gain of 45,000 K!

But this would not be the only issue. The portion of a gliding flight where the main heat generation takes place, lasts about eight minutes, which means an energy absorption of 66 kW/kg (in case this would be a ballistic entry, it would take about three minutes, so the energy absorption would be 178 kW/kg!). Even if the TPS could somehow divert the total amount of energy, it has to do this at a staggering rate!

Obviously, this is not how it happens in reality, otherwise no spacecraft could ever return to Earth in one piece, or, rather, at all, as it would be vaporised. Revisiting the flow field shown in Fig. 3.41, a bow shock at a stand-off distance is present in front of the (blunt) nose of the vehicle. Behind the shock, there is a subsonic bubble in which the entropy is rising. Along the wall of the vehicle a boundary layer is forming, and also the entropy layer is growing towards the rear of the vehicle. Due to the shock, there will be a significant drop in total pressure, and in combination with the growing entropy this acts as a kind of shield for the vehicle, such that only limited heat will reach—and enter—the vehicle wall. Earlier, in Fig. 3.42, the flow conditions behind the shock were shown. With increasing Mach number the drop in total pressure keeps increasing, as well as the entropy.

In this section, we will deal with the aerodynamic-heating problem of entry flight. Given the situation described above, one might suspect that entry flight is indeed the most demanding part of a spaceflight that requires a return to Earth. In a way this is true, but looking somewhat ahead into the future, where hopefully we have reusable spaceplanes launching payloads and passengers into space and returning them safely to Earth, it is actually the ascent flight that becomes driving for mission and vehicle design. Spaceplanes with air-breathing engines typically follow a shallow path to space, to make most use of the denser part of the atmosphere for engine efficiency. Tauber and Menees (1986) analysed the heat loads for an ascending and re-entering spaceplane, in its ascent flight flying at a constant dynamic pressure of 20 kPa and accelerating with a constant 0.4 g, and compared them with Space Shuttle data of its re-entry. It can be seen in Fig. 3.81 that the heating environment for the spaceplane is more demanding, and since the Space Shuttle's thermal protection system was already loaded close to its limits, completely new (and active) cooling systems will be required for these new launchers. In all fairness, the nose and wing-leading edges of the spaceplane were a lot smaller than those of the Space Shuttle, and as we will see later, this drives the heat load up significantly. However, to achieve its aerodynamic performance, this particular geometry was required for the spaceplane. Nonetheless, currently we will focus on the problem at hand, the re-entry flight. To get an appetite for the challenges at hand, Fig. 3.82 shows the stream-line pattern left behind by the

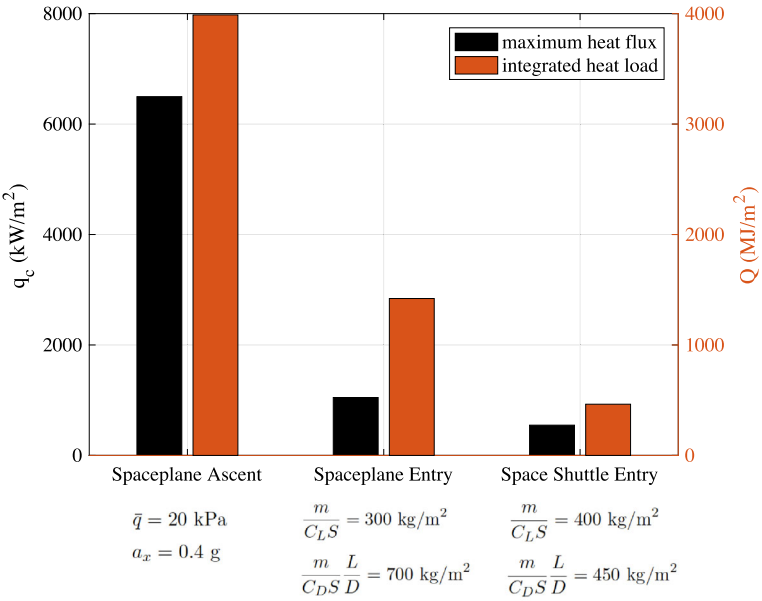


Fig. 3.81 Comparison of heat loads between a (future) spaceplane and the Space Shuttle

hot gasses flowing over the surface during re-entry. Note that the difference in tile colours indicates the difference in thermal loading, as well as the result of tiles being replaced during the maintenance cycle.

Three aspects of aerodynamic heating are important. The first aspect is the time rate and maximum rate of local stagnation region heat input per unit area. This stagnation area is an important local hot spot, e.g., the nose for (ballistic) vehicles, wing leading edges for gliding vehicles, etc. In general, the stagnation point is the hottest part of the vehicle. The second aspect is the time rate and maximum rate of average heat input per unit area, which determines the peak average flow-rates in case of fluid coolants, or the overall structural strength of the heated structure, when thermal stresses dominate. The third and last aspect is the total heat input, integrated over the duration of the entry flight and (parts of) the vehicle's surface area. The integrated heat load is a measure for the overall coolant weight, being solid, fluid, gas, or combination of those.

To treat aerodynamic heating, we need to consider the viscous flow field close to the body. This part of the flow field, the boundary layer, was discussed previously in Sect. 3.3.5, where an analytical solution to the incompressible boundary layer was derived, and briefly revisited in Sect. 3.4.2. In that section, compressibility effects and transition from laminar to turbulent flow was added to the picture.

For a proper treatment of wall friction and associated heating we need to go into more detail, and will do so in Sect. 3.6.1. The two sections thereafter will treat

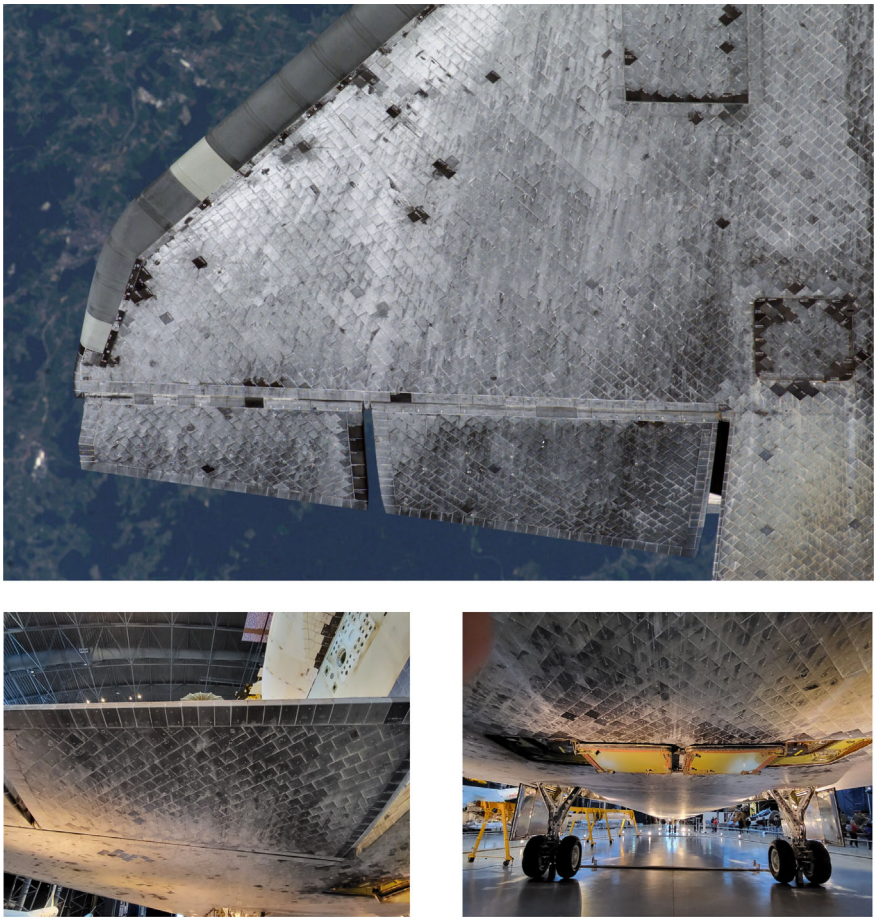


Fig. 3.82 Bottom side of the Space Shuttle Discovery's thermal protection system (right wing), with a clear indication of the stream-line pattern of the very hot gasses passing over its surface. Left insert: detail of the left, in-board elevon, right insert: detail of the aft fuselage. Top image: NASA, bottom images: author

two mechanisms of heat transfer, i.e., convection in Sect. 3.6.3 and radiation in Sect. 3.6.2.

3.6.1 Boundary-Layer Dynamics

From the upcoming Sect. 3.7, we get the general boundary-layer equations:

$$\frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} = 0 \tag{3.251}$$

$$\rho u \frac{\partial u}{\partial x} + \rho v \frac{\partial u}{\partial x} = -\frac{\partial p}{\partial x} + \frac{\partial}{\partial y}\left(\mu \frac{\partial u}{\partial y}\right) \tag{3.252}$$

$$\frac{\partial p}{\partial y} = 0 \tag{3.253}$$

$$\rho u \frac{\partial h}{\partial x} + \rho v \frac{\partial h}{\partial y} = u \frac{\partial p}{\partial x} + \mu \left(\frac{\partial u}{\partial y}\right)^2 + \frac{\partial}{\partial y}\left(k \frac{\partial T}{\partial y}\right) \tag{3.254}$$

Before we move on to the derivation of boundary-layer equations for a laminar, compressible flow over a flat plate, we will give the non-dimensional form of Eqs. (3.251) through (3.254). These equations will introduce some similarity parameters, which will give the reader insight under which flow conditions the boundary layers have similar shape and conditions, and solving the related Navier-Stokes equations will provide identical solutions.

*** Intermezzo 3.4 (Non-Dimensional Form of Boundary-Layer Equations)**
The variables appearing in Eqs. (3.251) through (3.254) can be non-dimensionalised by the following set of definitions:

$$\bar{x} = \frac{x}{L} \qquad \bar{y} = \frac{y}{L} \qquad \bar{u} = \frac{u}{V_\infty} \qquad \bar{v} = \frac{v}{V_\infty} \qquad \bar{\rho} = \frac{\rho}{\rho_\infty}$$

$$\bar{p} = \frac{p}{\rho V_\infty^2} \qquad \bar{T} = \frac{T}{T_\infty} \qquad \bar{\mu} = \frac{\mu}{\mu_\infty} \qquad \bar{k} = \frac{k}{k_\infty} \qquad \bar{h} = \frac{h}{h_\infty} = \frac{h}{c_p T_\infty}$$

and after substitution, it follows for Eqs. (3.251)–(3.254)

$$\frac{\partial \bar{\rho}\bar{u}}{\partial \bar{x}} + \frac{\partial \bar{\rho}\bar{v}}{\partial \bar{y}} = 0 \tag{3.255}$$

$$\bar{\rho}\bar{u} \frac{\partial \bar{u}}{\partial \bar{x}} + \bar{\rho}\bar{v} \frac{\partial \bar{u}}{\partial \bar{x}} = -\frac{\partial \bar{p}}{\partial \bar{x}} + \frac{\mu_\infty}{\rho_\infty V_\infty L} \frac{\partial}{\partial \bar{y}}\left(\bar{\mu} \frac{\partial \bar{u}}{\partial \bar{y}}\right) \tag{3.256}$$

$$\frac{\partial \bar{p}}{\partial \bar{y}} = 0 \tag{3.257}$$

$$\begin{aligned} \bar{\rho}\bar{u} \frac{\partial \bar{h}}{\partial \bar{x}} + \bar{\rho}\bar{v} \frac{\partial \bar{h}}{\partial \bar{y}} = \frac{V_\infty^2}{c_p T_\infty} \bar{u} \frac{\partial \bar{p}}{\partial \bar{x}} + \frac{\mu_\infty}{\rho_\infty V_\infty L} \frac{V_\infty^2}{c_p T_\infty} \bar{\mu} \left(\frac{\partial \bar{u}}{\partial \bar{y}}\right)^2 + \\ + \frac{\mu_\infty}{\rho_\infty V_\infty L} \frac{k_\infty}{\mu_\infty c_p} \frac{\partial}{\partial \bar{y}}\left(\frac{\partial \bar{T}}{\partial \bar{y}}\right) \end{aligned} \tag{3.258}$$

In the above equations some similarity parameters can be discerned, i.e., the Reynolds number, scaled for a characteristic length, L,

$$\frac{\rho_\infty V_\infty L}{\mu_\infty} \tag{3.259}$$

the Prandtl number,

$$Pr = \frac{\mu_\infty c_p}{k_\infty} \tag{3.260}$$

and the Mach number, through

$$\frac{V_\infty^2}{c_p T_\infty} = \frac{(\gamma - 1)V_\infty^2}{\gamma R T_\infty} = (\gamma - 1)M_\infty^2 \tag{3.261}$$

Thus, Eqs. (3.256) and (3.258) can be rewritten as:

$$\bar{\rho}\bar{u}\frac{\partial \bar{u}}{\partial \bar{x}} + \bar{\rho}\bar{v}\frac{\partial \bar{u}}{\partial \bar{x}} = -\frac{\partial \bar{p}}{\partial \bar{x}} + \frac{1}{Re}\frac{\partial}{\partial \bar{y}}\left(\bar{\mu}\frac{\partial \bar{u}}{\partial \bar{y}}\right) \tag{3.262}$$

$$\bar{\rho}\bar{u}\frac{\partial \bar{h}}{\partial \bar{x}} + \bar{\rho}\bar{v}\frac{\partial \bar{h}}{\partial \bar{y}} = (\gamma - 1)M_\infty^2\left[\bar{u}\frac{\partial \bar{p}}{\partial \bar{x}} + \frac{\bar{\mu}}{Re}\left(\frac{\partial \bar{u}}{\partial \bar{y}}\right)^2\right] + \frac{1}{Re}\frac{1}{Pr}\frac{\partial}{\partial \bar{y}}\left(\frac{\partial \bar{T}}{\partial \bar{y}}\right) \tag{3.263}$$

In normal boundary-layer theory, the Reynolds number is on the order of $10^5 - 10^8$. The Prandtl number is dependent on the composition of the fluid, and for air under standard conditions $Pr = 0.71$ (White 2005). When no chemical reactions take place, $Pr = f(T)$, because in that case also μ, c_p and k are a function of temperature only. For a large range of temperatures, $Pr = 0.71$ is a reasonable approximation. ■

The derivation of the boundary-layer equations for a laminar, compressible flow over a flat plate is illustrative and often used as the starting point of engineering analysis of aerodynamic heating. Let us, before we begin, introduce two parameters that play an important role in viscous flow with heat transport. The friction coefficient, c_f, is defined as

$$c_f = \frac{\tau_w}{\frac{1}{2}\rho_e u_e^2} \tag{3.264}$$

and the Stanton number, St, as

$$St = \frac{q_w}{\rho_e u_e (h_{aw} - h_w)} \tag{3.265}$$

where τ_w is the shear stress at the wall, q_w the heat flux per m^2, and h_w the enthalpy at the wall. The subscript "e" denotes that ρ and u are considered at the edge of the boundary layer, and h_{aw} is the enthalpy for an adiabatic wall, i.e., for which $q_w = 0$.

For a flat plate, p is constant, so Eqs. (3.252) and (3.254) simply reduce to

$$\rho u \frac{\partial u}{\partial x} + \rho v \frac{\partial u}{\partial y} = \frac{\partial}{\partial y}\left(\mu \frac{\partial u}{\partial y}\right) \tag{3.266}$$

$$\rho u \frac{\partial h}{\partial x} + \rho v \frac{\partial h}{\partial y} = \mu \left(\frac{\partial u}{\partial y}\right)^2 + \frac{\partial}{\partial y}\left(k \frac{\partial T}{\partial y}\right) \tag{3.267}$$

Summing Eqs. (3.266) and (3.267), and realising that $\frac{\partial u}{\partial ..} = \frac{\partial}{\partial ..}\left(\frac{u^2}{2}\right)$, yields

$$\rho u \frac{\partial}{\partial x}\left(h + \frac{u^2}{2}\right) + \rho v \frac{\partial}{\partial y}\left(h + \frac{u^2}{2}\right) = \frac{\partial}{\partial y}\left(k \frac{\partial T}{\partial y} + \mu u \frac{\partial u}{\partial y}\right) \tag{3.268}$$

Note that using the chain rule, it is easily shown that

$$\frac{\partial}{\partial y}\left(\mu u \frac{\partial u}{\partial y}\right) = \mu \frac{\partial u}{\partial y}\frac{\partial u}{\partial y} + \frac{\partial}{\partial y}\left(\mu \frac{\partial u}{\partial y}\right) \tag{3.269}$$

In Eq. (3.268), $h + \frac{u^2}{2}$ can be considered to be the *total* energy of the fluid, which is easily proven. Because the vertical velocity in the boundary layer much smaller than the horizontal one, $v \ll u$, in Eq. (3.268) we can write $h + \frac{u^2+v^2}{2}$ instead of $h + \frac{u^2}{2}$. And that expression represents the total energy of the fluid, i.e., the sum of enthalpy and kinetic energy.

The two terms in brackets on the right-hand side of Eq. (3.268) can be rewritten with the use of the Prandtl number, $Pr = \frac{\mu c_p}{k}$:

$$k \frac{\partial T}{\partial y} + \mu u \frac{\partial u}{\partial y} = \frac{\mu}{Pr}\frac{\partial h}{\partial y} + \mu \frac{\partial}{\partial y}\left(\frac{u^2}{2}\right) \tag{3.270}$$

which changes Eq. (3.268) into

$$\rho u \frac{\partial}{\partial x}\left(h + \frac{u^2}{2}\right) + \rho v \frac{\partial}{\partial y}\left(h + \frac{u^2}{2}\right) = \frac{\partial}{\partial y}\left[\mu \frac{\partial}{\partial y}\left(h + \frac{u^2}{2}\right) + \mu\left(\frac{1 - Pr}{Pr}\right)\frac{\partial h}{\partial y}\right] \tag{3.271}$$

A simplifying assumption that reduces the complexity and yet provides a good approximation is to set $Pr = 1$ (rather than $Pr = 0.72$ for air); in that case the last term on the right-hand side disappears:

$$\rho u \frac{\partial}{\partial x}\left(h + \frac{u^2}{2}\right) + \rho v \frac{\partial}{\partial y}\left(h + \frac{u^2}{2}\right) = \frac{\partial}{\partial y}\left[\mu \frac{\partial}{\partial y}\left(h + \frac{u^2}{2}\right)\right] \tag{3.272}$$

and the energy equation takes the same form as the momentum Eq. (3.266). This means that solutions to both equations are identical, apart from the constants of integration. For instance, if Eq. (3.266) has as a solution $u = f(x, y)$, then solving Eq. (3.272) results in

$$h + \frac{u^2}{2} = c_1 f(x, y) + c_2 = c_1 u + c_2 \tag{3.273}$$

The two constants c_1 and c_2 follow from the boundary conditions at the wall and the edge of the boundary layer, which, for a flat plate, is the free stream. So, for the wall we have: $y = 0$, $u = 0$, and $h = h_w$, and for the boundary-layer edge: $y = \delta$, $u = u_e$ ($= V_\infty$), and $h = h_e$ ($= h_\infty$). It then follows that

$$c_2 = h_w \qquad c_1 = \frac{1}{u_e}\left(h_e - h_w + \frac{u_e^2}{2}\right)$$

With the total enthalpy $h_0 = h_e + \frac{u_e^2}{2}$ it follows that $c_1 = \frac{1}{u_e}(h_0 - h_w)$, and Eq. (3.273) becomes

$$h + \frac{u^2}{2} = (h_0 - h_w)\frac{u}{u_e} + h_w \tag{3.274}$$

or, with $h = c_p T$:

$$T = T_w + (T_0 - T_w)\frac{u}{u_e} - \frac{u^2}{2c_p} \tag{3.275}$$

The conditions at the wall follows from taking the derivative of T to y:

$$\frac{\partial T}{\partial y} = (T_0 - T_w)\frac{1}{u_e}\frac{\partial u}{\partial y} - \frac{u}{c_p}\frac{\partial u}{\partial y} \tag{3.276}$$

which, for $y = 0$, evaluates to

$$\left(\frac{\partial T}{\partial y}\right)_w = (T_0 - T_w)\frac{1}{u_e}\left(\frac{\partial u}{\partial y}\right)_w \tag{3.277}$$

For an adiabatic wall it holds that $\left(\frac{\partial T}{\partial y}\right)_w = \frac{q_w}{k} = 0$, which means that $T_w = T_0$, under the assumption that $Pr = 1$. By definition, this temperature is equal to the adiabatic wall temperature, T_{aw}. The total enthalpy and temperature for this case follow from Eqs. (3.274) and (3.275)

$$h_0 = h_w = h + \frac{u^2}{2} \tag{3.278}$$

$$T_0 = T_w = T_{aw} = T + \frac{u^2}{2c_p} \tag{3.279}$$

and are constant throughout the boundary layer (of the flat plate), and amounts to the free-stream value. In case $Pr < 1$, which is the more realistic scenario, then $T_w < T_0$. Since $q_w = 0$ will still lead to the adiabatic wall temperature it can be expressed in terms of the so-called *recovery factor*, r:

$$T_{aw} = T_e + r\frac{u_e^2}{2c_p} \tag{3.280}$$

where r can be interpreted as a measure of how well the free-stream total temperature is reached. Earlier we found for $Pr = 1$ that $T_{aw} = T_0$, which means that $r = 1$. Equation (3.265) combined with the definition of total enthalpy, $h_0 = h_e + \frac{u_e^2}{2}$, gives us an expression for r:

$$r = \frac{T_{aw} - T_e}{T_0 - T_e} = \frac{h_{aw} - h_e}{h_0 - h_e} \tag{3.281}$$

For incompressible flow, the value of r is related to the Prandtl number as $r = \sqrt{Pr}$. Exact results for compressible flow show only a small difference, and also the effect of increasing Mach number that reduces r is small (2.4% for M_e going from 0 to 16). Hence, for all practical purposes we can assume for compressible laminar flow over a flat plate that $r = \sqrt{Pr}$ (Anderson 2019), and with $Pr = 0.71$ this yields $r = 0.84$.

Recalling the definition of the shear stress, $\tau_w = \mu\left(\frac{\partial u}{\partial y}\right)_w$, and the wall flux, $q_w = k\left(\frac{\partial T}{\partial y}\right)_w$, Eq. (3.277) can be rewritten as (with $T_0 = T_{aw}$ for $Pr = 1$):

$$\frac{q_w}{k} = \frac{T_{aw} - T_w}{u_e}\frac{\tau_w}{\mu} \tag{3.282}$$

From the definitions of the friction coefficient and the Stanton number, Eqs. (3.264)–(3.265) it then follows that

$$St = \frac{c_f}{2Pr} = \frac{c_f}{2} \tag{3.283}$$

This relation between heat transfer and friction is known as the *Reynolds Analogy*. In case one would use the more accurate value of $Pr = 0.71$, a more accurate result is Anderson (2019)

$$St = \frac{c_f}{2Pr^{\frac{2}{3}}} \tag{3.284}$$

3.6.2 Convection

3.6.2.1 Stagnation Heat Flux

To derive an expression for the convective heat flux into a re-entry vehicle we will, for practical reasons, consider a vehicle with a blunt nose. The flow will reach its highest temperature in the stagnation point, the point where theoretically the flow comes to a stand still. At this point, or in the vicinity thereof, the greatest transfer of heat to the body takes place. As we have discussed earlier, at hypersonic velocities a local subsonic flow with low Mach number is established behind the bow shock, in the region of the stagnation point.

The heat transfer rate that takes place between the flow and the wall is given by

$$\dot{Q} = \alpha A(T_2 - T_1) \tag{3.285}$$

with α being the heat-transfer coefficient in W/(m^2K), A in m^2 is the surface area under consideration, T_2 is the temperature of the fluid and T_1 is the temperature of the wall (the solid), both in K (or °C, as only the difference is important). For the convective heat flux (per unit area) from the boundary layer to the wall, one may write (Hankey 1988)

$$q_c = \frac{\mathrm{d}\dot{Q}}{\mathrm{d}A} = \alpha(T_{aw} - T_w) \tag{3.286}$$

From the definition of Stanton number, Eq. (3.265), the (convective) heat flux at the wall can also be written as

$$q_c = St\rho_e u_e(h_{aw} - h_w) = St\rho_e u_e c_p(T_{aw} - T_w) \tag{3.287}$$

As mentioned in the previous section, the subscript "e" indicates the conditions at the edge (or, equivalently, outside) the boundary layer. In the stagnation point, however, we cannot use the conditions from the edge of the boundary layer, but should take those behind the bow shock in the subsonic region.

Empirical correlations for local coefficients of skin friction in the stagnation region of a blunt nose have been developed over the past many years, and are similar to those for an incompressible boundary layer over a flat plate (cf. Eqs. (3.161) and (3.168) for the laminar and turbulent case, respectively). Hankey (1988) states that

$$c_f = \frac{c_{bl}}{Re^n} \tag{3.288}$$

with $Re = \rho_1 u_x x/\mu_1$ being the local Reynolds number at a distance x from the stagnation point; the values of the density and the viscosity in the subsonic region are taken as the values behind a straight shock wave (index 1). Further, for a laminar boundary layer, $c_{bl} = 0.664$ and $n = 0.5$ and for a turbulent boundary layer, $c_{bl} = 0.0592$ and $n = 0.2$.

Using the Reynolds analogy for $Pr = 1$, Eq. (3.283), the convective heat flux, Eq. (3.287) at a location x is now written as

$$q_c = \frac{c_f}{2}\rho_1 u_x c_{p,1}(T_{aw} - T_w) = \frac{c_{bl}}{2\left(\frac{\rho_1 u_x x}{\mu_1}\right)^n}\rho_1 u_x c_{p,1}(T_{aw} - T_w) \tag{3.289}$$

With Eq. (3.279), we can relate the adiabatic wall temperature to the undisturbed flow conditions,

$$T_{aw} = T_\infty + \frac{u_\infty^2}{2c_{p,\infty}} \tag{3.290}$$

and by realising that for hypersonic flow the "dynamic" temperature, i.e., the second term on the right-hand side, is a lot larger than the static temperature, T_∞, so

$$T_{aw} \approx \frac{u_\infty^2}{2c_{p,\infty}} \tag{3.291}$$

Since total temperature is (obviously) constant, it also holds for the temperature behind the shockwave that

$$T_0 = T_1 + \frac{u_1}{2c_{p,1}} \tag{3.292}$$

but since $u_1 \ll u_\infty$ it means that $T_0 \approx T_1$. Thus

$$T_{aw} \approx \frac{u_\infty^2}{2c_{p,\infty}} \approx T_1 \tag{3.293}$$

For the dynamic viscosity behind the shockwave, μ_1, we use Sutherlands equation, Eq. (3.34):

$$\frac{\mu_1}{\mu_\infty} = \frac{T_1^{1.5}}{T_1 + S}\frac{T_\infty + S}{T_\infty^{1.5}} \approx \frac{T_1^{1.5}}{T_1}\frac{T_\infty}{T_\infty^{1.5}} = \sqrt{\frac{T_1}{T_\infty}} \tag{3.294}$$

as both T_1 and T_∞ are substantially larger than $S(= 110.4\,\mathrm{K})$. Substituting Eq. (3.293) this yields:

$$\frac{\mu_1}{\mu_\infty} = \frac{u_\infty}{\sqrt{2c_{p,\infty}T_\infty}} \quad \text{or} \quad \mu_1 = \frac{u_\infty}{\sqrt{2c_{p,\infty}T_\infty}}\mu_\infty \tag{3.295}$$

Before we substitute Eqs. (3.293) and (3.295) into Eq. (3.289), we rewrite the latter, and normalise the distance x by dividing it with the nose radius, R_N:

$$q_c = \frac{c_{bl}}{2} \frac{\mu_1^n}{\left(\frac{x}{R_N}\right)^n} \frac{1}{R_N^n} \rho_1^{1-n} u_x^{1-n} c_{p,1} T_{aw} \left(1 - \frac{T_w}{T_{aw}}\right) \tag{3.296}$$

After substituting the approximations for T_{aw} and μ_1, and introducing an extra term $\frac{\rho_\infty^{1-n}}{\rho_\infty^{1-n}}$, we obtain:

$$q_c = \frac{c_{bl}}{4} \frac{c_{p,1}}{c_{p,\infty}} \left(\frac{\rho_1}{\rho_\infty}\right)^{1-n} \left(\frac{u_x}{u_\infty}\right)^{1-n} \frac{1}{\left(\frac{x}{R_N}\right)^n} \frac{\mu_\infty^n}{\left(2c_{p,\infty} T_\infty\right)^{\frac{n}{2}}} \frac{\rho_\infty^{1-n} u_\infty^3}{R_N^n} \left(1 - \frac{T_w}{T_{aw}}\right) \tag{3.297}$$

In the region of the stagnation point, it can be assumed for the velocity u_x:

$$\frac{u_x}{u_\infty} = c_1 \frac{x}{R_N} \tag{3.298}$$

with c_1 a constant. For hypersonic Mach numbers, Eq. (3.185) provides

$$\frac{\rho_1}{\rho_\infty} \approx \frac{\gamma + 1}{\gamma - 1} \approx \text{ constant} \tag{3.299}$$

which means that

$$\left(\frac{\rho_1}{\rho_\infty}\right)^{1-n} \left(\frac{u_x}{u_\infty}\right)^{1-n} \frac{1}{\left(\frac{x}{R_N}\right)^n} = c_2 \left(\frac{x}{R_N}\right)^{1-2n} \tag{3.300}$$

and reduces Eq. (3.297) to

$$q_c = c_3 \frac{c_{p,1}}{c_{p,\infty}} \left(\frac{x}{R_N}\right)^{1-2n} \frac{\mu_\infty^n}{\left(2c_{p,\infty} T_\infty\right)^{\frac{n}{2}}} \frac{\rho_\infty^{1-n} u_\infty^3}{R_N^n} \left(1 - \frac{T_w}{T_{aw}}\right) \tag{3.301}$$

If we assume $\frac{c_{p,1}}{c_{p,\infty}}$ to be constant and an isothermal atmosphere (T_∞, $c_{p,\infty}$ and μ_∞ are constant), the above equation reduces to

$$q_c = c_4 \left(\frac{x}{R_N}\right)^{1-2n} \frac{\rho_\infty^{1-n} u_\infty^3}{R_N^n} \left(1 - \frac{T_w}{T_{aw}}\right) \tag{3.302}$$

For a laminar boundary layer, $n = 0.5$, so the term $\frac{x}{R_N}$ in the above equation disappears, which means that the heat flux q_c near the stagnation point is independent of the distance x. For this heat flux it follows that:

$$q_{c,lam} = \frac{c_{lam}^*}{\sqrt{R_N}} \sqrt{\rho_\infty} u_\infty^3 \left(1 - \frac{T_w}{T_{aw}}\right) \tag{3.303}$$

At some distance from the stagnation point, the laminar boundary layer can become turbulent, which means that $n = 0.2$. The maximum heat flux then occurs at the location of the sonic line in the flow field, behind the bow shock wave (compare this to the throat of a rocket engine). For hypersonic flow, the location of the sonic line varies only slightly with Mach number, M_∞. Therefore, in Eq. (3.302) x/R_N can be assumed constant for the point of maximum heat flux. It is often assumed that $\varphi = x/R_N \approx 40°$ (Hayes and Probstein 1964); see Fig. 3.41 for the definition of φ.

With this, it follows for the turbulent boundary layer:

$$q_{c,turb} = \frac{c^*_{turb}}{R_N^{0.2}} \rho_\infty^{0.8} u_\infty^3 \left(1 - \frac{T_w}{T_{aw}}\right) \tag{3.304}$$

Equations (3.303) and (3.304) can only be solved when information about T_w is available. It is common to assume radiation equilibrium, which we will consider later in this section.

In case a so-called "cold wall" is considered, this means that $T_w \ll T_{aw}$ and can thus be neglected. For that situation it follows that

$$q_{c,lam} = \frac{c^*_{lam}}{\sqrt{R_N}} \sqrt{\rho_\infty} u_\infty^3 \qquad q_{c,turb} = \frac{c^*_{turb}}{R_N^{0.2}} \rho_\infty^{0.8} u_\infty^3 \tag{3.305}$$

The above derivation for the heat flux through convection, is, of course, only a rough approximation. The found expressions are used in most literature for vehicles re-entering the atmosphere (where more often than not the cold-wall approximation is applied), and they prove suitably useful, in comparison with experimental results. Inspecting the equations also illustrates that the vehicles benefit from blunt noses: a larger R_N limits the heat flux at hypersonic velocities.

Values for the constants c^*_{lam} and c^*_{turb} can be found in literature, but it is noted that different values can be found throughout, and that also the formulation of Eqs. (3.303) and (3.304) is sometimes slightly different (as we will also see in Chap. 5). From the book by Hankey (1988) it can be deduced that $c^*_{lam} \approx 18$ and $c^*_{turb} \approx 115$ in the case where the units used are W/cm^2 for q_c, kg/m^3 for ρ_∞, m for R_N and km/s for u_∞.

It can easily be verified that eventually $q_{c,turb} > q_{c,lam}$, if $\rho_\infty R_N > 0.0021$ kg/m^2. For $R_N = 0.5$ m, this implies that the turbulent heat transfer is greater than the laminar heat transfer for altitudes below approximately 40 km (standard atmosphere), assuming that the boundary layer is turbulent in the sonic point.

The above discussion is mainly related to convective heating during an Earth entry. In the following Intermezzo, a more general formulation is presented that takes the atmospheric composition into account. However, it is stressed that all models discussed are engineering models, to be used to get a first impression of the thermal loads. To obtain more accurate input for, for instance, structural design and analysis, one has to resort to numerical models. These are, unfortunately, beyond the scope of the current text.

Table 3.4 Molecular properties (Svehla 1962)

Molecule	M (g/mol)	σ_p (Å)	ϵ/σ (K)
H	1.008	2.708	37.0
N	14.008	3.298	71.4
O	16.000	3.050	106.7
H_2	2.016	2.827	59.7
N_2	28.020	3.798	71.4
O_2	32.000	3.467	106.7
CO	28.010	3.690	91.7
CO_2	44.010	3.941	195.2
Ar	39.944	3.542	93.3
He	4.003	2.551	10.22
H_2O	18.020	2.641	809.1

$1\ \text{Å} = 10^{-10}$ m

*** Intermezzo 3.5 (General Stagnation-Point Formulation)**
To account for different atmosphere compositions, one could consider to use the general stagnation-point convective-heating equation proposed by Sutton and Graves (1971). They suggest the expression for the convective heating to be

$$q_c = K\sqrt{\frac{p_s}{R_N}}\,(h_s - h_w) \tag{3.306}$$

with q_c being the convective heat flux in MW/m^2, p_s is the stagnation pressure in bars, R_N is the nose radius, h_s is the total enthalpy, and h_w is the wall enthalpy. The constant K is the heat-transfer coefficient in kg/s $\cdot$ m$^{-3/2}\cdot$ atm$^{-1/2}$, and is a function of the atmospheric composition, consisting of n_s species:

$$K = \frac{0.0885}{Pr^{0.6}}\left(\sum_{i=1}^{n_s}\frac{c_i}{M_i\gamma_i}\right)^{-1/2} \tag{3.307}$$

Here, Pr is the Prandtl number, c_i is the mass fraction of species i, γ_i is the transport parameter, and M_i is the molecular mass. The transport parameter γ in the above equation is a function of the collision diameter σ_p and the ratio of the maximum energy of attraction between colliding molecules over the Boltzmann constant, ϵ/σ:

$$\gamma_i = \frac{1}{\sigma_{p,i}^2\left(\frac{\epsilon_i}{\sigma}\right)^{0.15}} \tag{3.308}$$

In Table 3.4, the molecular mass M, the collision diameter σ_p and the ratio ϵ/σ are shown for the molecules that are most significantly present in the atmospheres of several planets. Data for other planets can be found in, for example, the report by

Svehla (1962) on molecular properties, combined with other, more recent, literature on atmospheric composition.

Note that Eq. (3.306) uses the stagnation pressure and enthalpy to calculate the convective heat flux. It is more common, though, to express the heat flux as a function of density and velocity, as is done by, for example, Casoliva et al. (2008), when simulating an aero-gravity manoeuvre around Mars. It is also the common approach for entry studies in the Earth atmosphere, hence it will be the preferred formulation in this text.

■

3.6.2.2 Local Heat Flow

For an arbitrary point on the wall of the vehicle, we may write:

$$q_{c,x} = \alpha(T_{aw} - T_w) = St_x \rho_x u_x c_{p,x} \frac{u_\infty^2}{2c_{p,\infty}} \tag{3.309}$$

where Eq. (3.291) has been substituted, together with the assumption that $T_w \ll T_{aw}$. The (local) Stanton number is yet again related to the local friction coefficient through the Reynolds analogy: $St_x = \frac{c_{f,x}}{2}$. Let us now define the heat supply to the vehicle per unit of time as follows:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \dot{Q} = \int_{S_w} q_{c,x} \mathrm{d}S \tag{3.310}$$

where S_w is the so-called wetted surface area, and $\mathrm{d}S$ being an element thereof. With Eq. (3.309) substituted, we write

$$\dot{Q} = \int_{S_w} St_x \rho_x u_x c_{p,x} \frac{u_\infty^2}{2c_{p,\infty}} \mathrm{d}S \tag{3.311}$$

For this particular heat flux, an average friction coefficient, C_F, is introduced:

$$C_F = \frac{1}{S_w} \int_{S_w} St_x \frac{\rho_x}{\rho_\infty} \frac{u_x}{u_\infty} \frac{c_{p,x}}{c_{p,\infty}} \mathrm{d}S \tag{3.312}$$

such that

$$\dot{Q} = \frac{C_F}{4} \rho_\infty u_\infty^3 S_w \tag{3.313}$$

To evaluate the integral of Eq. (3.312), the free-stream properties and the skin-friction coefficient have to be known at all points of the wetted area. Since these

elements depend mainly on the Mach number, the angle of attack, and the Reynolds number, C_F also depends on these parameters. It was found by Allen and Eggers (1957) and Eggers et al. (1957) that for ballistic missiles and skip vehicles entering the atmosphere at zero or constant angle of attack, the variation in C_F along the flight trajectory is very small, and can be neglected in conceptual design phases. So, for an engineering approach Eq. (3.313) can be used with constant C_F.

3.6.2.3 Heat-Flux Indicator

For aeromanoeuvring or aerobraking missions, where a major part of the atmospheric flight takes place in either the free-molecular or the transition flow-regime, a potential heat-flux *indicator* is typically used (Chambre and Schaaf 2017; McRonald and Nock 1977), which gives a good indication of the actual heat flux occurring in the system. More accurate predictions would require a thermal analysis in combination with free-molecular flow methods, the latter of which has been discussed in Sect. 3.5.

The heat-flux indicator can be derived as follows, starting with Eq. (3.287), and following certain simplifications that were introduced earlier. Here, the related steps will be repeated for the sake of clarity. Equation (3.287) reads:

$$q_{ab} = St \rho_e u_e (h_{aw} - h_w) \tag{3.314}$$

where the subscript "e" indicates the conditions at the edge of the boundary layer, "aw" stands for adiabatic wall, and the the subscript "w" indicates the spacecraft wall. The notation q_{ab} for *aerobraking* will be used to distinguish the heat flux from the "conventional" convective heat flux discussed earlier. The adiabatic wall enthalpy, h_{aw}, is related to the enthalpy at the edge of the boundary layer, h_e, and total enthalpy, h_0, as:

$$h_{aw} = h_e + r\,(h_0 - h_e) \tag{3.315}$$

with r being the so-called recovery factor, i.e., a measure of how well the free-stream total temperature is reached.

The first simplification that we apply is that the body is assumed to be a flat plate to which the incoming flow is parallel. This means that the conditions at the edge of the boundary layer are the same as the free-stream conditions. Thus, the term $\rho_e u_e$ can be replaced with $\rho_\infty u_\infty$:

$$q_{ab} = St \rho_\infty u_\infty (h_{aw} - h_w) \tag{3.316}$$

Second, a unitary recovery factor is assumed, i.e., $r \approx 1$. As a consequence, $h_{aw} \approx h_0$, which means that no change in enthalpy is experienced through the shock wave and boundary layer:

$$q_{ab} = St \rho_\infty u_\infty (h_0 - h_w) \tag{3.317}$$

By definition, the total enthalpy is given by:

$$h_0 = h_\infty + \frac{u_\infty^2}{2} \tag{3.318}$$

With a third assumption that $u_\infty^2 \gg h_\infty$ under hypersonic conditions, this definition of total enthalpy changes to

$$h_0 \approx \frac{u_\infty^2}{2} \tag{3.319}$$

and so

$$q_{ab} = St\,\rho_\infty u_\infty (h_0 - h_w) \tag{3.320}$$

The fourth assumption concerns a unitary Stanton number, i.e., $St \approx 1$. This means that the heat flux at the wall is fully absorbed by the vehicle. From free-molecular flow analyses done for the aerobraking of Mars Reconnaissance Orbiter, it was found that the average heat transfer was less than one, although its value does locally reach unity at the edges of its solar panels (Gallis et al. 2014; Liechty 2007).[27] Thus, using this value represents a more conservative scenario that provides some design margin.

The fifth and final assumption is the earlier introduced cold-wall approximation, i.e., $h_0 \gg h_w$. This approximation implies that the temperature of the wall is much lower than the temperature of the surrounding gas, providing the following simplification:

$$(h_{aw} - h_w) \approx h_0 - h_w \approx h_0 \left(\approx \frac{u_\infty^2}{2} \right) \tag{3.321}$$

With this result, Eq. (3.320) changes to

$$q_{ab} = St\,\rho_\infty u_\infty (h_0 - h_w) \approx St \frac{1}{2} \rho_\infty u_\infty^3 \tag{3.322}$$

or, with $St = 1$ substituted:

$$q_{ab} \approx \frac{1}{2} \rho_\infty u_\infty^3 \tag{3.323}$$

where q_{ab} keeps the unit of W/m^2, similar to Eqs. (3.303) and (3.304). This is easily found from unit analysis: 1 W/m^2 = kg m^2/s^3 per m^2 = kg/s^3. The unit of the right-hand side of Eq. (3.323) is kg/m^3 m^3/s^3 = kg/s^3.

[27] It is noted that in these references, one mentions the so-called heat-transfer coefficient, defined as $c_h = 2q/\rho u^3$, which is, in fact, the same as St in the current derivation.

Given the many simplifications above, one may think that the model from Eq. (3.323) will not perform well in a real-world scenario. It was proven, however, during the analysis of several aerobraking missions, that this indicator does indeed do a sufficient job at representing the heating of the spacecraft, and as such, its use is justified.

3.6.2.4 Integrated Heat Load

The integrated heat load is, as it names implies, a simple integration of either Eqs. (3.303), (3.304), (3.313), or (3.323). Note that the first two are point heat fluxes per unit of time and unit of surface, whereas the third one is a surface heat flow per unit of time, and the last one is similar to the first two. Therefore, for the stagnation-point heat fluxes the unit of the integrated heat load is J/m^2, whereas the integrated surface heat flow is simply in J.

Summarised:

$$Q_{lam} = \int_0^{t_f} q_{c,lam} \mathrm{d}t \quad Q_{turb} = \int_0^{t_f} q_{c,turb} \mathrm{d}t \quad Q_{loc} = \int_0^{t_f} \dot{Q} \mathrm{d}t \quad Q_{ab} = \int_0^{t_f} q_{ab} \mathrm{d}t \tag{3.324}$$

As a last remark for this subsection, it should be kept in mind that for an Apollo-like capsule flying at an angle of attack, as is usually the case, the point with maximum heat is usually not the stagnation point, but the *hot corner*, which is either the part of the edge of the heat-shield that is first impinged by the flow, or it has a relatively small corner radius that will also increase the heat flux. And last but not least, the nose radius should sometimes be substituted by an effective nose radius, if the heat-shield radius is not constant over its diameter and thus becomes a function of the angle of attack.

3.6.3 Radiation

Any body with a temperature T larger than absolute zero radiates heat, as formulated by Stefan-Boltzmann's law:

$$q_{rad} = \varepsilon \sigma T^4 \tag{3.325}$$

where σ is the Stephan-Boltzmann constant, $\sigma = 5.670367 \cdot 10^{-8}$ W/(m^2 K^4), and ε is the wall emissivity. The emissivity defines how closely the radiation of a body/medium resembles that of a black body. The higher the value (with an absolute maximum of 1 for a black body), the more the body radiates the heat outwards. The emissivity of a body is equal to its absorption, and in reality is a function of the wavelength. For wavelengths corresponding to the average temperature of a Lunar-return

entry, the Apollo Command Module had an emissivity of $\varepsilon \approx 0.8$ (Robinson et al. 2009). For a metallic TPS, one strives for values of $\varepsilon \approx 0.85$ or higher.

Normally, the vehicle will have a low wall temperature, T_w, upon entering the atmosphere. Because of aerodynamic heating, the wall temperature will increase during the atmospheric flight. Especially with a sufficiently long flight duration, which is the case for glide vehicles, T_w can reach such high values that the radiation heat emission of the vehicle to its surroundings should not be ignored. The highest wall temperature that the vehicle can thereby reach, occurs when the radiated heat, q_{rad}, equals the heat supplied through convection, q_c. This wall temperature is called the vehicle's equilibrium temperature, $T_{w,eq}$. In general it holds that

$$q_c = St\rho_\infty u_\infty c_{p,\infty}(T_{aw} - T_{w,eq}) = \varepsilon\sigma T_{w,eq}^4 = q_{rad} \tag{3.326}$$

For a "cold wall", $T_{aw} \gg T_{w,eq}$, the above equation is easily solved, i.e.,

$$T_{w,eq} = \sqrt[4]{\frac{St\rho_\infty u_\infty c_{p,\infty} T_{aw}}{\varepsilon\sigma}} \tag{3.327}$$

However, this would lead to a conservative, i.e., too high, equilibrium temperature. The difference with the "hot-wall" model will be around 10%. To use the hot-wall model instead is a bit more involved, as $T_{w,eq}$ appears on both sides of Eq. (3.326). Therefore, an iterative procedure is required. As an example, consider Eq. (3.303) for the heat flux in a laminar boundary layer (which is, of course, equal to Eq. (3.326)):

$$\frac{c_{lam}^*}{\sqrt{R_N}}\sqrt{\rho_\infty}u_\infty^3\left(1 - \frac{T_{w,eq}}{T_{aw}}\right) = \varepsilon\sigma T_{w,eq}^4 \tag{3.328}$$

In a step k, assume a value for $T_{w,eq}$ $(= T_{w,k})$, and evaluate the left-hand side of Eq. (3.328), which yields $q_{c,k}$. From the equilibrium condition

$$q_{c,k} = \varepsilon\sigma T_{w,eq}^4 \tag{3.329}$$

this gives us the new $T_{w,eq}$, which is equal to the updated temperature, $T_{w,k+1}$, in the iteration process. Use this new temperature value to evaluate q_c again to provide us with $q_{c,k+1}$, etc., until convergence. Usually, this procedure converges quickly, with at most up to five steps, and produces both q_c and $T_{w,eq}$.

So far, we considered the situation that only the vehicle radiates heat. However, at hypersonic speeds, the gas behind the bow shock wave reaches an extremely high temperature on the order of 6,000 K and will also radiate heat. Fortunately, the gas does not radiate like a black body, but has a rather low value for the emissivity coefficient. As a consequence, and to the vehicle's advantage, only a smaller fraction of the heat will be transferred to the vehicle wall.

The peak heating rate as a function of entry velocity for several missions is shown in Fig. 3.83. The peak non-ablative heating rate, i.e., the effect of ablation of the heat shield is excluded from these data, ranges from 300 kW/m^2 for the Viking mission

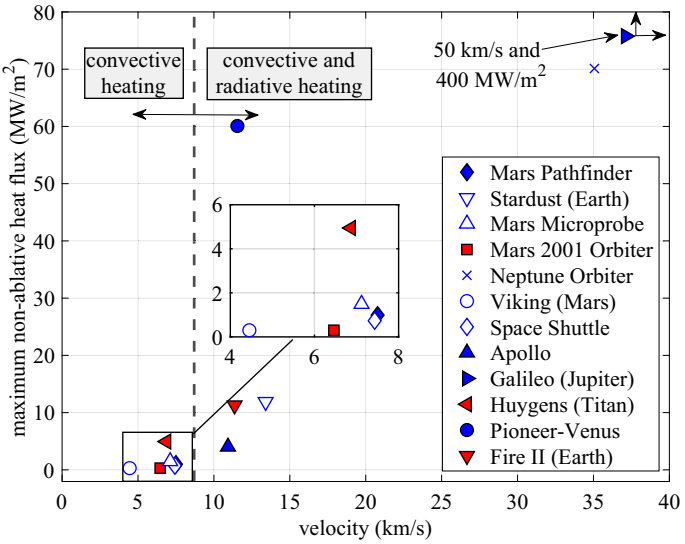


Fig. 3.83 Peak heating as a function of entry velocity for several entry vehicles. Data taken from Gnoffo et al. (1999)

(entry velocity about 4.5 km/s) to 400 MW/m^2 at 50 km/s for the Galileo Jupiter probe. It is also indicated when the heating is predominantly convective (below 9 km/s) and when the radiation becomes important. Of course, the larger the velocity the more effect radiation will have. For Galileo, for instance, 99% of the heating is from radiation. The heat shield of most vehicles indicated in this plot is an ablative one (more details about this in Chap. 10). For instance, for Viking a silicon based ablator was used (SLA-561), while carbon-phenolic was used on Pioneer-Venus and Galileo. The Phenolic Impregnated Carbon Ablator (PICA) family of materials was developed for Stardust, and was later also used for Mars Science Laboratory and Mars 2020, amongst others. Finally, a non-ablating Beryllium heat shield was used for Fire II, and non-ablating, silica-based tiles were used on the Space Shuttle Orbiter.

Yoshikawa and Wick (1961) analysed the radiative heating for the gas around a hemispheric nose in hypersonic flow at velocities below 10 km/s. For thermal-equilibrium conditions, in the neighbourhood of the stagnation point the radiative heat flux could be approximated by

$$q_{rad,gas} = C R_N \rho_\infty^{1.6} u_\infty^{8.5} \tag{3.330}$$

where $C = 1.23 \cdot 10^{-25}$. They concluded that for entry from low-Earth orbit radiant heating is negligible with respect to convective heating.

However, for a return flight from the Moon, aerocapture missions returning samples from asteroids or in a future mission from Mars, the arrival speed at Earth is a lot larger than the orbital speed of satellites. Such a speed is called super orbital, and the

Table 3.5 Radiative heating speed functions for Earth (E) and Mars (M) (Tauber and Sutton 1991)

Earth		Mars	
V (m/s)	$f_E(u_\infty)$	V (m/s)	$f_M(u_\infty)$
9,000	1.5	6000	0.2
9,250	4.3	6150	1.0
9,500	9.7	6300	1.95
9,750	19.5	6500	3.42
10,000	35	6700	5.1
10,250	55	6900	7.1
10,500	81	7000	8.1
10,750	115	7200	10.2
11,000	151	7400	12.5
11,500	238	7600	14.8
12,000	359	7800	17.1
12,500	495	8000	19.2
13,000	660	8200	21.4
13,500	850	8400	24.1
14,000	1065	8600	26.0
14,500	1313	8800	28.9
15,000	1550	9000	32.8
15,500	1780	–	–
16,000	2040	–	–

interaction with the surrounding atmosphere is intensified. Although for applications of direct entry the duration of the high-velocity flight is short, for an aerobraking mission the flight path is much shallower and the duration of flight through the atmosphere at super-orbital speed may be much longer.

At these super-orbital speeds also radiative heating from the gas becomes relevant. Tauber and Sutton (1991) proposed a simple analytical expression, comparable to Eq. (3.330), for the radiative heating in the stagnation point of a blunt body, assuming thermochemical equilibrium. The heating depends on the gas mixture, and thus on the planet of entry. It becomes almost zero for speeds lower than 9 km/s on Earth (and lower than 6 km/s on Mars). The radiative heating from the surrounding gas, $q_{rad,gas}$, in W/m^2 is given by:

$$q_{rad,gas} = C\, R_N^{a_1}\, \rho^{a_2}\, f(u_\infty) \tag{3.331}$$

where C is a constant depending on the gas mixture of the atmosphere, and $f(u_\infty)$ is an interpolated function, which depends on the atmospheric gas mixture. Its tabulated values for both Earth (air) and Mars (97% CO_2 and 3% N_2) can be found in the paper by Tauber and Sutton (1991), and have been reproduced in Table 3.5 (for speeds

larger than the maximum value this limit value shall be used). Values of a_1 and a_2 depend on the planet, but may as well be a function of nose radius or velocity.

The related coefficients for Earth (based on the composition of air) are:

$$\begin{aligned} C &= 4.736 \times 10^8 \\ a_1 &= 1.072 \times 10^6 \, V^{-1.88} \, \rho^{-0.325} \\ \text{if} \quad 1\,\text{m} \leq R_N \leq 2\,\text{m}, &\qquad a_1 \leq 0.6 \\ \text{if} \quad 2\,\text{m} < R_N \leq 3\,\text{m}, &\qquad a_1 \leq 0.5 \\ a_2 &= 1.22 \end{aligned} \tag{3.332}$$

The above values are only applicable in the speed range from 10 to 16 km/s and for free-stream densities varying between $6.66 \cdot 10^{-5}$ kg/m^3 and $6.31 \cdot 10^{-4}$ kg/m^3. This density variation corresponds to an altitude range of about 72 to 54 km, respectively. Finally, the nose radii can vary from 0.3 to 3 m. With values beyond the given ranges the validity decreases, and should only be used as indicative and with great care. In any case, the condition of $a_1 < 1$ must always be met.

The numerical values for the constants to be used in the Martian atmosphere are

$$\begin{aligned} C &= 2.35 \times 10^8 \\ a_1 &= 0.526 \\ a_2 &= 1.19 \end{aligned} \tag{3.333}$$

for an applicable speed range from about 6.5 to 9 km/s and for free-stream densities ranging from 10^{-4} kg/m^3 to 10^{-3} kg/m^3. These densities correspond to altitudes that vary between 51 down to 30 km, respectively. Within the acceptable ranges, for Earth Eq. (3.331) gives an error between -18% and $+20\%$ for the situations in which it was compared to more accurate numerical methods (with a minimum deviation of 8% at the 66 km altitude marker). In case of a Mars entry, the errors vary between -17% and $+12\%$.

For the radiative heat flux for Venusian entry, a method developed by Page and Woodward (1972) in the context of the Pioneer-Venus programme, and applied by, for instance, Craig and Lyne (2002), can be used. This method provides a single-base equation that is to be evaluated using different coefficients. The value of the coefficients depends on the speed. The base equation proposed is

$$q_{rad,gas} = C \, R_N^{0.49} \rho^{1.2} V^M \tag{3.334}$$

where the constant C and the exponent M are dependent on the current relative speed. Values for these parameters have been tabulated in Table 3.6 for different speed ranges. For a speed above 12 km/s, $q_{rad,gas}$ is approximated by the limit value of the equation at 12 km/s. Similarly, as mentioned earlier, for a speed above 16 km/s for Earth and 9 km/s for Mars, their corresponding radiative heat flux is approximated by using $f_E(16)$ and $f_M(9)$, respectively.

An alternative formulation is provided by Tauber et al. (2012), who fitted the equations to 10 values of more precise numerical simulations obtained from several

Table 3.6 Radiative heating functions for Venus entry (Craig and Lyne 2002)

Velocity range (km/s)	Constant C	Exponent M
≤ 8	$3.33 \cdot 10^{-34}$	10.0
8–10	$1.22 \cdot 10^{-16}$	5.5
10–12	$3.07 \cdot 10^{-34}$	1.2

sources, in combination with shock layer pressure and thickness dependence from Page and Woodward (1972). For speeds smaller than 10,028 m/s, the radiative heat flux is given by:

$$q_{rad,gas} = 9.497 \cdot 10^{-63} V^{18} \rho^{1.2} R_N^{0.49} \tag{3.335}$$

whereas for velocities larger than 10,028 m/s and smaller than 12,000 m/s it is defined by:

$$q_{rad,gas} = 2.195 \cdot 10^{22} V^{7.9} \rho^{1.2} R_N^{0.49} \tag{3.336}$$

Finally, to have an engineering equation for entering the atmosphere of the gas giants, Jupiter, Saturn, Uranus and Neptune, this section concludes by providing the radiative heat-transfer model for Jupiter, developed in the framework of the Galileo entry probe. For the other gas giants, not much information is available. Tauber et al. (1999) studied the entry in the Jovian atmosphere, and developed a code to support the design of the entry probe's ablative heat shield. They used a fast, tailor-made numerical codes for the calculation of the radiative heating, and validated the results with flight data from the Galileo probe. Although it seems reasonable that an engineering-level equation for $q_{rad,gas}$ in any of the gas giants' atmosphere will have the form of Eq. (3.331), the corresponding constants and tabulated values were not provided.

However, also Ritter et al. (2006) did a study for a Jovian entry probe, and produced an empirical relation for $q_{rad,gas}$ from several, yet unspecified data sets. The related equation for $q_{rad,gas}$ in W/m^2 is

$$q_{rad,gas} = 9.7632379^{-40} (2R_N)^{-0.17905} (\rho)^{1.763827469} (V)^{10.993852} \tag{3.337}$$

For the above equation, it was assumed that the Jovian atmosphere consists of 89% H_2 and 11% He. Similar to Tauber et al. (1999) also Ritter et al. (2006) used in-flight measurements from the Galileo entry probe in their validation process of this equation.

For hot-wall equilibrium, the radiative heat flux emitted from the atmosphere has to be added to the convective heat flux and again equated to the radiative heat flux emitted by the wall itself, so the equilibrium equation Eq. (3.326) changes to

$$q_{c,cold}\left(1-\frac{T_{w,eq}}{T_{w,ad}}\right)+q_{rad,gas}=\varepsilon\sigma T_{w,eq}^4 \tag{3.338}$$

Note that in theory, also the incoming radiative heat flux estimated by Eq. (3.331) should be multiplied by an emissivity/absorption coefficient, similar to the radiative heat flux emitted from the wall. This is not trivial, as this radiation occurs at a variety of wavelengths. To be conservative, in the above equation on the left-hand side the corresponding emissivity/absorption has been set equal to 1 (= black body).

Example 3.7 (*Heat Load for a Lunar Direct Return Flight*) Robinson et al. (2009) discuss the entry trajectory and aeroheating environment for capsule-shaped vehicles, in the context of the development of the Crew Exploration Vehicle (CEV) to return humans from lunar missions. In this example, we will present the heat load acting on the capsule for one of their test cases. This particular test case assumes an Apollo-shaped capsule, with a(n adapted) ballistic coefficient of 122 kg/m^2; note that for the same aerodynamics the original Apollo capsule had a ballistic coefficient of around $K = 356$ kg/m^2 ($m = 5{,}000$ kg, $S_{ref} = 12.017$ m^2 and $C_D = 1.2891$ at $M = 30$). For this new ballistic coefficient, keeping the aerodynamic properties the same, the vehicle mass decreases to $m = 1{,}890$ kg.

In anticipation of the discussion in Chap. 11, an aerocapture mission is defined here, representing a direct entry from a lunar mission. The entry velocity at the interface of $h = 120$ km is $V_E = 12{,}201$ m/s. The aerocapture is executed in a so-called full lift-down orientation, to maximise the pull into the atmosphere. This means, however, that the entry path cannot be too steep, otherwise the vehicle will not skip out of the atmosphere again and simply crash on the surface. The initial entry angle that will just result in the skip out, was found by trial and error, and is $\gamma_E = -5.282°$. The simulation is performed for a rotating Earth, with the US76 atmosphere model and a central gravity field. Focus in this example is only on the encountered thermal load; the aerocapture manoeuvre itself will be discussed at length in Sect. 11.2.

Figure 3.84 shows the experienced heat flux, separated in a (cold-wall) convective component and a component due to radiative heating of the atmospheric gas. Looking at the fluxes, the maximum convective flux is about $q_c \approx 930$ kW/m^2. However, not only should the radiative flux be taken into account, it is even larger than q_c: the peak value reaches $q_{rad,gas} \approx 1.6$ MW/m^2, about 75% more! The peak width is smaller, though, compared to that of q_c, because with the decreasing speed the effect of $q_{rad,gas}$ rapidly diminishes. It is noted that the radiative peak occurs a few seconds (about 5 s) earlier than the convective peak.

Looking at the right plot, the equilibrium wall temperature is $T_{w,eq} \approx 2{,}130$ K when only convection is considered (assuming an emissivity of $\varepsilon = 0.8$). However, including the radiative component increases the temperature to $T_{w,eq} \approx 2{,}741$ K, an increase close to 30%. It thus shows how important it is to include radiation for high-speed entry flight, as the thermal-protection system needs to cope with a lot more heat.

■

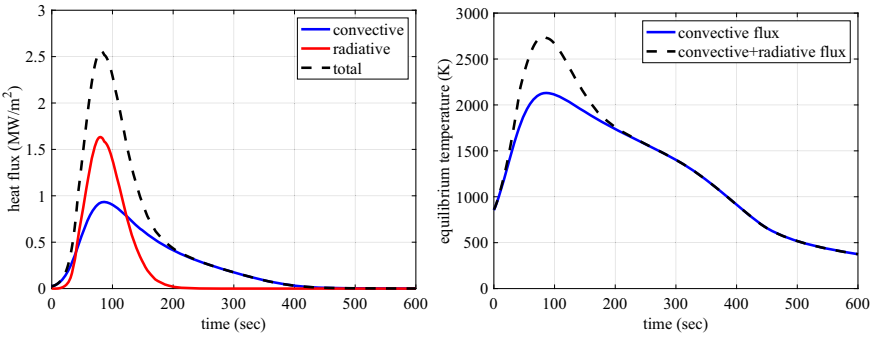


Fig. 3.84 Thermal loads during a full lift-down aerocapture: fluxes (left) and equilibrium temperature (right)

3.6.4 Heat Balance

So far, we discussed convective and radiative heat fluxes. The last component of the heat balance to consider is linked to the vehicle's structural design, i.e., the type of thermal protection system and the underlying structure (see Fig. 3.85). In case the vehicle has an ablative heat shield the ablation process results in heat loss, as it should, of course. Alternative, active cooling systems, transport heat away from the wall by the cooling liquid. In case of a hot-structure TPS design, conduction through the system may distribute the heat loading. Also, some conduction to the structural layers under the TPS should be considered. The complete picture is a very complex one, and cannot be solved without a detailed structural design. Chapter 10 discusses thermal-protection systems in more detail.

The heat balance to be solved is

$$q_c + q_{rad,gas} = q_{rad} + q_{loss} + q_{cond} \tag{3.339}$$

For conceptual analysis, it usually suffices to include the convective and radiative heat flux, and when the velocities are large, also the radiative heat flux from the gas.

3.7 *Navier-Stokes Equations

The section on hypersonic aerodynamics is concluded by a summary of the full Navier-Stokes equations that describe the motion of viscous fluids. It is a combination of Newton's second law applied to fluid motion (a pressure term), together with the assumption that the stress in the fluid is the sum of a diffusing viscous term (proportional to the gradient of velocity). The NS-equations were independently

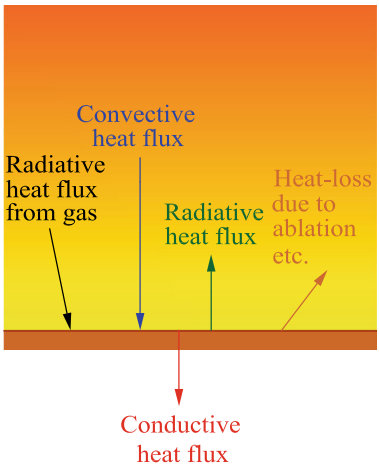


Fig. 3.85 Heat balance for vehicle wall

derived by Claude-Louis Navier[28] and George Stokes.[29] The objective of this subsection is not to give the full derivation—the reader is referred to the literature for that, e.g., the books by White (2005) and Anderson (2011)—but to have the equations that fully describe the viscous flow in one place. Hopefully, by studying them one not only understands the relation with the shock equations that we derived earlier, but also realises the fundamental importance of Newton's second law. It is noted that in the presented form, the equations are applicable to a chemically non-reacting and non-radiating flow.

In the formulation of the Navier-Stokes equations the so-called *material derivative* is used. This material derivative is defined as the following non-linear operator:

$$\frac{\mathrm{D}}{\mathrm{D}t} \triangleq \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$$

and describes the time rate of change of some physical quantity (like heat or momentum) of a material element, which is subjected to a space-and-time-dependent macroscopic velocity-field variation of that physical quantity. For instance, for a quantity $y = y(t, \mathbf{x})$, the material derivative is $\frac{\mathrm{D}y}{\mathrm{D}t} \triangleq \frac{\partial y}{\partial t} + \mathbf{u} \cdot \nabla y$, which is also written as $\frac{\mathrm{D}y}{\mathrm{D}t} \triangleq \frac{\partial y}{\partial t} + \nabla \cdot (y\mathbf{u})$.[30]

For a general solution of a compressible, viscous flow, the required equations are derived from the conservation of mass, momentum, and energy. These fundamental principles were introduced in Sect. 3.3, where we discussed supersonic flow and

[28] Claude-Louis Navier (10 February 1785–21 August 1836), was a French engineer and physicist who specialised in mechanics.

[29] George Gabriel Stokes (13 August 1819–1 February 1903) was an Anglo-Irish physicist and mathematician.

[30] In contrast, $\frac{\partial f}{\partial x_i}$ indicates the partial derivative of functional $f(x_1, x_2, ...)$ with respect to variable x_i. Finally, $\frac{\mathrm{d}f}{\mathrm{d}x}$ is the "normal" derivative of $f(x)$.

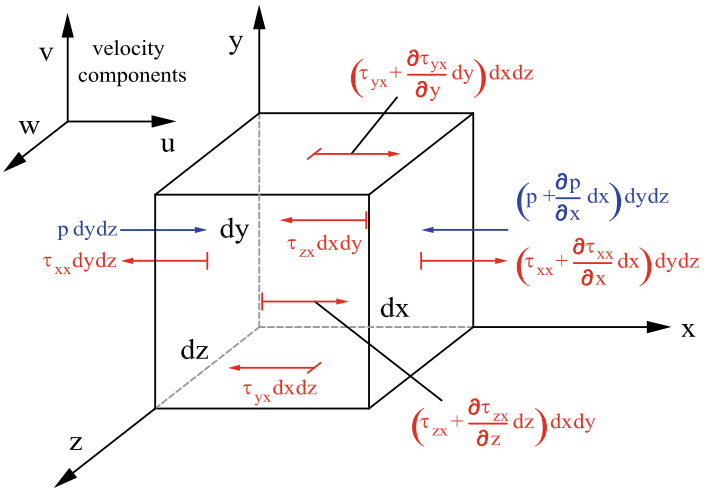


Fig. 3.86 Infinitesimally small, moving fluid element, with only the forces in x-direction shown

shock waves. To begin with the continuity equation, this is derived from conservation of mass. A unit volume is moving with velocity $\mathbf{v}$, and for conservation of mass the material derivative for the mass density per unit volume, ρ, should be zero:

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = \frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 \tag{3.340}$$

Note that for an incompressible flow, the density following the path of a flow element is constant. In that case, Eq. (3.340) reduces to $\nabla\cdot\mathbf{v} = 0$, which states that volume is conserved.

The actual Navier-Stokes equations are the momentum equations, taking into account the shear and normal stresses caused by viscous action on a fluid element. In Fig. 3.86, this element is shown, with all forces involved, but for clarity only in x-direction. This element has a Cartesian position $\mathbf{r} = (x, y, z)^{\mathrm{T}}$ and velocity $\mathbf{v} = (u, v, w)^{\mathrm{T}}$; the dimensions are dx, dy, and dz in its three dimensions. The external forces come from a combination of pressure, gravity, and viscous interaction. Relevant parameters are pressure, p, gravitational acceleration, g_i, and components of the stress tensor. The normal stresses are designated with σ_{ii} and the shear stresses with τ_{ij}. There are nine stress components: σ_{xx}, σ_{yy}, σ_{zz}, τ_{xy}, τ_{yx}, τ_{xz}, τ_{zx}, τ_{yz}, and τ_{zy}. If moments are taken about the x-axis, the y-axis, and the z-axis, respectively, the outcome would be that $\tau_{xy} = \tau_{yx}$, $\tau_{xz} = \tau_{zx}$, and $\tau_{yz} = \tau_{zy}$. Thus, there are six stress components left, which have to be related to the pressure and velocity components.

Derivation of the Navier-Stokes momentum equations follows from Newton's second law, where the viscous forces are expressed explicitly in terms of the relevant flow-field variables. White (2005) states the alternative formulation of Newton's second law as follows:

$$\rho \frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = \rho \mathbf{g} + \nabla \cdot \tau_{ij} \tag{3.341}$$

with the second component on the right-hand side the total vector surface force (pressure plus viscous stresses). The divergence of τ_{ij} should be interpreted in the tensor sense, which means that the result is a vector. To apply this, it remains only to express τ_{ij} in terms of the velocity, $\mathbf{v}$. Using Fig. 3.86 to evaluate the force equilibrium, and writing the end result in scalar form, the resulting equations are:

$$\rho \frac{\mathrm{D}u}{\mathrm{D}t} = \rho g_x - \frac{\partial p}{\partial x} + \frac{\partial}{\partial x}\overbrace{\left(2\mu \frac{\partial u}{\partial x} + \lambda \nabla \cdot \mathbf{v}\right)}^{\tau_{xx}} + \frac{\partial}{\partial y}\overbrace{\left[\mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\right]}^{\tau_{xy}} + \frac{\partial}{\partial z}\overbrace{\left[\mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)\right]}^{\tau_{xz}} \tag{3.342}$$

$$\rho \frac{\mathrm{D}v}{\mathrm{D}t} = \rho g_y - \frac{\partial p}{\partial y} + \frac{\partial}{\partial x}\overbrace{\left[\mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\right]}^{\tau_{yx}} + \frac{\partial}{\partial y}\overbrace{\left(2\mu \frac{\partial v}{\partial y} + \lambda \nabla \cdot \mathbf{v}\right)}^{\tau_{yy}} + \frac{\partial}{\partial z}\overbrace{\left[\mu\left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right)\right]}^{\tau_{yz}} \tag{3.343}$$

$$\rho \frac{\mathrm{D}w}{\mathrm{D}t} = \rho g_z - \frac{\partial p}{\partial z} + \frac{\partial}{\partial x}\overbrace{\left[\mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)\right]}^{\tau_{zx}} + \frac{\partial}{\partial y}\overbrace{\left[\mu\left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right)\right]}^{\tau_{zy}} + \frac{\partial}{\partial z}\overbrace{\left(2\mu \frac{\partial w}{\partial z} + \lambda \nabla \cdot \mathbf{v}\right)}^{\tau_{zz}} \tag{3.344}$$

where in addition to the already introduced parameters, μ is the dynamic viscosity and λ is the so-called volume or bulk viscosity (the volume viscosity becomes important only for such effects where fluid compressibility is essential, see for more details the paper by Sharma and Kumar (2019)). Note that we have kept the original notation for τ_{ij}, meaning that $\tau_{xx} = \sigma_{xx} = \sigma_x$, etc.

In vector form, the momentum equations are written as

$$\rho \frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = \rho \mathbf{g} - \nabla p + \frac{\partial}{\partial x_j}\left[\mu\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) + \delta_{ij}\lambda \nabla \cdot \mathbf{v}\right] \tag{3.345}$$

with δ_{ij} being the Kronecker delta function. This equation represents the complete NS-equations for an unsteady, compressible, three-dimensional viscous flow.[31] To solve for an incompressible viscous flow, this equation, together with Eq. (3.340), is sufficient. When the flow is compressible, the solution requires an additional equation, the energy equation.

The energy equation is derived from the first law of thermodynamics, and reads:

$$\rho \frac{\mathrm{D}h}{\mathrm{D}t} = \frac{\mathrm{D}p}{\mathrm{D}t} + \nabla \cdot (k \nabla T) + \Phi \tag{3.346}$$

[31] In the stated form, the NS-equations are suitable for the analysis of laminar flow. When the flow is turbulent, the flow variables can be assumed to be time-mean values over the turbulent fluctuations, and μ and k should be replaced by $\tilde{\mu} = \mu + \mu_T$ and $\tilde{k}k + k_T$, see, for instance, the book by Bradshaw et al. (1981). Here, μ_T and k_T are the eddy viscosity and eddy thermal conductivity, respectively.

where $h = e + \mathbf{v}^{\mathrm{T}}\mathbf{v}$ is the enthalpy (e is the internal energy, whereas the second term represents the kinetic energy), k is the thermal conductivity, T is the temperature (the temperature gradient, ∇T, represents the thermal conductivity), and Φ is the dissipation function, which for a Newtonian fluid is given by (see also Fig. 3.87):

$$\Phi = \mu \left[2\left(\frac{\partial u}{\partial x}\right)^2 + 2\left(\frac{\partial v}{\partial y}\right)^2 + 2\left(\frac{\partial w}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right)^2 + \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)^2 \right] + \lambda\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right)^2 \tag{3.347}$$

Note that an alternative formulation of Φ could involve the components τ_{ij}, as defined in Eqs. (3.342) through (3.344). Including the volumetric heating of the element in Fig. 3.87, i.e., $\rho q dx dy dz$, which means that q is the heat flux per unit mass, the final form of Eq. (3.346) is

$$\begin{aligned}\rho\frac{\mathrm{D}\left(e + \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v}\right)}{\mathrm{D}t} &= \rho q + \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) + \\ &-\nabla\cdot p\mathbf{v} + \frac{\partial(u\tau_{xx})}{\partial x} + \frac{\partial(u\tau_{yx})}{\partial y} + \frac{\partial(u\tau_{zx})}{\partial z} + \\ &+\frac{\partial(v\tau_{xy})}{\partial x} + \frac{\partial(v\tau_{yy})}{\partial y} + \frac{\partial(v\tau_{zy})}{\partial z} + \frac{\partial(w\tau_{xz})}{\partial x} + \frac{\partial(w\tau_{yz})}{\partial y} + \frac{\partial(w\tau_{zz})}{\partial z}\end{aligned} \tag{3.348}$$

Alternatively, the energy equation can be rewritten using the following thermodynamic identity, which is the relation for a change in enthalpy for a non-ideal gas (the second term vanishes for an ideal gas):

$$dh = c_p \mathrm{d}T + (1 - \beta T)\frac{\mathrm{d}p}{\rho} \tag{3.349}$$

resulting in

$$\rho c_p \frac{\mathrm{D}T}{\mathrm{D}t} = \rho q + \beta T \frac{\mathrm{D}p}{\mathrm{D}t} + \nabla\cdot(k\nabla T) + \Phi \tag{3.350}$$

where c_p is the specific heat for constant pressure and β is the thermal expansion coefficient.

There is no analytic solution to the general form of the Navier-Stokes equation and software incorporating this equation requires high computation power and time. For certain situations, there are several simplifications possible to allow for analysis and

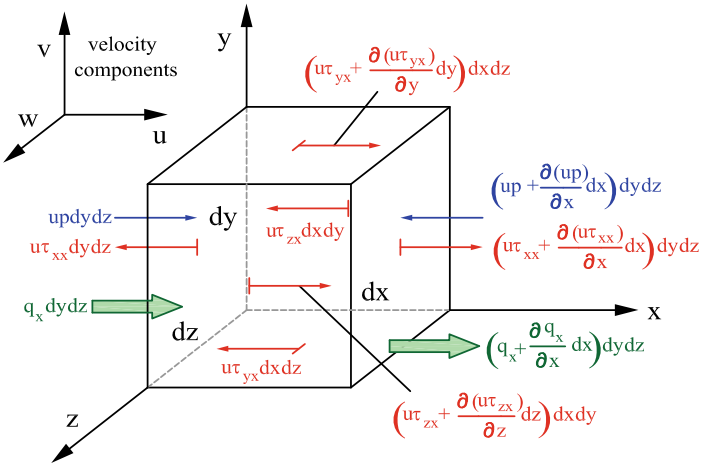


Fig. 3.87 Infinitesimally small, moving fluid element, with only the forces and energy flows in x-direction shown

(quicker) simulation, see below examples. Depending on the situation and boundary conditions, one or more additional equations is/are required to solve for all unknown variables. One such equation could be the ideal gas law.

Examples

3.1. *Steady versus unsteady*

An easy simplification to make is to assume steady flow, as an unsteady flow is only of interest for a non-equilibrium flow. During a disturbance and for a short instance after the disturbance, the flow will be unsteady, while it adjusts itself to a new equilibrium. For certain flight-dynamics applications, e.g., to study the vehicle's stability, a common approach is to use aerodynamic stability derivatives that only depend on the instantaneous state of the vehicle, so steady-flow conditions. The assumption of steady flow, i.e.,

$$\frac{\partial f}{\partial t} = 0, \quad \forall\, f(t) \tag{3.351}$$

reduces the equations to:

$$\nabla \cdot \mathbf{v} = 0 \tag{3.352}$$

$$\rho(\nabla \mathbf{v} \cdot \mathbf{v}) = \rho \mathbf{g} - \nabla \cdot p\mathbf{v} + \frac{\partial}{\partial x_j}\left[\mu\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right)\right] \tag{3.353}$$

$$\rho c_p(\nabla T \cdot \mathbf{v}) = \beta T(\nabla \cdot p\mathbf{v}) + \nabla \cdot (k \nabla T) + \Phi \tag{3.354}$$

where in Eq. (3.353), for the sake of compactness, we have used the notation x_i, x_j, v_i and v_j (with $i = 1, 2, 3$ and $j = 1, 2, 3$) to indicate x, y, z and u, v, w, respectively. Also, because of Eq. (3.352), the term $\delta_{ij}\lambda\nabla \cdot \mathbf{v}$ has dropped out of Eq. (3.353).

3.2. *Compressible versus incompressible*
When the flow is incompressible the density is constant, so $\frac{\partial \rho}{\partial t} = 0$, and also the viscosity can be assumed constant. Constant density reduces the continuity equation to Eq. (3.352), and, therefore, also $\delta_{ij}\lambda\nabla \cdot \mathbf{v} = 0$. The viscous terms in the momentum equations further reduce as follows, with as an example Eq. (3.342):

$$\begin{aligned}
&\frac{\partial}{\partial x}\left(2\mu\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left[\mu\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\right] + \frac{\partial}{\partial z}\left[\mu\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)\right] \\
&\quad = 2\mu\frac{\partial^2 u}{\partial x^2} + \mu\frac{\partial^2 u}{\partial y^2} + \mu\frac{\partial^2 v}{\partial y \partial x} + \mu\frac{\partial^2 u}{\partial z^2} + \mu\frac{\partial^2 w}{\partial z \partial x} \\
&\quad = \mu\frac{\partial^2 u}{\partial x^2} + \mu\frac{\partial^2 u}{\partial y^2} + \mu\frac{\partial^2 u}{\partial z^2} + \mu\frac{\partial^2 u}{\partial x^2} + \mu\frac{\partial^2 v}{\partial y \partial x} + \mu\frac{\partial^2 w}{\partial z \partial x} \\
&\quad = \mu\nabla^2 u + \mu\frac{\partial}{\partial x}\overbrace{\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right)}^{=0} = \mu\nabla^2 u
\end{aligned} \tag{3.355}$$

Equations (3.343) and (3.344) reduce in a similar way to $\mu\nabla^2 v$ and $\mu\nabla^2 w$. Further, since incompressible flow is limited to low velocity (and allows us to "linearise"), both ∇p and Φ are second-order terms and can thus be neglected. The resulting equations are:

$$\nabla \cdot \mathbf{v} = 0 \tag{3.356}$$

$$\rho(\nabla \mathbf{v} \cdot \mathbf{v}) = \rho\mathbf{g} - \nabla p + \mu\nabla^2 \mathbf{v} \tag{3.357}$$

$$\rho c_p(\nabla T \cdot \mathbf{v}) \approx \nabla \cdot (k\nabla T) \tag{3.358}$$

If the thermal conductivity is assumed to be constant, the energy equation can be further simplified to:

$$\rho c_p(\nabla T \cdot \mathbf{v}) \approx k\nabla^2 T \tag{3.359}$$

This is true when the Mach number is below 0.3, and as a result div $\mathbf{v}$ vanishes, as well as the bulk viscosity λ. This does not simplify the equations much, though:

$$\rho\frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = \rho\mathbf{g} - \nabla p + \frac{\partial}{\partial x_j}\left[\mu\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right)\right] \tag{3.360}$$

3.3. Viscous versus inviscid flow
Consider an inviscid, adiabatic (hence, isentropic) flow field. Inviscid flow means that $\mu = 0$ (and also $\Phi = 0$), whereas for an adiabatic flow $q = 0$. The mass continuity equation remains unchanged, whereas the momentum equation reduces to the Euler equation,[32] and also the energy equation simplifies. The resulting equations are:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{3.361}$$

$$\rho \frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = \rho \mathbf{g} - \nabla p \tag{3.362}$$

$$\rho c_p \frac{\mathrm{D}T}{\mathrm{D}t} = \beta T \frac{\mathrm{D}p}{\mathrm{D}t} + \nabla \cdot (k \nabla T) \tag{3.363}$$

3.4. Boundary-layer equations
The equations, which are generally used to represent a two-dimensional boundary layer, follow from the Navier-Stokes equations for steady flow, under the following approximations (Anderson 2019): (i) $\frac{\partial}{\partial t} = 0$, (ii) $\frac{\partial^2}{\partial x^2} = 0$, and (iii) $v \ll u$.

The continuity equation for the compressible boundary layer becomes:

$$\frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} = 0 \tag{3.364}$$

The momentum equation in x-direction is:

$$\rho u \frac{\partial u}{\partial x} + \rho v \frac{\partial u}{\partial x} = -\frac{\partial p}{\partial x} + \frac{\partial}{\partial y}\left(\mu \frac{\partial u}{\partial y}\right) \tag{3.365}$$

Because of the above approximations the momentum equation in y-direction reduces to:

$$\frac{\partial p}{\partial y} = 0 \tag{3.366}$$

This simplification is due to the fact that v and its change with x, $\frac{\partial v}{\partial x}$, is very small compared to u and $\frac{\partial u}{\partial x}$. It is emphasised here that for (very) large Mach numbers, the assumption that p is constant in the normal direction through a boundary layer is not always valid. For a proper treatment of the boundary layer, one should thus not make this assumption, and calculate the normal pressure gradient as part of the solution to the problem. This is discussed in quite some

[32] Often the combined mass continuity, momentum, and energy equations are called the Euler equations.

detail by Anderson (2019). However, for the purpose of a first analysis we assume this assumption to hold.
The energy equation is given by:

$$\rho u \frac{\partial e}{\partial x} + \rho v \frac{\partial e}{\partial y} = -p \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \mu \left(\frac{\partial u}{\partial y} \right)^2 + \frac{\partial q}{\partial y} \tag{3.367}$$

In the last equation, q is the heat flux per unit area. Equivalently—and more common in boundary-layer research—the latter equation is rewritten using the expression for enthalpy $h = e + \frac{p}{\rho}$:

$$\rho u \frac{\partial h}{\partial x} + \rho v \frac{\partial h}{\partial y} = u \frac{\partial p}{\partial x} + \mu \left(\frac{\partial u}{\partial y} \right)^2 + \frac{\partial q}{\partial y} \tag{3.368}$$

The heat flux per unit area is usually expressed by the Fourier equation of heat conduction:

$$q = k \frac{\partial T}{\partial y} \tag{3.369}$$

It is clear that a complete book can (and has been, at many occasions) be written about the Navier-Stokes Equations and their application in the field of computational fluid dynamics. That will not serve the purpose of this book, though. So, no matter how important a more extensive discussion is, unfortunately the current brief introduction has come to an end.

Problems

Judge whether the following statements are TRUE or FALSE (Problems 3.1–3.13).

3.1. In supersonic flight, the shock wave is not always attached to the nose.
3.2. The boundary layer thickness increases with an increase in Reynolds number.
3.3. In the rarefied regime, one will encounter both hypersonic and supersonic flow.
3.4. Generally, supersonic flow will only occur above a free-stream Mach of 1.
3.5. The Mach independence principle also applies to the pressure distribution on the vehicle.
3.6. A laminar boundary layer typically experiences higher skin friction compared to a turbulent boundary layer.
3.7. Hypersonic flow is primarily governed by compressibility effects, with negligible influence from viscosity.
3.8. The concept of boundary layers is only applicable to viscous flow regimes..
3.9. In general, hypersonic flow is not complex in nature.
3.10. The flow becomes isentropic in hypersonic regimes, meaning that there is no entropy generation.

3.11. The momentum thickness of a boundary layer represents the distance by which the external flow would need to be decelerated to eliminate the boundary layer.
3.12. Hypersonic vehicles experience significant heating due to compression of the air ahead of them.
3.13. In hypersonic flow, the aerodynamic force coefficients vary strongly with Mach number.

Shock waves (Problems 3.14–3.16).
Consider a conic-shaped re-entry capsule. The geometry of the capsule is determined by the so-called semi-cone angle, θ, which can be interpreted as a wedge angle in supersonic and hypersonic flow. In Fig. 3.25, the relation between the wedge angle, θ, shock angle, β, and free-stream Mach number, M_1, is shown. Assume that the capsule is flying at low hypersonic speed, with $M_1 = 5$.

3.14. For a semi-cone angle $\theta = 35°$, what can you say about the shock wave(s) (orientation, location, *etc.*)?
3.15. Assume that the shock wave is a so-called strong shock wave: what can you say about the flow velocity behind the shock? Be specific.
3.16. If the semi-cone-angle is increased to $\theta = 43°$, what happens with the shock wave?

Modified Newtonian flow (Problems 3.17–3.21).
In hypersonic flow, the gas around the vehicle is heated to very high temperatures. This causes a number of physical effects that are not present in flows with lower Mach numbers. As they are hard to model and time-consuming to calculate, rapid engineering methods have been developed that give a decent approximation of some aspects of hypersonic flow. The modified Newtonian method is one of them.

For this set of problems, two standard integrals are given that may be of use:

$$\int \cos^n ax \mathrm{d}x = \frac{\cos^{n-1} ax \sin ax}{na} + \frac{n-1}{n} \int \cos^{n-2} ax \mathrm{d}x + C \qquad (n > 0) \tag{3.370}$$

$$\int \cos^n ax \sin ax \mathrm{d}x = -\frac{1}{a(n+1)} \cos^{n+1} ax + C \tag{3.371}$$

3.17. Name one advantage and one disadvantage of high-temperature effects on vehicle design and/or performance.
3.18. In case of an axisymmetric entry capsule like Apollo, explain what the values of the lift and side force on this vehicle will be at zero angle of attack.
3.19. For a sphere, the drag coefficient in modified Newtonian flow should be $C_{p,s}/2$ (where $C_{p,s}$ is slightly smaller than 2 for a typical Earth entry). For a flat plate at zero angle of attack, the result should be 0 from modified Newtonian theory. In high Mach number continuum flow, how realistic do you expect these numbers to be and why?
3.20. Consider a cylinder with radius, R, and length, ℓ, which enters the atmosphere transverse to flow. Following Example 3.1, what is the Newtonian drag coefficient of the cylinder at zero angle of attack? Hint: divide the cylinder in rings

with width $\mathrm{d}\ell$, and a circle segment per ring with angle $\mathrm{d}\theta$. Set up the equations per element and integrate over ℓ and the cross section.

3.21. For continuum Newtonian flow, the air molecules predominately interact with shock waves. How would the analysis (and determination of drag and lift) change?

References

Allen HJ, Eggers AJ (1957) A study of the motion and aerodynamic heating of missiles entering the earth's atmosphere at high supersonic speeds, NACA-TN-4047. National Aeronautics and Space Administration

Anderson JD Jr (1995) Computational fluid dynamics: the basics with applications. McGraw-Hill

Anderson JD Jr (2011) Fundamentals of aerodynamics, 5th edn. McGraw-Hill

Anderson Jr JD (2019) Hypersonic and high-temperature gas dynamics, 3rd edn. American Institute of Aeronautics and Astronautics

Anderson JD Jr (2003) Modern compressible flow: with historical perspective, 3rd edn. McGraw-Hill

Anonymous (1965) Aerodynamic data manual for project Apollo, Technical report NAS 9-159 (U), North American Aviation

Anonymous (1980) Aerodynamic design data book, vol 1, Orbiter Vehicle, STS-1, Technical report SD72-SH-0060, Rockwell International

Arnal D, Delery J (2004) Laminar-turbulent transition and shockwave/boundary layer interaction, Paper presented at the RTO AVT lecture series on critical technologies for hypersonic vehicle development, held at the Von Kàrmàn Institute, Rhode-St-Genèse, Belgium, 10–14 May 2004. Published in RTO-EN-AVT-116 by NATO Research and Technology Organisation

Baals DD, Corliss WR (1981) Windtunnels of NASA. NASA-SP-440. National Aeronautics and Space Administration

Barbante P (2004) Fundamentals of hypersonic flight—properties of high temperature gasses. In: Critical technologies for hypersonic vehicle development. NATO Research and Technology Organisation

Bertin JJ (1994) Hypersonic aerothermodynamics, 3rd printing. American Institute of Aeronautics and Astronautics

Bertin JJ, Cummings RM (2006) Critical hypersonic aerothermodynamic phenomena. Annu Rev Fluid Mech 38:129–157

Bird G (2013) The DSMC method. CreateSpace Independent Publishing Platform

Blanchard RC, Buck GM (1986) Rarefied-flow aerodynamics and thermosphere structure for shuttle flight measurements. J Spacecraft and Rockets J Spacecraft and Rockets) 23(1):18–24

Bonner E, Clever W, Dunn K (1981) Aerodynamic preliminary analysis system II, Part I—theory, NASA-CR-165627. National Aeronautics and Space Administration

Bradshaw P, Cebeci T, Whitelaw J (1981) Engineering calculation methods for turbulent flow. Academic Press, New York

Brauckmann GJ, Paulson Jr JW, Weilmuenster KJ (1994) Experimental and computational analysis of the Space Shuttle Orbiter hypersonic "pitch-up anomaly", AIAA-1994-0632. In: 32nd aerospace sciences meeting & exhibit. Reno, NV, 10–13 Jan 1994

Casoliva J, Lyons DT, Wolf AA, Mease KD (2008) Robust guidance via a predictor-corrector algorithm with drag tracking for aero-gravity assist maneuvers, AIAA-2008-6816. In: AIAA guidance, navigation and control conference and exhibit. Honolulu, Hawaii, 18–21 Aug 2008

Cebeci T, Shao JP, Kafyeke F, Laurendeau E (2005) Computational fluid dynamics for engineers. In: From panel to Navier-Stokes methods with computer programs. Horizons Publishing Inc., Long Beach, California

Chambre PA, Schaaf SA (2017) Flow of rarefied gases. Princeton University Press, New Jersey, Chap, p 2

Chanetz B, Délery J, Gilliéron P, Gnemmi P, Gowree ER, Perrier P (2020) Experimental aerodynamics. an introductory guide. In: Springer tracts in mechanical engineering. Springer Nature Switzerland AG

Craidon C (1985) A description of the langley wireframe geometry standard (LaWGS) format, NASA-TM-85767. National Aeronautics and Space Administration

Craig S, Lyne JE (2002) A parametric study of aerocapture for missions to Venus, AIAA-2002-4500. In: AIAA atmospheric flight mechanics conference. Monterey, CA, 5–8 Aug 2002

Cremers GGJ (1996) Hypersonic aerothermodynamic module for the NLRAERO program. MSc thesis (also published as Memorandum M-748), Delft University of Technology

Cruz CI, Ware GM (1992) Predicted aerodynamic characteristics for HL-20 lifting-body using the aerodynamic preliminary analysis system (APAS), AIAA-1992-3941. In: AIAA 28th joint propulsion conference and exhibit. Nashville, TN, 6–8 July 1992

Cruz CI, Wilhite AW (1989) Prediction of high-speed aerodynamic characteristics using the aerodynamic preliminary analysis system (APAS), AIAA-1989-2173. In: AlAA 7th applied aerodynamics conference. Seattle, WA, July 31–August 2 1989

Dirkx D, Mooij E (2011) Continuous aerodynamic modelling of entry shapes, AIAA-2011-6575. In: AIAA atmospheric flight mechanics conference. Portland, OR, 8–11 Aug 2011

Dirkx D, Mooij E (2017) Conceptual shape optimization of entry vehicles: applied to capsules and winged fuselage vehicles. Springer, Berlin

Divan P, Sova G (1991) Aerodynamic preliminary analysis system II, Part II—user's manual. NASA CR-165628, National Aeronautics and Space Administration

Eggers AJ, Savin RC, Syvertson CA (1955) The generalized shock-expansion method and its application to bodies traveling at high supersonic air speeds. J Aeronaut Sci 22(4):231–238

Eggers AJ, Allen HJ, Neice SE (1957) A comparative analysis of the performance of long-range hypervelocity vehicles, NACA-TN-4046. National Aeronautics and Space Administration

Erickson L (1990) Panel methods—an introduction, NASA-TP-2995. National Aeronautics and Space Administration

Gabeaud A (1950) Base pressures at supersonic velocities. J Aeronaut Sci 7(8):525–526

Gallis M, Hermina W, Johnson M, Chapel J (2004) Aerothermal analysis for the Mars Reconnaissance Orbiter, AIAA-2004-2684. In: 37th AIAA thermophysics conference. Portland, Oregon, June 28–July 1 2004

Gallis M, Torczynski J, Plimpton S, Rader D, Koehler T (2014) Direct simulation Monte Carlo: the quest for speed. In: AIP conference proceedings, vol 1628, no 1. American Institute of Physics, pp 27–36

Gentry A, Smyth D, Oliver W (1973) The mark IV supersonic-hypersonic arbitrary body program, volume ii - program formulation. Report AFFDL-TR-73-159, Douglas Aircraft Company

Gnoffo PA, Weilmuenster KJ, Hamilton HH II, Olynick DR, Venkatapathy E (1990) Computational aerothermodynamic design issues for hypersonic vehicles. J Spacecr Rocket 36(1):21–43

Gomez G (1990) Influence of local inclination methods on the S/HABP aerodynamic analysis tool. Technical report EWP-1793, ESA

Hankey WL (1988) Re-entry aerodynamics. American Institute of Aeronautics and Astronautics

Hansen C (1958) Approximation for the thermodynamic and transport properties of high-temperature air, NASA-TN-4150. National Aeronautics and Space Administration

Hart K, Dutta S, Simonis K, Steinfeldt B, Braun R (2014) Analytically-derived aerodynamic force and moment coefficients of resident space objects in free-molecular flow, AIAA-2014-0728. In: SciTech forum, AIAA atmospheric flight mechanics conference. National Harbor, MD, 13–17 Jan 2014

Hayes WD, Probstein RF (1964) Hypersonic flow theory, 3rd printing. Academic Press, London

Hirschel EH (1991) Viscous effects. In: Space course 1991. RWTH Aachen, pp 12–1 to 12–35
Hirschel EH (2015) Basics of aerothermodynamics, 2nd, revised. Springer, Berlin
Hirschel EH, Weiland C (2009) Selected aerothermodynamic design problems of hypersonic flight vehicles. Springer/AIAA
Hoffmann KA, Chiang ST (2000) Computational fluid dynamics, vol 1, 4 edn. Engineering Education System
Hoffmann KA, Chiang ST (2000) Computational fluid dynamics, vol 2, 4th edn. Engineering Education System
Hoffmann KA, Chiang ST (2000) Computational fluid dynamics, vol 3, 4th edn. Engineering Education System
Houghton EL, Carpenter PW (2003) Aerodynamics for engineering students. 5th edn. Butterworth-Heinemann
Johnson J, Starkey R, Lewis M (2007) Aerothermodynamic optimization of reentry heat shield shapes for a crew exploration vehicle. J Spacecraft and Rockets 44(4):849–859
Kaushik M (2019) Theoretical and experimental aerodynamics. Springer, Berlin
Klothakis A, Nikolos I (2015) Modeling of rarefied hypersonic flows using the massively parallel DSMC kernel "SPARTAâŁž 8th GRACM international congress on computational mechanics. Volos, Greece
Krieger RJ (1981) Summary of design and performance characteristics of aerodynamic configured missiles, AIAA-1981-0286. In: AIAA 19th aerospace sciences meeting, St. Louis, MO, 12–15 Jan 1981
Lewis RW, Nithiarasu P, Seetharamu KN (2004) Fundamentals of the finite element method for heat and fluid flow. Wiley
Liechty D (2007) Aeroheating analysis for the Mars Reconnaissance Orbiter with comparison to flight data. J Spacecraft and Rockets 44(6):1226–1231
Lind CA (1997) Effect of geometry on the unsteady type-IV shock interaction. J Aircr 34(1):64–71
Maslen SH (1964) Inviscid hypersonic flow past smooth symmetric bodies. AIAA J 2(6):1055–1061
Maughmer M, Ozoroski L, Straussfogel D, Long L (1993) Validation of engineering methods for predicting hypersonic vehicle control forces and moments. J Guid, Control Dyn 16(4):762–782
Maus JR, Griffith BJ, Szema KY, Best JT (1983) Hypersonic mach number and real gas effects on space shuttle orbiter aerodynamics. J Spacecraft and Rockets 21(2):136–141
McGinley CB, Berry SA, Kinder GR, Barnwell M, Wang KC, Kirk BS (2006) Review of orbiter flight boundary layer transition data, AIAA 2006-2921. In: 9th AIAA/ASME joint thermophysics and heat transfer conference. San Francisco, California, 5–8 June 2006
McRonald A, Nock K (1977) Use of atmospheric drag to circularize an elliptic spacecraft orbit at Venus, AIAA/AAS. In: Astrodynamics specialist conference. Jackson Hole, Wyoming, pp 760–784, 7–9 Sept 1977
Moore M, Williams J (1989) Aerodynamic prediction rationale for analyses of hypersonic configurations, AIAA-1989-0525. In: 27th aerospace sciences meeting. Reno, NV, 9–12 Jan 1989
Moss JN (1992) Computation of flow fields for hypersonic flight at high altitudes. In: Bertin JJ, Periaux J, Ballmann J (eds) Advances in hypersonics, vol 3. Computing hypersonic flows. Birkhäuser, Boston, pp 371–427
Moss JN, Glass CE, Greene FA (2006) DSMC simulations of Apollo capsule aerodynamics for hypersonic rarefied conditions, AIAA-2006-3577. In: 9th AIAA/ASME joint thermophysics and heat transfer conference. San Fransisco, CA, 5–8 June 2006
Olivier H (2000) A theoretical model for the shock stand-off distance in Frozen and equilibrium flows. J Fluid Mech 413:345–353
Padilla JF, Boyd ID (2006) Assessment of rarefied hypersonic aerodynamics modeling and wind-tunnel data, AIAA-2006-3390. In: 9th AIAA/ASME joint thermophysics and heat transfer conference. San Fransisco, CA, 5–8 June 2006
Page WA, Woodward HT (1972) Radiative and convective heating during Venus entry. AIAA J 10(10):1379–1381

Pamadi BN, Brauckmann GJ (1999) Aerodynamic characteristics and development of the aerodynamic database of the X-34 reusable launch vehicle. In: International symposium on atmospheric reentry vehicles and systems. Arcachon, France, 16–18 Mar 16–18
Plimpton S, Gallis M (2017) SPARTA users manual. Sandia National Laboratories, United States Department of Energy
Potter J (1988) Procedure for estimating aerodynamics of three-dimensional bodies in transitional flow. Prog Astronaut Aeronaut 118:484–492
Potter M, Wiggert D (2008) Schaum's outline of fluid mechanics. McGraw-Hill
Rasmussen M (1994) Hypersonic flow. Wiley
Regan FJ, Anandakrishnan SM (1993) Dynamics of atmospheric re-entry. American Institute of Aeronautics and Astronautics
Riley CJ, DeJarnette FR (1990) An approximate method for calculating three-dimensional inviscid hypersonic flow fields. In: NASA-TP-3018, National Aeronautics and Space Administration
Ritter H, Mazoue F, Santovincenzo A, Atzei A (2006) Jupiter entry probe feasibility study from the ESTEC CDF team: heat flux evaluation and TPS definition. In: SP-631, Proceedings of the 5th European workshop on thermal protection systems and hot structures
Robinson JS, Wurster KE, Mills JC (2009) Entry trajectory and aeroheating environment definition for capsule-shaped vehicles. J Spacecraft and Rockets 46(1):74–86
Ronse ALAB, Mooij E (2014) Statistical impact prediction of decaying objects. J Spacecraft and Rockets 51(6):1797–1810
Schetz JA (2010) Boundary layer analysis. American Institute of Aeronautics and Astronautics
Schlichting H, Gersten K (2017) Boundary-layer theory. Springer, Berlin
Sharma B, Kumar R (2019) Estimation of bulk viscosity of dilute gases using a nonequilibrium molecular dynamics approach. Phys Rev E 100:013309
Sutton K, Graves Jr RA (1971) A general stagnation-point convective-heating equation for arbitrary gas mixtures, NASA-TR-R-376. National Aeronautics and Space Administration
Svehla R (1962) Estimated viscosities and thermal conductivities of gases at high temperatures, NASA-TR-R-132. National Aeronautics and Space Administration
Tauber ME, Menees GP (1986) Aerothermodynamics of transatmospheric vehicles, AIAA-1986-1257. In: AlAA/ASME 4th joint thermophysics and heat transfer conference. Boston, MA, 2–4 June 1986
Tauber ME, Sutton K (1991) Stagnation-point radiative heating relations for Earth and mars entries. J Spacecr Rocket 28(3):43–49
Tauber ME, Palmer GE, Prabhu D (2012) Stagnation point radiative heating relations for Venus entry, ARC-E-DAA-TN2887. National Aeronautics and Space Administration
Tauber ME, Wercinski P, Yang L, Chen Y-K (1999) A fast code for Jupiter atmospheric entry analysis, NASA/TM-1999-208796. National Aeronautics and Space Administration
Vos R, Farokhi S (2015) Introduction to transonic aerodynamics. Springer, Berlin
Watts JD (1968) Flight experience with shock impingement and interference heating on the X-15-2 research airplane, NASA-TM-X-1669. National Aeronautics and Space Administration
Weiland C (2014) Aerodynamic data of space vehicles. Springer, Berlin
White FM (2005) Viscous fluid flow, 3rd edn. McGraw-Hill
Whitmore SA (2007) Real gas extensions to tangent-wedge and tangent-cone analysis methods. AIAA J 45(8):2024–2032
Wilmoth R, Blanchard R, Moss J (1998) Rarefied transitional bridging of blunt body aerodynamics, 1st international symposium on rarefied gas dynamics. Marseille, France
Yoshikawa KK, Wick BH (1961) Radiative heat transfer during atmosphere entry at parabolic velocity, NASA-TN-D-1074. National Aeronautics and Space Administration

Chapter 4
Fundamentals of Motion



After setting the stage by discussing the flight environment and the external forces originating from that, i.e., the ones due to gravity and the aerothermodynamical loading, these forces will now be linked to the motion of the vehicle. This chapter approaches the topic from a more fundamental point of view, whereas in the next chapter the physical mechanisms of re-entry flight will be discussed. The velocities of re-entry flight are usually not so high that classical mechanics can no longer describe the vehicle's motion. Within this field of mechanics it is Newton's Laws of Motion that lie at the root of it. Originally postulated for motion in an inertial frame the concept of reference frames becomes an important aspect.

Therefore, this chapter kicks off with a summary of the reference frames that are commonly used for the study of re-entry dynamics. Next, in Sect. 4.2 the variables, which define the motion of a spacecraft are introduced, the so-called state variables. The use of multiple reference frames and different options necessitates frame and coordinate transformations, treated in Sects. 4.3 and 4.4, respectively. Then, the general form of the equations of translational and rotational motion is summarised in Sect. 4.5; a more detailed derivation can be found in Appendix A. This chapter concludes with a summary of the equations of motion in spherical coordinates (Sect. 4.6), as they will be intensely used throughout the remainder of this book.

4.1 Introduction

4.1.1 Classical Mechanics

Classical mechanics deals with the motion of bodies under the influence of forces or with the equilibrium of bodies when all forces are balanced. The subject of mechanics can be divided into statics, dynamics and kinematics. Of these three, statics is dealing with the state of equilibrium of a system under the influence of forces, dynamics is

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_4

concerned with the forces that change or produce the motion of bodies, and kinematics studies the motion of bodies without reference to mass or force.

Applying the above mentioned definition of mechanics to *flight* mechanics, we can restrict ourselves to dynamics and kinematics. Flight *dynamics* is then related to the motion of a vehicle subjected to forces (or moments) of, for instance, aerodynamic, gravitational and propulsive origin resulting in information about the linear (and angular) *velocity* of the vehicle. Flight *kinematics*, on the other hand, gives information about the *position* (and attitude) of the vehicle originating from these velocities.

Sir Isaac Newton in his *Philosophiae Naturalis Principia Mathematica* (1687), popularly called the *Principia*, was the first one who elaborated on the subject, which can be seen as an application of basic postulates. Nowadays, these are known as Newton's Laws of Motion, three physical laws that form the basis for classical mechanics. They are loosely formulated as:

1. In the absence of force, a body either is at rest or moves in a straight line with constant speed.
2. A body experiencing a force F experiences an acceleration a related to F by $F = ma$, where m is the mass of the body. Alternatively, force is proportional to the time derivative of momentum.
3. Whenever a first body exerts a force F on a second body, the second body exerts a force $-F$ on the first body. F and $-F$ are equal in magnitude and opposite in direction.

The first and second laws hold for a certain reference frame with respect to which a point mass, free of all external forces, moves in a straight line with a constant speed. Such a frame is called an inertial frame. However, Newton's laws can be modified to derive the equations of motion in a (rotating) non-inertial frame.

Newton used these laws to explain and investigate the motion of many physical objects and systems. For example, in the third volume of the text, Newton showed that these laws of motion, combined with his law of universal gravitation, explained Kepler's laws of planetary motion. Newton's law of universal gravitation states that every object in this universe attracts every other object with a force which is directly proportional to the product of their masses and inversely proportional to the square of distance between their centres:

$$F = G\frac{m_1 m_2}{r^2} \tag{4.1}$$

where

F = magnitude of the gravitational force between the two point masses (N)
G = universal gravity constant = $6.668 \cdot 10^{-11}$ m^3/kg s^2
m_1 = mass of the first point mass (kg)
m_2 = mass of the second point mass (kg)
r = distance between the two point masses (m).

The law of universal gravitation is a general physical law derived from empirical observations by what Newton called induction. Sometimes this law is called Newton's Fourth Law.[1]

Based on these Laws of Motion, differential equations of motion can be derived, which allow us to study, for instance, the flight of an aircraft through the atmosphere or a spacecraft orbiting the Earth. But before we can begin the discussion on re-entry flight dynamics, we have to cover some fundamental aspects. Not only from the physical side, i.e., the environment in which the re-entry vehicle will be flying needs to be explained, but also from the mathematical side. Motion can be described in different forms, taking a certain definition of the state variables, and even though Newton's laws assume an inertial reference frame as the basis, more reference frames are required, for instance to combine all external forces that may defined in different frames. Before moving on, the most commonly used reference frames will be introduced.

4.1.2 Reference Frames

One of the key elements in Newton's Laws of Motion is the frame in which this motion is defined. Originally defined for a so-called *inertial frame* it has a practical value to express motion in a rotating frame: in every-day life we are bound to the Earth's surface and we are used to express motion in a frame that is fixed to the Earth. Even though Newton's laws cannot be used directly in combination with rotating frames, they can be adapted for these *non-inertial* frames by introducing the concept of *apparent forces*. This will be discussed in more detail later in this chapter and in Appendix A.

Vectors are important entities in the equations of motion as described in this chapter. Also the use of vectors implies a definition with respect to some reference frame. Sometimes, it is convenient to express several vectors in different reference frames. However, when they are needed in one and the same set of equations (which are defined in only one reference frame) all vectors have to be transformed to this one frame. Common to all frames is that they are right handed.

Typically, the required reference frames for the study of atmospheric flight requires frames having their origin at the centre of the main attracting body, or at a fixed point of the vehicle, e.g., its centre of mass. For more generic astrodynamics problems typically a Sun-centred frame (the so-called *heliocentric* frame) can be used, or different orbital reference frames that allows for a convenient definition of the motion of a satellite around a planet or moon. Even though the primary focus of this book is the study of vehicles entering the sphere of influence of a celestial body and ultimately landing on it, in Chap. 11 the concept of aeromanoeuvring is discussed, i.e., aerobraking, aerocapture, and aero-gravity assists. This type of manoeuvring

[1] In Chap. 2 a practical formulation of gravitational acceleration for planetary bodies with (irregular) finite shape has been given. The original formulation of Newton is restricted to point masses only.

has a strong link between (interplanetary) orbital motion and atmospheric flight. For this reason, we will introduce a few reference frames that are specifically required to study those manoeuvres.

Inertial Planetocentric Reference Frame, Index I

The origin of the inertial reference frame, $\mathcal{F}_I$, is located at the centre of mass (c.o.m.) of the central body (the body around which the vehicle is moving).[2] The OX_IY_I-plane coincides with the equatorial plane of the central body. In general, the Z_I-axis is pointing north and the reference meridian which determines the direction of the X_I-axis, is defined by the zero-longitude or *prime* meridian at zero time. Note that the Z_I-axis is coincident with the rotational axis of the central body (the rotation of the central body is assumed to be constant in magnitude and direction). The Y_I-axis completes the right-handed system.

The above definition is not generally used for the Earth. In this case, the X_I-axis is pointing at a fixed point in inertial space at a specified time. For the so-called $J2000$ frame this is the Vernal Equinox, a point in the star sign Aries, at 12h (noon) on 1 January 2000. The orientation of the prime meridian, which passes through Greenwich, is called the right ascension of the Greenwich meridian at time t, also known as the sidereal time or Greenwich Hour Angle. It is indicated by the symbol Θ. Denoting by d the time in days since 12h on 1 January 2000, the angle $\Theta(t)$ is given by $\Theta(t) = 100.4606^\circ + 360.9856473^\circ d$ (Montebruck and Gill 2000).

Note that the inertial heliocentric reference frame could be defined similarly, albeit with the origin in the (geometric) centre of the Sun.

Orbital Reference Frame, Index O

The orbital reference frame, $\mathcal{F}_O$, results from rotating $\mathcal{F}_I$ about the line of nodes of the spacecraft orbit (line of intersection between the orbital plane and the $X_I - Y_I$ plane) by the inclination of the orbit, such that the resulting X_O-axis is still pointing toward the vernal equinox, but now the Z_O-axis is perpendicular to the orbital plane rather than the equatorial plane. The Y_O-axis complements the right-handed system and the frame origin is at the centre of the central body.

[2] In principle, this reference frame is pseudo-inertial because of the motion of the central body itself. When the central body is a planet, it orbits around the Sun and when we are dealing with a moon, it describes an orbit around a planet which in itself revolves around the Sun. However, since this rotation of the pseudo-inertial frame has only a marginal effect on the results, we can consider the frame to be inertial.

Local Orbital Reference Frame, Index L

The local orbital reference frame, $\mathcal{F}_L$, is also centred at the central body, but it is a rotating frame, whose X_L-axis is always pointing toward the spacecraft, the *radial* direction. The Z_L-axis is perpendicular to the orbital plane, positive along the the positive orbit normal (or orbital momentum vector). The Y_L-axis complements the right-handed system and is pointing in the direction of flight, in the so-called *along-track* direction.

Rotating Planetocentric Reference Frame, Index R

Rotating planetocentric reference frame, $\mathcal{F}_R$, is fixed to the central body and coincides with $\mathcal{F}_I$ at zero time (and once every full rotation of the central body after that). The Z_R-axis is pointing north, the X_R-axis intersects the equator at zero longitude and the Y_R-axis completes the right-handed system.

The origin of the following reference frames is located at the centre of mass (c.o.m.) of the vehicle. We assume that the vehicle has at least one plane of (geometrical) symmetry (in longitudinal direction).

Body Reference Frame, Index B

The body frame, $\mathcal{F}_B$, is fixed to the vehicle. The X_B-axis lies in the plane of symmetry and is positive in forward direction. The Z_B-axis also lies in the plane of symmetry and is positive in downward direction. The Y_B-axis completes the right-handed system.

Vertical Reference Frame, Index V

The Z-axis of the vertical reference frame, $\mathcal{F}_V$, is pointing towards the c.o.m. of the central body, along the radial component of the gravitational acceleration. The X_V-axis lies in a meridian plane, perpendicular to Z_V, and points to the northern hemisphere. The Y_V-axis is the right-handed supplement. The $X_V Y_V$-plane is referred to as the local horizontal plane. Note that strictly speaking, this is only true when the central body is a true sphere. A small error is introduced, when an elliptical shape is assumed.

The components along the axes of $\mathcal{F}_V$ are often referred to as being in north, east and down direction. Velocity components expressed in this frame are often used in navigation studies.

Trajectory Reference Frame (Groundspeed Based), Index TG

The three axes of $\mathcal{F}_{TG}$ are defined as follows. The X_{TG}- axis points in the direction of the velocity vector relative to $\mathcal{F}_R$, the Z_{TG}-axis lies in the vertical plane, pointing downwards, and the Y_{TG}-axis completes right-handed system.

Trajectory Reference Frame (Airspeed Based), Index TA

The definition of the axes of the airspeed-based trajectory reference frame, $\mathcal{F}_{TA}$, is in principle identical to that of $\mathcal{F}_{TG}$, with only one fundamental difference. The X_{TA}-axis is now positive along the velocity vector relative to the atmosphere rather than the groundspeed. Of course, this affects the absolute orientation of the Y_{TA}- and Z_{TA}-axis, although the definition remains the same: the Z_{TA}-axis lies in the vertical plane, pointing downwards, and the Y_{TA} completes the right-handed system.

Aerodynamic Reference Frame (Groundspeed Based), Index AG

The X-axis of $\mathcal{F}_{AG}$ is defined along the velocity vector of the vehicle relative to the rotating planetocentric frame (the groundspeed). This definition implies, that the X_{AG}-axis is collinear with the X_{TG}-axis. The Z_{AG}-axis is collinear with the aerodynamic lift force (based on groundspeed variables), but opposite in direction. The Y_{AG}-axis completes the right-handed system. Note that when the vehicle is not banking, $\mathcal{F}_{AG}$ and $\mathcal{F}_{TG}$ are coinciding.

Aerodynamic Reference Frame (Airspeed Based), Index AA

Also for the airspeed-based aerodynamic reference frame, $\mathcal{F}_{AA}$, the definition is similar to the previous one, but now the X_{AA}-axis is defined along the velocity vector of the vehicle relative to the atmosphere. This definition implies, that the X_{AA}-axis is collinear with the X_{TA}-axis. The Z_{AA}-axis is collinear with the aerodynamic lift force (based on airspeed variables), but opposite in direction. The Y_{AA}-axis completes the right-handed system. As before, when the vehicle is not banking, $\mathcal{F}_{AA}$ and $\mathcal{F}_{TA}$ are coincident.

Wind Reference Frame, Index W

The X_W-axis is collinear with the wind-velocity vector; for a southern wind, i.e., wind blowing from the south to the north, the X_W-axis is positive in northern direction. For a wind in the local horizontal plane, the Z_W-axis is positive pointing downward. The Y_W-axis is the right-handed supplement.

Propulsion Reference Frame, Index P

Even though we are usually dealing with unpowered re-entry vehicles, there are situations when propulsion systems are involved. One can think of, for instance, the sustained cruise flight of an experimental re-entry vehicle that has to fly with constant Mach number to do aerodynamic experiments. An even more common application is the use of reaction-control thrusters for attitude control, e.g., when the dynamic pressure is too low for the aerodynamic control surfaces to produce sufficiently large moments.

The propulsion frame, $\mathcal{F}_P$, is strongly related to the body frame, in the sense that commonly the three components of the thrust force are defined in $\mathcal{F}_B$. However, on occasion the thrust force is defined in spherical components, with a magnitude and two direction angles. So, for the definition of the propulsion frame we will usually rotate over the two direction angles to reach $\mathcal{F}_P$ (see for the proper transformation matrix Sect. 4.3). As a result, the X_P-axis is collinear with the thrust vector. The Y_P- and Z_P-axis follow from the transformation to the B-frame.

4.2 State-Variable Definitions

4.2.1 Position and Velocity

The state variables for position and velocity can, amongst others, be expressed in either Cartesian or spherical components. However, when the motion of a spacecraft orbiting a celestial body is studied, it is common to use the classical orbital elements. Even though we are mainly focussing on objects or vehicles in a direct approach to the surface of a celestial body, coming from space usually involves an exoatmospheric orbit at one stage or another. Therefore, we will start this section by introducing the orbital elements.

Orbital Elements

In case of a state defined by the classical orbital elements, we will restrict ourselves to elliptical orbits. The six parameters, defining the position and velocity of a spacecraft in an elliptical orbit, with respect to an inertial frame, are (see also Figs. 4.1 and 4.2):

e : the eccentricity ($0 \leq e < 1$)
a : the semi-major axis ($a > R_e$)
i : the inclination ($0° \leq i \leq 180°$)
ω : argument of pericentre ($0° \leq \omega < 360°$)
Ω : the longitude of the ascending node ($0° \leq \Omega < 360°$)
M : mean anomaly ($0° \leq M < 360°$)

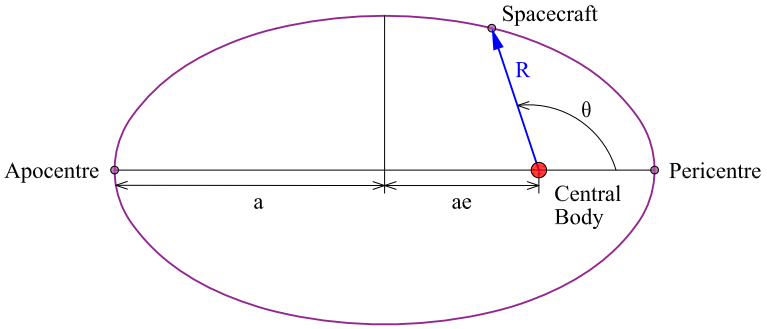


Fig. 4.1 Definition of semi-major axis a and the eccentricity e. The spacecraft is moving at a distance r; the true anomaly is indicated by θ

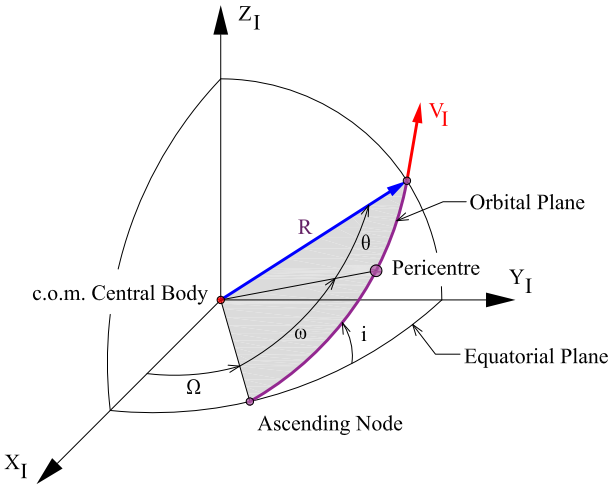


Fig. 4.2 Definition of the three orbital parameters Ω, ω and i. The spacecraft is moving at a distance r with a velocity $\mathbf{V_I}$ with respect to to the inertial planetocentric frame (index I). The true anomaly is indicated by θ

The orbit of a spacecraft in the orbital plane is a conic section, with the central body at one focus. The shape is defined by the eccentricity, e, and the size by the semi-major axis, a. The orientation of the conic in the orbital plane is specified by the argument of pericentre, ω, which is measured from the line of intersection of the orbital plane with the XY-plane of the inertial frame, here the equatorial plane of the central body towards the pericentre[3] the name of the planet. This intersection line is also called the line of nodes. It intersects the orbit at two points, the ascending

[3] Alternatively, the closest and farthest point are also know as periapsis and apoapsis, respectively. To specify these two points for a particular central body, the terminology includes a reference to that body, e.g., perigee and apogee (Earth), periareion and apoareion (Mars), and perihelion and

node (the z-coordinate of the spacecraft changes from negative to positive) and the descending node. The position of the ascending node in the XY-plane is specified by the angle Ω, which is called the longitude of the ascending node. It is measured positively from the X_I-axis eastward. The angle i between the XY-plane and the orbital plane is called the inclination. It is considered to be the (smallest) angle between the positive Z_I-axis and the orbital angular momentum vector. To determine the position of the spacecraft in its orbit at a given time, we need a sixth parameter, the time of pericentre passage τ. However, often τ is replaced by the mean anomaly at the epoch t_0, M_0, defined as

$$M_0 = n\,(t_0 - \tau) \tag{4.2}$$

in which

$$n = \sqrt{\frac{\mu}{a}^3} \tag{4.3}$$

Note that μ is the gravitation parameter of the central body. The mean anomaly M is then

$$M = M_0 + n\,(t - t_0) = n\,(t - \tau) \tag{4.4}$$

which is, in fact, the product of a mean angular velocity and the time, which has passed since the last pericentre passage. Alternatively, in literature the true anomaly θ is often used as running parameter.

Kepler's equation relates the mean anomaly to the eccentric anomaly E:

$$M = E - e \sin E \tag{4.5}$$

Given M, this equation does not provide a closed-loop solution for E, so one has to resort to numerical methods to iterate a solution, for instance, the Newton-Raphson method (see Appendix C.6).

Cartesian Components

Using Cartesian components, we will express the position and velocity with respect to either $\mathcal{F}_I$ or $\mathcal{F}_R$. In both cases we will use the same variables, but with different subscripts (a subscript I or R). The position and velocity vectors are defined with the following components:

Position : x, y, z
Velocity : $\dot{x}, \dot{y}, \dot{z}$

---

aphelion (Sun). See https://en.wikipedia.org/wiki/Apsis for a complete overview of the naming convention.

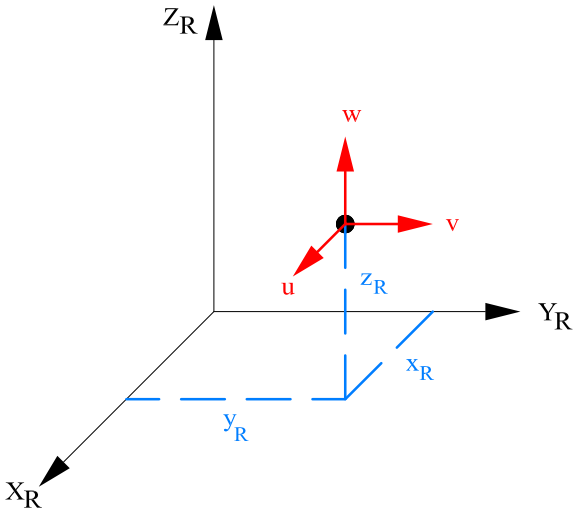


Fig. 4.3 Definition of the Cartesian position and velocity components with respect to $\mathcal{F}_R$

Note that when the velocity components are defined in $\mathcal{F}_R$, instead of $\dot{x}_R$, $\dot{y}_R$ and $\dot{z}_R$, the notation u, v, w is commonly used (see also Fig. 4.3).

Spherical Components

With spherical components, position and velocity (in the current context, we will only define them with respect to $\mathcal{F}_R$) can be expressed by (Fig. 4.4):

Position : distance R, longitude τ and latitude δ
Velocity : groundspeed V_g, flight-path angle γ_g and heading χ_g

The longitude is measured positively to the east ($0° \leq \tau < 360°$). The latitude is measured along the appropriate meridian starting at the equator, positive in north direction ($-90° \leq \delta \leq 90°$) and negative to the south. The distance R, finally, is the distance from the c.o.m. of the central body to the c.o.m. of the vehicle. The relative velocity V_g (i.e., the modulus of the velocity vector **v**) is expressed with respect to the rotating planetocentric frame. γ_g is the angle between **v** and the local horizontal plane; it ranges from –90° to +90° and is negative when **v** is below the local horizon. χ_g defines the direction of the projection of **v** in the local horizontal plane with respect to the local north and ranges from –180° to +180°. When χ_g = +90°, the vehicle is moving parallel to the equator to the east.

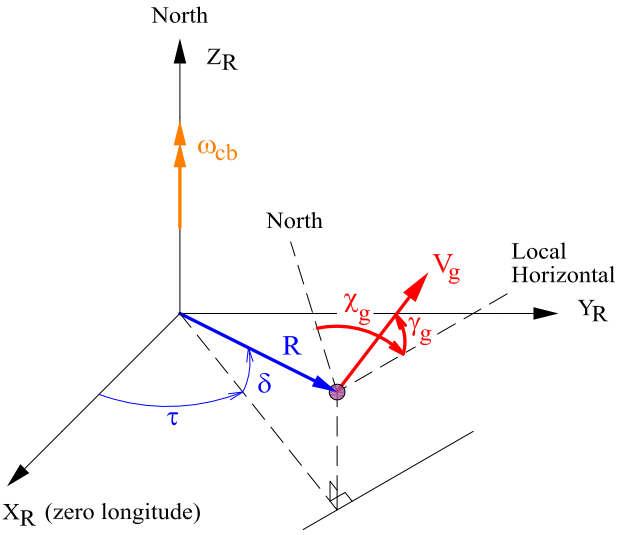


Fig. 4.4 Definition of the six spherical flight parameters, the position (R,τ,δ) and velocity (V_g,γ_g,χ_g). Here, both τ, δ, γ_g and χ_g are positive

4.2.2 Attitude and Angular Rates

Attitude

The attitude of a vehicle, or in more mathematical terms, the orientation of a body-fixed reference frame with respect to another one, can be expressed in several ways. Here, we restrict ourselves to three definitions:

1. *Classical attitude angles*, i.e., the roll angle φ, the pitch angle θ and the yaw angle ψ. For re-entry applications these angles usually define the attitude of the body frame with respect to inertial space, but it is also possible to have them define the attitude of the body with respect to the local horizontal plane. In the current text, we will consider both options.
   Note that it is important to specify the order in which the transformation is executed, and about which axes. In total there are 12 possible combinations, i.e., six so-called *Type-1* representations, where all three axes are involved, and six *Type-2* representations (Wertz 1978). For the second type, only two axes are involved, e.g., the sequence 1-2-1 begins with a rotation about the X-axis, is followed by one about the (local) Y-axis, and concluded by a rotation about the (local) X-axis. The commonly used sequence in aerospace applications is the *Type-1* sequence 3-2-1: a yaw rotation about the Z-axis, followed by a pitch rotation about the (local) Y-axis and concluded by a roll rotation about the (local) X-axis. As an example, in Fig. 4.5, the Euler transformation from $\mathcal{F}_I$ to $\mathcal{F}_B$ is shown.

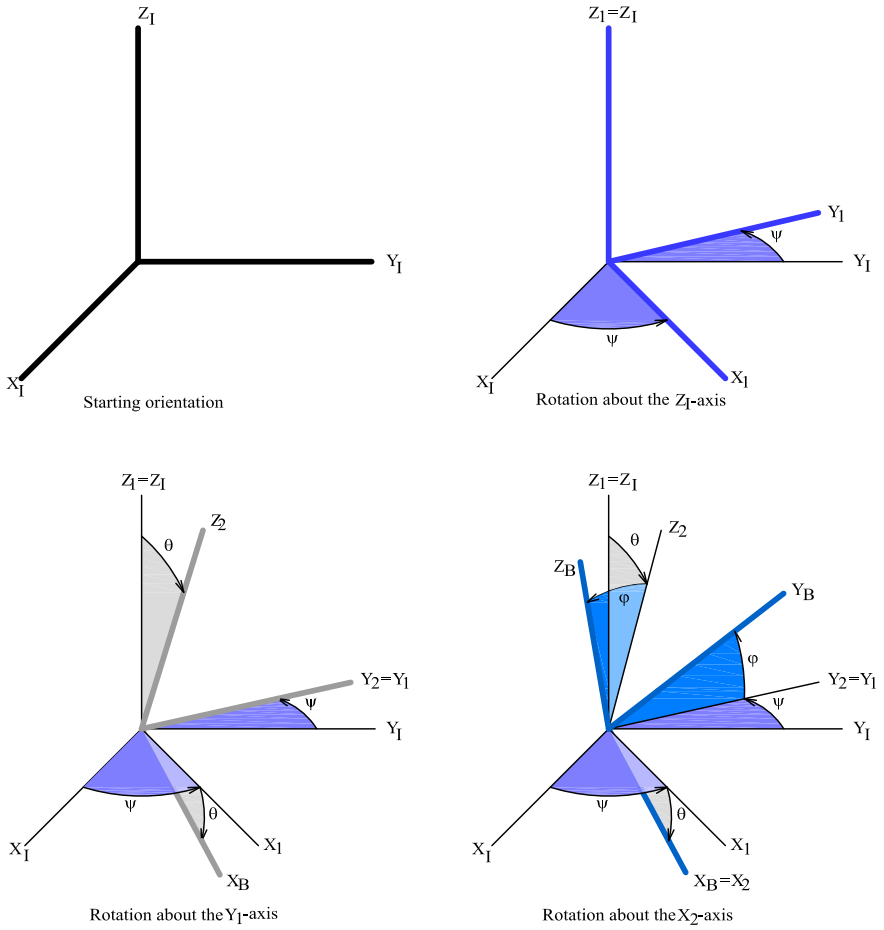


Fig. 4.5 The Euler transformation: successive rotations about the Z_I-axis (yaw angle ψ), the Y_1-axis (pitch angle θ) and the X_2-axis (roll angle ϕ) brings us from $\mathcal{F}_I$ to $\mathcal{F}_B$

2. *Aerodynamic angles*, i.e., the angle of attack α ($-180° \leq \alpha < 180°$, for a 'nose-up' attitude $\alpha > 0°$), the angle of sideslip β ($-90° \leq \beta \leq 90°$, β is positive for a 'nose-left' attitude), and the bank angle σ ($-180° \leq \sigma < 180°$, σ is positive when banking to the right), see also Fig. 4.6. When these angles are used in the equations of motion, they define the attitude of the vehicle with respect to the groundspeed and will get the subscript 'g'. Otherwise, they can also define the orientation of the body with respect to the airspeed, and will have the subscript 'a' in that case. It should be noted that α, β and σ form a set of Euler angles: a *Type-1* sequence with order 2-3-1.
3. *Quaternions*, or the four parameters of Euler: q_1, q_2, q_3 and q_4. A quaternion is a 4-dimensional hyper-complex number, which was invented by Sir William

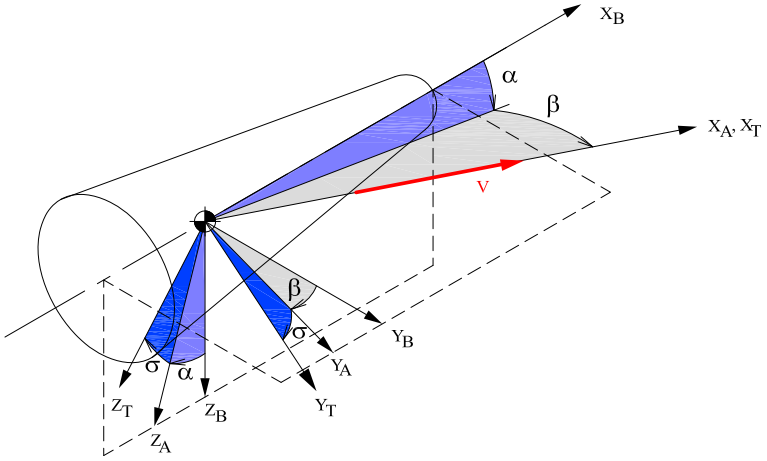


Fig. 4.6 Definition of the aerodynamic attitude angles α, β and σ. The three related reference frames are the body frame (index B), the aerodynamic frame (index A) and the trajectory frame (index T). Here, both α, β and σ are positive

Rowan Hamilton in 1843. A quaternion consists of one real and three imaginary numbers. The imaginary numbers are different square roots of -1 and obey the following constraint:

$$i^2 = j^2 = k^2 = ijk = -1$$

The quaternions can be defined in terms of the Euler axis, $\mathbf{a}$, and the Euler angle, Φ. The vector part is defined as

$$\mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \mathbf{a} \sin \frac{\Phi}{2} \tag{4.6}$$

and the scalar part as

$$q_4 = \cos \frac{\Phi}{2} \tag{4.7}$$

The rotation specified by the quaternions is expressed as $(\mathbf{q}, q_4)$. The 4 quaternions are not mutually independent because they satisfy the following constraint:

$$\mathbf{q}^{\mathrm{T}}\mathbf{q} + {q_4}^2 = {q_1}^2 + {q_2}^2 + {q_3}^2 + {q_4}^2 = 1 \tag{4.8}$$

This shows that the quaternions describe a unit 3-sphere, i.e., a sphere in 4-dimensional space. Therefore, any rotation is a trajectory on the surface of the 3-sphere. Also, the orientation specified by $(\mathbf{q}, q_4)$ is the same as the orientation

specified by $(-\mathbf{q}, -q_4)$, since the latter describes the Euler rotation with Euler axis -**a** and Euler angle $-\Phi$. This means that if $(\mathbf{q}, q_4)$ describes the shortest rotation then $(-\mathbf{q}, -q_4)$ describes the longest rotation. To enforce the shortest rotation the necessary condition is that $q_4 > 0$.
Quaternions form a very vast topic, with their own special algebra and properties making an exhaustive treatment beyond the scope of these lecture notes. The interested reader is referred to the many textbooks covering this topic, see, for instance, the book by Kuipers (2002). On a final note, since the quaternions form a set of four coordinates, there are no singularities present for any rotation.

Angular Rates

The angular rate of the body is here defined as the rotational velocity of the body frame with respect to the inertial frame, expressed in components along the body axes. The rotation vector, $\boldsymbol{\omega}$, is defined by the roll rate p, the pitch rate q, and the yaw rate r (Fig. 4.7).

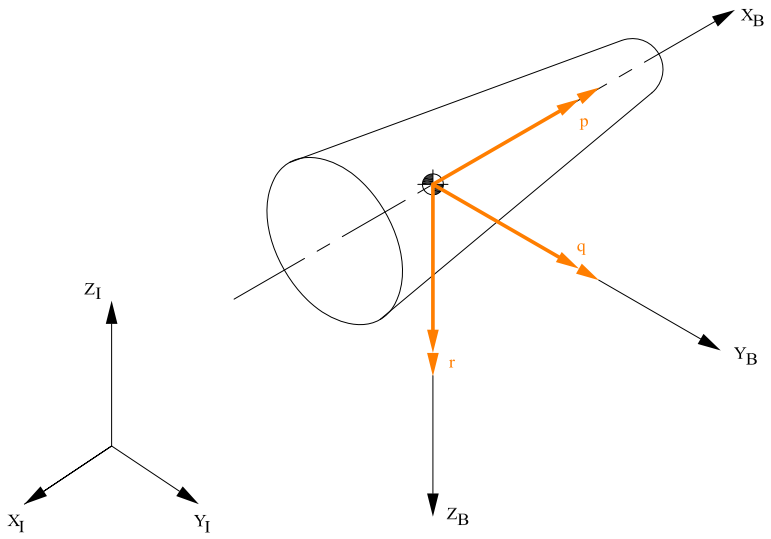


Fig. 4.7 Definition of the angular rate of the vehicle, $\boldsymbol{\omega} = (p, q, r)^{\mathrm{T}}$. Relevant frames are $\mathcal{F}_I$, which has its origin in the c.o.m. of the central body, and the body-fixed frame $\mathcal{F}_B$, with its origin in the c.o.m. of the vehicle

4.3 Frame Transformations

In general, an axis transformation consists of a translation and a rotation (Fig. 4.8). So if we want to express vector $\mathbf{v_A}$, defined in axis system A, in axis system B, we get:

$$\mathbf{v_B} = \mathbf{T} + \mathbf{C_{B,A}}\mathbf{v_A} \tag{4.9}$$

where

$\mathbf{T}$ = translation vector = $\mathbf{O_A O_B}$
$\mathbf{C_{B,A}}$ = transformation matrix defining the rotation from frame A ($\mathcal{F}_A$) to B ($\mathcal{F}_B$).

4.3.1 Unit-Axis Rotations

The transformation from one (right-handed Cartesian) frame to another one can be expressed by means of, for instance, unit-axis rotations, Directional Cosine Matrices or quaternions. In this section we will concentrate on unit-axis rotations (positive rotations according to the right-hand rule). This way of transforming from one frame to another is the most intuitive and straightforward process. However, a more generic approach to frame transformations is provided by Euler's rotation theorem. For this reason, it will be discussed hereafter.

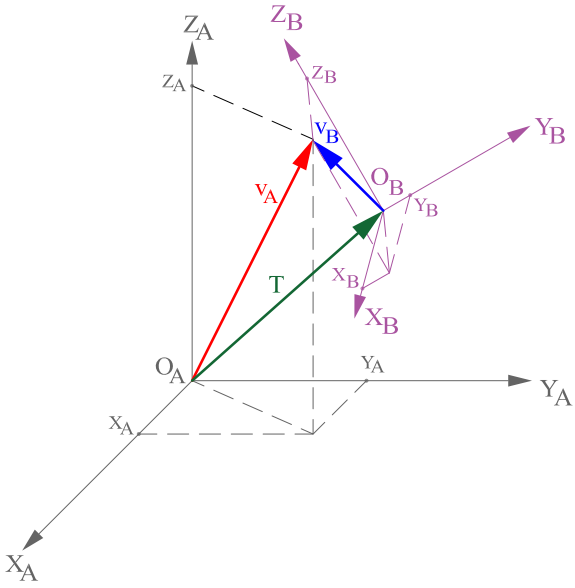


Fig. 4.8 Axis transformation from frame A to frame B, composed of a translation and a rotation

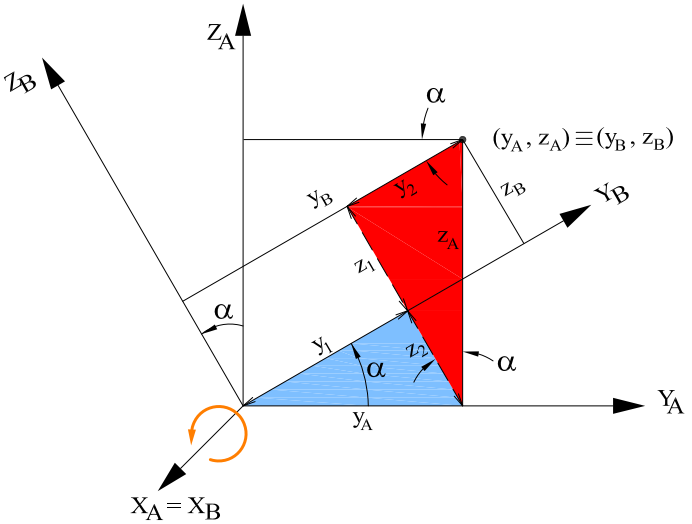


Fig. 4.9 Rotation from reference frame A to reference frame B

Let us have a look at the transformation from $\mathcal{F}_A$ to $\mathcal{F}_B$. Frame $\mathcal{F}_B$ can be obtained from $\mathcal{F}_A$ by rotating this frame about its unit axis X_A by the angle α. This rotation is shown in Fig. 4.9.

From this figure, the following equations can be derived:

$$x_B = x_A \tag{4.10a}$$

$$y_B = y_A \cos\alpha + z_A \sin\alpha \tag{4.10b}$$

$$z_B = -y_A \sin\alpha + z_A \cos\alpha \tag{4.10c}$$

In matrix form this is:

$$\mathbf{x_B} = \begin{pmatrix} x_B \\ y_B \\ z_B \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \begin{pmatrix} x_A \\ y_A \\ z_A \end{pmatrix} \tag{4.11}$$

From this equation, we can define the basic transformation matrix $\mathbf{C_x}(\alpha)$:

$$\mathbf{C_x}(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \tag{4.12}$$

Similarly, we can find the unit rotation matrices about the Y and Z-axes for a rotation about an arbitrary angle α:

$$\mathbf{C_y}(\alpha) = \begin{bmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{bmatrix} \tag{4.13}$$

$$\mathbf{C_z}(\alpha) = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.14}$$

Any rotation from $\mathcal{F}_A$ to $\mathcal{F}_B$ can always be decomposed into a number of sequential unit-axis rotations. This implies that the resulting transformation matrix can be written as a combination of $\mathbf{C_x}$, $\mathbf{C_y}$ and $\mathbf{C_z}$. Each of the matrices $\mathbf{C_x}$, $\mathbf{C_y}$ and $\mathbf{C_z}$ is orthonormal. Also the product of orthonormal matrices is orthonormal again. The inverse of this kind of matrix is simply its transpose.

For the sake of convenience, we will introduce a short-hand notation for the resulting transformation matrices. A matrix written as $\mathbf{C_{B,A}}$ means that the matrix defines the transformation from $\mathcal{F}_A$ to $\mathcal{F}_B$. Using the same convention, we can write for the inverse transformation matrix $\mathbf{C_{A,B}}$:

$$\mathbf{C_{A,B}} = \mathbf{C_{B,A}^{-1}} = \mathbf{C_{B,A}^{T}} \tag{4.15}$$

As an example we will give the transformation matrices from an arbitrary frame A to B and vice versa. The transformation comprehends three transformations about different axes (we use the angles α_1, α_2 and α_3 for the rotation angles about the corresponding axes). Successively, we rotate around the negative X-axis, the negative Z-axis and the positive Y-axis. The resulting transformation matrix $\mathbf{C_{B,A}}$ is obtained as follows:

$$\mathbf{C_{B,A}} = \mathbf{C_y}(\alpha_2)\mathbf{C_z}(-\alpha_3)\mathbf{C_x}(-\alpha_1) \tag{4.16}$$

The inverse matrix, $\mathbf{C_{A,B}}$, is now given by

$$\mathbf{C_{A,B}} = \mathbf{C_{B,A}^{-1}} = \mathbf{C_x}(\alpha_1)\mathbf{C_z}(\alpha_3)\mathbf{C_y}(-\alpha_2). \tag{4.17}$$

*4.3.2 *Euler's Rotation Theorem*

The Swiss mathematician Leonard Euler (1707–1783) proved in 1776 that a rotation between any two frames can be expressed by a rotation of the angle Φ about an axis through their common origin, $\hat{\mathbf{a}}$. The angle Φ is known as the Euler angle, whereas the rotation axis has the direction of a unit vector and is known as the Euler axis. Both Φ and $\hat{\mathbf{a}}$ have been visualised in Fig. 4.10.

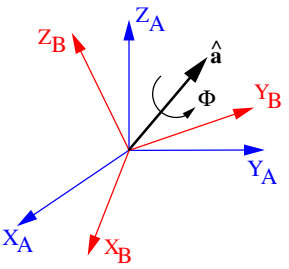


Fig. 4.10 Definition of the Euler axis and angle

Let us now define an arbitrary transformation matrix $\mathbf{C}$:

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

To find the Euler axis and angle, it is necessary to carry out an eigenvalue analysis on $\mathbf{C}$. Let $\mathbf{C}$ have the eigenvalue λ and corresponding eigenvector $\boldsymbol{\mu}$, which, by definition, obeys:

$$\mathbf{C}\boldsymbol{\mu} = \lambda\boldsymbol{\mu} \tag{4.18}$$

Both sides of Eq. (4.18) are now left-multiplied by their respective Hermitian conjugates,[4] which yields

$$(\mathbf{C}\boldsymbol{\mu})^{\mathrm{H}}(\mathbf{C}\boldsymbol{\mu}) = \bar{\lambda}\lambda\boldsymbol{\mu}^{\mathrm{H}}\boldsymbol{\mu} \quad \Rightarrow \quad \boldsymbol{\mu}^{\mathrm{H}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\mu} = \bar{\lambda}\lambda\boldsymbol{\mu}^{\mathrm{H}}\boldsymbol{\mu} \quad \Rightarrow \quad \boldsymbol{\mu}^{\mathrm{H}}\boldsymbol{\mu} = \bar{\lambda}\lambda\boldsymbol{\mu}^{\mathrm{H}}\boldsymbol{\mu} \tag{4.21}$$

resulting in

$$\left(\bar{\lambda}\lambda - 1\right)\boldsymbol{\mu}^{H}\boldsymbol{\mu} = 0 \tag{4.22}$$

Because $\mathbf{C}$ is orthonormal, it will have three different eigenvalues and eigenvectors. With this property and the fact that $\det \mathbf{C} = 1$, it can be deduced that one of the eigenvalues in Eq. (4.22) is 1, and the other two are complex conjugates of each other. Hughes (1986) states that these can be expressed as

[4] The Hermitian conjugate of a matrix $\mathbf{A}$ is found by transposing it and then by taking the complex conjugate of each element. To illustrate this, let $\mathbf{A}$ be defined with complex components as:

$$\mathbf{A} = \begin{bmatrix} a + bi & c - di \\ e - fi & g + hi \end{bmatrix} \tag{4.19}$$

The Hermitian conjugate of $\mathbf{A}$ is then

$$\mathbf{A}^{\mathrm{H}} = \begin{bmatrix} a - bi & e + fi \\ c + di & g - hi \end{bmatrix} \tag{4.20}$$

.

$$\lambda_1 = 1 \quad \wedge \quad \lambda_{2,3} = e^{\pm i\Phi} \tag{4.23}$$

Using $\lambda_1 = 1$, Eq. (4.22) reduces to

$$\mathbf{C}\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1 \tag{4.24}$$

which means that $\boldsymbol{\mu}_1$ remains invariant through the rotation expressed by $\mathbf{C}$, which means that $\boldsymbol{\mu}_1$ is the Euler axis, $\hat{\mathbf{a}}$.

Similarly, substituting the other two eigenvalues yields

$$\mathbf{C}\boldsymbol{\mu}_{2,3} = e^{\pm i\Phi}\boldsymbol{\mu}_{2,3} \tag{4.25}$$

which means that the vectors undergo a rotation of $+\Phi$ and $-\Phi$ about $\hat{\mathbf{a}}$. Because all eigenvectors are mutually orthogonal, Φ is the angle of rotation and therefore the Euler angle.

To actually use Euler's theorem, we need to be able to extract the Euler axis and angle from a given rotation matrix. With the trace of a matrix being equal to the sum of its eigenvalues, Hughes (1986) shows that the trace of a rotation matrix can be formulated as

$$\sigma = \text{Tr}\,\mathbf{C} = c_{1,1} + c_{2,2} + c_{3,3} = \lambda_1 + \lambda_2 + \lambda_3 = 1 + e^{i\Phi} + e^{-i\Phi} = 1 + 2\cos\Phi \tag{4.26}$$

The cosine of the Euler angle can thus be found from the trace to be

$$\cos\Phi = \frac{1}{2}(\sigma - 1) \tag{4.27}$$

As is common with trigonometric equalities, there are (two) cases where special care has to be taken. These are

$\sigma = -1$: this is the case when $\Phi = \pm\pi, \pm 3\pi, \ldots$
$\sigma = 3$: this is the case when $\Phi = 0, \pm 2\pi, \pm 4\pi, \ldots$

Otherwise, outside these values when $\sigma \neq 3, -1$, the Euler axis is found to be

$$\hat{\mathbf{a}} = \frac{1}{2\sin\Phi}\begin{pmatrix} c_{2,3} - c_{3,2} \\ c_{3,1} - c_{1,3} \\ c_{1,2} - c_{2,1} \end{pmatrix} \tag{4.28}$$

Note that in case $\sigma = 3$, either no rotation or numerous full rotations occur, and the Euler axis cannot be computed. However, for the other special case when $\sigma = -1$, the Euler axis can be calculated as follows:

$$a_1 = \pm\sqrt{\frac{1 + c_{1,1}}{2}} \quad a_2 = \pm\sqrt{\frac{1 + c_{2,2}}{2}} \quad a_3 = \pm\sqrt{\frac{1 + c_{3,3}}{2}} \tag{4.29}$$

$$a_1 a_2 = \frac{c_{1,2}}{2} \quad a_2 a_3 = \frac{c_{2,3}}{2} \quad a_3 a_1 = \frac{c_{3,1}}{2} \tag{4.30}$$

where the latter three equalities allow for solving the disambiguity of the signs.

The direction cosine matrix can now be expressed in terms of $\hat{\mathbf{a}}$ and Φ, using "c" and "s" as short-hand notation for *cosine* and *sine*, respectively:

$$\mathbf{C} = \begin{bmatrix} \mathrm{c}\Phi + a_1^2(1-\mathrm{c}\Phi) & a_1 a_2(1-\mathrm{c}\Phi) + a_3\mathrm{s}\Phi & a_1 a_3(1-\mathrm{c}\Phi) - a_2\mathrm{s}\Phi \\ a_1 a_2(1-\mathrm{c}\Phi) - a_3\mathrm{s}\Phi & \mathrm{c}\Phi + a_2^2(1-\mathrm{c}\Phi) & a_2 a_3(1-\mathrm{c}\Phi) + a_1\mathrm{s}\Phi \\ a_1 a_3(1-\mathrm{c}\Phi) + a_2\mathrm{s}\Phi & a_2 a_3(1-\mathrm{c}\Phi) - a_1\mathrm{s}\Phi & \mathrm{c}\Phi + a_3^2(1-\mathrm{c}\Phi) \end{bmatrix} \tag{4.31}$$

and in a more compact form as

$$\mathbf{C} = \mathbf{I_3}\cos\Phi + (1-\cos\Phi)\hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathrm{T}} - \left[\hat{\mathbf{a}}\times\right]\sin\Phi \tag{4.32}$$

In Eq. (4.32), $\mathbf{I_3}$ is a 3×3 identity matrix and the term $\left[\hat{\mathbf{a}}\times\right]$ is the skewsymmetric form of the Euler axis, and is defined as

$$\left[\hat{\mathbf{a}}\times\right] = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \tag{4.33}$$

Subsequent rotations are easily obtained as well. As was discussed earlier, linking unit-axis transformations will allow for obtaining a complex transformation. This will also be the case now. Suppose that two transformations $\mathbf{C_{B,A}}$ and $\mathbf{C_{C,B}}$ are defined by:

$$\mathbf{C_{B,A}} = \mathbf{I_3}\cos\Phi_1 + (1-\cos\Phi_1)\hat{\mathbf{a}}_\mathbf{1}\hat{\mathbf{a}}_\mathbf{1}^{\mathrm{T}} - \left[\hat{\mathbf{a}}_\mathbf{1}\times\right]\sin\Phi_1 \tag{4.34}$$

and

$$\mathbf{C_{C,B}} = \mathbf{I_3}\cos\Phi_2 + (1-\cos\Phi_2)\hat{\mathbf{a}}_\mathbf{2}\hat{\mathbf{a}}_\mathbf{2}^{\mathrm{T}} - \left[\hat{\mathbf{a}}_\mathbf{2}\times\right]\sin\Phi_2 \tag{4.35}$$

then the compound transformation matrix, $\mathbf{C_{C,A}}$, simply follows from

$$\mathbf{C_{C,A}} = \mathbf{C_{C,B}}\mathbf{C_{B,A}} \tag{4.36}$$

For these transformations, one needs to identify the Euler axes $\hat{\mathbf{a}}_\mathbf{1}$ and $\hat{\mathbf{a}}_\mathbf{2}$, as well as the corresponding rotation angles Φ_1 and Φ_2.

Example 4.1 (*Euler's Rotation Theorem and Unit-Axis Transformations*) To illustrate the use of Euler's rotation theorem, let us consider the rotation from $\mathcal{F}_A$ to $\mathcal{F}_B$, considering a rotation about the Z_A-axis over an angle α. This is, of course, a unit-axis transformation as discussed in Sect. 4.3.1, with the result for this example given by Eq. (4.14). The rotation axis for the current example is, as specified, the Z_A-axis. Thus, the Euler axis is given by

$$\hat{\mathbf{a}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{4.37}$$

and the Euler angle is defined by $\Phi = \alpha$.

Substituting $\hat{\mathbf{a}}$ and Φ into Eq. (4.34) gives

$$\mathbf{C_{B,A}} = \begin{bmatrix} \cos\alpha & 0 & 0 \\ 0 & \cos\alpha & 0 \\ 0 & 0 & \cos\alpha \end{bmatrix} + (1 - \cos\alpha) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} - \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \sin\alpha \tag{4.38}$$

which simplifies to

$$\mathbf{C_{B,A}} = \begin{bmatrix} \cos\alpha & 0 & 0 \\ 0 & \cos\alpha & 0 \\ 0 & 0 & \cos\alpha \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & (1 - \cos\alpha) \end{bmatrix} - \begin{bmatrix} 0 & -\sin\alpha & 0 \\ \sin\alpha & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.39}$$

and finally to

$$\mathbf{C_{B,A}} = \mathbf{C_z}(\alpha) = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.40}$$

which is the sought-for unit-axis transformation matrix. ■

4.3.3 Standard Frame Transformations

In this section, a number of basic frame transformations is introduced. These transformations are commonly used to derive the equations of translational and rotational motion.

Inertial Planetocentric Frame to Orbital Frames, $\mathbf{C_{O,I}}$ and $\mathbf{C_{L,I}}$

To go from the inertial planetocentric frame to the local orbital frame, i.e., from $\mathcal{F}_I$ to $\mathcal{F}_L$, three successive rotations are required. The first rotation will have the positive X_I-axis intersect with the ascending node (see Fig. 4.2). To reach the line of nodes, starting in $\mathcal{F}_I$, one rotates around the Z_I axis over an angle Ω. Next, a rotation over i takes place around the line of nodes, the X-axis of an intermediate frame, and will align the equatorial with the orbital plane. After that, a combined rotation around the local Z-axis over $\omega + \theta$ will bring the local frame to $\mathcal{F}_L$. Thus:
Related variables:

Ω = longitude of ascending node (rad)
i = inclination (rad)

ω = argument of pericentre (rad)
θ = true anomaly (rad)

Transformation matrix:

$$\mathbf{C}_{\mathbf{L,I}} = \mathbf{C}_{\mathbf{z}}(\omega + \theta)\mathbf{C}_{\mathbf{x}}(i)\mathbf{C}_{\mathbf{z}}(\Omega) \tag{4.41}$$

However, during these individual rotations $\mathcal{F}_O$ was never reached. To achieve this, the equatorial plane needs to be rotated *before* the rotation over Ω. Therefore, a different approach will be used, as we are missing intermediate angles. In Sect. 4.3.2, Euler's Rotation Theorem was discussed. This theorem lies at the root of all frame transformations, albeit a bit less transparent compared with the unit-axis transformations that we will predominantly use. To be able to use this theorem, the Euler axis, $\hat{\mathbf{a}}$, and Euler angle, Φ, need to be established.

To align the two planes, a single rotation around the line of nodes is required (see also Fig. 4.11), which is thus the Euler axis for the rotation:

$$\hat{\mathbf{a}} = \begin{pmatrix} \cos\Omega \\ \sin\Omega \\ 0 \end{pmatrix} \tag{4.42}$$

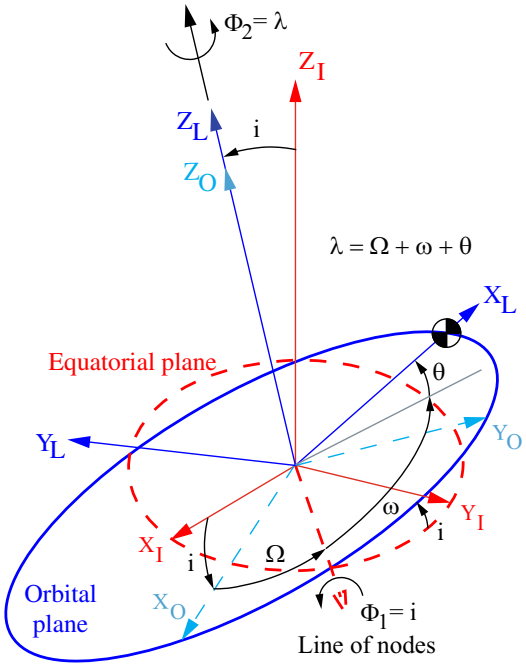


Fig. 4.11 Relation between the inertial planetocentric frame (index I) and the two orbital frames (index O and L, respectively)

The corresponding Euler angle is, as before, $\Phi = i$. Thus, substituting both into Eq. (4.34) results in:

$$\mathbf{C_{O,I}} = \begin{bmatrix} \cos i & 0 & 0 \\ 0 & \cos i & 0 \\ 0 & 0 & \cos i \end{bmatrix} + (1 - \cos i) \begin{pmatrix} \cos\Omega \\ \sin\Omega \\ 0 \end{pmatrix} \begin{pmatrix} \cos\Omega & \sin\Omega & 0 \end{pmatrix} + \\ - \begin{bmatrix} 0 & 0 & \sin\Omega \\ 0 & 0 & -\cos\Omega \\ -\sin\Omega & \cos\Omega & 0 \end{bmatrix} \sin i \tag{4.43}$$

which is rewritten as:

$$\mathbf{C_{O,I}} = \begin{bmatrix} \cos i & 0 & 0 \\ 0 & \cos i & 0 \\ 0 & 0 & \cos i \end{bmatrix} + (1 - \cos i) \begin{bmatrix} \cos^2\Omega & \cos\Omega\sin\Omega & 0 \\ \sin\Omega\cos\Omega & \sin^2\Omega & 0 \\ 0 & 0 & 0 \end{bmatrix} + \\ + \begin{bmatrix} 0 & 0 & -\sin\Omega\sin i \\ 0 & 0 & \cos\Omega\sin i \\ \sin\Omega\sin i & -\cos\Omega\sin i & 0 \end{bmatrix} \tag{4.44}$$

Combining all terms, the end result is

$$\mathbf{C_{O,I}} = \begin{bmatrix} \cos i + \cos^2\Omega(1 - \cos i) & \cos\Omega\sin\Omega(1 - \cos i) & -\sin\Omega\sin i \\ \cos\Omega\sin\Omega(1 - \cos i) & \cos i + \sin^2\Omega(1 - \cos i) & \cos\Omega\sin i \\ \sin\Omega\sin i & -\cos\Omega\sin i & \cos i \end{bmatrix} \tag{4.45}$$

Once the equatorial and orbital planes are aligned, only one rotation will suffice to go from $\mathcal{F}_O$ to $\mathcal{F}_L$, see again Fig. 4.11. This will be rotation about the Z_O-axis (which is thus equal to the Z_L-axis) over the sum of angles $\Omega + \omega + \theta$. The first angle will bring the intermediate X-axis to the line of nodes, the second one will point it to the pericentre and, finally, the third one to the location of the spacecraft.

For ease of notation, we will introduce the auxiliary angle λ:

$$\lambda = \Omega + u \tag{4.46}$$

with u being the argument of latitude, i.e., the sum of the argument of perigee and the true anomaly:

$$u = \omega + \theta \tag{4.47}$$

Thus, the rotation from $\mathcal{F}_O$ to $\mathcal{F}_L$ is simply:

$$\mathbf{C_{L,O}} = \mathbf{C_z}(\lambda) = \begin{bmatrix} \cos\lambda & \sin\lambda & 0 \\ -\sin\lambda & \cos\lambda & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.48}$$

It may be clear that equivalent of the transformation matrix given by Eq. (4.41) is also found from

$$\mathbf{C_{L,I}} = \mathbf{C_{L,O}}\mathbf{C_{O,I}} \tag{4.49}$$

with the individual matrices defined by Eqs. (4.45) and (4.48).

It is noted that the two currently discussed transformation matrices are not commonly used for the study of re-entry flight. However, in Chap. 11 some aero manoeuvres are discussed, most notably in Sect. 11.1, where the concept of aerobraking is introduced. This manoeuvre is a combination of orbital flight outside the atmosphere and an atmospheric braking phase, and discussing this requires these two transformation matrices.

Rotating Planetocentric to Inertial Planetocentric Frame, $\mathbf{C_{I,R}}$

By definition, the two frames are coinciding at $t_0 = 0$. $\mathcal{F}_R$ is fixed to the central body and is therefore rotating (about the Z_R-axis) with the same rotational rate as the central body. The transformation matrix for an arbitrary time t can easily be derived.

Related variables (Fig. 4.12):

ω_{cb} = rotational rate of the central body (rad/s)
t = time from epoch (s)

Transformation matrix:

$$\mathbf{C_{I,R}} = \mathbf{C_z}(-\omega_{cb}t) = \begin{bmatrix} \cos\omega_{cb}t & -\sin\omega_{cb}t & 0 \\ \sin\omega_{cb}t & \cos\omega_{cb}t & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.50}$$

Vertical to Rotating Planetocentric Frame, $\mathbf{C_{R,V}}$

Related variables (Fig. 4.12):

τ = planetocentric longitude (rad)
δ = planetocentric latitude (rad)

Transformation matrix:

$$\mathbf{C_{R,V}} = \mathbf{C_z}(-\tau)\mathbf{C_y}\left(\frac{\pi}{2}+\delta\right) = \begin{bmatrix} -\mathrm{c}\tau\mathrm{s}\delta & -\mathrm{s}\tau & -\mathrm{c}\tau\mathrm{c}\delta \\ -\mathrm{s}\tau\mathrm{s}\delta & \mathrm{c}\tau & -\mathrm{s}\tau\mathrm{c}\delta \\ \mathrm{c}\delta & 0 & -\mathrm{s}\delta \end{bmatrix} \tag{4.51}$$

In the above expression, "c" has been used for *cosine* and "s" for *sine*. This notation is occasionally used for clarity and compactness.

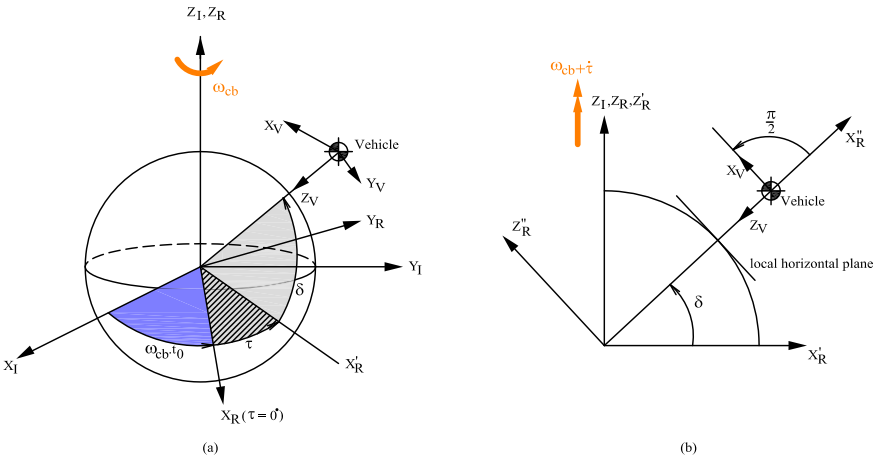


Fig. 4.12 Relation between the inertial and rotating planetocentric frames (index I and R, respectively) and the vertical frame (index V)

Wind to Vertical Frame, $\mathbf{C_{V,W}}$

Related variables (Fig. 4.13):

γ_w = flight-path angle for wind vector (rad)
χ_w = heading for wind vector (rad)

Transformation matrix:

$$\mathbf{C_{V,W}} = \mathbf{C_z}(-\chi_w)\mathbf{C_y}(-\gamma_w) = \begin{bmatrix} \mathrm{c}\chi_w\mathrm{c}\gamma_w & -\mathrm{s}\chi_w & \mathrm{c}\chi_w\mathrm{s}\gamma_w \\ \mathrm{s}\chi_w\mathrm{c}\gamma_w & \mathrm{c}\chi_w & \mathrm{s}\chi_w\mathrm{s}\gamma_w \\ -\mathrm{s}\gamma_w & 0 & \mathrm{c}\gamma_w \end{bmatrix} \tag{4.52}$$

(Groundspeed-Based) Trajectory to Vertical Frame, $\mathbf{C_{V,TG}}$

Related variables (Fig. 4.14):

γ_g = flight-path angle for groundspeed (rad)
χ_g = heading for groundspeed (rad)

Transformation matrix:

$$\mathbf{C_{V,TG}} = \mathbf{C_z}(-\chi_g)\mathbf{C_y}(-\gamma_g) = \begin{bmatrix} \mathrm{c}\chi_\mathrm{g}\mathrm{c}\gamma_\mathrm{g} & -\mathrm{s}\chi_\mathrm{g} & \mathrm{c}\chi_\mathrm{g}\mathrm{s}\gamma_\mathrm{g} \\ \mathrm{s}\chi_\mathrm{g}\mathrm{c}\gamma_\mathrm{g} & \mathrm{c}\chi_\mathrm{g} & \mathrm{s}\chi_\mathrm{g}\mathrm{s}\gamma_\mathrm{g} \\ -\mathrm{s}\gamma_\mathrm{g} & 0 & \mathrm{c}\gamma_\mathrm{g} \end{bmatrix} \tag{4.53}$$

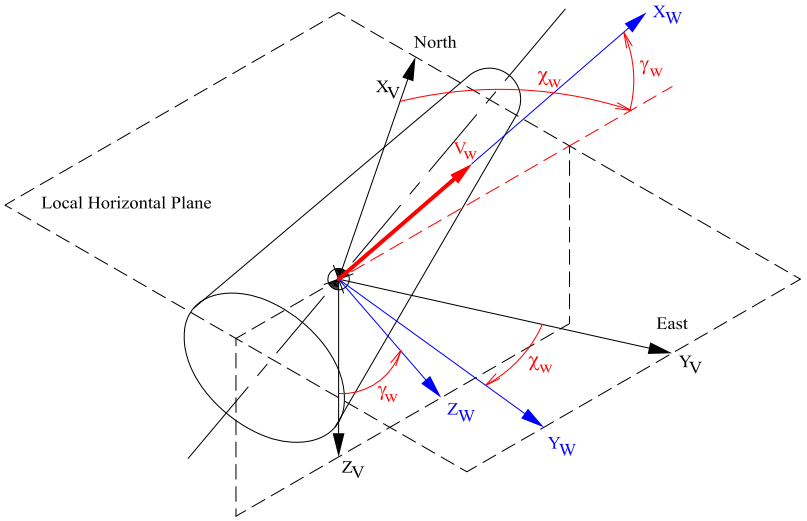


Fig. 4.13 The relation between the wind frame (index W) and the vertical frame (index V)

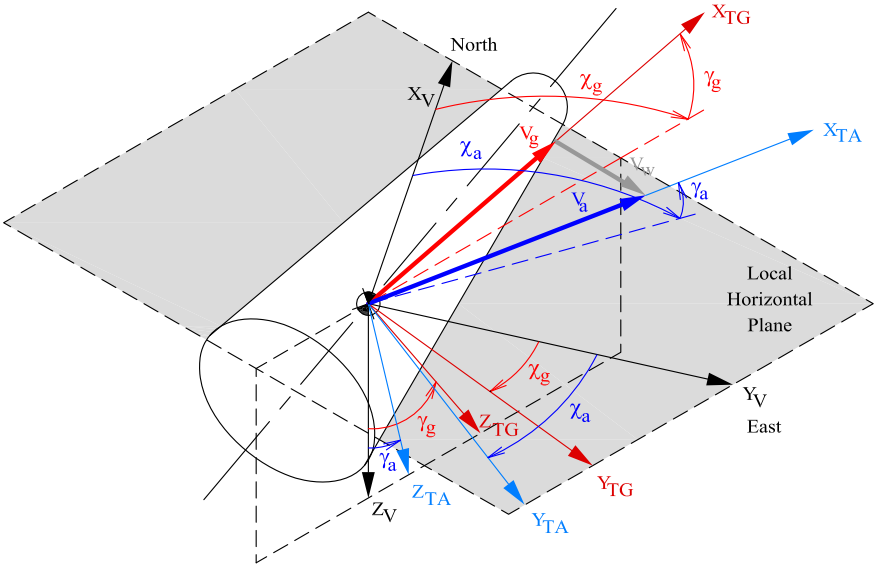


Fig. 4.14 The relation between the groundspeed- and airspeed-based trajectory frame (index TG and TA, respectively) and the vertical frame (index V)

(Airspeed-Based) Trajectory to Vertical Frame, $\mathbf{C}_{\mathbf{V,TA}}$

Related variables (Fig. 4.14):

γ_g = flight-path angle for groundspeed (rad)
χ_g = heading for groundspeed (rad)

Transformation matrix:

$$\mathbf{C}_{\mathbf{V,TA}} = \mathbf{C}_{\mathbf{z}}(-\chi_a)\mathbf{C}_{\mathbf{y}}(-\gamma_a) = \begin{bmatrix} \mathrm{c}\chi_a\mathrm{c}\gamma_a & -\mathrm{s}\chi_a & \mathrm{c}\chi_a\mathrm{s}\gamma_a \\ \mathrm{s}\chi_a\mathrm{c}\gamma_a & \mathrm{c}\chi_a & \mathrm{s}\chi_a\mathrm{s}\gamma_a \\ -\mathrm{s}\gamma_a & 0 & \mathrm{c}\gamma_a \end{bmatrix} \tag{4.54}$$

(Airspeed-Based) Aerodynamic to (Airspeed-Based) Trajectory Frame, $\mathbf{C}_{\mathbf{TA,AA}}$

Related variable (Fig. 4.15):

σ_a = bank-angle based on airspeed (rad)

Transformation matrix:

$$\mathbf{C}_{\mathbf{TA,AA}} = \mathbf{C}_{\mathbf{x}}(\sigma_a) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\sigma_a & \sin\sigma_a \\ 0 & -\sin\sigma_a & \cos\sigma_a \end{bmatrix} \tag{4.55}$$

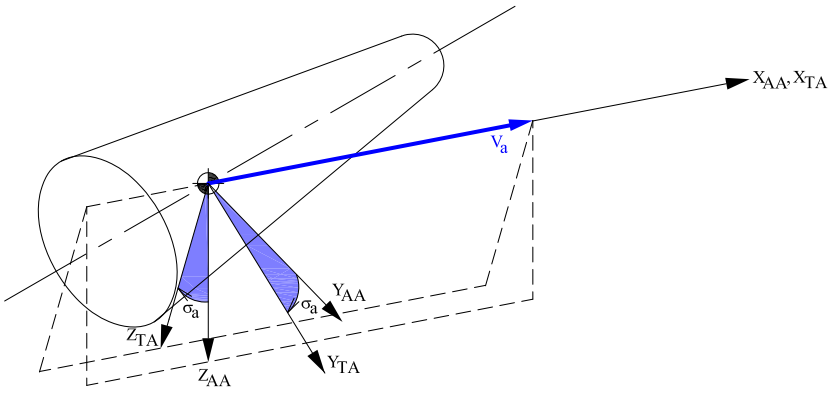


Fig. 4.15 The relation between the airspeed-based aerodynamic frame (index AA) and the airspeed-based trajectory frame (index TA)

(Airspeed-Based) Aerodynamic to Vertical Frame, $\mathbf{C}_{\mathbf{V,AA}}$

Based on the transformation matrices given by Eqs. (4.54)–(4.55), the transformation matrix from the vertical to the aerodynamic frame, $\mathbf{C}_{\mathbf{V,AA}}$, can easily be derived.

$$\begin{aligned}\mathbf{C}_{\mathbf{V,AA}} &= \mathbf{C}_{\mathbf{V,TA}}\mathbf{C}_{\mathbf{TA,AA}} = \mathbf{C}_{\mathbf{z}}(-\chi_a)\mathbf{C}_{\mathbf{y}}(-\gamma_a)\mathbf{C}_{\mathbf{x}}(\sigma_a)\\ &= \begin{bmatrix} \mathrm{c}\chi_a\mathrm{c}\gamma_a & -\mathrm{s}\chi_a\mathrm{c}\sigma_a - \mathrm{c}\chi_a\mathrm{s}\gamma_a\mathrm{s}\sigma_a & -\mathrm{s}\chi_a\mathrm{s}\sigma_a + \mathrm{c}\chi_a\mathrm{s}\gamma_a\mathrm{c}\sigma_a \\ \mathrm{s}\chi_a\mathrm{c}\gamma_a & \mathrm{c}\chi_a\mathrm{c}\sigma_a - \mathrm{s}\chi_a\mathrm{s}\gamma_a\mathrm{s}\sigma_a & \mathrm{c}\chi_a\mathrm{s}\sigma_a + \mathrm{s}\chi_a\mathrm{s}\gamma_a\mathrm{c}\sigma_a \\ -\mathrm{s}\gamma_a & -\mathrm{c}\gamma_a\mathrm{s}\sigma_a & \mathrm{c}\gamma_a\mathrm{c}\sigma_a \end{bmatrix}\end{aligned} \tag{4.56}$$

(Airspeed-Based) Aerodynamic to (Groundspeed-Based) Trajectory Frame, $\mathbf{C}_{\mathbf{TG,AA}}$

Using the above result together with the inverse of Eq. (4.53), we obtain for $\mathbf{C}_{\mathbf{TG,AA}}$:

$$\begin{aligned}\mathbf{C}_{\mathbf{TG,AA}} &= \mathbf{C}_{\mathbf{TG,V}}\mathbf{C}_{\mathbf{V,AA}}\\ &= \mathbf{C}_{\mathbf{y}}(\gamma_g)\mathbf{C}_{\mathbf{z}}(\chi_g)\mathbf{C}_{\mathbf{z}}(-\chi_a)\mathbf{C}_{\mathbf{y}}(-\gamma_a)\mathbf{C}_{\mathbf{x}}(\sigma_a)\\ &= \mathbf{C}_{\mathbf{y}}(\gamma_g)\mathbf{C}_{\mathbf{z}}(\Delta\chi_{ga})\mathbf{C}_{\mathbf{y}}(-\gamma_a)\mathbf{C}_{\mathbf{x}}(\sigma_a)\end{aligned} \tag{4.57}$$

with $\Delta\chi_{ga} = \chi_g - \chi_a$. Carrying out the matrix multiplications, we find for the above transformation matrix:

$$\mathbf{C}_{\mathbf{TG,AA}} = \begin{bmatrix} \mathrm{c}\gamma_g\mathrm{c}\Delta\chi_{ga}\mathrm{c}\gamma_a + \mathrm{s}\gamma_g\mathrm{s}\gamma_a & -\mathrm{c}\gamma_g\mathrm{c}\Delta\chi_{ga}\mathrm{s}\gamma_a\mathrm{s}\sigma_a + \mathrm{c}\gamma_g\mathrm{s}\Delta\chi_{ga}\mathrm{c}\sigma_a + \mathrm{s}\gamma_g\mathrm{c}\gamma_a\mathrm{s}\sigma_a & \cdots \\ -\mathrm{s}\Delta\chi_{ga}\mathrm{c}\gamma_a & \mathrm{s}\Delta\chi_{ga}\mathrm{s}\gamma_a\mathrm{s}\sigma_a + \mathrm{c}\Delta\chi_{ga}\mathrm{c}\sigma_a & \cdots \\ \mathrm{s}\gamma_g\mathrm{c}\Delta\chi_{ga}\mathrm{c}\gamma_a - \mathrm{c}\gamma_g\mathrm{s}\gamma_a & -\mathrm{s}\gamma_g\mathrm{c}\Delta\chi_{ga}\mathrm{s}\gamma_a\mathrm{s}\sigma_a + \mathrm{s}\gamma_g\mathrm{s}\Delta\chi_{ga}\mathrm{c}\sigma_a - \mathrm{c}\gamma_g\mathrm{c}\gamma_a\mathrm{s}\sigma_a & \cdots \end{bmatrix}$$

$$\begin{bmatrix} \cdots & \mathrm{c}\gamma_g\mathrm{c}\Delta\chi_{ga}\mathrm{s}\gamma_a\mathrm{c}\sigma_a + \mathrm{c}\gamma_g\mathrm{s}\Delta\chi_{ga}\mathrm{s}\sigma_a - \mathrm{s}\gamma_g\mathrm{c}\gamma_a\mathrm{c}\sigma_a \\ \cdots & -\mathrm{s}\Delta\chi_{ga}\mathrm{s}\gamma_a\mathrm{c}\sigma_a + \mathrm{c}\Delta\chi_{ga}\mathrm{s}\sigma_a \\ \cdots & \mathrm{s}\gamma_g\mathrm{c}\Delta\chi_{ga}\mathrm{s}\gamma_a\mathrm{c}\sigma_a + \mathrm{s}\gamma_g\mathrm{s}\Delta_{ga}\mathrm{s}\sigma_a + \mathrm{c}\gamma_g\mathrm{c}\gamma_a\mathrm{c}\sigma_a \end{bmatrix} \tag{4.58}$$

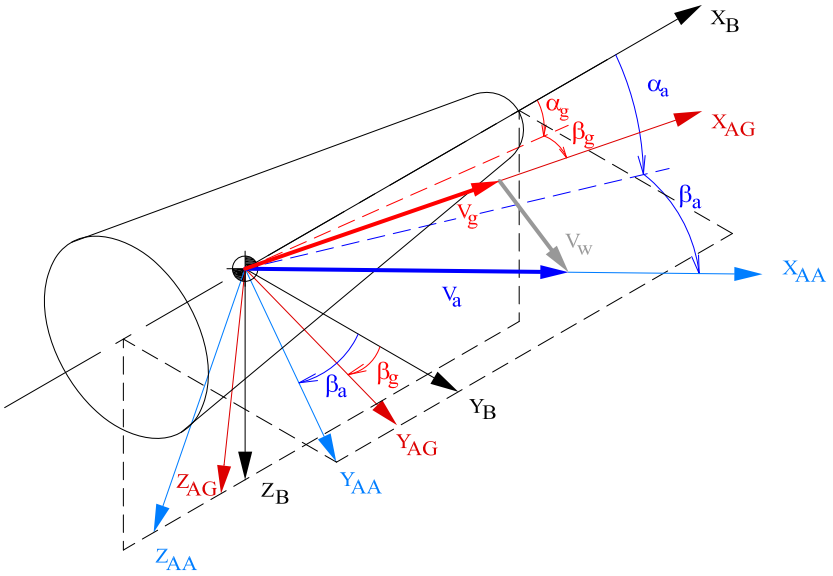


Fig. 4.16 The relation between the body frame (index B) and the airspeed- and groundspeed-based aerodynamic frames (index AA and AG, respectively). Here, all angles are plotted in the corresponding positive directions

Body to (Airspeed-Based) Aerodynamic Frame, $\mathbf{C_{AA,B}}$

Related variables (Fig. 4.16):

α_a = angle of attack with respect to airspeed (rad)
β_a = angle of sideslip with respect to airspeed (rad)

Transformation matrix:

$$\mathbf{C_{AA,B}} = \mathbf{C_z}(\beta_a)\mathbf{C_y}(-\alpha_a) = \begin{bmatrix} \mathrm{c}\alpha_a\mathrm{c}\beta_a & \mathrm{s}\beta_a & \mathrm{s}\alpha_a\mathrm{c}\beta_a \\ -\mathrm{c}\alpha_a\mathrm{s}\beta_a & \mathrm{c}\beta_a & -\mathrm{s}\alpha_a\mathrm{s}\beta_a \\ -\mathrm{s}\alpha_a & 0 & \mathrm{c}\alpha_a \end{bmatrix} \tag{4.59}$$

Body to (Groundspeed-Based) Aerodynamic Frame, $\mathbf{C_{AG,B}}$

Related variables (Fig. 4.16):

α_g = angle of attack with respect to groundspeed (rad)
β_g = angle of sideslip with respect to groundspeed (rad)

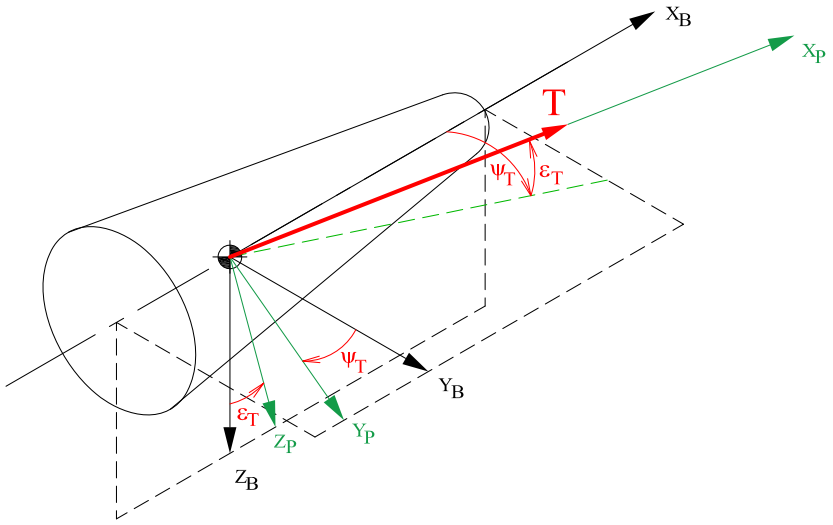


Fig. 4.17 The relation between the propulsion frame (index P) and the body frame (index B). The indicated angles ϵ_T and ψ_T are positive

Transformation matrix:

$$\mathbf{C_{AG,B}} = \mathbf{C_z}(\beta_g)\mathbf{C_y}(-\alpha_g) = \begin{bmatrix} \mathrm{c}\alpha_g\mathrm{c}\beta_g & \mathrm{s}\beta_g & \mathrm{s}\alpha_g\mathrm{c}\beta_g \\ -\mathrm{c}\alpha_g\mathrm{s}\beta_g & \mathrm{c}\beta_g & -\mathrm{s}\alpha_g\mathrm{s}\beta_g \\ -\mathrm{s}\alpha_g & 0 & \mathrm{c}\alpha_g \end{bmatrix} \tag{4.60}$$

Body to Propulsion Frame, $\mathbf{C_{P,B}}$

Related variables (Fig. 4.17):

ϵ_T = elevation angle of thrust force (rad)
ψ_T = azimuth angle of thrust force (rad)

Note that the definition of ϵ_T is such, that a positive angle will contribute to a positive change in the flight-path angle, which results in an increase of the altitude. Similarly, ψ_T is defined such, that a positive value of this angle will tend to increase the heading of the vehicle, e.g., when it is flying to the north, a positive azimuth of the thrust gives rise to a more eastern trajectory.

When the thrust force is large compared with the other forces, and this thrust is assumed to act through the c.o.m. of the vehicle, the vehicle tends to take such an attitude, that the thrust vector will be aligned with the velocity vector. For the aerodynamic angles this means, that a positive ψ_T results in a positive slip angle, and a positive ϵ_T in a negative angle of attack.

Transformation matrix:

$$\mathbf{C_{P,B}} = \mathbf{C_y}(\epsilon_T)\mathbf{C_z}(\psi_T) = \begin{bmatrix} \mathrm{c}\psi_T\mathrm{c}\epsilon_T & \mathrm{s}\psi_T\mathrm{c}\epsilon_T & -\mathrm{s}\epsilon_T \\ -\mathrm{s}\psi_T & \mathrm{c}\psi_T & 0 \\ \mathrm{c}\psi_T\mathrm{s}\epsilon_T & \mathrm{s}\psi_T\mathrm{s}\epsilon_T & \mathrm{c}\epsilon_T \end{bmatrix} \tag{4.61}$$

With all the transformations given above, some combined transformations can be easily derived.

Body to Trajectory Frame, $\mathbf{C_{T,B}}$

The transformation from the body to the trajectory frame comes in two variations, i.e., one based on groundspeed variables and one based on airspeed variables. The structure is the same for both, only the indices will be different. Here, we will show the generic form. The matrix is composed of:

$$\mathbf{C_{T,B}} = \mathbf{C_{T,A}}\mathbf{C_{A,B}} = \mathbf{C_x}(\sigma)\mathbf{C_z}(\beta)\mathbf{C_y}(-\alpha) \tag{4.62}$$

with the airspeed variant of $\mathbf{C_{T,A}}$ shown in Fig. 4.15, and both variants of $\mathbf{C_{A,B}}$ shown in Fig. 4.16. Substituting Eqs. (4.55) and (4.60), without the subscripts, into Eq. (4.62) yields:

$$\mathbf{C_{T,B}} = \begin{bmatrix} \mathrm{c}\alpha\mathrm{c}\beta & \mathrm{s}\beta & \mathrm{s}\alpha\mathrm{c}\beta \\ -\mathrm{c}\alpha\mathrm{s}\beta\mathrm{c}\sigma - \mathrm{s}\alpha\mathrm{s}\sigma & \mathrm{c}\beta\mathrm{c}\sigma & \mathrm{c}\alpha\mathrm{s}\sigma - \mathrm{s}\alpha\mathrm{s}\beta\mathrm{c}\sigma \\ \mathrm{c}\alpha\mathrm{s}\beta\mathrm{s}\sigma - \mathrm{s}\alpha\mathrm{c}\sigma & -\mathrm{c}\beta\mathrm{s}\sigma & \mathrm{c}\alpha\mathrm{c}\sigma + \mathrm{s}\alpha\mathrm{s}\beta\mathrm{s}\sigma \end{bmatrix} \tag{4.63}$$

Vertical to Inertial Planetocentric Frame, $\mathbf{C_{I,V}}$

$$\begin{aligned} \mathbf{C_{I,V}} &= \mathbf{C_{I,R}}\mathbf{C_{R,V}} = \mathbf{C_z}(-\omega_{cb}t)\mathbf{C_z}(-\tau)\mathbf{C_y}(\tfrac{\pi}{2}+\delta) = \mathbf{C_z}(-\tilde{\tau})\mathbf{C_y}(\tfrac{\pi}{2}+\delta) \\ &= \begin{bmatrix} -\mathrm{c}\tilde{\tau}s\delta & -\mathrm{s}\tilde{\tau} & -\mathrm{c}\tilde{\tau}c\delta \\ -\mathrm{s}\tilde{\tau}s\delta & \mathrm{c}\tilde{\tau} & -\mathrm{s}\tilde{\tau}c\delta \\ \mathrm{c}\delta & 0 & -\mathrm{s}\delta \end{bmatrix} \end{aligned} \tag{4.64}$$

with the celestial longitude $\tilde{\tau}$ defined as $\tilde{\tau} = \tau + \omega_{cb}t$.

(Airspeed-Based) Aerodynamic to Rotating Planetocentric Frame, $\mathbf{C_{R,AA}}$

$$\begin{aligned} \mathbf{C_{R,AA}} &= \mathbf{C_{R,V}}\mathbf{C_{V,AA}} \\ &= \mathbf{C_z}(-\tau)\mathbf{C_y}(\tfrac{\pi}{2}+\delta)\mathbf{C_z}(-\chi_a)\mathbf{C_y}(-\gamma_a)\mathbf{C_x}(\sigma_a) \end{aligned} \tag{4.65}$$

Body to Rotating Planetocentric Frame, $\mathbf{C}_{\mathbf{R,B}}$

$$\begin{aligned}\mathbf{C}_{\mathbf{R,B}} &= \mathbf{C}_{\mathbf{R,V}}\mathbf{C}_{\mathbf{V,AG}}\mathbf{C}_{\mathbf{AG,B}} \\ &= \mathbf{C}_{\mathbf{z}}(-\tau)\mathbf{C}_{\mathbf{y}}(\tfrac{\pi}{2}+\delta)\mathbf{C}_{\mathbf{z}}(-\chi_g)\mathbf{C}_{\mathbf{y}}(-\gamma_g)\mathbf{C}_{\mathbf{x}}(\sigma_g)\mathbf{C}_{\mathbf{z}}(\beta_g)\mathbf{C}_{\mathbf{y}}(-\alpha_g)\end{aligned} \tag{4.66}$$

or, if we base the transformation on airspeed variables,

$$\begin{aligned}\mathbf{C}_{\mathbf{R,B}} &= \mathbf{C}_{\mathbf{R,V}}\mathbf{C}_{\mathbf{V,AA}}\mathbf{C}_{\mathbf{AA,B}} \\ &= \mathbf{C}_{\mathbf{z}}(-\tau)\mathbf{C}_{\mathbf{y}}(\tfrac{\pi}{2}+\delta)\mathbf{C}_{\mathbf{z}}(-\chi_a)\mathbf{C}_{\mathbf{y}}(-\gamma_a)\mathbf{C}_{\mathbf{x}}(\sigma_a)\mathbf{C}_{\mathbf{z}}(\beta_a)\mathbf{C}_{\mathbf{y}}(-\alpha_a)\end{aligned} \tag{4.67}$$

(Groundspeed-Based) Trajectory to Wind Frame, $\mathbf{C}_{\mathbf{W,TG}}$

$\mathbf{C}_{\mathbf{W,TG}}$ follows from multiplying the inverse of $\mathbf{C}_{\mathbf{V,W}}$, Eq. (4.52), and $\mathbf{C}_{\mathbf{V,TG}}$, Eq. (4.53):

$$\begin{aligned}\mathbf{C}_{\mathbf{W,TG}} &= \mathbf{C}_{\mathbf{W,V}}\mathbf{C}_{\mathbf{V,TG}} \\ &= \mathbf{C}_{\mathbf{y}}(\gamma_w)\mathbf{C}_{\mathbf{z}}(\chi_w)\mathbf{C}_{\mathbf{z}}(-\chi_g)\mathbf{C}_{\mathbf{y}}(-\gamma_g) \\ &= \mathbf{C}_{\mathbf{y}}(\gamma_w)\mathbf{C}_{\mathbf{z}}(\Delta\chi_{wg})\mathbf{C}_{\mathbf{y}}(-\gamma_g)\end{aligned} \tag{4.68}$$

with $\Delta\chi_{wg} = \chi_w - \chi_g$. Carrying out the matrix multiplications yields for $\mathbf{C}_{\mathbf{W,TG}}$:

$$\mathbf{C}_{\mathbf{W,TG}} = \begin{bmatrix} \mathrm{c}\gamma_w\mathrm{c}\Delta\chi_{wg}\mathrm{c}\gamma_g + \mathrm{s}\gamma_w\mathrm{s}\gamma_g & \mathrm{c}\gamma_w\mathrm{s}\Delta\chi_{wg} & \cdots \\ -\mathrm{s}\Delta\chi_{wg}\mathrm{c}\gamma_g & \mathrm{c}\Delta\chi_{wg} & \cdots \\ \mathrm{s}\gamma_w\mathrm{c}\Delta\chi_{wg}\mathrm{c}\gamma_g - \mathrm{c}\gamma_w\mathrm{s}\gamma_g & \mathrm{s}\gamma_w\mathrm{s}\Delta\chi_{wg} & \cdots \\ \cdots & \mathrm{c}\gamma_w\mathrm{c}\Delta\chi_{wg}\mathrm{s}\gamma_g - \mathrm{s}\gamma_w\mathrm{c}\gamma_g \\ \cdots & -\mathrm{s}\Delta\chi_{wg}\mathrm{s}\gamma_g \\ \cdots & \mathrm{s}\gamma_w\mathrm{c}\Delta\chi_{wg}\mathrm{s}\gamma_g + \mathrm{c}\gamma_w\mathrm{c}\gamma_g \end{bmatrix} \tag{4.69}$$

(Airspeed-Based) Aerodynamic to Wind Frame, $\mathbf{C}_{\mathbf{W,AA}}$

$\mathbf{C}_{\mathbf{W,AA}}$ follows from multiplying the inverse of $\mathbf{C}_{\mathbf{V,W}}$, Eq. (4.52), and $\mathbf{C}_{\mathbf{V,AA}}$, Eq. (4.56):

$$\begin{aligned}\mathbf{C}_{\mathbf{W,AA}} &= \mathbf{C}_{\mathbf{W,V}}\mathbf{C}_{\mathbf{V,AA}} \\ &= \mathbf{C}_{\mathbf{y}}(\gamma_w)\mathbf{C}_{\mathbf{z}}(\chi_w)\mathbf{C}_{\mathbf{z}}(-\chi_a)\mathbf{C}_{\mathbf{y}}(-\gamma_a)\mathbf{C}_{\mathbf{x}}(\sigma_a) \\ &= \mathbf{C}_{\mathbf{y}}(\gamma_w)\mathbf{C}_{\mathbf{z}}(\Delta\chi_{wa})\mathbf{C}_{\mathbf{y}}(-\gamma_a)\mathbf{C}_{\mathbf{x}}(\sigma_a)\end{aligned} \tag{4.70}$$

with $\Delta\chi_{wa} = \chi_w - \chi_a$. The result after the matrix multiplications is:

$$\mathbf{C_{W,AA}} = \begin{bmatrix} \mathrm{c}\gamma_w\mathrm{c}\Delta\chi_{wa}\mathrm{c}\gamma_a + & -\mathrm{c}\gamma_w\mathrm{c}\Delta\chi_{wa}\mathrm{s}\gamma_a\mathrm{s}\sigma_a + & \cdots \\ +\,\mathrm{s}\gamma_w\mathrm{s}\gamma_a & +\mathrm{c}\gamma_w\mathrm{s}\Delta\chi_{wa}\mathrm{c}\sigma_a + & \\ & +\mathrm{s}\gamma_w\mathrm{c}\gamma_a\mathrm{s}\sigma_a & \\ -\mathrm{s}\Delta\chi_{wa}\mathrm{c}\gamma_a & \mathrm{s}\Delta\chi_{wa}\mathrm{s}\gamma_a\mathrm{s}\sigma_a + & \cdots \\ & +\,\mathrm{c}\Delta\chi_{wa}\mathrm{c}\sigma_a & \\ \mathrm{s}\gamma_w\mathrm{c}\Delta\chi_{wa}\mathrm{c}\gamma_a + & -\,\mathrm{s}\gamma_w\mathrm{c}\Delta\chi_{wa}\mathrm{s}\gamma_a\mathrm{s}\sigma_a + & \cdots \\ -\,\mathrm{c}\gamma_w\mathrm{s}\gamma_a & +\,\mathrm{s}\gamma_w\mathrm{s}\Delta\chi_{wa}\mathrm{c}\sigma_a + & \\ & -\,\mathrm{c}\gamma_w\mathrm{c}\gamma_a\mathrm{s}\sigma_a & \end{bmatrix} \begin{bmatrix} \cdots & \mathrm{c}\gamma_w\mathrm{c}\Delta\chi_{wa}\mathrm{s}\gamma_a\mathrm{c}\sigma_a + \\ & +\,\mathrm{c}\gamma_w\mathrm{s}\Delta\chi_{wa}\mathrm{s}\sigma_a + \\ & -\,\mathrm{s}\gamma_w\mathrm{c}\gamma_a\mathrm{c}\sigma_a \\ \cdots & -\,\mathrm{s}\Delta\chi_{wa}\mathrm{s}\gamma_a\mathrm{c}\sigma_a + \\ & +\,\mathrm{c}\Delta\chi_{wa}\mathrm{s}\sigma_a \\ \cdots & \mathrm{s}\gamma_w\mathrm{c}\Delta\chi_{wa}\mathrm{s}\gamma_a\mathrm{c}\sigma_a + \\ & +\,\mathrm{s}\gamma_w\mathrm{s}\Delta_{wa}\mathrm{s}\sigma_a + \\ & +\,\mathrm{c}\gamma_w\mathrm{c}\gamma_a\mathrm{c}\sigma_a \end{bmatrix} \tag{4.71}$$

4.4 Coordinate Transformations

4.4.1 Position and Velocity

The relation between the Cartesian position in $\mathcal{F}_I$ and $\mathcal{F}_R$ is given by:

$$\mathbf{r_R} = \mathbf{C_{R,I}}\mathbf{r_I} \tag{4.72}$$

with $\mathbf{C_{R,I}}$ given by the inverse of Eq. (4.50). Doing the matrix multiplication, this results in:

$$\begin{aligned} x_R &= \cos(\omega_{cb}t)x_I + \sin(\omega_{cb}t)y_I \\ y_R &= \sin(\omega_{cb}t)x_I + \cos(\omega_{cb}t)y_I \\ z_R &= z_I \end{aligned} \tag{4.73}$$

The velocity cannot directly be transformed, because $\mathcal{F}_R$ rotates with respect to $\mathcal{F}_I$, which creates an additional velocity component. This means that before we transform the velocity components from $\mathcal{F}_I$ to $\mathcal{F}_R$ the velocity contribution due to the rotation has to be subtracted. Starting with the inertial velocity $\mathbf{v_I}$, we get:

$$\mathbf{v_R} = \mathbf{C_{R,I}}\left(\mathbf{v_I} - \boldsymbol{\omega_{cb}} \times \mathbf{r_I}\right) \tag{4.74}$$

or

$$\begin{pmatrix} \dot{x}_R \\ \dot{y}_R \\ \dot{z}_R \end{pmatrix} = \begin{bmatrix} \cos(\omega_{cb}t) & \sin(\omega_{cb}t) & 0 \\ -\sin(\omega_{cb}t) & \cos(\omega_{cb}t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \left\{ \begin{pmatrix} \dot{x}_I \\ \dot{y}_I \\ \dot{z}_I \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \omega_{cb} \end{pmatrix} \times \begin{pmatrix} x_I \\ y_I \\ z_I \end{pmatrix} \right\} \tag{4.75}$$

The corresponding inverse relations are given by:

$$\mathbf{r_I} = \mathbf{C_{I,R}}\mathbf{r_R} \tag{4.76}$$

$$\mathbf{v_I} = \mathbf{C_{I,R}}\mathbf{v_R} + \boldsymbol{\omega_{cb}} \times \mathbf{r_I} \tag{4.77}$$

Now that the position and velocity are available in $\mathcal{F}_R$ components, we can convert both to spherical components. The relation between spherical and Cartesian position is given by the following well-known equations (see Fig. 4.18):

$$\begin{aligned} R &= \sqrt{x_R^2 + y_R^2 + z_R^2} \\ \tau &= \arctan\left(\frac{y_R}{x_R}\right) \\ \delta &= \arcsin\left(\frac{z_R}{\sqrt{x_R^2 + y_R^2 + z_R^2}}\right) \end{aligned} \tag{4.78}$$

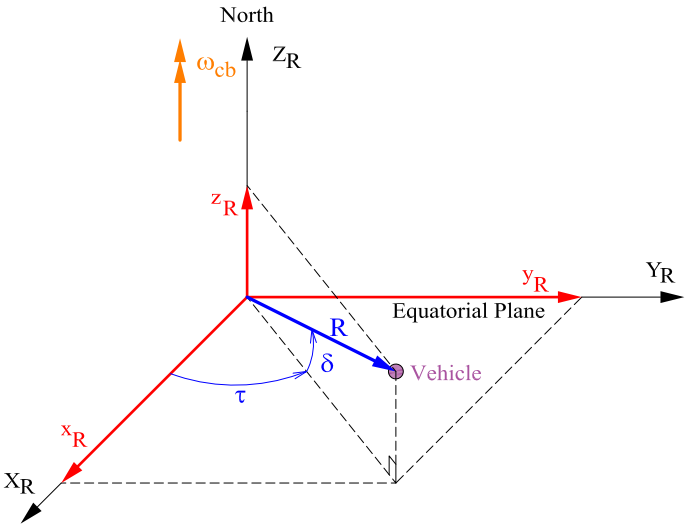


Fig. 4.18 The relation between Cartesian and spherical position

and the inverse relations by

$$\begin{aligned} x_R &= R\cos\delta\cos\tau \\ y_R &= R\cos\delta\sin\tau \\ z_R &= R\sin\delta \end{aligned} \tag{4.79}$$

The relation between Cartesian and spherical velocity components is not as straightforward as for the position. The definition of the velocity in spherical elements is one that is closely related to the vertical frame: the flight-path angle gives the elevation of the velocity with respect to the local horizontal-plane and the heading angle gives the flight direction towards the north, which is the X-axis of $\mathcal{F}_V$ (see also Fig. 4.19). The logical choice for the intermediate step between the Cartesian velocity in $\mathcal{F}_R$ and its spherical counterparts is the Cartesian velocity in $\mathcal{F}_V$, defined as $\mathbf{v_V} = (v_n, v_e, v_d)^{\mathrm{T}}$. These three components are in the north, east, and down directions, respectively.

Let us start with the Cartesian velocity with respect to $\mathcal{F}_R$, $\mathbf{v_R} = (u, v, w)^{\mathrm{T}}$. For now we will deal with the ground velocity only; the influence of wind will be discussed later, in Sect. 4.6. The modulus of the (ground) velocity vector, which is by definition aligned with the X-axis of the groundspeed-based trajectory frame, $\mathcal{F}_{TG}$, can directly be computed from the three Cartesian components,

$$V_g = \sqrt{u^2 + v^2 + w^2} \tag{4.80}$$

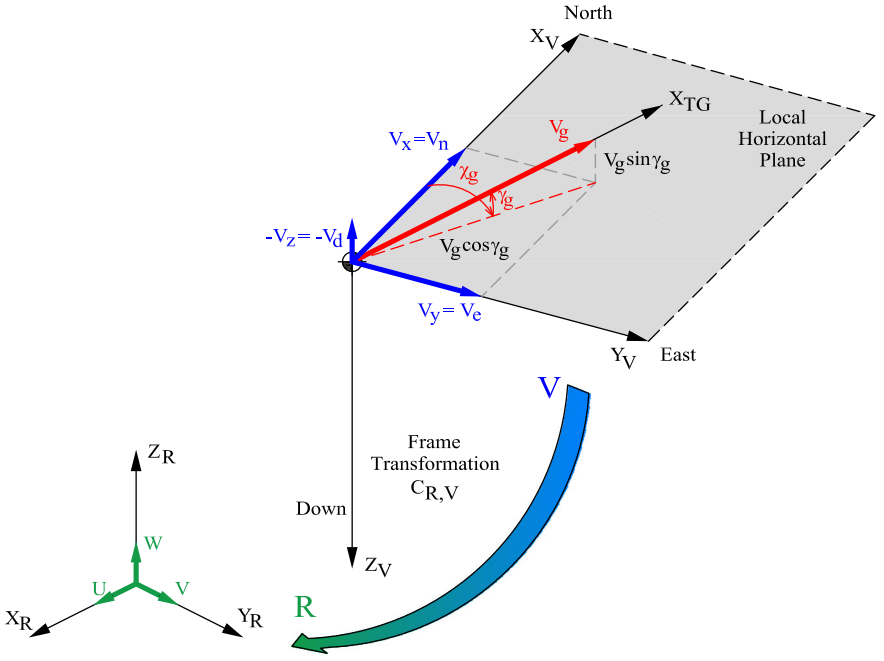


Fig. 4.19 The relation between the Cartesian velocity in $\mathcal{F}_R$, $(u, v, w)^{\mathrm{T}}$ and $\mathcal{F}_V$, $(v_n, v_e, v_d)^{\mathrm{T}}$, and the spherical velocity components V_g, γ_g, and χ_g

For the two direction angles, γ_g and χ_g, we transform the velocity from $\mathcal{F}_R$ to $\mathcal{F}_V$:

$$\mathbf{v_V} = \mathbf{C_{V,R}}\mathbf{v_R} \tag{4.81}$$

with $\mathbf{C_{V,R}}$ given by the inverse of Eq. (4.51):

$$\mathbf{C_{V,R}} = \begin{bmatrix} -c\tau s\delta & -s\tau s\delta & c\delta \\ -s\tau & c\tau & 0 \\ -c\tau c\delta & -s\tau c\delta & -s\delta \end{bmatrix}$$

The projection of V_g in the local vertical-plane gives (see Fig. 4.19)

$$V_g \sin\gamma_g = -v_d \tag{4.82}$$

Because γ_g is restricted to values between –90° and +90°, and the arcsine function is defined for the same interval, the flight-path angle is then obtained with:

$$\gamma_g = -\arcsin\left(\frac{v_d}{V_g}\right) \tag{4.83}$$

For the heading χ_g it is found that

$$\tan\chi_g = \frac{v_e}{v_n} \Rightarrow \chi_g = \arctan\left(\frac{v_e}{v_n}\right) \tag{4.84}$$

The inverse relation can quite easily be obtained, using above results. From Fig. 4.19, we get the three components of $\mathbf{v_V}$:

$$\begin{aligned} v_n &= V_g \cos\gamma_g \cos\chi_g \\ v_e &= V_g \cos\gamma_g \sin\chi_g \\ v_d &= -V_g \sin\gamma_g \end{aligned} \tag{4.85}$$

$\mathbf{v_R}$ is now computed using

$$\mathbf{v_R} = \mathbf{C_{R,V}}\mathbf{v_V} \tag{4.86}$$

with $\mathbf{C_{R,V}}$ given by Eq. (4.51).

4.4.2 Attitude

In this section we will restrict ourselves to extracting attitude angles from transformation matrices, also known as Direction Cosine Matrices. That means the coordinate transformation can also be done from one set of Euler angles to the other, or from

quaternions to Euler angles, as long as we construct the DCM first. The procedure to convert attitude angles from one set to the other is always the same, i.e., though comparing transformation matrices. Let us consider the following expression:

$$\mathbf{C}_{\mathbf{A},\mathbf{B}} = \mathbf{C}^{*}_{\mathbf{A},\mathbf{B}} \tag{4.87}$$

where the matrix on the left-hand side is one composed of the unknown attitude angles, and $\mathbf{C}^{*}_{\mathbf{A},\mathbf{B}}$ is an equivalent, but *known* matrix. By comparing the individual matrix elements and isolating the unknown angles we will be able to write them as a function of the known values. As a very simple example, suppose we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 5 & 2 \end{bmatrix} \tag{4.88}$$

then it is immediately clear that $a = 1$, $b = 4$, $c = 5$, and $d = 2$.

In a similar way we can find relations between the Euler angles φ, θ, and ψ, and the aerodynamic angles α, β, and σ. The same procedure applies to both groundspeed-based and airspeed-based variables, so there is no need to include the subscripts in the aerodynamic angles. Depending on the angles considered, we need to use the corresponding transformation matrix, e.g., $\mathbf{C}_{\mathbf{B},\mathbf{TG}}$ or $\mathbf{C}_{\mathbf{B},\mathbf{TA}}$. For now, $\mathbf{C}_{\mathbf{B},\mathbf{T}}$ would be used. As an example transformation, the (unknown) aerodynamic angles will be obtained from a set of known Euler angles and position data.

In Fig. 4.5, the Euler angles define the orientation of $\mathcal{F}_B$ with respect to $\mathcal{F}_I$. The transformation matrix from $\mathcal{F}_I$ to $\mathcal{F}_B$, given by $\mathbf{C}_{\mathbf{B},\mathbf{I}}$, is obtained from three successive unit axis rotations:

$$\mathbf{C}_{\mathbf{B},\mathbf{I}} = \mathbf{C}_{\mathbf{x}}(\varphi)\mathbf{C}_{\mathbf{y}}(\theta)\mathbf{C}_{\mathbf{z}}(\psi) \tag{4.89}$$

Fully written out, we obtain for the known starting point $\mathbf{C}_{\mathbf{B},\mathbf{I}}$

$$\mathbf{C}_{\mathbf{B},\mathbf{I}} = \begin{bmatrix} \mathrm{c}\theta\mathrm{c}\psi & \mathrm{c}\theta\mathrm{s}\psi & -\mathrm{s}\theta \\ \mathrm{s}\varphi\mathrm{s}\theta\mathrm{c}\psi - \mathrm{s}\psi\mathrm{c}\varphi & \mathrm{s}\psi\mathrm{s}\theta\mathrm{s}\varphi + \mathrm{c}\psi\mathrm{c}\varphi & \mathrm{s}\varphi\mathrm{c}\theta \\ \mathrm{c}\psi\mathrm{c}\varphi\mathrm{s}\theta + \mathrm{s}\psi\mathrm{s}\varphi & \mathrm{s}\psi\mathrm{c}\varphi\mathrm{s}\theta - \mathrm{c}\psi\mathrm{s}\varphi & \mathrm{c}\varphi\mathrm{c}\theta \end{bmatrix} \tag{4.90}$$

The aerodynamic angles α, β and σ define the orientation of the body frame with respect to the trajectory frame. The corresponding transformation matrix is given by $\mathbf{C}_{\mathbf{T},\mathbf{B}}$, Eq. (4.63):

$$\mathbf{C}_{\mathbf{T},\mathbf{B}} = \begin{bmatrix} \mathrm{c}\alpha\mathrm{c}\beta & \mathrm{s}\beta & \mathrm{s}\alpha\mathrm{c}\beta \\ -\mathrm{c}\alpha\mathrm{s}\beta\mathrm{c}\sigma - \mathrm{s}\alpha\mathrm{s}\sigma & \mathrm{c}\beta\mathrm{c}\sigma & \mathrm{c}\alpha\mathrm{s}\sigma - \mathrm{s}\alpha\mathrm{s}\beta\mathrm{c}\sigma \\ \mathrm{c}\alpha\mathrm{s}\beta\mathrm{s}\sigma - \mathrm{s}\alpha\mathrm{c}\sigma & -\mathrm{c}\beta\mathrm{s}\sigma & \mathrm{c}\alpha\mathrm{c}\sigma + \mathrm{s}\alpha\mathrm{s}\beta\mathrm{s}\sigma \end{bmatrix} \tag{4.91}$$

For each of the angles α, β and σ we find, after inspecting the matrix elements in Eq. (4.91):

$$\alpha = \arctan\left(\frac{\mathbf{C}_{\mathbf{T},\mathbf{B}}(1,3)}{\mathbf{C}_{\mathbf{T},\mathbf{B}}(1,1)}\right) \tag{4.92}$$

$$\beta = \arcsin\left(\mathbf{C_{T,B}}(1,2)\right) \tag{4.93}$$

$$\sigma = -\arctan\left(\frac{\mathbf{C_{T,B}}(3,2)}{\mathbf{C_{T,B}}(2,2)}\right) \tag{4.94}$$

taking quadrant information into account. The angles are confined to the regions $-180° \leq \alpha < 180°$, $-90° \leq \beta \leq 90°$ and $-180° \leq \sigma < 180°$, which uniquely describe meaningful vehicle orientations. It is stressed again that the above equations can be used for both groundspeed-based variables (α_g, β_g and σ_g), using $\mathbf{C_{TG,B}}$, and $\mathbf{C_{TG,V}}$ with γ_g and χ_g, and airspeed-based variables, involving α_a, β_a, σ_a, γ_a and χ_a.

What remains to be done is to find the equivalent expression for $\mathbf{C_{T,B}}$. Having available the "inertial" Euler angles, and thus $\mathbf{C_{B,I}}$, or equivalently its transpose, $\mathbf{C_{I,B}}$, $\mathbf{C_{T,B}}$ is written as:

$$\mathbf{C_{T,B}} = \mathbf{C_{T,I}}\mathbf{C_{I,B}} \tag{4.95}$$

with $\mathbf{C_{T,I}}$ defined by the multiplication of several individual transformation matrices:

$$\mathbf{C_{T,I}} = \mathbf{C_{T,V}}\mathbf{C_{V,R}}\mathbf{C_{R,I}} = \mathbf{C_y}(\gamma)\mathbf{C_z}(\chi)\mathbf{C_y}(-\frac{\pi}{2} - \delta)\mathbf{C_z}(\tau + \omega_{cb}t) \tag{4.96}$$

Here, $\mathbf{C_{T,V}}$, $\mathbf{C_{V,R}}$, and $\mathbf{C_{R,I}}$ are given by the inverse matrices of Eqs. (4.53) (without the subscripts), (4.51), and (4.50), respectively.

It is noted once more that the above discussed conversion works for all combinations of Euler angles.

4.5 General Formulation of the Equations of Motion

With the relatively low speeds encountered by present-day spacecraft (that is, when compared to the speed of light), we can use classical or Newtonian mechanics to describe the motion of a vehicle flying in a planetary atmosphere. This kind of mechanics is based on the three Laws of Motion of Newton and Galileo's principle of relativity.

The motion of a non-elastic body can be divided into the motion *of* the c.o.m. and the one *around* this c.o.m. The equations describing the first part are called the equations of translational motion, giving information about position and velocity in three directions (a position coordinate is also called a degree of freedom). The second part of the motion is described by the equations of rotational motion, resulting in information on the angular rates around three axes and the corresponding attitude of the body (similarly, an "attitude coordinate" is a degree of freedom). The complete set of equations incorporates therefore 6 degrees of freedom (6 d.o.f.), i.e., three translations and three rotations.

Let us assume, that an arbitrary vehicle with variable mass m is moving about a celestial body and is subjected to a number of external forces. The total sum of the

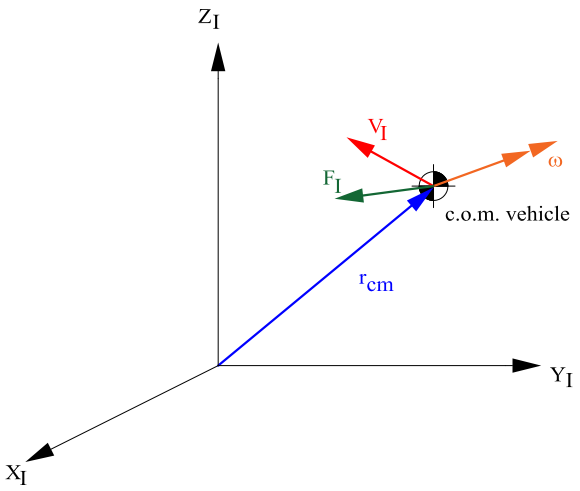


Fig. 4.20 The motion of a body with respect to inertial space

external forces is denoted by $\mathbf{F_I}$. The position vector of the c.o.m. of the vehicle is given by $\mathbf{r_{cm}}$, relative to an arbitrary inertial frame, whereas its velocity with respect to this frame is given by $\mathbf{V_I}$. Furthermore, the vehicle is rotating with an angular velocity $\boldsymbol{\omega}$ with respect to the inertial frame. The situation is depicted in Fig. 4.20. Starting with Newton's second law of motion, in Appendix A the translational motion of this mass-varying body under the influence of an external force is derived, the result of which can be described by:

$$\mathbf{F_I} + \mathbf{F_C} + \mathbf{F_{rel}} = \tilde{\mathbf{F}}_\mathbf{I} = m\frac{\mathrm{d}^2\mathbf{r_{cm}}}{\mathrm{d}t^2} \tag{4.97}$$

with

$\mathbf{F_I}$	= total of the external forces expressed in components of the inertial frame (N)
$\tilde{\mathbf{F}}_\mathbf{I}$	= pseudo external-force vector in inertial frame (N)
$\mathbf{F_C} = -2\boldsymbol{\omega} \times \int_m \frac{\delta\tilde{\mathbf{r}}}{\delta t}\mathrm{d}m$	= Coriolis force, due to time variations in mass distribution (N)
$\mathbf{F_{rel}} = -\int_m \frac{\delta^2\tilde{\mathbf{r}}}{\delta t^2}\mathrm{d}m$	= relative force, due to time variations in mass distribution (N)
$\frac{\mathrm{d}^2\mathbf{r_{cm}}}{\mathrm{d}t^2}$	= acceleration of the c.o.m. with respect to the inertial frame (m/s^2)
$\boldsymbol{\omega} = (p, q, r)^\mathrm{T}$	= the rotation vector of the body frame with respect to the inertial frame, expressed in components along the body axes
$\tilde{\mathbf{r}}$	= the location of a mass element with respect to the c.o.m. of the vehicle (m)

The notation $\frac{\delta_{..}}{\delta t}$ denotes a derivative of a vector relative to the local (body) frame, as compared with $\frac{\mathrm{d}_{..}}{\mathrm{d}t}$, which denotes a derivative relative to the inertial frame.

Example 4.2 (*External Force in the Inertial Frame*) In Eq. (4.97), $\mathbf{F_I}$ is the external force vector that can have contributions from aerodynamics, gravity, and possibly propulsion. Each of these individual forces are usually defined in there own reference frame: the gravitational force in $\mathcal{F}_V$, the aerodynamic force in $\mathcal{F}_{AA}$ or $\mathcal{F}_{AG}$, and the propulsion force in either $\mathcal{F}_P$ or $\mathcal{F}_B$. To add these components together and use them in the equations of motion, they have to be transformed to the inertial frame.

So

$$\mathbf{F_I} = \mathbf{C_{I,V}}\mathbf{F_G} + \mathbf{C_{I,AA}}\mathbf{F_A} + \mathbf{C_{I,P}}\mathbf{F_P} \tag{4.98}$$

with $\mathbf{C_{I,V}}$ given by Eq. (4.64),

$$\mathbf{C_{I,V}} = \mathbf{C_{I,R}}\mathbf{C_{R,V}} \tag{4.99}$$

$$\mathbf{C_{I,P}} = \mathbf{C_{I,R}}\mathbf{C_{R,B}}\mathbf{C_{B,P}} \tag{4.100}$$

with $\mathbf{C_{I,R}}$ and $\mathbf{C_{R,B}}$ given by Eqs. (4.50) and (4.67), respectively, and $\mathbf{C_{B,P}}$ by the inverse of Eq. (4.61).

Finally,

$$\mathbf{C_{I,AA}} = \mathbf{C_{I,V}}\mathbf{C_{V,AA}} \tag{4.101}$$

with $\mathbf{C_{V,AA}}$ given by Eq. (4.56).

Transformation matrices are used extensively in any flight simulation software, it is therefore advised that once developed and verified, this functionality is put in a software library that can be reused in any new application. ■

The general form of the equation of rotational motion is expressed as (see again Appendix A for a complete derivation):

$$\mathbf{M_{cm}} + \mathbf{M_C} + \mathbf{M_{rel}} = \tilde{\mathbf{M}}_{\mathbf{cm}} = \int_m \tilde{\mathbf{r}} \times \left(\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \tilde{\mathbf{r}}\right) \mathrm{d}m + \int_m \tilde{\mathbf{r}} \times \left[\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}})\right] \mathrm{d}m \tag{4.102}$$

with

$\mathbf{M_{cm}}$	=	total external moment about the vehicle's c.o.m. (Nm)
$\tilde{\mathbf{M}}_{\mathbf{cm}}$	=	pseudo external moment about c.o.m. (Nm)
$\mathbf{M_C} = -2\int_m \tilde{\mathbf{r}} \times \left(\boldsymbol{\omega} \times \frac{\delta \tilde{\mathbf{r}}}{\delta t}\right) dm$	=	Coriolis moment due to time variations in mass distribution (Nm)
$\mathbf{M_{rel}} = -\int_m \tilde{\mathbf{r}} \times \frac{\delta^2 \tilde{\mathbf{r}}}{\delta t^2} dm$	=	relative moment due to time variations in mass distribution (Nm)
$\int_m \tilde{\mathbf{r}} \times \left(\frac{d\boldsymbol{\omega}}{dt} \times \tilde{\mathbf{r}}\right) dm$	=	apparent moment due to the angular acceleration of the vehicle with respect to the inertial frame (Nm)
$\int_m \tilde{\mathbf{r}} \times \left[\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}})\right] dm$	=	apparent moment due to the angular velocity of the vehicle with respect to the inertial frame (Nm)

Equations (4.97) and (4.102) describe the motion of a body under the influence of external forces and moments. The corresponding change in position and attitude can be described by:

$$\frac{\mathrm{d}\mathbf{r}_{\mathbf{cm}}}{\mathrm{d}t} = \mathbf{V}_{\mathbf{I}} \tag{4.103}$$

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = f(\Phi)\boldsymbol{\omega} \tag{4.104}$$

In Eq. (4.104), we have used the vector Φ to denote the attitude of the body in the inertial space, even though this does not have to be a vector quantity. This is the case, for instance, when three angles define the attitude.

We note that the dynamic equations of motion for an elastic mass-varying body have the same form as those for a rigid body by introducing a pseudo external force and moment vector, that includes the Coriolis and relative force and moment vector. An important principle is involved here, the *Principle of Solidification.* This principle states (Cornelisse et al. 1979) that *in general, equations of translational and rotational motion of an arbitrary variable mass system at time* t *can be written as the translational and rotational equations for a rigid body with mass* M *equal to the mass of the system at time* t*, while in addition to the true external forces and moments, two apparent forces and moments are applied: the Coriolis and relative forces and moments, respectively.*

For a rigid body, i.e., a vehicle that has a constant mass and is also non-elastic, the equations of rotational motion can be derived in an alternative way, again see Appendix A. Expanding the dynamic equations of rotational motion and expressing the components along the body axes, gives us the so-called Euler equations that give information about the angular accelerations:

$$\dot{\boldsymbol{\omega}} = \mathbf{I}^{-1}\left(\tilde{\mathbf{M}}_{\mathrm{cm}} - \boldsymbol{\omega} \times \mathbf{I}\boldsymbol{\omega}\right) \tag{4.105}$$

where $\tilde{\mathbf{M}}_{\mathrm{cm}} = \left(M_x, M_y, M_z\right)^{\mathrm{T}}$ is the sum of external, Coriolis and relative moments about the c.o.m., expressed in components along the body axes, $\boldsymbol{\omega} = (p, q, r)^{\mathrm{T}}$ is the rotation vector of the body frame with respect to the inertial frame, expressed in components along the body axes, and

$$\mathbf{I} = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{xy} & I_{yy} & -I_{yz} \\ -I_{xz} & -I_{yz} & I_{zz} \end{bmatrix} \tag{4.106}$$

is the inertia tensor of the re-entry vehicle, referenced to the body frame.

Solving the above equations asks for the inversion of the inertia tensor. However, in literature also explicit expressions can be found for $\dot{p}, \dot{q}$ and $\dot{r}$ (Duke et al. 1988).

In a number of cases, these equations can be simplified. In the case of a (mass-symmetrical) vehicle, two out of three products of inertia are equal to zero. When the plane of symmetry is the $X_B Z_B$-plane, then $I_{xy} = I_{yz} = 0$. Equation (4.105) then simplifies to

$$\begin{bmatrix} I_{xx} & 0 & -I_{xz} \\ 0 & I_{yy} & 0 \\ -I_{xz} & 0 & I_{zz} \end{bmatrix} \begin{pmatrix} \dot{p} \\ \dot{q} \\ \dot{r} \end{pmatrix} = \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} - \begin{pmatrix} p \\ q \\ r \end{pmatrix} \times \begin{bmatrix} I_{xx} & 0 & -I_{xz} \\ 0 & I_{yy} & 0 \\ -I_{xz} & 0 & I_{zz} \end{bmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix} \tag{4.107}$$

Solving this system of equations for $\dot{p}$, $\dot{q}$ and $\dot{r}$ leads to

$$\dot{p} = \frac{I_{zz}}{I^*} M_x + \frac{I_{xz}}{I^*} M_z + \frac{\left(I_{xx} - I_{yy} + I_{zz}\right) I_{xz}}{I^*} pq + \frac{\left(I_{yy} - I_{zz}\right) I_{zz} - I_{xz}^2}{I^*} qr \tag{4.108a}$$

$$\dot{q} = \frac{M_y}{I_{yy}} + \frac{I_{xz}}{I_{yy}} \left(r^2 - p^2\right) + \frac{I_{zz} - I_{xx}}{I_{yy}} pr \tag{4.108b}$$

$$\dot{r} = \frac{I_{xz}}{I^*} M_x + \frac{I_{xx}}{I^*} M_z + \frac{\left(I_{xx} - I_{yy}\right) I_{xx} + I_{xz}^2}{I^*} pq + \frac{\left(-I_{xx} + I_{yy} - I_{zz}\right) I_{xz}}{I^*} qr \tag{4.108c}$$

with $I^* = I_{xx} I_{zz} - I_{xz}^2$.

Even simpler expressions can be obtained, when we have two planes of symmetry. This is the case for rotational symmetric bodies. In that case, all three products of inertia are zero and Eqs. (4.108a)–(4.108c) simplify to

$$\dot{p} = \frac{M_x}{I_{xx}} + \frac{I_{yy} - I_{zz}}{I_{xx}} qr \tag{4.109a}$$

$$\dot{q} = \frac{M_y}{I_{yy}} + \frac{I_{zz} - I_{xx}}{I_{yy}} pr \tag{4.109b}$$

$$\dot{r} = \frac{M_z}{I_{zz}} + \frac{I_{xx} - I_{yy}}{I_{zz}} pq \tag{4.109c}$$

It is stressed that by *rotational symmetric* we mean rotational symmetric in *mass*. If we move the centre of mass downwards, for instance, the vehicle can still be rotational symmetric in *geometry*, but no longer in mass.

4.6 *Equations of Motion in Spherical Coordinates

4.6.1 General

In the next chapter the different entry mechanisms will be discussed using a simplified version of the equations of motion expressed in spherical coordinates (see also Fig. 4.4). Even though those equations have a multitude of uses, they are usually not used to obtain accurate trajectories. Simulating the re-entry problem is usually done with the equations of motion discussed in the previous section, most notably Eqs. (4.97) and (4.103) for the (inertial) translational motion, and Eqs. (4.104) and (4.105) for the (body-frame related) rotational motion.

However, there are a few reasons to provide the reader with the full set of equations of translational motion based on spherical coordinates. The first one is to put the simplified equations of the next chapter in perspective, and to understand what the effects of the simplifying assumptions are. The second reason is that these equations are the starting point to obtain a linearised system of equations, which is used in Chap. 6 to study the characteristic motion of re-entry vehicles. Third, when treating the flight in the terminal area (Chap. 8), i.e., the last phase of a re-entry vehicle before descending into the approach and landing phase, starting point of the discussion is yet again the full set of equations. The final reason is that we want to show how to include the effect of wind in the equations of motion. Of all formulations, the implementation for equations of motion with spherical coordinates is by far the most complicated. Because similar steps would be required for other formulations of the equations, it will be a useful exercise to go through these steps.

To begin with, we will simply state the full set of equations. This will be done without derivation, as that will be quite lengthy and is also not the main purpose of this section. The interested reader is referred to the books by, for instance, Vinh et al. (1981) and Mooij (1997) for more details.

The equations are valid for a rotating central body with a central gravity field with zonal harmonics (see Sect. 2.2.1, Eq. (2.32) through (2.34)). Furthermore, the vehicle can either be asymmetric or flying with an angle of sideslip $\beta \neq 0$, for instance, due to wind, leading to a non-zero side force, S. However, typical simplifications that one can find in literature have only a central gravity field ($g_n = 0$) and no sideslip ($S = 0$).

$$\begin{aligned} \dot{V}_g &= \frac{F_V}{m} + \omega_{cb}^2 R \cos\delta \left(\sin\gamma_g \cos\delta - \cos\gamma_g \sin\delta \cos\chi_g \right) \\ \text{with } F_V &= -D - mg_d \sin\gamma_g - mg_n \cos\gamma_g \cos\chi_g \end{aligned} \tag{4.110}$$

$$\begin{aligned} V_g\dot{\gamma}_g = \;& \tfrac{F_\gamma}{m} + 2\omega_{cb}V_g\cos\delta\sin\chi_g + \tfrac{V_g^2}{R}\cos\gamma_g + \\ & + \omega_{cb}^2 R\cos\delta\left(\cos\delta\cos\gamma_g + \sin\gamma_g\sin\delta\cos\chi_g\right) \\ \text{with } F_\gamma = \;& -S\sin\sigma_g + L\cos\sigma_g - mg_d\cos\gamma_g + mg_n\sin\gamma_g\cos\chi_g \end{aligned} \tag{4.111}$$

$$\begin{aligned} V_g\cos\gamma_g\dot{\chi}_g = \;& \tfrac{F_\chi}{m} + 2\omega_{cb}V_g\left(\sin\delta\cos\gamma_g - \cos\delta\sin\gamma_g\cos\chi_g\right) + \\ & + \tfrac{V_g^2}{R}\cos^2\gamma_g\tan\delta\sin\chi_g + \omega_{cb}^2 R\cos\delta\sin\delta\sin\chi_g \\ \text{with } F_\chi = \;& -S\cos\sigma_g - L\sin\sigma_g + mg_n\sin\chi_g \end{aligned} \tag{4.112}$$

$$\dot{R} = V_g\sin\gamma_g \tag{4.113}$$

$$\dot{\tau} = \frac{V_g\cos\gamma_g\sin\chi_g}{R\cos\delta} \tag{4.114}$$

$$\dot{\delta} = \frac{V_g\cos\gamma_g\cos\chi_g}{R} \tag{4.115}$$

The above equations describe the motion of the vehicle with respect to ground; the subscript "g" indicates that all state variables are ground-speed related. This means that also the aerodynamic forces D, S and L are defined with respect to the groundspeed-based aerodyanamic reference frame, $\mathcal{F}_{AG}$, even though their magnitude is based on airspeed:

$$\begin{pmatrix} D \\ S \\ L \end{pmatrix} = \bar{q}_a S_{ref}\begin{pmatrix} C_D \\ C_S \\ C_L \end{pmatrix} \tag{4.116}$$

with the dynamic pressure being $\bar{q}_a = \frac{1}{2}\rho V_a^2$, and in case the aerodynamic coefficients are dependent on Mach number, this evaluation is based on airspeed, $M = \frac{V_a}{a}$. The airspeed is different from the groundspeed when a wind field is present, which brings

us to the last topic of this chapter: namely how to include wind in the above equations of motion.

4.6.2 Wind Equations

Background

In case the central body has an atmosphere, there may be a wind field that has to be included in the simulation for more accurate predictions. It is easy to understand that when there is a crosswind, a payload under a parachute may start drifting, especially when the descent duration is long. But also at a smaller scale wind can be important, e.g., high-frequency turbulence or discrete wind gusts can affect the vehicle's stability and controllability. As an example, in Fig. 4.21, we have depicted the situation of vertical wind in the surroundings of a vehicle flying with a positive angle of attack.

From this figure, it follows:

$$V_w^2 = V_g^2 + V_a^2 - 2V_g V_a \cos \Delta\alpha \tag{4.117}$$

or

$$\Delta\alpha = \arccos\left(\frac{V_g^2 + V_a^2 - V_w^2}{2V_g V_a}\right) \tag{4.118}$$

With a ground velocity V_g = 1000 m/s at an initial angle of attack of 10°, and a wind velocity V_w = 40 m/s, the air speed V_a becomes 993.8 m/s at an angle of attack of 7.7°. However, with V_g = 1,000 m/s at α = 2° and V_w = 40 m/s, V_a becomes 999.4 m/s at α = –2.3°. One can imagine that if we want to compensate for this vertical wind, we must deflect the elevator to maintain equilibrium. This deflection can result in a smaller controllability margin. On the other hand, when there is a strong side wind, the vehicle will deviate from its ground track, which has to be corrected. It is noted,

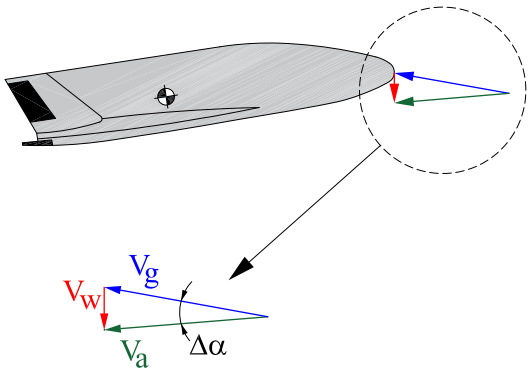


Fig. 4.21 The influence of a vertical wind on the angle of attack

though, that due to the high velocities during entry only close to the end of the hypersonic descent the effect may be noticeable.

In this section we assume that there is a wind vector, without going into detail where it comes from. Similar to the ground speed, wind is typically defined with respect to the rotating central body, and can conveniently be expressed in $\mathcal{F}_\mathcal{V}$, either by a modulus and two direction angles, i.e., V_w, γ_w and χ_w (see Fig. 4.13), or in Cartesian components, depending in which format these (input) data are available. Between the two sets there is a simple relationship, as has been discussed in Sect. 4.4:

$$\begin{aligned} v_{w,n} &= V_w \cos\gamma_w \cos\chi_w \\ v_{w,e} &= V_w \cos\gamma_w \sin\chi_w \\ v_{w,d} &= -V_w \sin\gamma_w \end{aligned} \tag{4.119}$$

and vice versa

$$\begin{aligned} V_w &= \sqrt{v_{w,n}^2 + v_{w,e}^2 + v_{w,d}^2} \\ \gamma_w &= -\arcsin\left(\frac{v_{w,d}}{V_w}\right) \\ \chi_w &= \arctan\left(\frac{v_{w,e}}{v_{w,n}}\right) \end{aligned} \tag{4.120}$$

with $\gamma_w \in [-90°, +90°]$ and $\chi_w \in [-180°, +180°)$. As before, the ned subscripts indicate a north, east, and down wind direction, respectively.

The velocity of a vehicle with respect to the planetary surface (or ground) can be expressed by the following vector equation:

$$\mathbf{v}_{vehicle/ground} = \mathbf{v}_{vehicle/air} + \mathbf{v}_{air/ground} \quad \Rightarrow \quad \mathbf{v_g} = \mathbf{v_a} + \mathbf{v_w} \tag{4.121}$$

or, in words, the velocity of the vehicle with respect to the ground is the vectorial summation of the velocity of the vehicle with respect to the air and the velocity of the air with respect to the ground. The latter component is the wind velocity.

While using spherical components for both position and velocity, the computational scheme of solving the equations of motion becomes rather complex, as can be seen below.

1. Start with the initial conditions $V_{g,0}$, $\gamma_{g,0}$, $\chi_{g,0}$, R_0, τ_0, δ_0, $\alpha_{g,0}$, $\beta_{g,0}$ and $\sigma_{g,0}$[5];
2. Depending on the format available, compute the wind parameters V_w, γ_w and χ_w (see Eqs. (4.119) and (4.120));

[5] We will consider translational motion only, so the angular-rate components p_0, q_0, r_0 are not relevant at this moment. We have assumed that the attitude is output from, for instance, the guidance system, and given by aerodynamic angles. To keep the discussion general, we assume a non-zero β, although from a guidance point of view zero sideslip would be commanded. It is noted that when the rotational dynamics are simulated as well, α_g, β_g and σ_g come from the related kinematic equations. In case the attitude is defined by quaternions, then a transformation needs to take place to derive the angles. However, this is beyond the scope of the current text.

3. Compute the airspeed components V_a, γ_a and χ_a;
4. Compute the aerodynamic attitude with respect to the airspeed, α_a, β_a and σ_a;
5. Compute aerodynamic forces based on airspeed variables;
6. Transform these forces to groundspeed-related forces;
7. Evaluate the equations of motion Eqs. (4.110) through (4.115), and get the new state vector V_g, γ_g, χ_g, R, τ, δ, and from the guidance system α_g, ($\beta_g = 0$) and σ_g;
8. Continue with the next step until final conditions have been reached (go to 2).

Step 3—Airspeed Vector

Starting from Eq. (4.121), we can express the airspeed vector $\mathbf{v_a}$ as

$$\mathbf{v_a} = \mathbf{v_g} - \mathbf{v_w} \tag{4.122}$$

This vector is defined by its length V_a and the two directional angles γ_a and χ_a. These two angles define in principle the orientation of $\mathcal{F}_{TA}$ with respect to $\mathcal{F}_V$, i.e.,

$$\mathbf{v_{a,V}} = \mathbf{C_{V,TA}}\mathbf{v_{a,TA}} \tag{4.123}$$

with $\mathbf{C_{V,TA}}$ given by Eq. (4.54), $\mathbf{C_{V,TA}} = \mathbf{C_z}(-\chi_a)\mathbf{C_y}(-\gamma_a)$.

Similarly, $\mathbf{v_g}$ and $\mathbf{v_w}$ can be defined by their respective length and two angles. In that case, the two related angles define the orientation of $\mathcal{F}_{TG}$ and $\mathcal{F}_W$, respectively, with respect to $\mathcal{F}_V$, i.e.,

$$\mathbf{v_{g,V}} = \mathbf{C_{V,TG}}\mathbf{v_{g,TG}} \tag{4.124}$$

$$\mathbf{v_{w,V}} = \mathbf{C_{V,W}}\mathbf{v_{w,W}} \tag{4.125}$$

with $\mathbf{C_{V,TG}} = \mathbf{C_z}(-\chi_g)\mathbf{C_y}(-\gamma_g)$ and $\mathbf{C_{V,W}} = \mathbf{C_z}(-\chi_w)\mathbf{C_y}(-\gamma_w)$. Substituting Eqs. (4.123)–(4.125) into Eq. (4.122), we get

$$\cos\chi_a \cos\gamma_a V_a = \cos\chi_g \cos\gamma_g V_g - \cos\chi_w \cos\gamma_w V_w \tag{4.126a}$$

$$\sin\chi_a \cos\gamma_a V_a = \sin\chi_g \cos\gamma_g V_g - \sin\chi_w \cos\gamma_w V_w \tag{4.126b}$$

$$\sin\gamma_a V_a = \sin\gamma_g V_g - \sin\gamma_w V_w \tag{4.126c}$$

To solve the three airspeed components (V_a, γ_a and χ_a) out of these three equations, we will first square the three equations and add them together, which leaves us with an expression for V_a^2, i.e.,

$$V_a^2 = V_g^2 + V_w^2 - 2V_g V_w \left(\cos\gamma_g \cos\gamma_w \cos(\chi_w - \chi_g) + \sin\gamma_g \sin\gamma_w\right) \tag{4.127}$$

Taking the positive root from the above equation gives us V_a.

Dividing Eq. (4.126b) by Eq. (4.126a), acknowledging the fact that V_a is always larger than zero and that the equations hold a singularity for $\gamma_a = \pm 90°$, yields

$$\chi_a = \arctan\left(\frac{\sin\chi_g \cos\gamma_g V_g - \sin\chi_w \cos\gamma_w V_w}{\cos\chi_g \cos\gamma_g V_g - \cos\chi_w \cos\gamma_w V_w}\right) \tag{4.128}$$

γ_a simply follows from Eq. (4.126c):

$$\gamma_a = \arcsin\left(\frac{\sin\gamma_g V_g - \sin\gamma_w V_w}{V}_a\right) \tag{4.129}$$

with V_a as given above.

Step 4—Aerodynamic Attitude

The attitude of the vehicle with respect to the airspeed is given by the three aerodynamic angles α_a, β_a and σ_a. We assume here that the equations of motion give us α_g, β_g and σ_g, the aerodynamic angles with respect to the groundspeed–these equations are provided in the Intermezzo below. Using the velocity parameters γ_g and χ_g, $\mathbf{C_{B,V}}$, the transformation matrix from the vertical to the body frame, can be computed. This matrix can be expressed (and computed) by

$$\mathbf{C_{B,V}} = \mathbf{C_y}(\alpha_g)\mathbf{C_z}(-\beta_g)\mathbf{C_x}(-\sigma_g)\mathbf{C_y}(\gamma_g)\mathbf{C_z}(\chi_g) \tag{4.130}$$

Another way to express this matrix is to decompose it into

$$\mathbf{C_{B,V}} = \mathbf{C_{B,TA}}\mathbf{C_{TA,V}} \tag{4.131}$$

In this case $\mathbf{C_{B,TA}}$ is given by

$$\mathbf{C_{B,TA}} = \mathbf{C_y}(\alpha_a)\mathbf{C_z}(-\beta_a)\mathbf{C_x}(-\sigma_a) \tag{4.132}$$

which includes the three variables we want to calculate. To isolate them, we post-multiply $\mathbf{C_{B,V}}$ with $\mathbf{C_{V,TA}}$, i.e.,

$$\mathbf{C_{B,V}}\mathbf{C_{V,TA}} = \mathbf{C_{B,TA}}\mathbf{C_{TA,V}}\mathbf{C_{V,TA}} = \mathbf{C_{B,TA}} \tag{4.133}$$

Using $\mathbf{C_{B,V}}$ as computed according to Eq. (4.130), we only need to calculate $\mathbf{C_{V,TA}}$, which is defined by

$$\mathbf{C_{V,TA}} = \mathbf{C_z}(-\chi_a)\mathbf{C_y}(-\gamma_a) \tag{4.134}$$

with γ_A and χ_A given by Eqs. (4.128)–(4.129). $\mathbf{C_{B,TA}}$ fully written out gives us

$$\mathbf{C}_{\mathbf{B},\mathbf{TA}} = \begin{bmatrix} \mathrm{c}\alpha_a\mathrm{c}\beta_a & -\mathrm{c}\alpha_a\mathrm{s}\beta_a\mathrm{c}\sigma_a + \mathrm{s}\alpha_a\mathrm{s}\sigma_a & \mathrm{c}\alpha_a\mathrm{s}\beta_a\mathrm{s}\sigma_a - \mathrm{s}\alpha_a\mathrm{c}\sigma_a \\ \mathrm{s}\beta_a & \mathrm{c}\beta_a\mathrm{c}\sigma_a & -\mathrm{c}\beta_a\mathrm{s}\sigma_a \\ \mathrm{s}\alpha_a\mathrm{c}\beta_a & \mathrm{c}\alpha_a\mathrm{s}\sigma_a - \mathrm{s}\alpha_a\mathrm{s}\beta_a\mathrm{c}\sigma_a & \mathrm{c}\alpha_a\mathrm{c}\sigma_a + \mathrm{s}\alpha_a\mathrm{s}\beta_a\mathrm{s}\sigma_a \end{bmatrix} \tag{4.135}$$

From this matrix, we can derive

$$\beta_a = \arcsin(\mathbf{C}_{\mathbf{B},\mathbf{TA}}(2,1)), \text{ with } \beta_a \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \tag{4.136a}$$

$$\alpha_a = \arctan\left(\frac{\mathbf{C}_{\mathbf{B},\mathbf{TA}}(3,1)}{\mathbf{C}_{\mathbf{B},\mathbf{TA}}(1,1)}\right), \text{ with } \alpha_a \in [-\pi, \pi) \tag{4.136b}$$

$$\sigma_a = \arctan\left(-\frac{\mathbf{C}_{\mathbf{B},\mathbf{TA}}(2,3)}{\mathbf{C}_{\mathbf{B},\mathbf{TA}}(2,2)}\right), \text{ with } \sigma_a \in [-\pi, \pi) \tag{4.136c}$$

So if we take for the elements of $\mathbf{C}_{\mathbf{B},\mathbf{TA}}$ the corresponding elements given by Eq. (4.133), we can compute α_a, β_a and σ_a.

*** Intermezzo 4.1 (Kinematic Attitude Equations Using Aerodynamic Angles)** The aerodynamic angles α_g, β_g, and σ_g, shown in Fig. 4.6, define a *Type-1* sequence of Euler angles with order 2-3-1. Following a tedious derivation process (Mooij 1997), beyond the scope of the current text, leads to the following set of kinematic attitude equations:

$$\begin{aligned} \dot{\alpha}_g\mathrm{c}\beta_g = &- p\mathrm{c}\alpha_g\mathrm{s}\beta_g + q\mathrm{c}\beta_g - r\mathrm{s}\alpha_g\mathrm{s}\beta_g + \\ &+ \mathrm{s}\sigma_g\left[\dot{\chi}_g\mathrm{c}\gamma_g - \dot{\delta}\mathrm{s}\chi_g\mathrm{s}\gamma_g + (\dot{\tau}+\omega_{cb})\left(\mathrm{c}\delta\mathrm{c}\chi_g\mathrm{s}\gamma_g - \mathrm{s}\delta\mathrm{c}\gamma_g\right)\right] + \\ &- \mathrm{c}\sigma_g\left[\dot{\gamma}_g - \dot{\delta}\mathrm{c}\chi_g - (\dot{\tau}+\omega_{cb})\mathrm{c}\delta\mathrm{s}\chi_g\right] \end{aligned} \tag{4.137}$$

$$\begin{aligned} \dot{\beta}_g = &\ p\mathrm{s}\alpha_g - r\mathrm{c}\alpha_g + \mathrm{s}\sigma_g\left[\dot{\gamma}_g - \dot{\delta}\mathrm{c}\chi_g - (\dot{\tau}+\omega_{cb})\mathrm{c}\delta\mathrm{s}\chi_g\right] + \\ &+ \mathrm{c}\sigma_g\left[\dot{\chi}_g\mathrm{c}\gamma_g - \dot{\delta}\mathrm{s}\chi_g\mathrm{s}\gamma_g + (\dot{\tau}+\omega_{cb})\left(\mathrm{c}\delta\mathrm{c}\chi_g\mathrm{s}\gamma_g - \mathrm{s}\delta\mathrm{c}\gamma_g\right)\right] \end{aligned} \tag{4.138}$$

$$\begin{aligned} \dot{\sigma}_g = &- p\mathrm{c}\alpha_g\mathrm{c}\beta_g - q\mathrm{s}\beta_g - r\mathrm{s}\alpha_g\mathrm{c}\beta_g + \\ &+ \dot{\alpha}_g\mathrm{s}\beta_g - \dot{\chi}_g\mathrm{s}\gamma_g - \dot{\delta}\mathrm{s}\chi_g\mathrm{c}\gamma_g + (\dot{\tau}+\omega_{cb})\left(\mathrm{c}\delta\mathrm{c}\chi_g\mathrm{c}\gamma_g + \mathrm{s}\delta\mathrm{s}\gamma_g\right) \end{aligned} \tag{4.139}$$

In these equations, $\dot{\gamma}_g$, $\dot{\chi}_g$, $\dot{\tau}$ and $\dot{\delta}$ are given by, for instance, Eqs. (4.111), (4.112), (4.114) and (4.115), or can be obtained numerically by sampling the angle history. It

is noted that there is a singularity in the equations for $\beta_g = \pm 90°$, a situation that will not occur in normal flight. However, simulating the free-fall entry of, for instance, tumbling rocket stages cannot be done using these equations. In that case, one should use, for instance, the singularity-free quaternion attitude equations.

The above set of equations will, amongst others, be used in Chap. 6, as they provide a natural interpretation of the eigenmotion of the vehicle.

Step 5 and 6—Aerodynamic Forces

To compute the aerodynamic forces and moments, we need the aerodynamic coefficients, which are usually a function of the Mach number, and the angles of attack and sideslip (apart from any possible damping terms, which include in addition p, q and r), and the dynamic pressure. M and $\bar{q}$ are defined by

$$M = \frac{V_a}{a} \qquad \bar{q} = \frac{1}{2}\rho V_a^2 \tag{4.140}$$

The resulting aerodynamic force vector is now defined with respect to $\mathcal{F}_{AA}$. In the equations of translational motion, however, the aerodynamic-force components are expressed in components of $\mathcal{F}_{AG}$. To use the airspeed-based aerodynamic force vector, it has thus to be transformed. If we define the aerodynamic force vector in $\mathcal{F}_{AA}$ to be $\mathbf{F_{A,AA}}$, then we have to solve the following equation

$$\mathbf{F_{A,AG}} = \mathbf{C_{AG,AA}}\mathbf{F_{A,AA}} \tag{4.141}$$

in which $\mathbf{C_{AG,AA}}$ is defined by $\mathbf{C_{AG,AA}} = \mathbf{C_{AG,TG}}\mathbf{C_{TG,AA}}$ with $\mathbf{C_{AG,TG}} = \mathbf{C_x}(-\sigma_g)$ (which is the inverse of Eq. (4.55), albeit with a groundspeed-based σ_g) and $\mathbf{C_{TG,AA}}$ given by Eq. (4.58)). Substituting $\mathbf{F_{A,AG}} \equiv (D, S, L)_g$ into Eqs. (4.110)–(4.112), gives us a new state vector for position and velocity.

The aerodynamic moments usually consist of two parts: a moment which is computed using the moment coefficients, and a moment due to the aerodynamic forces, in case the aerodynamic reference centre is not coinciding with the c.o.m.. For the latter computation, we need the aerodynamic force vector in body-frame components. These are obtained by

$$\mathbf{F_{A,B}} = \mathbf{C_{B,AA}}\mathbf{F_{A,AA}} \tag{4.142}$$

with $\mathbf{C_{B,AA}} = \mathbf{C_y}(\alpha_a)\mathbf{C_z}(-\beta_a)$.

Problems

Judge whether the following statements are TRUE or FALSE (Problems 4.1 to 4.11*).*

4.1 For a transformation between two reference frames you always need three unit-axis rotations.
4.2 A Coriolis force can occur when mass is expelled from a rotating body.
4.3 A launch vehicle that is infinitely stiff can be considered a rigid body.
4.4 In the context of entry vehicles, state variables typically include parameters such as position, velocity, attitude (orientation), and mass.
4.5 The selection of state variables for position and velocity is trivial, and always leads to Cartesian components.
4.6 In its most general form, Newton's Second Law states that force is equal to mass times acceleration.
4.7 Matrix orthonormality is easily checked by calculating the cross product of each combination of matrix columns; the result should be zero.
4.8 Because a transformation matrix is orthonormal, the inverse transformation is simply given by the transposed matrix.
4.9 The equations of motion of an arbitrary mass-varying system can be derived from those of a rigid system by introducing two apparent forces (moments): the Coriolis and relative force (moment).
4.10 Angular momentum is a scalar property of a rotating body.
4.11 To compare planetary entry probes entering atmospheres of different celestial bodies, it is more convenient to define the ballistic parameter to be mass-referenced rather than weight-referenced.
4.12 Euler angles give the preferred attitude representation because they are singularity free.

Frame transformations (*Problems* 4.13 to 4.14).

4.13 Given the reference frames A, B, C and D with corresponding rotations in Fig. 4.22 set up the static transformation from:

(a) Frame B to C, $\mathbf{C_{C,B}}$, and the inverse $\mathbf{C_{B,C}}$
(b) Frame B to D, $\mathbf{C_{D,B}}$, and the inverse $\mathbf{C_{B,D}}$

In the above questions, *first* give the sequence of transformations in terms of individual unit-axis transformations $\mathbf{C}$, and *second* carry out the matrix multiplications. Explain what mathematical principle you have used to determine the inverse matrix.

4.14 In a dynamic rotation we assume that the angular displacement $d\alpha$ (the rotation) takes place in a certain time dt. Given the four frames A, B, C and D derive the angular rate of frame D with respect to frame A, $\boldsymbol{\omega}_{D/A}$, from the sequence of transformations. Make sure to express all components of $\boldsymbol{\omega}_{D/A}$ in components of frame D.

Coordinate transformations (*Problems* 4.15 to 4.19).

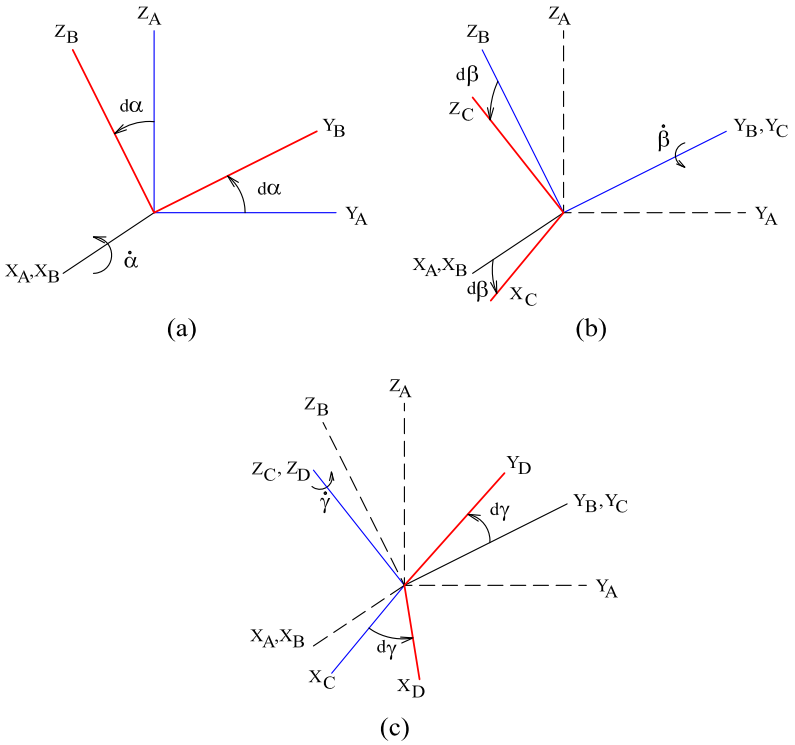


Fig. 4.22 Transformation from *A* to *D* frame. All indicated angles and rotations are positive; frames are right-handed

A Project Manager responsible for the mission analysis of a chaser satellite performing a rendezvous with a tumbling target, has decided to use Euler angles for the attitude definition of the target satellite. Unknown to him, the tumbling motion would lead to a singular attitude, crashing the simulator. To avoid this, the System Engineer has suggested to use two different sets of Euler angles and switch between them when necessary.

4.15 In the aerospace field, the common definition of the Euler angles is based on the so-called 3-2-1 transformation sequence: a rotation ψ_1 over Z, followed by a rotation θ_1 over Y, and concluded by a rotation φ_1 over X. This sequence uniquely defines the 3D orientation of one frame with respect to another. In the current situation, this is used to define the satellite's attitude (= body frame) with respect to the local-horizontal, local-vertical (the so-called vertical frame), represented by the transformation matrix $\mathbf{C}_{\mathbf{B},\mathbf{V_1}}$. You are tasked to derive this matrix.

4.16 When using three angles to define the attitude, the kinematic equations have a singularity originating from the second rotation, in the 3-2-1 sequence this is $\theta_1 = \pm\frac{1}{2}\pi$. An alternative rotation sequence is found in 1-3-2: a rotation θ_2

over Y, followed by a rotation ψ_2 over Z, and concluded by a rotation φ_2 over X. Derive the corresponding transformation matrix, $\mathbf{C}_{\mathbf{B},\mathbf{V_2}}$.

4.17 To avoid the singularity in θ_1, one could switch between the two sets of kinematic equations by converting φ_1, θ_1, and ψ_1 to φ_2, θ_2, and ψ_2. Set up this coordinate transformation.

4.18 Can this coordinate transformation always be used?

4.19 Is the method proposed by the System Engineer a fail-safe solution? If not, what would be a better alternative?

References

Cornelisse JW, Schöyer HFR, Wakker KF (1979) Rocket propulsion and spaceflight dynamics. Pitman, London

Duke EL, Antoniewicz RF, Krambeer KD (1988) Derivation and definition of a linear aircraft model. NASA Reference Publication 1207

Hughes PC (1986) Spacecraft attitude dynamics. John Wiley & Sons Inc, New York

Kuipers J (2002) Quaternions and rotation sequences. Princeton University Press. Aug 19, Fifth printing and first paperback printing

Montebruck O, Gill E (2000) Satellite orbits. models, methods and applications. Springer Verlag

Mooij E (1997) The motion of a vehicle in a planetary atmosphere. Series 08: astrodynamics and satellite systems 01. Delft University Press

Vinh NX (1981) Optimal trajectories in atmospheric flight. Elsevier Scientific Publishing Company, Amsterdam, The Netherlands

Wertz J (1978) Spacecraft attitude determination and control. In: Astrophysics and space science library, vol 73

Chapter 5
Planar Motion



5.1 Introduction

A space vehicle orbiting the Earth has large potential and kinetic energy—for each kg in a low Earth orbit the total energy amounts to about 30 MJ. This energy content will be even higher for vehicles entering the Earth's sphere of influence at hyperbolic velocities, typically associated with return flights from the Moon or interplanetary trajectories from Mars. When the vehicle needs to return to Earth, it has to reduce this large energy to the level of potential energy that the vehicle has after landing on the surface of Earth, and—what is even more constraining—with zero relative velocity. The only mechanism to dissipate the energy is by means of aerodynamic friction with the vehicle wall during its flight through the atmosphere. Satellites and other space vehicles that travel through the denser layers of the atmosphere with high velocities are subject to large aerodynamic forces and reach high temperatures as a result of the aerodynamic heating.

In general, by allowing these objects to fall freely through the atmosphere they will disintegrate and burn up completely, except, maybe, for very large and very heavy parts. This type of entry is beyond the scope of this text, so in this chapter we focus on the return of space vehicles, which, by their design and thermal protection, are able to do a full re-entry and remain unscathed. Of course this is an essential requirement for manned space vehicles, but is also relevant for other space missions.

The discussion in this chapter will be centred around re-entry trajectories into the Earth's atmosphere, although the derived theory is also applicable to descent trajectories into the atmospheres of other planets (e.g., Mars), obviously with a change of the relevant atmospheric parameters. Generally, it is assumed that the atmosphere starts at an altitude of 120 km, although the official definition for the beginning of space (or the boundary of the Earth's atmosphere) is an altitude of 100 km. However, for high-velocity entries the noticeable presence of the atmosphere can be in the range of 200–300 km.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_5

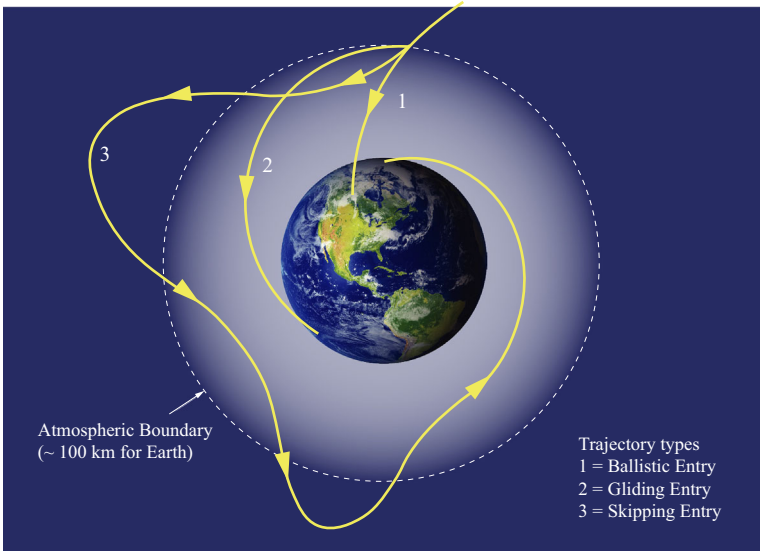


Fig. 5.1 Three different entry mechanisms

There are generally three distinguishable types of re-entry trajectories (Fig. 5.1):

1. The ballistic trajectory (zero-lift)
2. The glide trajectory (with lift)
3. The skip trajectory, whereby the vehicle repeatedly enters and then exits the atmosphere (again with lift)

These trajectories are characterised by different durations, flight-range covered, and a large variation in thermo-mechanical loads. What they have in common, though, is that the description of the corresponding flight mechanics is highly non-linear, not only due to the large velocity range, but also due to the strongly variable environment. In summary, one could say that the re-entry problem is typically a three-dimensional problem, where one can start from almost arbitrary positions, with the objective to reach some landing point, while fulfilling constraints on thermo-mechanical loads in an environment that is not (so) friendly. The exact equations of entry dynamics are not solvable analytically, so one has to rely on computer models to analyse the re-entry problem.

In the many years leading to the space age, when computers were not commonly available—and even if they were, they were too slow to allow for extensive mission planning and analysis—the need to develop analytical models was obvious. Major progress on analytical models was made in the 1950s, when the advantage of hypersonic missiles was acknowledged. The works of Allen and Eggers (1958) on aerodynamic heating of ballistic missiles entering the Earth's atmosphere at high supersonic speeds, of Allen et al. (1958) on the performance of long-range hypervelocity vehicles, as well as that of Chapman (1959) on an approximate analytical

method for studying entry into planetary atmospheres, were pioneering in that sense. However, they built further on the work of Sänger (1933), who was among the first to recognise the favourable (and strategic) connection between speed and range, and was perhaps the first to exploit the speed factor in designing a long-range bomber (Sänger and Bredt 1944), developing theories on gliding and skipping flight in the process.

Miele (1962), in his book on the analysis of trajectories of aircraft, missiles, satellites and spacecraft subjected to gravitational, aerodynamic and thrust forces, summarised many of the earlier (analytical) findings on ballistic, gliding and skipping flight. This was also done by several others, amongst them Regan and Anandakrishnan (1993) in their book on the dynamics of atmospheric re-entry.

The common aspect of all analytical formulations developed in the early years is that they are so-called first-order solutions, only applicable to certain parts of the re-entry domain. However, these models formed the basis for improved methods to more accurately describe the motion of atmospheric entry vehicles. Loh (1968, 1969) developed a second-order solution that gives an accurate analytical description valid for the entire region of entry in terms of lift-to-drag ratio and initial flight-path angle. Note that his theory is empirical, and is based on data from extensive numerical integration of entry trajectories.

Moreover, his two books give an overview of the many significant advances in the field of re-entry and planetary entry, divided into (i) entry dynamics, thermodynamics, physics and radiation, (ii) entry ablation and heat transfer, (iii) entry experimentation, (iv) entry concepts and technology, and (v) advanced entry programs. Despite the many years that have passed, these books still provide a valuable source of information.

Vinh et al. (1980) summarise the first- and second-order solutions and give some practical applications. Also, they include discussions on the (powered) launch phase, the ballistic flight over the Earth outside the atmosphere, and the interface with atmospheric entry. The main contribution of the book, though, is the presentation of a unified theory that is valid for all types of entry, including those with lift and bank modulation. Whereas Loh's second-order solution is empirical, this unified theory is analytical of nature. Starting from the model of Chapman (1959) the restrictive assumptions of, amongst others, planar motion are removed and the results are extended to three-dimensional entry trajectories, where only minor assumptions about atmosphere and aerodynamic characteristics need to be made. From the resulting exact equations, all the known first- and second-order solutions can be obtained, thus confirming the universality of this theory.

It can easily be understood that a complete non-linear description of the "re-entry problem" is too difficult for a fundamental understanding. Therefore, in this chapter we will simplify the 3D re-entry problem to a 2D planar motion, and assume a simplified environment. The corresponding simplified equations of motion will be derived in Sect. 5.2. Furthermore, starting from these equations, in Sects. 5.3 up to 5.5 the three different trajectory types of ballistic, gliding and skipping entry will be discussed in more detail. Suitable simplifications will be made for each trajectory type, allowing for an analytical solution of the equations of motion. It is noted that these analytical solutions are the earlier mentioned first-order solutions.

It is beyond the scope of this book to cover the second-order solution and its applications in detail, as they will not provide the fundamental understanding that the first-order solutions give. It is noted, though, that when second-order effects are disregarded, the second-order solution reduces either analytically or numerically to each one of the first-order solutions. Therefore, all first-order solutions simply become different special cases of the second-order solution (Loh 1968; Vinh et al. 1980). Yet, for the sake of completeness, a summary the second-order solution has been added at the end of this chapter, see Sect. 5.6. The model description includes a limited verification for a skipping re-entry flight, using the analytical first-order and second-order solutions on one hand, and the results from a numerical simulation on the other.

5.2 Equations of Motion

A re-entry mission typically starts in orbit, well away from the atmosphere, with a thrusted manoeuvre to target the vehicle for interception with the so-called atmospheric interface. After this manoeuvre the descending vehicle is in free flight and approaches the atmosphere on a Keplerian orbit. When the atmospheric density builds up also the aerodynamic force increases and slowly the vehicle begins to deviate from its Keplerian orbit: the atmospheric entry phase begins. For a return from low Earth-orbit this usually takes place between 100–120 km altitude, but, for instance, for high-velocity return flights from the Moon or one of the planets this may be well above 200 km.

The entry velocity from an elliptical orbit is near (local) circular velocity at the entry altitude, while an entry from a high-altitude orbit is effected at super-circular velocity. In particular, for a Moon-mission return to the Earth's atmosphere, the entry velocity is near parabolic speed, and in the case of an entry following an interplanetary flight, the vehicle enters the planetary atmosphere at hyperbolic velocity.

By intuition one will understand that given the large range of possible entry velocities a safe re-entry and landing on the planet's surface is not trivially guaranteed. In terms of velocity not only the magnitude V is an important factor, but also the direction, i.e., the so-called flight-path angle, γ. For a safe recovery of a ballistic vehicle, for an initial altitude h_E one must select the velocity and flight-path angle, V_E and γ_E, such that during entry the associated heat load and deceleration acting on the vehicle remain within acceptable limits. For a lifting hypervelocity vehicle, the entry position must also be selected such that the vehicle can make a safe approach to and landing on the selected airfield.

The range of trajectories that meets with the constraints encloses a three-dimensional spatial tube called the *entry corridor*. This concept has been extensively discussed in literature of the 1960s and 1970s, most notably by Chapman (1960), and in a study about manoeuvring performance of lifting re-entry vehicles by Mandell (1962). Later on in this chapter we will look into somewhat more detail of the entry corridor (Sect. 5.4.3).

Here, we will establish the basic set of equations with which the entry in a planetary atmosphere can be studied. In Chap. 4 (Sect. 4.5), with supporting material in Appendix A, the derivation of the non-linear equations of motion has been presented. In general, these equations hold for any mass-varying body and describe the 3D-motion of the vehicle with respect to inertial space. To express position and velocity with respect to any rotating central body, e.g., the Earth, one can either reformulate the equations of motion (as shown in Appendix A, Sect. A.4.2), or apply transformations taking the rotation of the central body into account.

For the upcoming discussion, though, we will assume that the Earth is spherical, but not rotating. The atmosphere is considered to be at rest, which means that there are no winds, and that the airspeed is equal to the ground speed. A further assumption is that the vehicle's trajectory is restricted to a flat plane, determined by the velocity vector at atmospheric entry and the position vector from the centre of the Earth. This implies that the resulting aerodynamic forces are also constrained to this plane, or in other words, there are no transverse forces. These transverse forces are of importance when a certain distance perpendicular to the trajectory plane needs to be covered, for instance, to reach the desired landing site. This distance is called the cross-range and is the result of a process called banking. Although important, it will only briefly be considered in Sect. 5.4.4. A detailed discussion is left for Chap. 7 (Sect. 7.6).

The situation for a vehicle with (constant) mass m and no thrust at an arbitrary moment t is shown in Fig. 5.2. The motion of the vehicle is driven by the aerodynamic lift, L, and drag, D, as well as the weight, $W = mg$. The gravitational acceleration, g, is, in general, a function of position (see Chap. 2, Sect. 2.2), but in this chapter we will use the approximation of constant gravity. The orientation of the velocity vector with respect to the local horizontal plane is given by the flight-path angle γ. Both in aircraft and re-entry performance studies, this angle is commonly defined to be

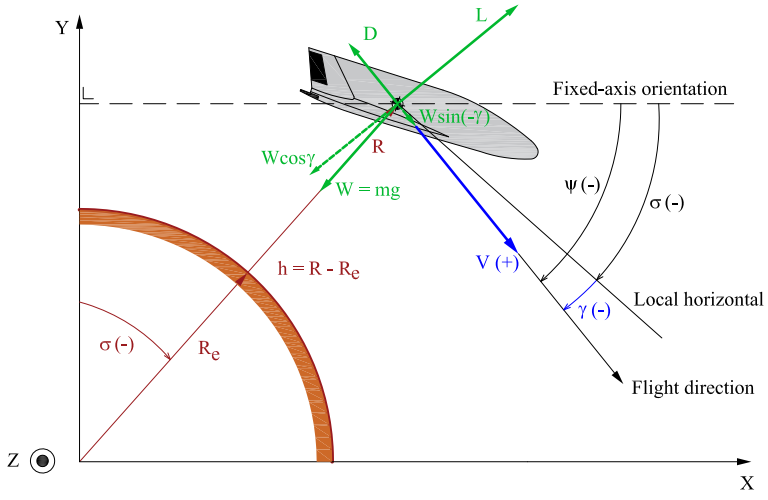


Fig. 5.2 Two-dimensional planar motion. Note that ψ, γ and σ are negative angles

negative for a velocity below the horizon. The velocity is, in this case, the velocity relative to the central body, i.e., the groundspeed defined in the previous chapter. Because we will not be discussing the influence of wind, and thus have to distinguish between groundspeed and airspeed, there is no need to use the subscript "g" for the velocity variables.

For the sake of the derivation of the equations of motion we will introduce the auxiliary angle ψ that gives the orientation of velocity V with respect to the fixed axis orientation of the inertial frame. Further, the complimentary angle σ is introduced, given by the relation

$$-\psi = -\sigma - \gamma \quad \Rightarrow \quad \psi = \sigma + \gamma \tag{5.1}$$

In line with the definition of γ, both ψ and σ are negative, as shown in Fig. 5.2, as they are both defined starting from the fixed axis orientation going downwards (note that the direction of the axis of rotation is negative, pointing inside the page).

The equations of motion will be expressed in the flight-path reference frame, i.e., parallel and perpendicular to V. Application of Newton's second law in velocity direction yields:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{5.2}$$

The normal acceleration is given by $V\frac{\mathrm{d}\psi}{\mathrm{d}t}$[1], so reformulation of Newton's second law gives

$$mV\frac{\mathrm{d}\psi}{\mathrm{d}t} = +L - mg\cos\gamma \tag{5.3}$$

From Eq. (5.1) we find

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\mathrm{d}\sigma}{\mathrm{d}t} + \frac{\mathrm{d}\gamma}{\mathrm{d}t} \tag{5.4}$$

so to rewrite Eq. (5.3) into an equation of motion ($\frac{\mathrm{d}\gamma}{\mathrm{d}t} = ...$) we need to find an expression for $\frac{\mathrm{d}\sigma}{\mathrm{d}t}$. Looking again at Fig. 5.2, we see that the position vector $\mathbf{R}$ describes an instantaneous circular motion, yielding a tangential velocity of the centre of mass of $R\frac{\mathrm{d}\sigma}{\mathrm{d}t}$. This tangential velocity is, of course, equal to the projection of $\mathbf{V}$ on the local horizontal, so:

$$R\frac{\mathrm{d}\sigma}{\mathrm{d}t} = -V\cos\gamma \tag{5.5}$$

Substitution of Eqs. (5.4) and (5.5) into Eq. (5.3) and rearranging terms results in

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{gR}\right) = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.6}$$

[1] Actually, the normal acceleration is $-V\frac{\mathrm{d}(-\psi)}{\mathrm{d}t}$, caused by a negative angular acceleration around the negative Z-axis.

where we have introduced the local circular velocity, V_c:

$$V_c = \sqrt{gR} \tag{5.7}$$

The term $\frac{V^2}{V_c^2}$ shows the influence of the Earth's curvature on the equation of motion.

To complete the equations of motion we have to add the kinematic equation for the rate of change of distance R; for a spherical Earth this is equivalent to the rate of change of altitude h. Thus, summarised, the equations of motion describing the (2D) planar motion of a mass point around a spherical central body read:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{5.8}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.9}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma \tag{5.10}$$

Example 5.1 (*Influence of planetary rotational rate*) In general, the three-dimensional motion of entry vehicles is governed by six first-order differential equations that include the sphericity of the central body, rotating with angular rate ω_{cb}. These equations are shown in Chap. 4, Eqs. (4.110)–(4.115). We highlight the equation for the rate of change of the flight-path angle:

$$\begin{aligned} V\dot{\gamma} = \frac{L\cos\sigma}{m} - g\cos\gamma + 2\omega_{cb}V\cos\delta\sin\chi + \frac{V^2}{R}\cos\gamma + \\ + \omega_{cb}^2 R\cos\delta(\cos\delta\cos\gamma + \sin\gamma\sin\delta\cos\chi) \end{aligned} \tag{5.11}$$

For a flight along the equator, the latitude $\delta = 0$, and the heading of the vehicle is due east, i.e., $\chi = 90°$. Substituted above, we obtain:

$$V\dot{\gamma} = \frac{L\cos\sigma}{m} - g\cos\gamma + 2\omega_{cb}V + \frac{V^2}{R}\cos\gamma + \omega_{cb}^2 R\cos\gamma \tag{5.12}$$

or, rewriting it in the form of Eq. (5.9):

$$V\dot{\gamma} = \frac{L}{m}\cos\sigma - g\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) + 2\omega_{cb}V + \omega_{cb}^2 R\cos\gamma \tag{5.13}$$

Apart from the $\cos\sigma$ term, which reduces the vertical component of the lift when banking the vehicle, there are two extra acceleration terms on the right-hand side: the Coriolis acceleration, $2\omega_{cb}V$, and the centripetal acceleration, $\omega_{cb}^2 R\cos\gamma$ ($\approx \omega_{cb}^2 R$ for small γ).

For the initial flight phase, density is still so small that despite the large velocity the lift acceleration is much smaller than the acceleration due to gravity. This means that we are dealing with three components of the acceleration, i.e., $g \approx 9.45\,\text{m/s}^2$, $2\omega_{cb}V \approx 1.08\,\text{m/s}^2$ ($\omega_{cb} \approx 7.29{\cdot}10^{-5}$ rad/s for the Earth, and $V_E \approx 7{,}450\,\text{m/s}$ for an orbital entry), and $\omega_{cb}^2 R \approx 0.03\,\text{m/s}^2$ ($R_e = 6{,}378.137\,\text{km}$ and $h_E = 120\,\text{km}$). The main error due to ignoring the rotational rate of the Earth stems from neglecting the Coriolis force. However, this is only true at the beginning of entry: the moment the aerodynamic forces become large, they will overshadow the Coriolis force and the simplified equations will be accurate enough. This means, that all approximations we will make later in this chapter will hold as long as the aerodynamic forces are the dominating components.

To conclude this example, we will show the effect of ignoring the angular rate of the Earth on the flight-path angle for a re-entry vehicle returning from Earth orbit. Using $V_E = 7{,}438\,\text{m/s}$ and $\gamma_E = -1.43°$ (for $h_E = 120\,\text{km}$) we find the flight-path angle profile shown in Fig. 5.3. The calculations were done until a final altitude of $h_F = 60\,\text{km}$ was reached. It is clear that omitting the angular rate of the Earth will allow the vehicle to dive deeper into the atmosphere, rather than being deflected right from the beginning. However, reaching lower altitudes with a relatively larger velocity will create a larger push by the atmosphere (= lift), so the vehicle will enter into a much more vehement skip (larger positive γ) with much larger thermo-mechanical loads (not shown here). Due to the larger drag deceleration, conform Eq. (5.8) the vehicle will loose its velocity (= kinetic energy) more rapidly.

In Fig. 5.4 the corresponding h-V diagram is shown, confirming the previous discussion. It is noted, though, that because of the slower motion (*with* planetary angular rate) the time that γ is positive is larger, in the end resulting in a larger first-skip altitude. During the subsequent skips this effect has reduced and the skip altitudes have become lower.

In conclusion, Eqs. (5.8)–(5.10) could be used as the starting point for an analytical analysis—we will see much more of that later on in this chapter. However, they can *only* be used for a numerical simulation when applied in the *right flight regime* with the *proper initial conditions*. ■

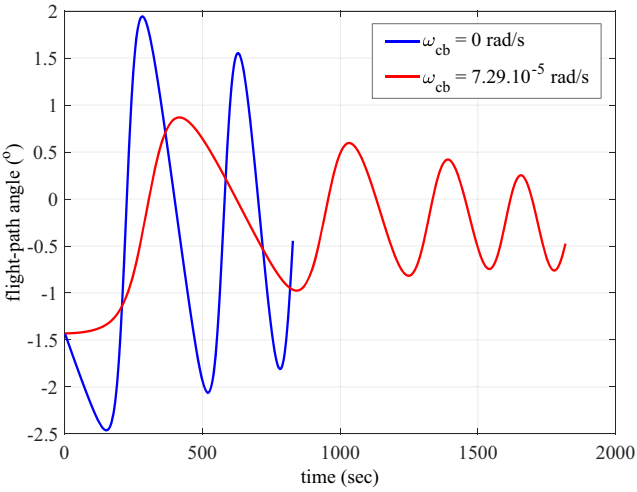


Fig. 5.3 The influence of planetary angular rate on the flight-path angle

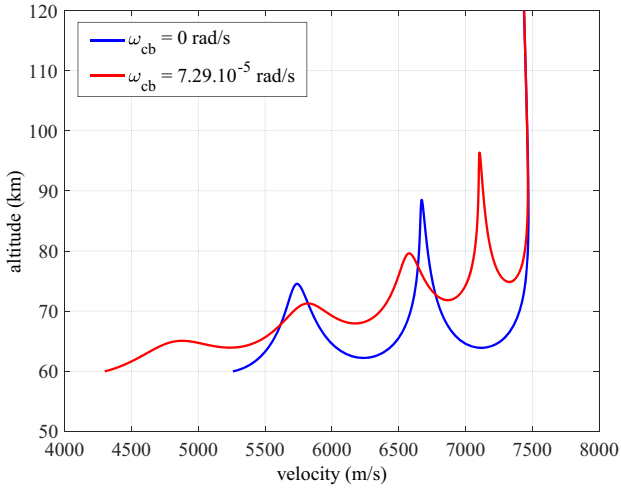


Fig. 5.4 The influence of planetary angular rate on the *h*-*V* diagram

5.3 Ballistic Entry

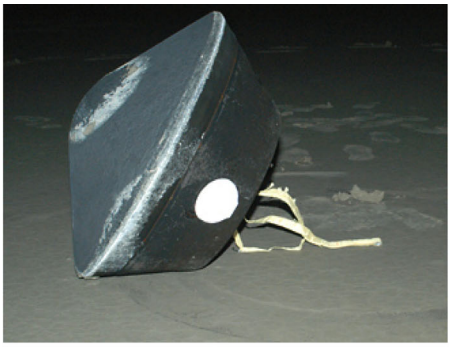
The ballistic entry is by definition a flight where no lift is acting on the vehicle, so for an atmospheric flight the trajectory is governed by drag and gravity only. This trajectory type typically applies to axisymmetric vehicles that fly with a zero angle of attack. Examples of such vehicles are the Mars entry capsules, sample return vehicles and the Huygens entry probe (Fig. 5.5), and in a way also the (manned) Apollo

capsule. However, we specifically say *in a way*, because although geometrically symmetric the Apollo capsule flew with a non-zero angle of attack to reduce the loads on the vehicle and to introduce some manoeuvrability. A small shift in its centre of mass in Z-direction gave the capsule an equilibrium angle of attack between $-23°$ and $-24°$, sufficient to create a lift-to-drag ratio of $L/D \approx 0.3$. Due to the large velocity during entry, part of the flight resembled a ballistic trajectory, though.

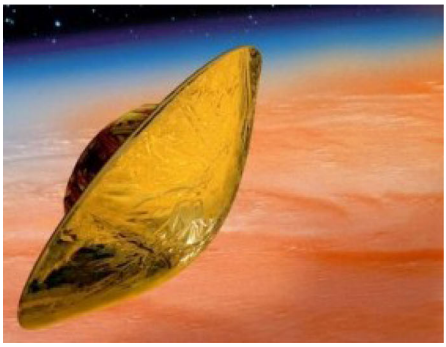
Subsequently, in Sect. 5.3.1 we will discuss flight-mechanics related issues of the ballistic trajectory. The trajectory shape is introduced and we also derive a relation between velocity and altitude. We will conclude this section by looking at the deceleration profile and finding an expression for the maximum deceleration that occurs, as well as the corresponding altitude at which this happens. In Sect. 5.3.2 we move to the thermal loads and look at both the maximum heat flux and the altitude at which this occurs. The altitudes at which the two peak values of mechanical and thermal load occur will also be related to each other.



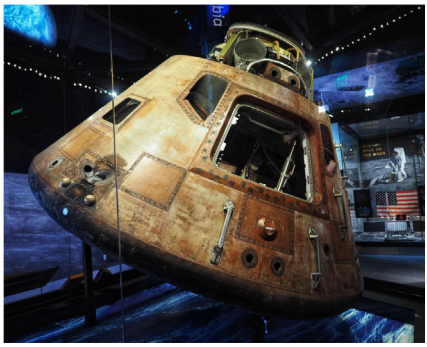
(a) Mars Science Laboratory. Image: NASA.



(b) Stardust sample return capsule. Image: NASA.



(c) Huygens entry probe. Image: ESA.



(d) Apollo-11 Command Module. Image: author.

Fig. 5.5 Examples of (semi-)ballistic entry capsules

5.3.1 Flight Mechanics

5.3.1.1 Flight Path

We first consider the equation of motion Eq. (5.9), perpendicular to the trajectory, and repeated here for convenience:

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.14}$$

By definition of a ballistic trajectory, the lift, L, is zero, and as a consequence also the mass of the vehicle disappears:

$$V\frac{\mathrm{d}\gamma}{\mathrm{d}t} = -g\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.15}$$

Let us now introduce the altitude h as independent variable, instead of t, such that:

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \frac{\mathrm{d}\gamma}{\mathrm{d}h}\frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma\frac{\mathrm{d}\gamma}{\mathrm{d}h} \tag{5.16}$$

where we have used the kinematic equation Eq. (5.10).

With Eq. (5.16) and assuming that V_c is (much) larger than V, Eq. (5.15) can be rewritten as:

$$\tan\gamma\frac{\mathrm{d}\gamma}{\mathrm{d}h} = -\frac{g}{V^2} \tag{5.17}$$

For the most important part of the hypersonic trajectory, i.e., the initial phase where the largest mechanical and thermal loads occur, the velocity is large, so $V^2 >> g$. For medium to large flight-path angles γ, one can thus use the approximation:

$$\frac{\mathrm{d}\gamma}{\mathrm{d}h} \approx 0 \tag{5.18}$$

or, in other words, $\gamma =$ constant. This implies that the vehicle describes a rectilinear path during the initial segment of the descent.

Example 5.2 (*Travelled distance*) At the end of its exoatmospheric flight, a ballistic vehicle generally enters the atmosphere at a rather steep angle, so that the travelled distance over the Earth is small and the influence of the Earth's curvature can be neglected (Fig. 5.6). The travelled distance, d, from an altitude of $h = 100\,\mathrm{km}$ for a flat Earth under an entry angle $\gamma = -20°$, amounts to a mere 275 km ($\tan|\gamma| = \frac{h}{d}$, so $d = \frac{h}{\tan|\gamma|}$).

■

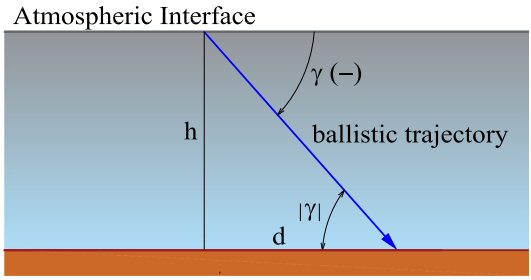


Fig. 5.6 Travelled distance in the ballistic flight

5.3.1.2 Altitude-Velocity Profile

To find a relation between altitude and velocity we begin again from the equations of motion, notably:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{5.19}$$

Rather than taking the time as independent variable, we change to altitude, so

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{\mathrm{d}V}{\mathrm{d}h}\frac{\mathrm{d}h}{\mathrm{d}t} \tag{5.20}$$

and with Eq. (5.10) this is rewritten as

$$\frac{\mathrm{d}V}{\mathrm{d}t} = V\sin\gamma\frac{\mathrm{d}V}{\mathrm{d}h} \tag{5.21}$$

Substituting the above result into Eq. (5.19) and rearranging terms yields

$$\frac{V}{g}\frac{\mathrm{d}V}{\mathrm{d}h} = \frac{-D}{mg\sin\gamma} - 1 = -\frac{C_D\frac{1}{2}\rho V^2 S}{mg\sin\gamma} - 1 \tag{5.22}$$

By defining the ballistic parameter K, a commonly used characteristic for ballistic vehicles, as

$$K = \frac{mg}{C_D S} = \frac{W}{C_D S} \tag{5.23}$$

and realising that at high velocities and steep trajectories, the first term on the right-hand side of Eq. (5.22), $D/mg\sin\gamma$, is much larger than 1, Eq. (5.22) can be approximated by:

$$\frac{\mathrm{d}V}{V} = -\frac{1}{2}\frac{g\rho\mathrm{d}h}{K\sin\gamma} \tag{5.24}$$

By assuming an exponential atmospheric-density profile (see Chap. 2, Sect. 2.3.1),

$$\frac{\rho}{\rho_0} = e^{-\frac{h}{H_s}} = e^{-\beta h} \tag{5.25}$$

we have a relation between ρ and h. However, there is an easier way to solve Eq. (5.24) then by direct substitution of Eq. (5.25). The basis for the derivation of the exponential atmosphere is the hydrostatic equation for the atmosphere:

$$\mathrm{d}p = -\rho g \mathrm{d}h \tag{5.26}$$

This equation gives a direct relation between altitude and pressure, and by substituting this we can eliminate ρ:

$$\frac{\mathrm{d}V}{V} = \frac{1}{2}\frac{\mathrm{d}p}{K \sin\gamma} \tag{5.27}$$

For a constant flight-path angle $\gamma = \gamma_E$ and a constant K, Eq. (5.24) can be integrated directly between the initial boundaries V_E and p_E, and V and p in a random point of the descent trajectory:

$$\int_{V_E}^{V} \frac{\mathrm{d}V^*}{V^*} = \frac{1}{2\,K \sin\gamma_E} \int_{P_E}^{p} \mathrm{d}p^* \tag{5.28}$$

which yields

$$\ln\frac{V}{V_E} = \frac{1}{2}\frac{p - p_E}{K \sin\gamma_E} \tag{5.29}$$

For the entry altitude h_E, it holds that $p_E \approx 0$ (or $\rho_E \approx 0$, for that matter), so Eq. (5.29) can also be written in the form:

$$\frac{V}{V_E} = e^{\frac{1}{2}\frac{p}{K \sin\gamma_E}} \tag{5.30}$$

The above equation gives the change of normalised velocity V/V_E with altitude, for a given K and γ_E in any atmosphere, wherein $p = p(h)$ is fixed. Equivalently, realising that p follows the same exponential profile as the density, or

$$\frac{p}{p_0} = \frac{\rho}{\rho_0} = e^{-\beta h} \tag{5.31}$$

With $p = \rho RT = \frac{g}{\beta}\rho$ it thus follows that

$$\frac{V}{V_E} = \exp\left(\frac{1}{2}\frac{g\rho}{K\beta \sin\gamma_E}\right) \tag{5.32}$$

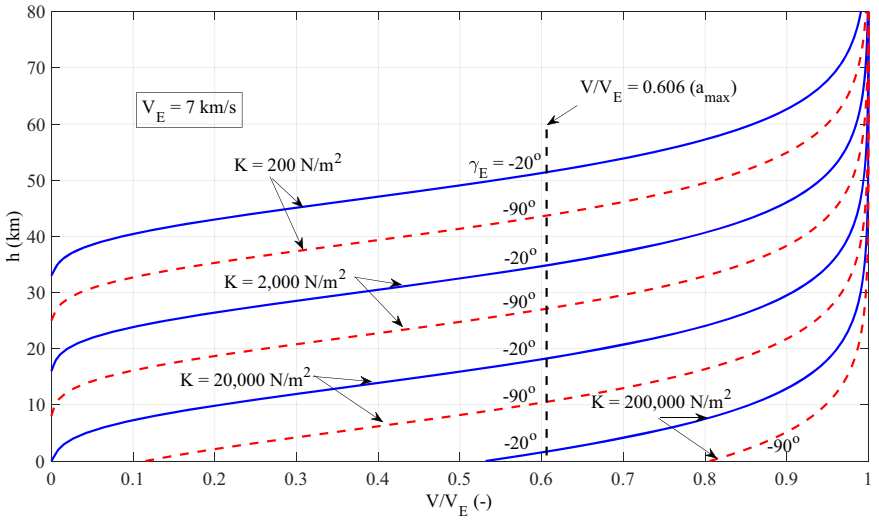


Fig. 5.7 Descent in the ballistic flight, for an Earth exponential atmosphere with $\rho_0 = 1.225\,\text{kg/m}^3$, and $H_s = 7{,}200\,\text{m}$

It is clear that the change of V/V_E with altitude is completely determined by the ballistic parameter K and the entry angle γ_E (Fig. 5.7). As K and/or γ_E get(s) larger (i.e., more negative), the vehicle will penetrate deeper into the atmosphere before the vehicle's velocity is reduced significantly.

For a given descent angle γ_E the curves for different values of K (increasing with factor 10) are equidistant and independent of V/V_E. Suppose we look at a given value for V/V_E, and substitute $K_2 = 10K_1$

$$\exp\left(\frac{1}{2}\frac{g\rho_0 e^{-\beta h_1}}{K_1 \beta \sin\gamma_E}\right) = \exp\left(\frac{1}{2}\frac{g\rho_0 e^{-\beta h_2}}{10K_1 \beta \sin\gamma_E}\right) \tag{5.33}$$

or

$$e^{-\beta h_1} = \frac{e^{-\beta h_2}}{10} \Rightarrow \tag{5.34}$$

$$\beta h_1 = \beta h_2 + \ln 10 \Rightarrow \tag{5.35}$$

$$\Delta h = h_1 - h_2 = \frac{\beta}{\ln 10} = H_s \ln 10 \tag{5.36}$$

With a scale height of $H_s = 7{,}200\,\text{m}$ this means that the altitude difference is $\Delta h = 16.6\,\text{km}$.

The approximation $D > mg\sin\gamma$ is not applicable at the beginning and the end of the ballistic trajectory due to the extremely low density and low velocity, respectively. More precise analysis of the motion indicate that in between the extremities the

trajectory is for the greater part rectilinear. If an extremely low velocity is reached, then the descent angle will increase significantly as a result of the weight influence, approaching a vertical free-fall. The significant points of the trajectory treated below, i.e., the points of maximum deceleration and maximum heat flux, are all found in this rectilinear part of the trajectory.

Example 5.3 (*Ballistic parameter*) The ballistic parameter, $K = \frac{mg}{C_D S}$ is a characteristic parameter that enables us to compare different ballistic vehicles with each other better. However, sometimes in literature one finds an alternative expression, notably

$$K_m = \frac{m}{C_D S} \tag{5.37}$$

which is "mass" referenced (hence the subscript 'm') rather than weight-referenced.

To get a feeling for some numerical values, we will give the related values for the Apollo Command Module and the Huygens probe (Fig. 5.5).

For Apollo (NAA 1965), with a mass of $m = 4{,}976\,\text{kg}$, a hypersonic drag coefficient of $C_D = 1.45$ ($\alpha = 0°$, for a Mach number $M > 3$) and a reference area of $S_{ref} = 12\,\text{m}^2$ (given the diameter $d = 3.9\,\text{m}$), $K = 2{,}806\,\text{N/m}^2$. Dividing this number by $g = 9.81\,\text{m/s}^2$, we get $K_m = 286\,\text{kg/m}^2$.

The Huygens probe (Mooij 1992) is a much smaller vehicle, it has a mass of $m = 318\,\text{kg}$, a hypersonic drag coefficient of $C_D = 1.54$ ($\alpha = 0°$, $M > 4$) and a reference area of $S_{ref} = 5.7\,\text{m}^2$ (given a diameter $d = 2.7\,\text{m}$). Thus, $K = 49\,\text{N/m}^2$ and $K_m = 36\,\text{kg/m}^2$. The latter value is referenced to Titan, with $g = 1.354\,\text{m/s}^2$. To put Huygens in perspective with Earth-referenced vehicles: $K_m = 5\,\text{kg/m}^2$.

■

5.3.1.3 Deceleration

The deceleration during the ballistic re-entry follows from Eq. (5.8), i.e.,

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{5.38}$$

With the deceleration being defined as $\bar{a} = -\frac{\mathrm{d}V}{\mathrm{d}t}$ and neglecting the influence of the vehicle weight term ($D >> mg\sin\gamma$) it follows that

$$\bar{a} = -\frac{\mathrm{d}V}{\mathrm{d}t} = +\frac{C_D \frac{1}{2}\rho V^2 Sg}{W} = \frac{g}{2K}\rho V^2 \tag{5.39}$$

To proceed we need an expression for ρ, which we have with Eq. (5.32)

$$\frac{V}{V_E} = \exp\left(\frac{1}{2}\frac{g\rho_0 e^{-\beta h}}{K\beta\sin\gamma_E}\right) \Rightarrow \tag{5.40}$$

$$\rho = \frac{2\,K\beta \sin\gamma_E}{g} \ln\frac{V}{V_E} \tag{5.41}$$

Substituting Eq. (5.41) into Eq. (5.39) yields:

$$\bar{a} = -\frac{\mathrm{d}V}{\mathrm{d}t} = \beta \sin\gamma_E V_E^2 \left(\frac{V}{V_E}\right)^2 \ln\frac{V}{V_E} \tag{5.42}$$

For a given entry angle γ_E, the maximum deceleration is achieved for the velocity that follows from:

$$\frac{\mathrm{d}\bar{a}}{\mathrm{d}\,(V/V_E)} = 0 \tag{5.43}$$

or

$$2\beta \sin\gamma_E V_E^2 \left(\frac{V}{V_E}\right) \ln\frac{V}{V_E} + \beta \sin\gamma_E V_E^2 \left(\frac{V}{V_E}\right)^2 \frac{1}{\frac{V}{V_E}} = 0 \tag{5.44}$$

This constraint yields the simple solution:

$$\ln\frac{V'}{V_E} = -0.5 \tag{5.45}$$

or

$$\frac{V'}{V_E} = \frac{1}{\sqrt{e}} = 0.606 \tag{5.46}$$

with V' being the velocity for which maximum deceleration occurs. Substitution of this condition for V'/V_E into Eq. (5.32) yields a maximum deceleration of:

$$\bar{a}_{max} = -\left(\frac{\mathrm{d}V}{\mathrm{d}t}\right)_{max} = -\frac{\beta \sin\gamma_E}{2e} V_E^2 \tag{5.47}$$

Using the velocity condition Eq. (5.45), we can use Eq. (5.41) to get the corresponding density—and thus altitude—at which this occurs. We find:

$$\rho' = -\frac{K\beta \sin\gamma_E}{g} = \rho_0 e^{-\beta h'} \Rightarrow \tag{5.48}$$

$$h' = \frac{1}{\beta} \ln\left(-\frac{\rho_0 g}{K\beta \sin\gamma_E}\right) \tag{5.49}$$

The maximum deceleration thus consistently occurs for a fixed value of V/V_E, Eq. (5.46), and increases with increasing values of V_E and γ_E, Eq. (5.47), but is nevertheless independent of the ballistic parameter. However, a larger value of K

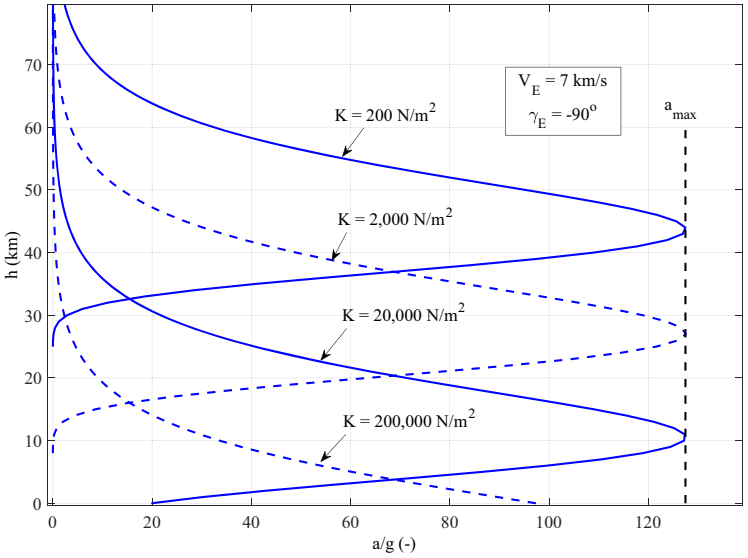


Fig. 5.8 Deceleration in the ballistic flight with $V_E = 7\,\text{km/s}$ and $\gamma_E = -20°$, for an Earth exponential atmosphere with $\rho_0 = 1.225\,\text{kg/m}^3$, and $H_s = 7{,}200\,\text{m}$

does result in a lower altitude for the maximum deceleration, Eq. (5.49); this altitude is dependent on γ_E, but not on V_E!

The above characteristics are illustrated in Figs. 5.8 and 5.9 for $V_E = 7\,\text{km/s}$ and $\gamma_E = -20°$ and $-90°$, respectively. To further analyse the relation between K and a_{max}, we solve Eq. (5.49) for $h' = 0$:

$$\ln\left(-\frac{\rho_0 g}{K\beta\sin\gamma_E}\right) = 0 \Rightarrow \tag{5.50}$$

$$-\rho_0 g = K\beta\sin\gamma_E \tag{5.51}$$

or

$$K\sin\gamma_E = -\frac{\rho_0 g}{\beta} \tag{5.52}$$

Filling in the numerical values shows that the maximum deceleration is achieved exactly at sea level when $|K\sin\gamma_E| = 86{,}500\,\text{N/m}^2$. If $|K\sin\gamma_E|$ is larger than this value, the maximum deceleration will thus not be achieved. Note that even for $\gamma_E = -20°$ the maximum deceleration is very large ($a_{max} = 43\,\text{g}$), so ballistic flights under these entry conditions are not suitable for manned space flight.

From Eqs. (5.42) and (5.47) it follows directly that:

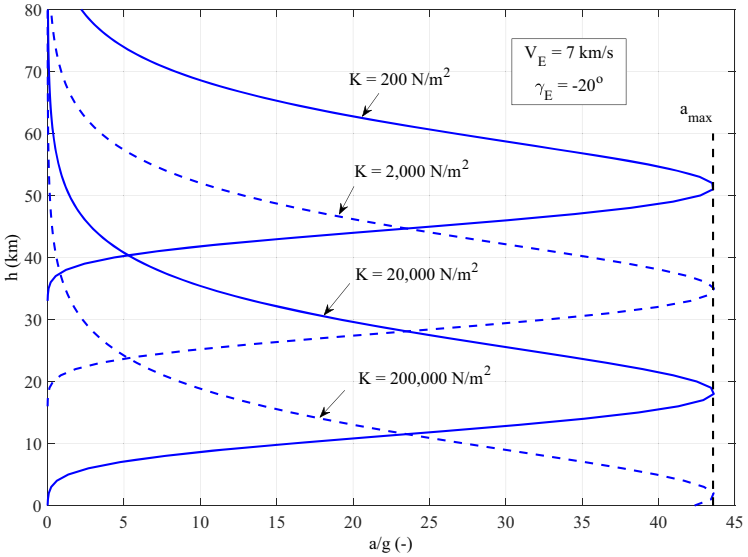


Fig. 5.9 Deceleration in the ballistic flight with $V_E = 7\,\text{km/s}$ and $\gamma_E = -90°$, for an Earth exponential atmosphere with $\rho_0 = 1.225\,\text{kg/m}^3$, and $H_s = 7{,}200\,\text{m}$

$$\frac{\bar{a}}{\bar{a}_{max}} = \frac{\mathrm{d}V/\mathrm{d}t}{(\mathrm{d}V/\mathrm{d}t)_{max}} = -2e\left(\frac{V}{V_E}\right)^2 \ln\frac{V}{V_E} = -5.44\left(\frac{V}{V_E}\right)^2 \ln\frac{V}{V_E} \tag{5.53}$$

This relationship is shown in Fig. 5.10 and holds for all entry conditions, independent of K and γ_E. From this figure it is clear that during a large part of the descent trajectory high decelerations occur, as a_{max} is generally very large (see again Figs. 5.8 and 5.9). Furthermore, from Eq. (5.39) it appears that for a constant C_D the maximum deceleration is driven by a maximum dynamic pressure $\bar{q} = \frac{1}{2}\rho V^2$. This is obvious from the fact that since we neglected the influence of weight the deceleration is only driven by the drag force, D.

Example 5.4 (*Maximum deceleration*) Let us consider the vertical return of a massive steel sphere with $V_E = 9\,\text{km/s}$ and $\gamma_E = -90°$. With $H_s = 1/\beta = 7{,}200\,\text{m}$ it follows directly from Eq. (5.47) that $a_{max} = 211\,\text{g}$! The velocity at which a_{max} occurs is given by $V' = 0.606V_E$ (Eq. (5.46)) and gives us $V' = 4.46\,\text{km/s}$. These numerical values are independent of the dimensions and the mass of the sphere.

The ballistic parameter of the steel sphere can be calculated according to

$$K = \frac{W}{C_D S} = \frac{\frac{4}{3}\pi R^3 \rho_{steel} g}{\pi R^2} = \frac{4}{3} R \rho_{steel} g \tag{5.54}$$

For a sphere diameter of 30 cm, a specific mass of $\rho_{steel} = 8{,}000\,\text{kg/m}^3$ and $C_D = 1$ (Newtonian hypersonic theory), it follows that $K = 1{,}596\,\text{N/m}^2$. With this

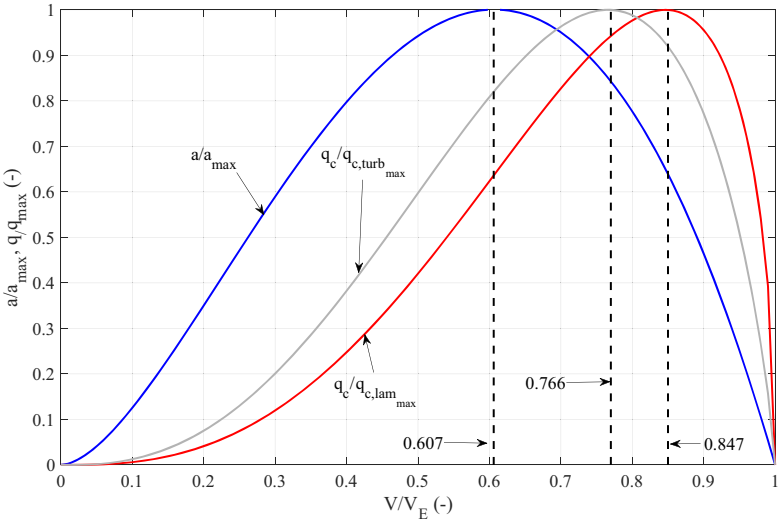


Fig. 5.10 Generalised curves for the ballistic flight

value of K we can calculate with Eq. (5.49) that the maximum deceleration occurs at an altitude of $h' = 28.7\,\text{km}$.

Note that the the obtained values are of course dependent on the selected value for H_s. A careful selection of this parameter for the selected atmosphere model and altitude range under consideration is important, as we will see in the next example. For instance, with $H_s = 6{,}950\,\text{m}$ $a_{max} = 218.5\,\text{g}$ with a corresponding altitude of h' = 27.5 km. ■

Example 5.5 (*Ballistic Apollo entry*) In this example we will compare the results of a numerical simulation of a ballistic entry capsule with the analytical expressions for trajectory and deceleration that we found earlier. To do so, we will use the Apollo capsule as our reference vehicle. We define the initial velocity to be the result of leaving a low Earth-orbit, notably $V_E = 7{,}200\,\text{m/s}$ and $\gamma_E = -10°$. The mass properties and aerodynamic characteristics as given in Appendix B, Sect. B.1, will be used, although in terms of aerodynamics the hypersonic drag coefficient is more or less constant (for $\alpha = 0°$, the required condition for a ballistic flight). The atmosphere is defined according to the United States Standard Atmosphere 1976 (Sect. 2.3.2.2). Finally, the ballistic parameter is $K = 2{,}806\,\text{N/m}^2$, as shown in Example 5.3.

To calculate the analytical altitude-velocity profile, Eq. (5.32), we select a scale height for the exponential atmosphere of $H_s = 7{,}200\,\text{m}$, and a zero density of ρ_0 = 1.225 kg/m^3. This leads to the profile shown in the left plot of Fig. 5.11. The corresponding deceleration profile conform Eq. (5.42) can be found in the right plot of Fig. 5.11. The value of maximum deceleration, Eq. (5.47), is $\frac{a_{max}}{g} = 23.45$, and, with Eq. (5.49), occurs at an altitude of $h' = 37{,}300\,\text{m}$.

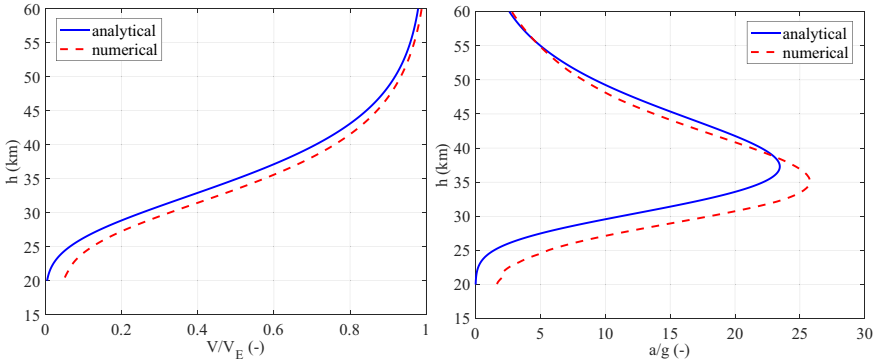


Fig. 5.11 Altitude-velocity profile (left) and acceleration versus altitude (right) for ballistic Apollo entry from LEO, $V_E = 7{,}200\,\text{m/s}$ and $\gamma_E = -10°$ ($H_s = 7{,}200\,\text{m}$)

In Fig. 5.11 also the corresponding curves from the numerical simulation have been plotted. It is clear there is a visible difference in both the trajectory and the deceleration profile, for a given velocity ratio the corresponding altitude is off by about 2 km. The maximum deceleration is $\frac{a_{max}}{g} = 25.77$ (a difference of about 10%), occurring at an altitude $h' = 35{,}152\,\text{m}$, about 5% lower than the corresponding analytical altitude. Given the assumptions that we made these difference may be acceptable, although one option remains to be explored.

The scale height and density at sea level are two parameters that affect the fit of the atmospheric properties. In Example 2.2 we found that $H_s = 7{,}050\,\text{m}$ resulted in a reasonable fit of the US76 atmosphere. However, reducing this value to $H_s = 6{,}950\,\text{m}$ gives the results as shown in Fig. 5.12. It is obvious that the differences have very much reduced, the two h-V profiles are even overlapping for the major part of the flight. Also the difference between the two deceleration profiles has reduced, with a peak value of $\frac{a_{max}}{g} = 24.29$ (the numerical value is only 4% larger) at an altitude of $h' = 35{,}800\,\text{m}$, which is very close to the numerical value. In any mission analysis we are usually only interested in the maximum-deceleration value, so the difference between the two curves is not the most important, since the peak value has a good match.

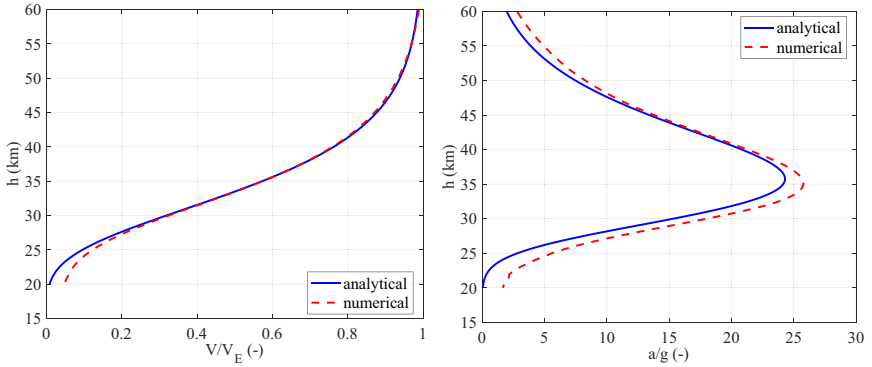


Fig. 5.12 Altitude-velocity profile (left) and acceleration versus altitude (right) for ballistic Apollo entry from LEO, $V_E = 7{,}200\,\text{m/s}$ and $\gamma_E = -10°$ ($H_s = 6{,}950\,\text{m}$)

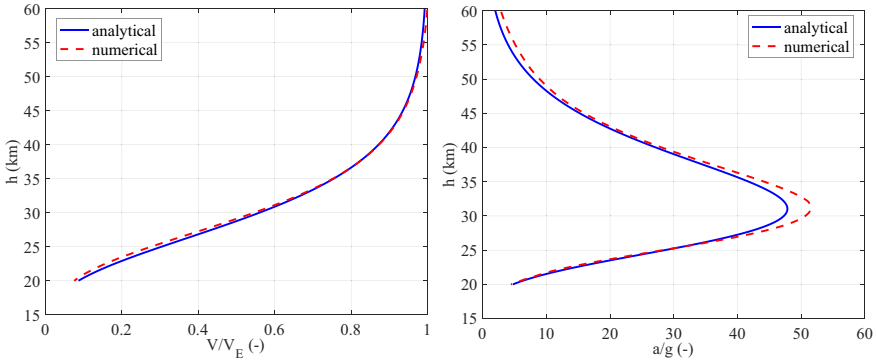


Fig. 5.13 Altitude-velocity profile (left) and acceleration versus altitude (right) for ballistic Apollo entry from LEO, $V_E = 7{,}200\,\text{m/s}$ and $\gamma_E = -20°$ ($H_s = 6{,}950\,\text{m}$)

A final comparison is done for a flight with an initial entry angle of $\gamma_E = -20°$. The corresponding curves can be found in Fig. 5.13 and show an excellent agreement. Inspecting the analytical and numerical maximum deceleration gives $\frac{a_{max}}{g} = 47.85$ and 51.54, respectively, with corresponding altitudes of $h' = 31{,}040\,\text{m}$ and $31{,}178\,\text{m}$. These differences are comparable to those we found for $\gamma_E = -10°$.

In conclusion, the analytical expressions for velocity and deceleration versus altitude give a good approximation of their numerical counterparts, provided that a correct fit of the atmosphere model is used, in terms of H_s and ρ_0. It is advised to do a preliminary analysis of the results before the analytical results are accepted as a valid approximation.

■

5.3.2 Thermal Loads

In Chap. 3, Sect. 3.6, the heat flux due to convection at the nose of the vehicle has been expressed as:

$$q_c = c_1 \frac{1}{R_N^n} \rho^{1-n} V^m \tag{5.55}$$

wherein $n = 0.5$ for a laminar boundary layer and $n = 0.2$ for a turbulent boundary layer, R_N is the nose radius and m is an empirical constant, for Earth entry commonly taken as $m = 3$. This equation is also known as Chapman's Equation.[2]

In the previous section we derived an expression for the atmospheric density ρ,

$$\rho = \frac{2\,K\beta \sin\gamma_E}{g} \ln\frac{V}{V_E} \tag{5.56}$$

which is, for a given K, β and γ_E, proportional to $\left|\ln\frac{V}{V_E}\right|$. Note that due to the negative value of the logarithm, we need to take the absolute value, as in Eq. (5.56) ρ is taken to a power smaller than 1.

If we furthermore assume a fixed value for the nose radius R_N, it then follows that q_c correlates to the ratio of velocities in the following manner:

$$q_c \propto \left|\ln\frac{V}{V_E}\right|^{1-n} \left(\frac{V}{V_E}\right)^3 \tag{5.57}$$

By differentiating this expression with respect to V/V_E, it is found that the heat flux is maximal for:

$$\ln\frac{V''}{V_E} = -\frac{1-n}{3} \Rightarrow \tag{5.58}$$

$$\frac{V''}{V_E} = e^{-\frac{1-n}{3}} \tag{5.59}$$

For $n = 0.5$ and $n = 0.2$ it is found that $\frac{V''}{V_E} = 0.85$ and 0.77, respectively. With Eq. (5.57) it can also be shown that

$$\frac{q_c}{q_{c,max}} = \left(\frac{3e}{1-n}\right)^{1-n} \left|\ln\frac{V}{V_E}\right|^{1-n} \left(\frac{V}{V_E}\right)^3 \tag{5.60}$$

[2] Dean Roden Chapman (8 March 1922–4 October 1995) was a mechanical engineer at NASA, and professor of aeronautics and astronautics and mechanical engineering at Stanford University. Internationally recognised for his pioneering research in fluid dynamics, as well as being the architect for the development of modern computational fluid dynamics, he was also known for his research and controversial theories on the origin of tektites, which are aerodynamically-shaped glassy objects originating from meteorite impacts and found scattered over the Earth's surface.

For both $n = 0.5$ (laminar boundary layer) and $n = 0.2$ (turbulent boundary layer) this relationship has been illustrated in Fig. 5.10.

By substituting Eq. (5.58) into Eq. (5.56) we find for the atmospheric density at which the maximum heat flux is reached:

$$\rho'' = -\frac{2}{3}(1-n)\frac{K\beta}{g}\sin\gamma_E \tag{5.61}$$

If we compare this with the equivalent expression for maximum deceleration, Eq. (5.45), we find for the ratio of densities for maximum heat flux and maximum deceleration:

$$\frac{\rho''}{\rho'} = \frac{2}{3}(1-n) \tag{5.62}$$

Relating the above expression to altitude, we find by substituting the exponential relation between ρ and h:

$$\frac{e^{-h''\beta}}{e^{-h'\beta}} = \frac{2}{3}(1-n) \Rightarrow \tag{5.63}$$

$$h'' - h' = -\frac{\ln\left\{\frac{2}{3}(1-n)\right\}}{\beta} \tag{5.64}$$

For the isothermal atmosphere with $H_s = 1/\beta = 7{,}200\,\text{m}$ and $n = 0.5$ it follows with Eq. (5.64) that the maximum heat flux occurs at an altitude 7,910 m above the altitude for maximum deceleration. For $n = 0.2$ the difference between these two altitudes is 4,526 m.

Note: the isolated expression for h'' easily follows from Eq. 5.62 by substituting the exponential density:

$$\rho_0 e^{-h''\beta} = -\frac{2}{3}(1-n)\frac{K\beta}{g}\sin\gamma_E \tag{5.65}$$

$$e^{-h''\beta} = -\frac{2}{3}(1-n)\frac{K\beta}{g\rho_0}\sin\gamma_E \tag{5.66}$$

$$-h'' = \frac{1}{\beta}\ln\left[-\frac{2}{3}(1-n)\frac{K\beta}{g\rho_0}\sin\gamma_E\right] \tag{5.67}$$

or

$$h'' = -H_s \ln\left[-\frac{2}{3}(1-n)\frac{K\beta}{g\rho_0}\sin\gamma_E\right] \tag{5.68}$$

The above discussion has been focussed on the stagnation point of the vehicle. If we want to do a similar exercise for an arbitrary point on the wall of the vehicle, our

starting point is (again, see Sect. 3.6 for the theoretical background):

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{C_F}{4}\rho V^3 S_W \tag{5.69}$$

with S_W being the wetted surface area and C_F a friction coefficient averaged over the surface.

Following the same approach as with q_c, by substituting Eq. (5.56) it is easily shown that for given K and γ_E and realising that $\frac{C_F}{4}$ and S_W are constant

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \propto \ln \frac{V}{V_E}\left(\frac{V}{V_E}\right)^3 \tag{5.70}$$

To find the velocity ratio for which $\frac{\mathrm{d}Q}{\mathrm{d}t}$ is maximum we need to solve

$$\frac{\mathrm{d}}{\mathrm{d}(V/V_E)}\left(\frac{\mathrm{d}Q}{\mathrm{d}t}\right) = 0 \tag{5.71}$$

which has as a solution

$$\ln \frac{V'''}{V_E} = -\frac{1}{3} \tag{5.72}$$

or, alternatively

$$\frac{V'''}{V_E} = e^{-\frac{1}{3}} = 0.72 \tag{5.73}$$

Substituting this result in Eq. (5.56) yields

$$\rho''' = -\frac{2}{3}\frac{K\beta}{g}\sin\gamma_E \tag{5.74}$$

As before, since we already have the density for which the maximum deceleration occurs, Eq. (5.48), i.e.,

$$\rho' = -\frac{K\beta\sin\gamma_E}{g} \tag{5.75}$$

we find for the ratio of densities for maximum heat flux and maximum deceleration:

$$\frac{\rho'''}{\rho'} = \frac{2}{3} \tag{5.76}$$

With $\rho = \rho_0 e^{-\beta h}$, the accompanying altitude difference is

Table 5.1 Characteristic points for the ballistic re-entry trajectory ($H_s = 7{,}200\,\text{m}$)

		$\frac{V}{V_E}$	h (m)
Max. heat flux in stagnation point q_c	Laminar	0.85	$h' + 7{,}910$
	Turbulent	0.77	$h' + 4{,}526$
Max. heatflow $\frac{\mathrm{d}Q}{\mathrm{d}t}$		0.72	$h' + 2{,}919$
Max. deceleration $\bar{a}_{max}$		0.606	h'

$$h''' - h' = -\frac{\ln \frac{2}{3}}{\beta} \tag{5.77}$$

which means that the altitude for maximum heat flux at an arbitrary location lies 2,919 m above that for maximum deceleration. In Table 5.1 one can find a summary of the special points for the ballistic re-entry trajectory (isothermal atmosphere).

Example 5.6 (*Apollo maximum heat flux*) To continue Example 5.4, we will have a look at the stagnation heat flux for a ballistic entry of the Apollo capsule. In our numerical model we use the same Chapman Equation as given by Eq. (5.55), although given by a different formulation:

$$q_c = c^* \frac{1}{R_N^n} \left(\frac{\rho}{\rho_0}\right)^{1-n} \left(\frac{V}{V_c}\right)^m \tag{5.78}$$

with $c^* = 1.1097 \cdot 10^8 \sqrt{\text{m}}$, $\rho_0 = 1.225\,\text{kg/m}^3$ and $V_c = 7{,}905.4\,\text{m/s}$. For Apollo, the nose radius in the stagnation point is $R_N = 4.7\,\text{m}$. Finally, as mentioned earlier, commonly used values for n and m are $n = 0.5$ and $m = 3$.

For the analytical model, we need to compute c_1 from c^*, ρ_0 and V_c, i.e.,

$$c_1 = c^* \frac{1}{\sqrt{\rho_0}} \frac{1}{V_c^3}$$

yielding $c_1 = 2.0294 \cdot 10^{-4}\ \frac{\text{s}^2}{\sqrt{\text{Nm}}}$.

In the simulation $H_s = 6{,}950\,\text{m}$ is used again, as this value gave the best approximation of the altitude-velocity profile, see Figs. 5.12 and 5.13. In line with Fig. 5.10, the ratio of the convective heat flux and its maximum value is shown in Fig. 5.14. The numerical and analytical show an excellent match, which was to be expected, since the stagnation heat flux is a parameter directly derived from h and V. These curves have been obtained for $\gamma_E = -10°$, but the equivalent curves for $\gamma_E = -20°$ show a similar match.

With Eq. (5.64) the altitude difference between the peak heat flux and peak deceleration is $-\frac{\ln \frac{2}{3}(1-n)}{\beta} = 7{,}635\,\text{m}$ ($H_s = 1/\beta = 6{,}950\,\text{m}$ and $n = 0.5$). The correspond-

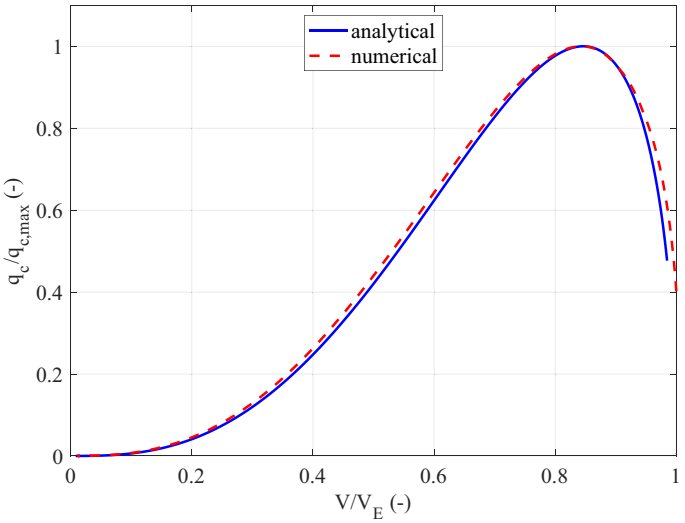


Fig. 5.14 Normalised convective heat-flux profile for ballistic Apollo flight (V_E = 7,200 m/s and γ_E = -10°); exponential atmosphere with H_s = 6,950 m

ing numerical result is, with $h'_{num} = 31{,}178$ m and $h''_{num} = 38{,}667$ m, $\Delta h = 7{,}489$ m. For $\gamma_E = -10°$ we find $\Delta h = 8{,}329$ m ($h'_{num} = 35{,}152$ m and $h''_{num} = 43{,}481$ m). The larger difference for $\gamma_E = -10°$ stems from the fact that the analytical h' is 648 m higher than the numerical one, whereas for $\gamma_E = -20°$ the analytical h' is 138 m lower. Since the analytical h'' shifts with the same amount (the difference is given with respect to h'), the error in Δh is larger for $\gamma_E = -10°$ and matches the obtained value. All things considered, the match between analytical and numerical is good.

The last parameter to consider is the actual value of $q_{c,max}$. Also in this case we expect a close match. Starting with Eq. (5.52) we need to substitute the values for V'' and ρ'' at which the heat flux peaks. Equations (5.59) and (5.61) give us

$$\frac{V''}{V_E} = e^{-\frac{0.5}{3}} \Rightarrow (V'')^3 = \frac{V_E^3}{\sqrt{e}} \tag{5.79}$$

and

$$\rho'' = -\frac{1}{3}\frac{K\beta}{g}\sin\gamma_E \tag{5.80}$$

so

$$q_{c,max} = c_1 V_E^3 \sqrt{-\frac{1}{3e}\frac{1}{R_N}\frac{K\beta}{g}\sin\gamma_E} \tag{5.81}$$

Substituting the corresponding parameters we find for $\gamma_E = -10°$ an analytical value of $q_{c,max} = 1034.2$ kW/m^2 (numerical value: $q_{c,max} = 1031.7$ kW/m^2), and

for $\gamma_E = -20°$ $q_{c,max} = 1451.5\ \text{kW/m}^2$ (numerical value: $q_{c,max} = 1480.8\ \text{kW/m}^2$). In both cases, an excellent agreement. ■

The last topic to be discussed in this section is the heat supplied to the vehicle during the whole descent. This quantity, expressed by Q, follows from integrating the heat flux at an arbitrary location of the vehicle:

$$Q = \int_0^t \frac{\mathrm{d}Q}{\mathrm{d}t}\mathrm{d}t = \int_{V_E}^{V_F} \frac{\mathrm{d}Q}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}V}\mathrm{d}V \tag{5.82}$$

with V_F being the final velocity. Substituting the expressions for $\frac{\mathrm{d}Q}{\mathrm{d}t}$ according to Eq. (5.69) and $\frac{\mathrm{d}V}{\mathrm{d}t}$ according to Eq. (5.39) gives

$$Q = \int_0^t \frac{\mathrm{d}Q}{\mathrm{d}t}\mathrm{d}t = \int_{V_E}^{V_F} \frac{C_F}{2} S_W \frac{K}{g} V \mathrm{d}V \tag{5.83}$$

which gives, after integration and substituting $K = W/C_D S$:

$$Q = \frac{1}{4}\frac{C_F S_W}{C_D S} m V_E^2 \left\{1 - \left(\frac{V_F}{V_E}\right)^2\right\} \tag{5.84}$$

In this expression $\mathcal{E}_k = \frac{1}{2} m V_E^2$ is the kinetic energy of the vehicle when entering the atmosphere, so that:

$$\frac{Q}{\mathcal{E}_k} = \frac{1}{2}\frac{C_F S_W}{C_D S}\left\{1 - \left(\frac{V_F}{V_E}\right)^2\right\} \tag{5.85}$$

For the velocity ratio V_F/V_E it holds true according to Eq. (5.32):

$$\frac{V_F}{V_E} = \exp\left(\frac{1}{2}\frac{g\rho_F}{K\beta \sin\gamma_E}\right) \tag{5.86}$$

so

$$\frac{Q}{\mathcal{E}_k} = \frac{1}{2}\frac{C_F S_W}{C_D S}\left\{1 - \exp\left(\frac{g\rho_F}{K\beta \sin\gamma_E}\right)\right\} \tag{5.87}$$

The following conclusions can now be drawn. For values of $K < 20{,}000\ \text{N/m}^2$ it holds: $V_F/V_E << 1$ (Fig. 5.7). In that case it follows:

$$\frac{Q}{\mathcal{E}_k} = \frac{1}{2}\frac{C_F S_W}{C_D S} \tag{5.88}$$

The heat supply Q to the body thus becomes less, when the total drag surface area $C_D S$ becomes larger in relation to the frictional drag surface area, $C_F S_W$. This occurs when the shockwave drag of the vehicle is large, or in other words, when the body is supplied with a blunt forebody. This is in agreement with the condition for a relatively low K, provided that the vehicle's weight is not too large. A blunt nose (large radius R_N) also limits the heat transfer in the stagnation point. A vehicle with a value for $K > 20{,}000\,\mathrm{N/m^2}$ in a vertical descent, will loose little velocity (Fig. 5.7). In that case it follows with $1 - e^{-x} \approx x$ and $\rho_F = \rho_0$:

$$\frac{Q}{\mathcal{E}_k} = -\frac{1}{2}\frac{C_F S_W}{W}\frac{g\rho_0}{\beta \sin\gamma_E} \tag{5.89}$$

Now, for given entry conditions, Q becomes as low as possible when the frictional resistance surface area is minimal, which is achieved for a vehicle with a slim shape. This also implies a relatively high value for K (i.e., a large weight and due to the slim shape, also a small drag area).

5.3.3 Summary

In the previous sections several aspects of the ballistic flight have been discussed. We will summarise the major findings here.

- During a major part of the flight, the ballistic vehicle will fly in a straight line as $\gamma = \gamma_E \approx$ constant.
- The altitude-velocity relation is given by

$$\frac{V}{V_E} = \exp\left(\frac{1}{2}\frac{g\rho_0 e^{-\beta h}}{K\beta \sin\gamma_E}\right) \tag{5.90}$$

- The maximum deceleration during a ballistic flight is given by

$$\bar{a}_{max} = -\left(\frac{\mathrm{d}V}{\mathrm{d}t}\right)_{max} = -\frac{\beta \sin\gamma_E}{2e}V_E^2 \tag{5.91}$$

  It is clear that a_{max} is independent of the ballistic parameter (so both weight and drag properties), but only changes with the initial conditions, V_E and γ_E.
- The altitude at which the maximum deceleration occurs is found to be

$$h' = \frac{1}{\beta}\ln\left(-\frac{\rho_0 g}{K\beta \sin\gamma_E}\right) \tag{5.92}$$

  which *does* show a dependency on the ballistic parameter. The initial velocity, V_E, has no impact on the altitude for a_{max}.
- The maximum convective heat flux in the stagnation point is given by

$$q_{c,max} = c_1 V_E^3 \sqrt{-\frac{1}{3e}\frac{1}{R_N}\frac{K\beta}{g}\sin\gamma_E} \tag{5.93}$$

which is proportional to V_E, γ_E and K.

- The altitude at which the maximum convective heat flux in the stagnation point occurs, in relation to the one of maximum deceleration, is:

$$h'' - h' = -\frac{\ln\frac{2}{3}(1-n)}{\beta} \tag{5.94}$$

This indicates that the altitude for $q_{c,max}$ is larger than the altitude for a_{max}. The altitude difference for a laminar boundary layer is larger than the one for a turbulent boundary layer.

- The altitude difference between the one at which the maximum convective heat flux in an arbitrary point on the vehicle occurs, and the one for maximum deceleration, is:

$$h''' - h' = -\frac{\ln\frac{2}{3}}{\beta} \tag{5.95}$$

Comparing numerical values shows that this altitude is in between the one for maximum flux in the stagnation point and the altitude for maximum deceleration.

5.3.4 Case Study: Huygens Entry

Huygens was an atmospheric entry probe carried to Saturn's moon Titan as part of the Cassini-Huygens mission. Cassini-Huygens is one of the four flagship-class missions to the outer planets, the other planetary flagships being Galileo, Voyager, and Viking. The Saturn Orbiter Titan Probe, as it is commonly called, was supplied by ESA and named for the Dutch 17th century astronomer Christiaan Huygens. The combined Cassini-Huygens spacecraft was launched from Cape Canaveral, Florida, on October 15, 1997, with a Titan IV-B/Centaur. Its purpose: to explore the Saturnian system in all its aspects, i.e., the planet itself, its rings and its multitude of moons.

After an interplanetary travel of more than seven years, Huygens separated from the Cassini orbiter on December 25, 2004, and landed on Titan on January 14, 2005, near the Xanadu region after a parachuted descent of close to 1.5 h. It was the first landing ever accomplished in the outer solar system and although it touched down on land, the possibility that it would touch down in an ocean (of liquid methane) was also taken into account in its design. Originally designed for a surface mission of only three minutes—the actual mission was in fact the descent under the parachute,

where many measurements of the atmosphere were taken—the probe continued to send data for about 90 min after reaching the surface.

Titan is the largest moon of Saturn, with an equatorial radius of $R_e = 2{,}575$ km (compare: the Earth's Moon has a radius of $R_e = 1{,}738$ km). The gravitational acceleration at the surface is $g_0 = 1.354\,\text{m/s}^2$, which is some 16% lower than the Earth's Moon, due to its lower-density composition. Titan has a very dense atmosphere with a surface value of $\rho_0 = 5.3\,\text{kg/m}^3$, over four times larger than the corresponding Earth value.

Conform Example 5.3, the Huygens probe entry configuration, with a diameter of $d = 2.7$ m (reference area of $S_{ref} = 5.7\,\text{m}^2$), a mass of $m = 318$ kg and a hypersonic drag coefficient of $C_D \approx 1.54$ ($M > 4$), has a ballistic parameter of $K = 49\,\text{N/m}^2$. More details on mass properties and aerodynamic characteristics can be found in Appendix B, Sect. B.3.

Huygens performed a ballistic entry from 1,200 km altitude, and in this case study we will look at the nominal entry mission down to the conditions for heat-shield jettisoning and parachute deployment, which takes place at an altitude of $h \approx 180$ km. We will look at the maximum occurring *g-load* and see how a change in entry conditions (V_E and γ_E) will affect the peak value. Also, by applying the simplified Chapman model for calculating the convective heat flux, we can analyze at which altitude the maximum flux will occur and compare this with the altitude for maximum deceleration. For the current study, we will use

$$q_{c,nose} = c_1 \rho^n V^m \tag{5.96}$$

with $c_1 = 1.76{\cdot}10^{-6}$, $n = 0.48$ (laminar flow) and $m = 3.54$, providing a heat flux in W/m^2. These numerical values have been obtained from curve fitting CFD results (Wright et al. 2006), and correspond with a peak design heat flux of $q_{c,max} = 450$ kW/m^2 for $V_E = 6.19$ km/s and $\gamma_E = -64°$.

In this case study we will first have a look at the nominal mission, notably the altitude-velocity profile and the corresponding deceleration and heat flux. Subsequently, we will have a look how the initial conditions, V_E and γ_E, as well as the mass (i.e., ballistic parameter) influence the thermo-mechanical loads. Finally, we will use the analytical expressions derived earlier to see how well they can predict the numerical results. It is noted that in Chap. 9, Sect. 9.8, the second part of this case study can be found, focusing on the parachute-descent phase.

5.3.4.1 Numerical Approach

The nominal mission of Huygens is characterised by an initial velocity and flight-path angle of $V_E = 6{,}190$ m/s and $\gamma_E = -64°$, respectively, defined at an altitude of $h = 1{,}270$ km. The entry latitude is $\delta = 45°$ (north) with a heading due east ($\chi = 90°$). The entry of Huygens is ballistic ($\alpha = 0°$). Titan is considered to be a sphere, with a central gravity field and a nominal atmosphere modelled according to Yelle and the so-called Flasar wind model—see for more detail on the atmosphere Sect. 2.3.4.

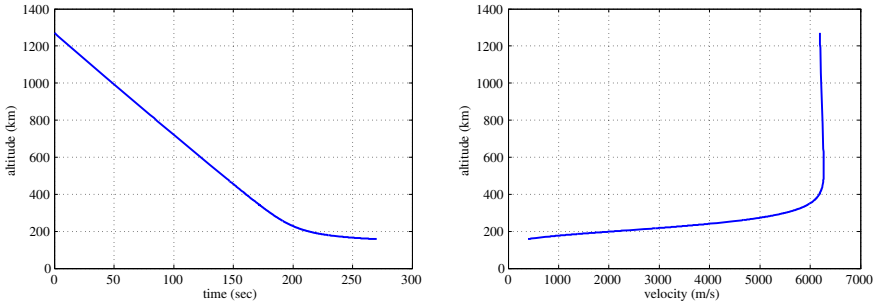


Fig. 5.15 Altitude versus time (left) and altitude-velocity profile (right) for the nominal Huygens mission ($V_E = 6{,}190$ m/s and $\gamma_E = -64°$)

The simulation will run until a Mach number of $M = 1.5$ is reached, which is an approximate condition for parachute deployment, just before the probe becomes statically unstable.

Figure 5.15, left, shows us the height as a function of flight time. We see that at lower altitudes the aerodynamic drag is forcing the probe to follow a less steep path. This is called rebound or skip, although this does not necessarily mean that the probe is leaving the atmosphere again. Due to this rebound, the time to parachute deployment is somewhat larger than for steeper entries (i.e., more negative γ). We will discuss this in more detail, when we are doing the sensitivity analysis and study the effect of different entry conditions. Looking at the right plot of Fig. 5.15, we see that in the first phase of re-entry the relative velocity is increasing. Huygens is accelerated down to an altitude of about 500 km, because gravitational forces are dominating the aerodynamic forces at high altitudes (due to the low atmospheric density). From $h = 400$ km, Huygens' speed is steadily decreasing. It is clear that the main deceleration takes place between $h = 350$ km and $h = 200$ km.

More or less the same pattern can be observed in Fig. 5.16, left, where h versus the *g-load* is given. The *g-load* is the deceleration (or acceleration) of the probe with respect to a reference value of the gravitational acceleration (the surface value for Titan is $g_0 = 1.354\,\text{m/s}^2$). For altitudes above $h = 500$ km the *g-load* would have been negative due to the acceleration. However, the acceleration due to gravity is not included in the definition of the *g-load*, so even in this case the *g-load* is (marginally) positive. The maximum *g-load* (or maximum deceleration) is about 93.7 ($126.9\,\text{m/s}^2$, normalised with Earth's $g_0 = 9.81\,\text{m/s}^2$ this would mean a structural load of 12.94 g). The altitude for maximum *g-load*, $h' = 233.8$ km, also corresponds with the altitude for maximum drag (the retarding force).

In the right plot of Fig. 5.16 the convective heat flux in the stagnation point is shown. The maximum value indeed corresponds with the design value of $450\,\text{kW/m}^2$; the actual value is $q_c = 467\,\text{kW/m}^2$, due to the small error made with the curve fit of the CFD data. The altitude for maximum flux is $h'' = 297.4$ km. We find that indeed the peak of maximum heat flux occurs well before the one for maximum deceleration. What is also noteworthy is that the peak of deceleration is somewhat sharper than the

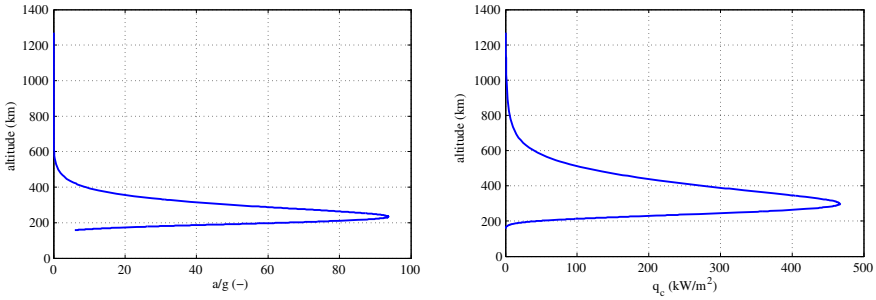


Fig. 5.16 *g-load* (left), normalised with $g_0 = 1.354$ m/s^2, and stagnation-point heat flux (right) as a function of altitude for the nominal Huygens mission ($V_E = 6{,}190$ m/s and $\gamma_E - 64°$)

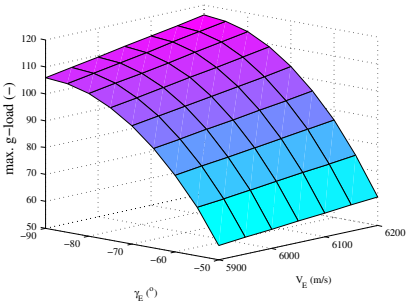
one for heat flux. This is a characteristic for ballistic flight, where indeed the main velocity reduction takes place in just a small altitude range.

Next we will study the effect of different entry conditions on maximum occurring *g-load* and maximum heat flux, and the corresponding altitudes.[3] A total of 63 simulations were executed with varying values of entry velocity V_E and entry path angle γ_E V_E varied from 5,900 m/s to 6,200 m/s with steps of 50 m/s. The range for γ_E was $[-50°, -90°]$, with steps of $-5°$.

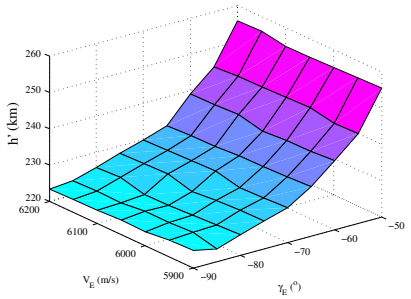
In Fig. 5.17a the variation in maximum *g-load* as a function of V_E and γ_E is shown. At first sight, γ_E seems to have a greater effect than V_E. However, we must not forget that for γ_E the range is $-30°$, which means a variation of about 33% (for a maximum value of $\gamma_E = -90°$). For V_E, on the other hand, the variation is 300 m/s which corresponds with a 5% variation. For this reason, one *may* expect a larger visible influence due to γ_E. The analytical expression for the maximum deceleration, $\bar{a}_{max} = -\frac{\beta \sin\gamma_E}{2e} V_E^2$, indicates that it varies as a sine function with γ_E, which, in fact, is obvious from the plot. It should also vary proportionally with V_E^2, but due to the small variation of V_E this is not really clear. However, the bottom line is that both an increasing V_E and an increasing γ_E will increase the maximum *g-load*.

In the adjoining Fig. 5.17b the altitude at which the maximum *g-load* occurs, h', has been plotted. It shows that the variation is slightly more irregular, i.e., non-linear, than the actual *g-load*. The explanation may be found in the fact that h' is more dependent on the atmospheric properties than $\bar{a}_{max}$, i.e., $h' = -\frac{1}{\beta} \ln \frac{\rho_0 g}{K \beta \sin\gamma_E}$ with a dependency on both fitting parameters β and ρ_0. As we already saw in Example 2.4—and will also see later in this case study—the Titan atmosphere is not easily fit by an

[3] More details on the influence on, for instance, the flight time to and altitude of parachute deployment, as well as the sensitivity to atmospheric conditions and aerodynamic conditions can be found in Mooij (1991, 1992).

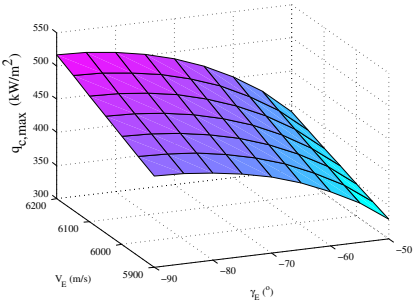


(a) Normalised maximum *g-load*, $g_0 = 1.354\ \mathrm{m/s^2}$

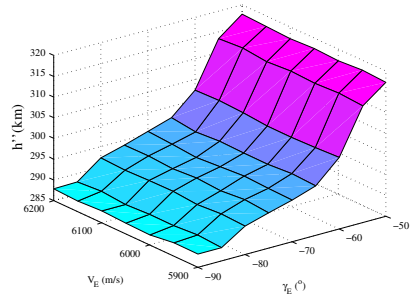


(b) Altitude for maximum *g-load*

Fig. 5.17 Maximum mechanical load as a function of initial conditions V_E and γ_E (vehicle mass = 318 kg)



(a) Maximum q_c



(b) Altitude for maximum q_c

Fig. 5.18 Maximum thermal load as a function of initial conditions V_E and γ_E (m = 318 kg)

exponential model. The atmosphere is actually more non-linear, which may show as a non-linear effect on h'. However, the sine-variation with γ_E and the independence on V_E is still visible.

The dependency on V_E and γ_E for the maximum stagnation heat flux is plotted in Fig. 5.18a with the corresponding altitude, h'', shown in Fig. 5.18b. As shown in Example 5.6, Eq. (5.81), $q_{c,max}$ is proportional to V_E^3 and $\sin\gamma_E$. Again, due to the small variation in V_E only a linear increase is seen although the absolute effect is significant ($\Delta q_c = 60\ \mathrm{kW/m^2}$ for $\gamma_E = -50°$, a variation of 13% of the design value of 450 kW/m^2). The variation over the full range of γ_E is $\Delta q_c = 121\ \mathrm{kW/m^2}$ ($V_E = 6{,}100$ m/s), almost 27%. A similar non-linear effect as for h' is observed in h'', which seems logical as h'' shares the same dependency on atmospheric density. Also in this case, only a variation with $\sin\gamma_E$ shows, as h'' is not dependent on V_E.

So far, we only looked at a dependency of the thermo-mechanical loads on the initial conditions V_E and γ_E. But what about the ballistic parameter? We have seen in Sects. 5.3.1 and 5.3.2 that $\bar{a}_{max}$ is independent of K, but $q_{c,max}$ is proportional to

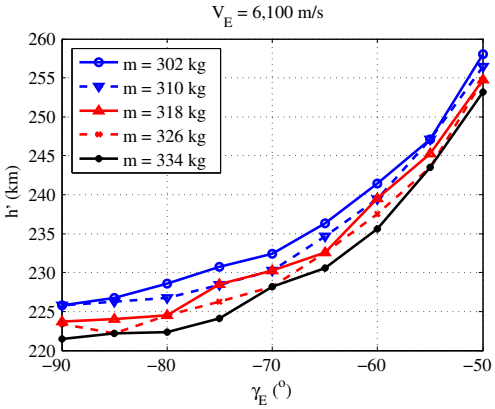
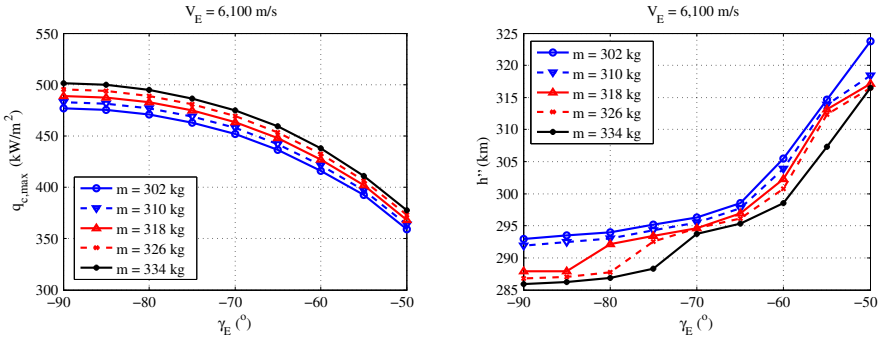


Fig. 5.19 Altitude for maximum *g-load* as a function of m and γ_E ($V_E = 6{,}100$ m/s)



(a) Maximum q_c (b) Altitude for maximum q_c

Fig. 5.20 Maximum thermal load as a function of m and γ_E ($V_E = 6{,}100$ m/s)

K, whereas both h' and h'' are inversely proportional. To briefly study the effect of a variation in K we will vary the vehicle mass from $m = 302$–334 kg, with steps of 8 kg, which changes K by $\pm 5\%$.

A first observation is that indeed the maximum *g-load* does not change visibly, only a marginal change due to some of the non-linearities. In Fig. 5.19 the corresponding changes in h' are shown. For visualisation purposes we have selected the curves for $V_E = 6{,}100$ m/s to show the dependency on m and γ_E. In Fig. 5.19 the curves for h' are shown—note that the curve for $m = 318$ kg is the cross-sectional curve for $V_E = 6{,}100$ m/s in Fig. 5.17b. Despite the effect of atmospheric non-linearities it is clear that with increasing mass (so increasing K) h' decreases. For each of the curves the inverse proportionality on $\sin\gamma_E$ is clear: a steeper entry (γ_E more negative) results in a lower h'. The same variation, i.e., proportional with K and inversely proportional with $\sin\gamma_E$, is shown in Fig. 5.20b for h''. In the adjoining Fig. 5.20a,

the curves for $q_{c,max}$ have been plotted. With increasing mass (thus K) the maximum heat flux increases.

So, in conclusion, the thermo-mechanical loads and corresponding altitudes behave the way we have analytically derived in the previous sections. How well the actual numerical values compare, we will see in the remaining part of this case study.

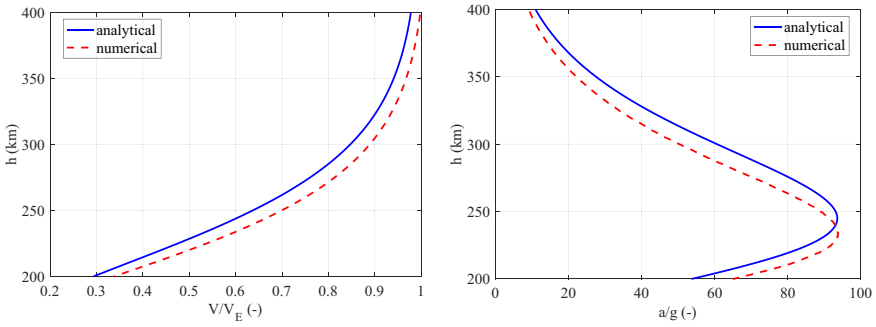
5.3.4.2 Analytical Approach

We saw in Example 5.5 the quality of the analytical solution is very much dependent on the goodness-of-fit of the exponential atmosphere with the actual model of the atmosphere used for the numerical simulation. As has been mentioned earlier in this case study, in Example 2.4 an exponential atmosphere for the Yelle atmosphere model has been derived. Due to the strong non-linearity the exponential atmosphere we already had to split the altitude range into two parts, i.e., from $h = 0$ to 140 km and from $h = 140$ to 1000 km. For the altitude range under study, we can thus suffice to use

$$\rho = \rho_0 e^{-\frac{h-140}{H_s}} \tag{5.97}$$

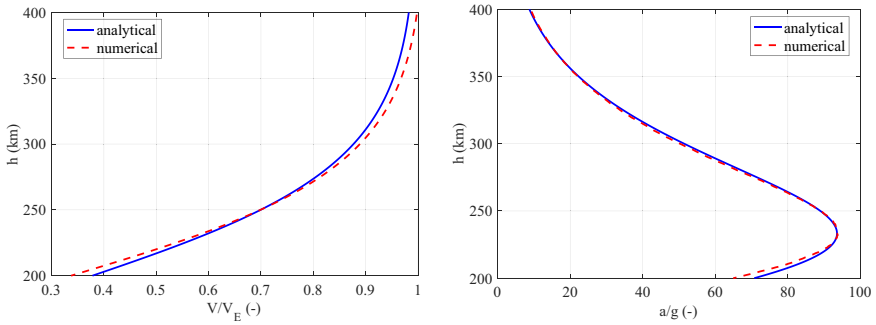
with $H_s = 50$ km and $\rho_0 = 0.0053$ kg/m^3 (for $h > 140$ km). However, we can expect some differences with the numerical values, because this particular curve fit may not be accurate enough.

The approach that we follow is that for the nominal mission ($V_E = 6{,}190$ m/s, $\gamma_E = -64°$ and $m = 318$ kg or $K = 49$ N/m^2) we will look at the profiles for altitude-velocity and thermo-mechanical loads, including the maximum values and corresponding altitudes. The corresponding numerical values for the nominal mission are $\bar{a}_{max} = 93.7$, $h' = 233.8$ km, $q_c = 467$ kW/m^2 and $h'' = 297.4$ km.



(a) h-V diagram (b) a/g versus h

Fig. 5.21 Comparison between numerical and analytical flight mechanics ($H_s = 50$ km and $\rho_0 = 0.0053$ kg/m^3)



(a) h-V diagram (b) a/g versus h

Fig. 5.22 Comparison between numerical and analytical flight mechanics (H_s = 50 km and ρ_0 = 0.0042 kg/m^3)

In Fig. 5.21a the h-V diagram is shown for the relevant altitude range, where the maximum thermal and mechanical loads occur. It is clear that there is a significant shift between the analytical and numerical curves, i.e., for any given altitude the difference in velocity is around 0.06 V_E. A similar shift is observed in the deceleration (Fig. 5.21b). This shift has resulted in a much larger analytical altitude for maximum deceleration (h' = 244.9 km versus h'_{num} = 233.8 km), but the peak value is almost identical ($\bar{a}_{max}$ = 93.6 versus $\bar{a}_{max,num}$ = 93.7). That means that the scale height has the proper value, as a change in H_s will lead to both a shift in the curves and a change in maximum value. A change in ρ_0, though, will give us the results of Fig. 5.22a, b. These figures have been obtained with ρ_0 = 0.0042 kg/m^3 and the difference between the analytical and numerical solution has reduced significantly. The deceleration profile almost gives a perfect match, with h' = 233.3 km versus h'_{num} = 233.8 km, and, as said, a peak value that is unchanged. However, from the h-V profile it is clear that the exponential atmosphere is not really a good fit of the atmosphere, because a perfect match cannot be obtained.

Before we come to discuss the convective heat flux in the stagnation point, it should be mentioned that Eqs. (5.59) and (5.64) cannot be used. These equations were derived under the assumption that the velocity exponent is m = 3. However, in the case for the Titan entry the heat-flux model given by Eq. (5.96) should be evaluated for n = 0.48 and m = 3.54. So, to generalise, the velocity ratio at which the maximum heat flux occurs changes into

$$\frac{V''}{V_E} = e^{-\frac{1-n}{m}} \tag{5.98}$$

with a corresponding altitude difference with respect to the peak deceleration of

$$h'' - h' = -\frac{\ln \frac{2}{m}(1-n)}{\beta} \tag{5.99}$$

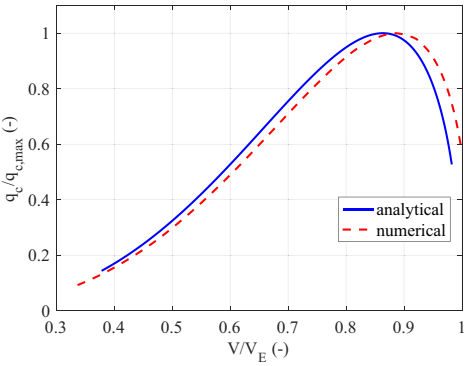


Fig. 5.23 Normalised convective heat-flux profile for Huygens flight (V_E = 6,190 m/s and $\gamma_E = -64°$); exponential atmosphere with H_s = 50 km and ρ_0 = 0.0043 kg/m^3

In Fig. 5.23 the curves for the normalised convective heat-flux profile are shown. The difference between the two curves is not so large, which means that the analytical expression can be used as an accurate approximation. Equation (5.99) gives an analytical altitude difference of $\Delta h = 61.25$ km, or, in other words, $h'' = 294.5$ km, at a corresponding velocity ratio of $\frac{V}{V_E} = 0.8634$. The numerical values are $h''_{num} = 297.4$ km and $\frac{V}{V_{E,num}} = 0.8825$. The actual value of $q_c = 455$ kW/m^2 compares well with $q_{c,num} = 467$ kW/m^2. So, for each of these parameters the analytical values are close to the numerical ones.

In summary, after some tweaking of the applied exponential atmosphere the analytical expressions of this section can be used to predict the deceleration and heat flux for Huygens' ballistic entry in the atmosphere of Titan. However, the exponential atmosphere is not the most accurate approximation of the Yelle atmosphere model due to some non-linearities that cannot be properly fitted. For different mission specifications it may very well be possible that first a new fit has to be established. Of course, this is not really a problem, but should not be overlooked.

5.3.5 *Case Study: Impact of Atmosphere-Model Choice on Mission Design*

AURORA is the European Space Agency programme that is preparing for the next step in exploration of the solar system (Cawthorne et al. 2004; ESA 2001). Currently in its preparatory phase, its goal is to define and implement the strategy for the human and robotic exploration of those solar system planets that may have traces of life. The objective: to map the way for a manned mission to Mars. Apart from the many technological capabilities required to achieve such an ambitious programme, mission design and analysis by means of trajectory simulation is an essential part of each of the AURORA missions. Detailed knowledge about Mars in all its aspects, e.g., gravitational field, topography and atmospheric properties, is mandatory for accurate

results. State-of-the-art atmospheric databases, such as NASA's Mars Global Reference Atmosphere Model (GRAM)—amongst others used to support the American Mars Pathfinder (Braun et al. 1995; Spencer and Braun 1996) and Mars Exploration Rover projects (Desai and Knocke 2004)—and ESA's Mars Climate Database (EMCD), e.g., used within the Beagle-2 project, are available to the mission designer.

With the release of Version 4.1 of the EMCD, an activity was initiated to review this database with respect to its ease of implementation in trajectory-analysis codes, its use and functionality, and the impact on mission design and analysis. The goal of the current case study is simply to compare the results obtained with EMCD 4.1, the previous major release EMCD 3.1 and Mars GRAM 2001.[4] Or, to formulate it as a research question:

How does the choice of atmosphere model influence the mission performance of a Mars lander?

It is noted that this study has also been published by Mooij et al. (2006). Even though the databases are slightly outdated, this case study serves as a good example to show the effect of proper atmosphere-model selection. For more information on the actual implementation of the atmospheric databases, the reader is referred to Chap. 2, Sect. 2.3.3.

5.3.5.1 Reference Vehicle and Mission

The selected vehicle is a Beagle-2 like entry vehicle (Fig. 5.24), with aerodynamics as given by Fig. 5.25. Note that only a Mach dependency is given, which is, for the ballistic entry studied here, also sufficient. However, when the influence of wind is considered an angle-of-attack dependency should in principle be used, due to the possible large variations in angle of attack. The forebody of the original Beagle-2 aero shell is a 60° half-angle cone with a maximum vehicle diameter of 0.9 m. It has a spherically blunted nose of radius 0.417 m and a shoulder radius of 0.029 m. The current vehicle has been scaled such that the diameter of the aeroshell is 4.4 m. The (fictitious) mission time line is given by:

1. Entry point at T_0
2. Mach detection at 1.8, mortar firing and drogue-chute deployment at T_1
3. Back-cover separation with drogue-chute release, and main-chute deployment: $T_2 = T_1 + 10\,\mathrm{s}$
4. Front shield separation: $T_3 = T_2 + 10\,\mathrm{s}$
5. Touch-down.

The corresponding entry conditions are listed in Table 5.2. Three target landing sites—the same as for which the atmospheric conditions were evaluated in

[4] Note that we restrict to the use of Mars GRAM 2001, despite newer versions, because of license availability. Of course, for a proper comparison also the state-of-the-art Mars GRAM Version 2010 should be used.

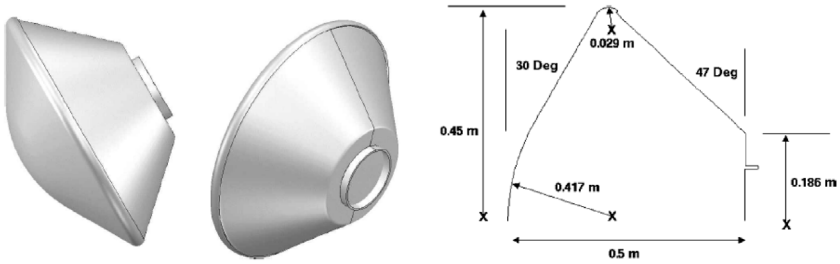
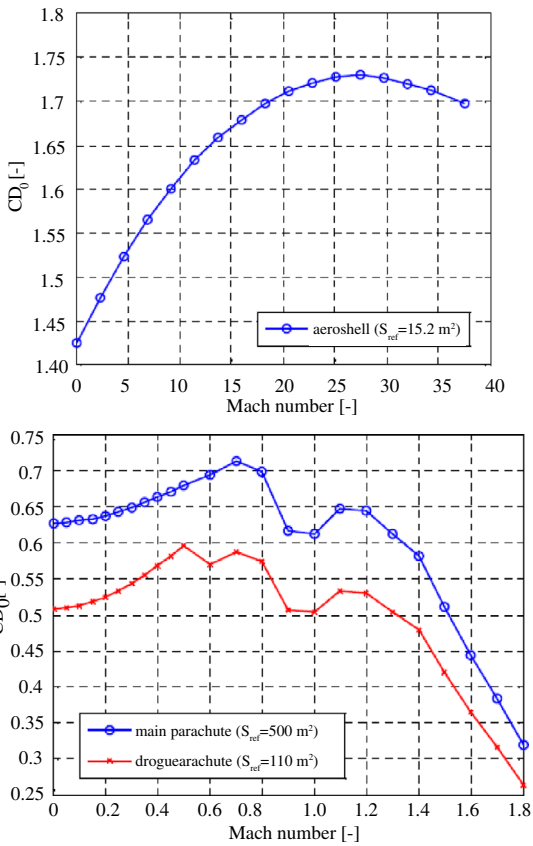


Fig. 5.24 Baseline vehicle (Liever et al. 2003)

Fig. 5.25 Aerodynamic characteristics of baseline vehicle and parachute system



Sect. 2.3.3—are considered, see Fig. 5.26. For these sites, the nominal entry conditions have been computed by backward integration. These landing sites were selected to address several regional aspects of the Mars climate database. The mass breakdown of the vehicle is the following. The entry mass is m = 1,900 kg, with a mass

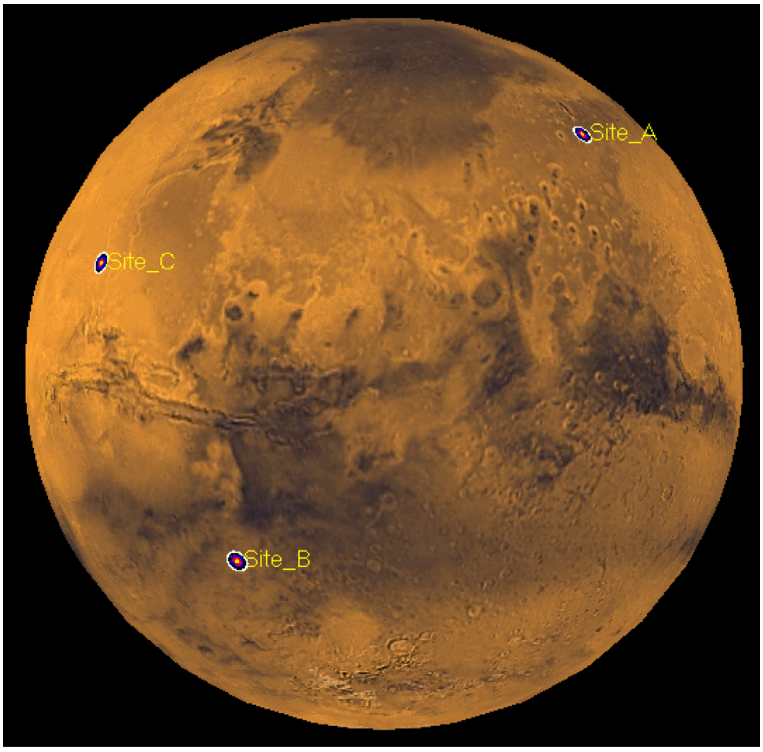


Fig. 5.26 Target landing sites. Mars image: NASA

Table 5.2 Entry conditions for three target landing sites

Mission	τ_0	δ_0	h_0	V_0	γ_0	χ_0
	[°]	[°]	[km]	[km/s]	[°]	[°]
CASE A	−14.8	28.6	120.0	5.980	−12.0	80.0
CASE B	−75.2	−38.3	120.0	5.980	−12.0	80.0
CASE C	−93.0	8.0	120.0	5.980	−12.0	80.0

of the lander of m_{lander} = 1,640 kg, the mass of the back cover and drogue chute of m_{cover} = 80 kg, and the mass of the heat shield of m_{shield} = 180 kg.

To accommodate the Mach-dependency of the aerodynamic properties of both vehicle and parachutes, the speed of sound, a, is calculated using the well-known formula of $a = \sqrt{\gamma RT}$, with γ the ratio of heat coefficients C_p and C_V, R the gas constant and T the temperature. For the EMCD 4.1, these properties are calculated internally in the database interface, using the actual atmospheric composition that has been modelled in the database. For the other two databases, a 100% CO_2 atmosphere is assumed, where γ is calculated from temperature-dependent tables.

The calculation of the convective heat-flux for the stagnation point, $q_{c,nose}$ is based on a simple engineering model, assuming a vehicle wall at room temperature. The previously introduced Chapman model is given by

$$q_{c,nose} = c_1 \rho^n V^m$$

with $c_1 = 1.5588 \cdot 10^{-4}$, $n = 0.5$ (laminar flow) and $m = 3.04$, providing a heat flux in W/m^2 for this particular vehicle. The vehicle and parachute systems have been implemented as described above, as well as the mission time line. All simulations will be carried out in three degrees of freedom, where the parachute drag is taken into account as an increase in drag area.

The nominal altitude-velocity profiles for the three baseline missions are shown in Fig. 5.27a, whereas the corresponding ground tracks are shown in Fig. 5.27b. These results have been obtained with the mean conditions of the EMCD 4.1 Mars Climate database. The influence of wind is not taken into account, to prevent obscuring the results with different wind profiles for the three reference missions.

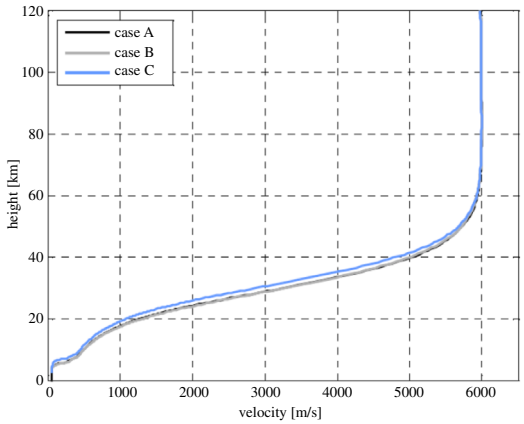
Due to the differences in density profiles and Mars Orbiter Laser Altimeter (MOLA) altitude of the landing sites (the stop criterion for the simulation), the flight time until touch down for the three reference missions is not the same: for case A the flight time $t_f = 356$ s, for case B, $t_f = 260$ s and for case C, $t_f = 319$ s. It is clear from Fig. 5.27a, that the difference in density profile also shows in the altitude-velocity profile. To anticipate on the results discussed below, it may be clear that the density differences in the three databases will lead to similar discrepancies.

5.3.5.2 Mars GRAM2001 and EMCD 3.1 Versus EMCD 4.1

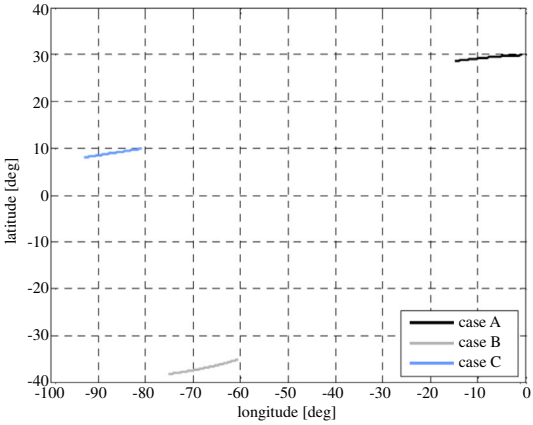
In this section trajectories originating from the use of the three databases will be compared, and to that end this section is divided into two parts: (i) the nominal mission and (ii) a sensitivity analysis by means of an extensive Monte Carlo simulation. The impact of simulating wind on the trajectory results is not done, due to the unavailability of an angle-of-attack dependency in the aerodynamic database. The angle of attack will vary significantly, but this will not reflect in any drag or lift changes. Considering the large differences in wind vector (see Fig. 2.43), no real conclusions could be drawn based on the results of such an analysis. Finally, it is stressed that the current case study covers only a first assessment of using EMCD 4.1. An extension of this work could focus on other dust scenarios, and possibly dust storms and gravity waves.

Nominal Mission

For each of the three target landing sites three trajectory simulations will be executed, i.e., one for each climate database. It is noted that despite the fact that the databases provide a wind profile, this is not taken into account right away, so that the focus can be on the differences in density profiles.



(a) Altitude versus velocity.



(b) Ground track.

Fig. 5.27 Nominal mission profile for the three reference missions (mean EMCD 4.1 database conditions)

As evaluation approach, it is proposed to evaluate the three reference missions Case A, B and C for the mean database settings. To allow for a comparison with the results of the Monte-Carlo analysis later on in this section, in Table 5.3 several selected performance indices are listed, i.e., the maximum dynamic pressure, $\bar{q}_{max}$, the maximum *g-load*, the maximum stagnation heat flux q_{nose}, the integrated heat load in the stagnation point, Q_{nose}, the dynamic pressure at drogue chute deployment, $\bar{q}_{drogue}$, and the dynamic pressure at main chute deployment, $\bar{q}_{main}$. It is noted that the latter two parameters are particularly useful for parachute sizing analysis, whereas the former four can serve as input for the design of the thermal protection system. Despite the differences in atmospheric properties, the differences in performance indices are not too great, in all cases less than 5%. However, it is noted again that at this moment no vehicle and mission uncertainties are taken into account, which is commonly done when providing input to the design teams. To conclude this part of the discussion, in Fig. 5.28 the stagnation heat flux and *g-load* are shown for the

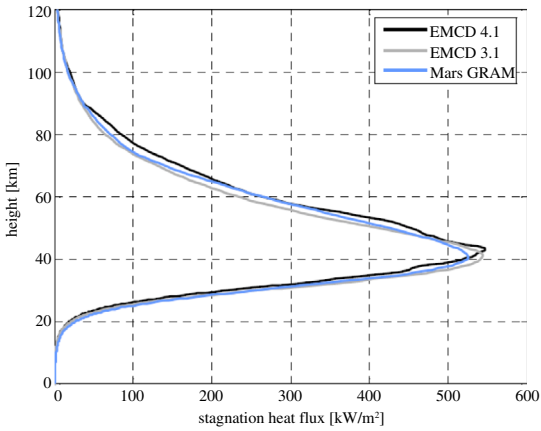
Table 5.3 Performance indices nominal mission

	Unit	EMCD41	EMCD31	MarsGRAM
Case A				
$\bar{q}_{max}$	N/m^2	5583	5719	5421
$g\text{-}load_{max}$	–	20.5	21.0	19.9
$q_{nose,max}$	kW/m^2	546	544	525
Q_{nose}	MJ/m^2	31.7	31.3	32.0
$\bar{q}_{drogue}$	N/m^2	650	621	639
$\bar{q}_{main}$	N/m^2	341	329	340
Case B				
$\bar{q}_{max}$	N/m^2	5935	6004	5615
$g\text{-}load_{max}$	–	21.8	22.1	20.6
$q_{nose,max}$	kW/m^2	539	545	533
Q_{nose}	MJ/m^2	31.5	31.0	31.5
$\bar{q}_{drogue}$	N/m^2	639	618	638
$\bar{q}_{main}$	N/m^2	334	327	339
Case C				
$\bar{q}_{max}$	N/m^2	5578	5856	5405
$g\text{-}load_{max}$	–	20.5	21.5	19.8
$q_{nose,max}$	kW/m^2	550	545	519
Q_{nose}	MJ/m^2	31.4	31.2	31.9
$\bar{q}_{drogue}$	N/m^2	643	617	624
$\bar{q}_{main}$	N/m^2	343	333	335

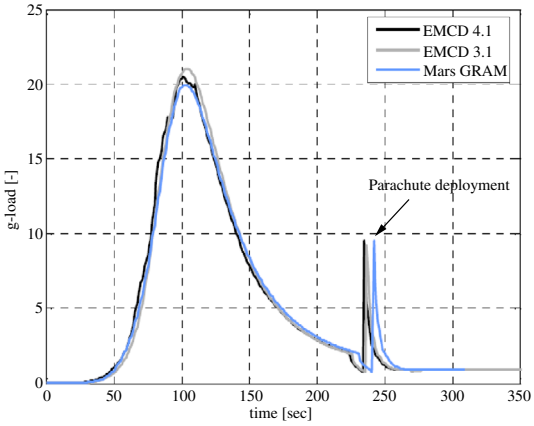
case-A reference mission. These curves support the results of Table 5.3, in the sense that in spite of the visible differences, these are not very large and the shape of the curves is similar.

Detailed mission analysis

This part will present the results of Monte-Carlo analyses concerning the three benchmark missions. The same database conditions as before are chosen. The parameters selected for variation in the Monte Carlo analysis are the vehicle drag coefficient C_D, the drogue and main parachute drag coefficient $C_{D,drogue}$ and $C_{D,main}$, the vehicle mass m, the initial longitude, τ_0, latitude, δ_0, velocity, V_0, flight-path angle, γ_0 and heading, χ_0, and the mission sequence times T_1 (initiated by the Mach number, M_1 = 1.8), T_2 and T_3. In Table 5.4, the numerical values of these dispersions are listed. During the simulations, the actual dispersions will be taken from a uniform distribution. The influence on the maximum *g-load*, the maximum heat flux, the dynamic pressure at parachute deployment, the touch-down velocity, and the landing position will be studied. We mention here that the dispersion values have been chosen quite arbitrarily, only to show the difference in results due to the use of the three databases.



(a) Stagnation heat flux.



(b) G-load.

Fig. 5.28 Case-A nominal trajectory results for the three databases

For an actual mission analysis, these dispersions have to be reviewed and possibly adjusted to the mission and system specifications.

For each of the three reference missions, 250 simulations per selected database are executed. The resulting footprints including the nominal landing sites are shown in Fig. 5.29. For the considered landing sites, 1°-difference in longitude corresponds with a surface distance between 50 and 60 km, depending on the latitude. Similarly, a 1°-difference in latitude corresponds with a surface distance of about 60 km. Thus, the maximum observed shift of the landing sites is between 10 and 15 km. It can be seen in Fig. 5.29 that the global shapes of the footprints is not affected by the choice of climate database. The footprints are shifted with respect to each other to the same extent as the nominal landing sites. This gives reason to assume that the differences between atmospheric conditions gives a constant shift of the individual trajectories.

Table 5.4 Monte-Carlo dispersions

nr	Parameter	Uniform dispersion	nr	Parameter	Uniform dispersion
1	C_D	10%	7	V_0	10 m/s
2	$C_{D,drogue}$	10%	8	γ_0	0.1°
3	$C_{D,main}$	10%	9	χ_0	2°
4	m	25 kg	10	M_1	0.1
5	τ_0	1°	11	T_2	1 s
6	δ_0	12°	12	T_3	1 s

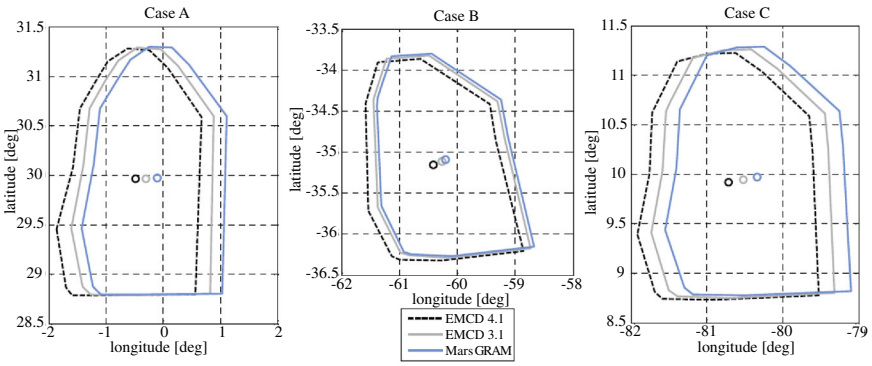


Fig. 5.29 Landing footprints from Monte-Carlo analysis for case A (left), B (middle) and C (right). The circles indicate the nominal landing sites

However, studying the corresponding histograms of longitude and latitude shows much more diversity of the interior points, which contradicts the former observation. However, a more detailed analysis should be carried out for a better understanding.

Finally, it is worth mentioning that although the stop condition for each of the simulations is the moment the vehicle reaches the actual surface, there is a discrepancy in at least the resolution of the MOLA topography of the related data files. Although this will not directly affect the landing point, since the final descent is close to vertical, it does affect the atmospheric properties that use the MOLA altitude as a reference for computation. It is therefore wise to carefully review the implemented MOLA topography in each of the databases before doing detailed analyses.

In Table 5.5, the minimum, maximum and mean values of the performance indices have been listed. It is clear that the differences are significant, and larger than the corresponding values of the nominal mission (Table 5.3). To appreciate the differences in the performance indices, in Fig. 5.30 the relative difference of the mean and standard deviation for the EMCD 3.1 and Mars-GRAM results with respect to the EMCD 4.1 results are shown. Although the difference in mean values stay well below 10%, the difference in standard deviation exceeds the 100% mark for several

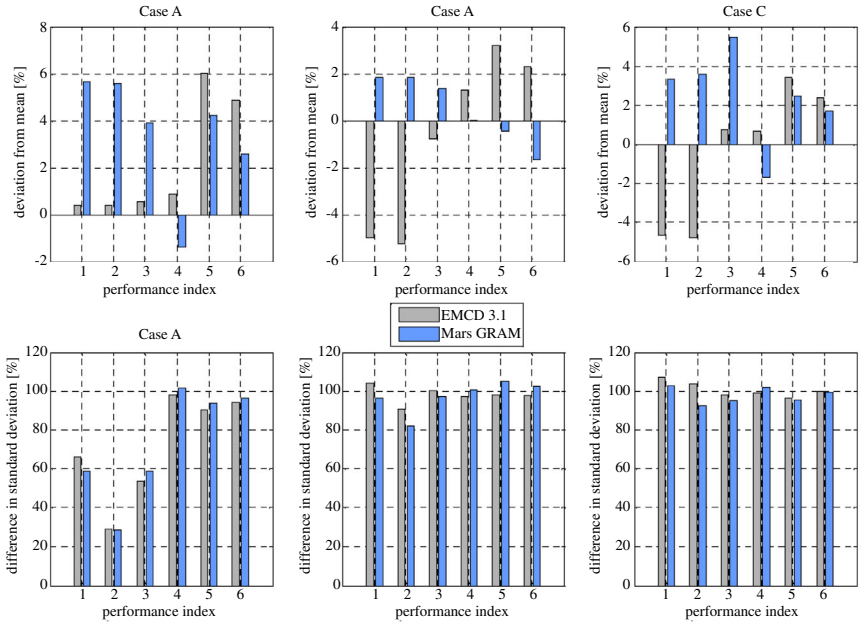


Fig. 5.30 Relative deviation from EMCD 4.1 mean (top) and standard deviation (bottom) for the six performance indices of Table 5.5

performance indices. Note also, that the difference in mean value is in some cases opposite for the EMCD 3.1 results and Mars-GRAM results. These results show, that if design criteria are based on 3σ-values, e.g., parachute deployment criteria based on dynamic pressure, great care should be taken when using a particular climate database. In fact, since the EMCD 4.1 is obligatory for use in AURORA projects, all previous results obtained with the EMCD 3.1 should be discarded.

In Fig. 5.31, the histograms of several case-A performance indices have been plotted. Despite the uniform distribution of the dispersions, the performance indices basically follow a normal distribution, albeit somewhat skewed in some cases (maximum stagnation heat flux and g-load for the EMCD 4.1 runs). The shift in mean value and difference in standard deviation is confirmed by these plots. We point out that the final-velocity distributions are almost adjoining. Since the final velocity is basically determined by the density profile of a limited altitude range close to the surface, this does not seem to be strange in light of the surface-density differences between the three databases. Overall, using the EMCD 4.1 leads to higher loads (conservative input for the thermo-mechanical design) and lower impact velocities (possibly an optimistic input for the parachute design). Of course, more detailed analyses would be required using actual mission and system specifications to sustain these (preliminary) conclusions.

Table 5.5 Performance indices Monte-Carlo analysis

		Unit	EMCD41			EMCD31			MarsGRAM		
	Case A		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
1	$\bar{q}_{max}$	N/m^2	4967	5757	7166	5162	5733	6370	4924	5432	5988
2	$g\text{-}load_{max}$	–	18.7	21.1	25.1	20.0	21.0	21.9	18.9	19.9	20.9
3	$q_{nose,max}$	kW/m^2	499	547	617	517	544	573	497	526	557
4	Q_{nose}	MJ/m^2	29.5	31.6	34.0	29.4	31.4	33.5	30.1	32.1	34.3
5	$\bar{q}_{drogue}$	N/m^2	578	667	794	542	627	718	557	639	731
6	$\bar{q}_{main}$	N/m^2	294	340	389	281	323	370	288	331	381
	Case B		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
1	$\bar{q}_{max}$	N/m^2	5146	5733	6537	5457	6019	6660	5105	5628	6218
2	$g\text{-}load_{max}$	–	19.9	21.0	22.4	21.0	22.1	22.9	19.7	20.6	21.4
3	$q_{nose,max}$	kW/m^2	512	541	584	516	545	576	506	534	564
4	Q_{nose}	MJ/m^2	29.6	31.5	33.6	29.2	31.1	33.1	29.5	31.5	33.6
5	$\bar{q}_{drogue}$	N/m^2	556	640	731	538	620	708	556	643	735
6	$\bar{q}_{main}$	N/m^2	284	327	377	278	320	368	288	333	382
	Case C		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
1	$\bar{q}_{max}$	N/m^2	5108	5600	6188	5277	5861	6484	4883	5415	6005
2	$g\text{-}load_{max}$	–	19.5	20.5	21.4	20.5	21.5	22.5	18.9	19.8	20.6
3	$q_{nose,max}$	kW/m^2	517	550	581	515	546	576	491	520	550
4	Q_{nose}	MJ/m^2	29.5	31.5	33.6	29.4	31.2	33.4	30.0	32.0	34.2
5	$\bar{q}_{drogue}$	N/m^2	556	644	736	537	622	710	543	628	716
6	$\bar{q}_{main}$	N/m^2	289	334	383	283	326	374	286	329	376

5.3.5.3 Conclusions

In this study, the ESA Mars Climate Database Version 4.1 has been implemented in a flight-dynamics software and used to do a simplified mission analysis. To compare the results with those obtained with other databases, the previous release of the ESA database (Version 3.1) and NASA's Mars GRAM 2001 have been linked with the same software. The interfacing of the three databases is similar, and relatively straightforward. Whereas the simulation of three reference missions for a Beagle-2 shaped vehicle gave more or less similar results (only differences up to 5% in selected performance indices), the results of a Monte Carlo analysis for each of the three target landing sites showed a larger difference. The mean values of the performance objectives differed up to 10%, but for the standard deviation differences larger than 100% were observed in some cases. The shift in landing point was found to be about 10 to 15 km. Although the mean and standard deviations of all the footprints (i.e., the combination of three landing locations and three databases) were different, the global shape of the footprints per landing location was the same. This should be analysed in more detail to find a suitable explanation.

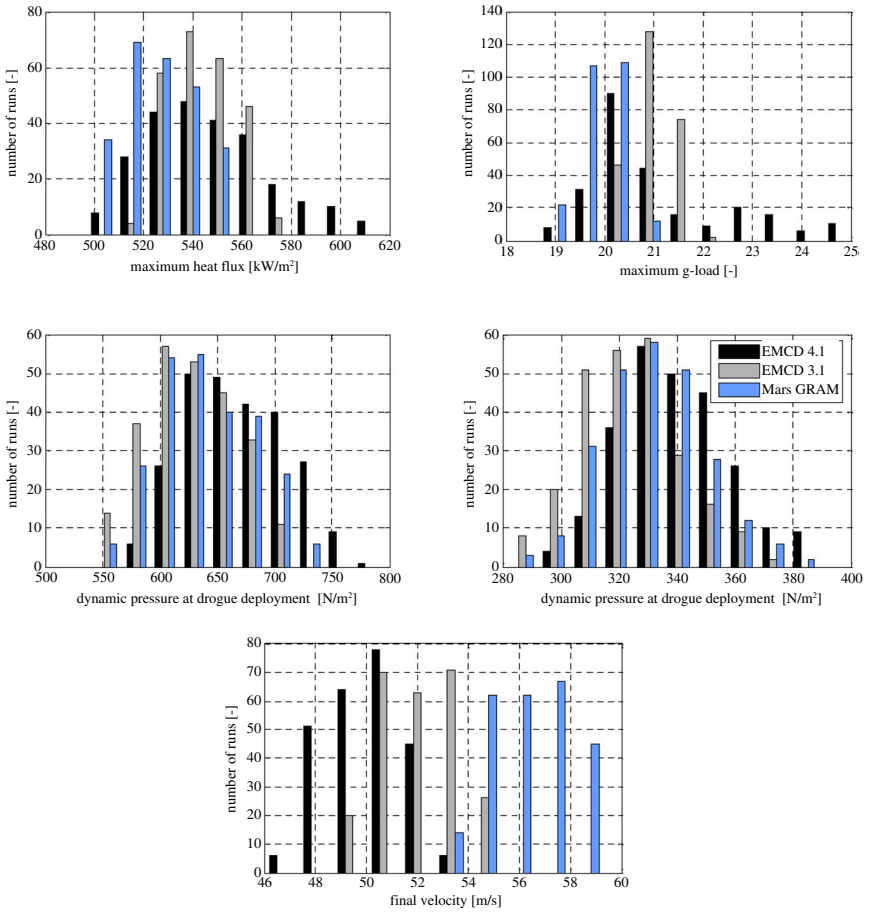


Fig. 5.31 CASE-A performance indices from Monte-Carlo analysis

Overall, using the EMCD 4.1 leads to higher loads (conservative input for the thermo-mechanical design) and lower impact velocities (possibly an optimistic input for the parachute design). Of course, more detailed analyses would be required using actual mission and system specifications to sustain these (preliminary) conclusions.

5.4 Gliding Entry

Important aspects of the problem of re-entry are the possible severe decelerations for human occupants of the re-entry vehicle, the intense aerodynamic heating and the tactical aspect of having satisfactory control over both the time and location of

landing. The first two aspects came up for discussion in the previous section about the ballistic entry. Due to the absence of lift, time and location of the landing cannot be influenced. Moreover, the large thermo-mechanical loads make the ballistic entry unsuitable for manned space flight. To minimise heating rates and decelerations the re-entry vehicle should follow a very shallow entry path, but fixing the time and location of landing is most difficult with these very shallow entries. Also the total absorbed heat is much greater than for steep entries, as there is a direct relation between the (much) longer flight time for shallow entries and the total heat load.

To maintain the shallow path and to avoid that the thermo-mechanical loads are exceeded, the guidance and control system can manipulate the lift force, either by changing the magnitude with angle-of-attack modulation or the direction with bank-angle control. The boundaries defined by the maximum allowable thermo-mechanical loads define the entry corridor that we discussed at the beginning of this chapter. For gliding entry we will go into more detail of the entry corridor in Sect. 5.4.3. Changing the direction of the lift results in an out-of-plane motion that will be covered at the end of this section, since we will again assume 2D planar motion at first. The boundaries of the out-of-plane motion define the so-called landing footprint, which we will discuss in Sect. 5.4.4.

Several examples of lifting vehicles that will follow a gliding trajectory are shown in Fig. 5.32. The most well-known example is, of course, the American Space Shuttle



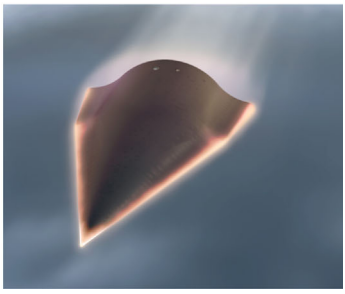
(a) Space Shuttle. Image: NASA.



(b) X-38. Image: NASA.



(c) SpaceShip 2. Image: Luke Colby/Virgin Galactic.



(d) HTV-2. Image: DARPA.

Fig. 5.32 Examples of lifting entry vehicles

(a). In service for 30 years, after 135 flights from its maiden flight in 1981 onwards the Shuttle fleet retired on July 21, 2011. The X-38 (b) is an example of the so-called lifting-body shape that came to development in the 1960s. The X-38 served as a test vehicle for the development of a wingless lifting-body reentry vehicle that was to be used as a Crew Return Vehicle for the International Space Station (ISS). Development started in 1995, but the project was cancelled in 2002 for budgetary reasons. Virgin Galactic's SpaceShip Two (c) is a suborbital, air-launched spaceplane designed for space tourism. It had its first powered test flight on 29 April 2013. The Virgin Galactic Spaceline planned to commercially fly the spaceplane in 2014, but had to reschedule their plans several times. The Hypersonic Technology Vehicle 2 (HTV-2), (d), was a crewless, experimental rocket glider developed as part of the DARPA Falcon Project, capable of flying at 21,000 km/h. The first test flight was on 22 April 2010, but contact was lost soon after booster separation. Also the second test (11 August 2011) failed, when nine minutes into its planned 30-minute Mach-20 glide flight contact was lost.

Subsequently, in Sect. 5.4.1 we will discuss the flight-mechanics aspects of the gliding trajectory, and subsequently derive equations for the flight range, the flight duration, the equilibrium descent angle, and the maximum deceleration and maximum heat flux as a function of altitude. As mentioned earlier, Sects. 5.4.3 and 5.4.4 introduce the entry corridor and landing footprint, respectively. Finally, a case study concludes this section on gliding entry.

5.4.1 Flight Mechanics

5.4.1.1 Equilibrium Glide Trajectories

Starting point for analysing several aspects of the gliding trajectory is the set of equations derived in Sect. 5.2, for convenience repeated below:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{5.100}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.101}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma \tag{5.102}$$

In the gliding descent it can safely be assumed that flight-path angle γ changes slowly with time, such that $\frac{\mathrm{d}\gamma}{\mathrm{d}t} \approx 0$. Additionally, the actual value of γ is assumed to be so small that $\cos\gamma \approx 1$. With these two assumptions Eq. (5.101) reduces to

$$L = mg\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.103}$$

or, the lift balances the weight minus the centrifugal force as a result of the Earth's curvature.

With $L = C_L \frac{1}{2}\rho V^2 S$ it follows directly from Eq. (5.103) that a relationship exists between the air density and the velocity:

$$\rho = 2\frac{W/S}{C_L}\left(\frac{1}{V^2} - \frac{1}{V_c^2}\right) \tag{5.104}$$

where we have introduced the lift parameter $\frac{W/S}{C_L}$, which can be seen as the equivalent of the ballistic parameter for vehicles with lift. The (local) circular velocity V_c is in principle a function of altitude, although the change is small. Realising that $h << R_e$ (the radius of the Earth) we can write:

$$V_c^2 = gR = g_0\frac{R_e^2}{R^2}R = g_0 R_e\frac{R_e}{R} = V_{c,0}^2\frac{R_e}{R_e + h} = V_{c,0}^2\frac{1}{1 + h/R_e} \approx V_{c,0}^2$$

so that from here on we can assume a constant value of V_c. Note that in the above expression we have used the expression of $g(h)$ for a spherical Earth, i.e., $g(h) = g_0\frac{R_e^2}{(R_e+h)^2}$.

Going back to Eq. (5.104) we see that for a given lift parameter and atmosphere $\rho = \rho(h)$, a relation between h and V is given. If ρ follows the exponential profile that we have consistently used in this chapter, and rewriting Eq. (5.104) then

$$\rho = \rho_0 e^{-\beta h} = 2\frac{W/S}{C_L}\left(\frac{1}{V^2} - \frac{1}{V_{c,0}^2}\right)\left(= \frac{c_1}{V^2} - c_2\right) \Rightarrow$$

$$\frac{1}{2}\frac{\rho}{\frac{W/S}{C_L}} = \frac{1}{V^2} - \frac{1}{V_{c,0}^2} \Rightarrow$$

$$\frac{\frac{1}{2}\rho V_{c,0}^2 + \frac{W/S}{C_L}}{\frac{W/S}{C_L}} = \frac{V_{c,0}^2}{V^2} \Rightarrow$$

$$\frac{V^2}{V_{c,0}^2} = \frac{\frac{W/S}{C_L}}{\frac{1}{2}\rho V_{c,0}^2 + \frac{W/S}{C_L}} \Rightarrow$$

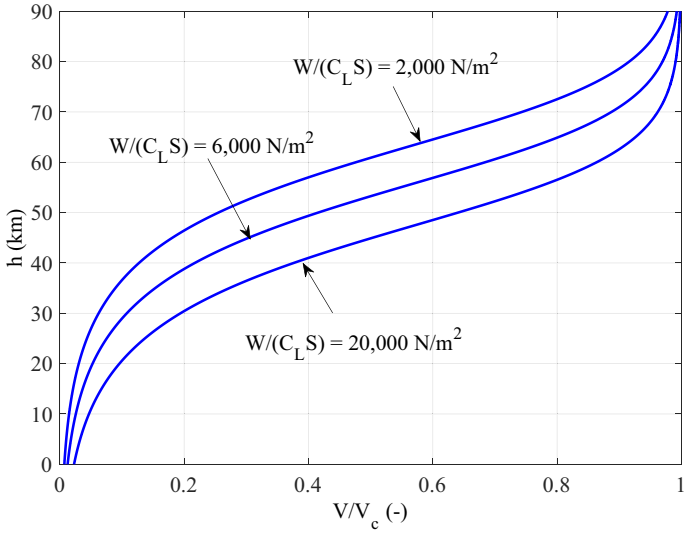


Fig. 5.33 Descent in gliding flight, exponential atmosphere ($V_c = V_{c,0} = 7{,}908\,\text{m/s}$)

$$\frac{V}{V_{c,0}} = \sqrt{\frac{\frac{W/S}{C_L}}{\frac{1}{2}\rho_0 e^{-\beta h} V_{c,0}^2 + \frac{W/S}{C_L}}} = f(h) \tag{5.105}$$

The (minimum) velocity with with the vehicle reaches the surface ($h = 0$) easily follows, if we assume that $\frac{1}{2}\rho V_{c,0}^2 > \frac{W/S}{C_L}$, so we can ignore the latter term[5]:

$$\frac{V_{min}}{V_{c,0}} = \sqrt{\frac{\frac{W/S}{C_L}}{\frac{1}{2}\rho_0 V_{c,0}^2 + \frac{W/S}{C_L}}} \approx \frac{1}{V_{c,0}}\sqrt{\frac{2}{\rho_0}\frac{W/S}{C_L}} \tag{5.106}$$

In Fig. 5.33, several flight paths in the h-V plane are shown for distinct values of $W/C_L S$, using the isothermal atmosphere with a scale height of $H_s = 6{,}950\,\text{m}$. The trajectories have been plotted using Eq. (5.105), and are known as equilibrium flight paths. These were first pointed out by the German Eugen Sänger (Sänger 1933); see also Sect. 1.2 for more biographical data. On a side note: it may be obvious by inspecting Eq. (5.105) that the choice of atmosphere model (i.e., in this case the selection of ρ_0 and H_s) will affect the curves of Fig. 5.33. For instance, when using larger values for H_s the trajectory curves shift upwards.

Inspecting the trajectories in Fig. 5.33 one can see that for a large portion of the flight the altitude decreases at a constant rate with the velocity. This corresponds

[5] With $V_{c,0} = 7{,}908\,\text{m/s}$ and $\rho_0 = 1.225\,\text{kg/m}^3$ it is easily verified that at low altitudes the lift parameter hardly contributes to the denominator of Eq. (5.105).

with a constant flight-path angle. In the following we will derive an expression for this equilibrium flight-path angle, $\bar{\gamma}$.

To find this expression, we begin with the notion that we need a relation between γ, h and V, which is provided by Eq. (5.102):

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} = V \sin\gamma$$

To eliminate the time dependency we separate the derivative, introducing the flight range, s, for which holds $\frac{\mathrm{d}s}{\mathrm{d}t} = V$:

$$\frac{\mathrm{d}h}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}s}V = V \sin\gamma \tag{5.107}$$

so

$$\frac{\mathrm{d}h}{\mathrm{d}s} = \sin\gamma \tag{5.108}$$

From the exponential atmosphere model, $\rho = \rho_0 e^{-\beta h}$, we have

$$\mathrm{d}\rho = -\beta\rho_0 e^{-\beta h}\mathrm{d}h = -\beta\rho\mathrm{d}h \tag{5.109}$$

or

$$\mathrm{d}h = -\frac{\mathrm{d}\rho}{\beta\rho} \tag{5.110}$$

Substituted, this gives for $\sin\gamma$:

$$\sin\gamma = \frac{\mathrm{d}h}{\mathrm{d}s} = -\frac{1}{\beta\rho}\frac{\mathrm{d}\rho}{\mathrm{d}s} = -\frac{1}{\beta\rho}\frac{\mathrm{d}\rho}{\mathrm{d}V}\frac{\mathrm{d}V}{\mathrm{d}s} \tag{5.111}$$

which means we need to find expressions for $\frac{\mathrm{d}\rho}{\mathrm{d}V}$ and $\frac{\mathrm{d}V}{\mathrm{d}s}$. To begin with, we can use the earlier found expression for the equilibrium trajectories, Eq. (5.104),

$$\rho = 2\frac{W/S}{C_L}\left(\frac{1}{V^2} - \frac{1}{V_{c,0}^2}\right)$$

to derive $\frac{\mathrm{d}\rho}{\mathrm{d}V}$:

$$\frac{\mathrm{d}\rho}{\mathrm{d}V} = -4\frac{W/S}{C_L}\frac{1}{V^3} \tag{5.112}$$

From the equation of motion $m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma$ we get, realising γ is small so the weight component can be neglected with respect to drag:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D \tag{5.113}$$

and thus

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{\mathrm{d}V}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = V\frac{\mathrm{d}V}{\mathrm{d}s} = -\frac{D}{m} = -\frac{D}{L}g\frac{L}{mg} \tag{5.114}$$

In the above equation we have introduced the lift-to-drag ratio L/D. Note that for a given vehicle, C_L and L/D are correlated to each other by the aircraft characteristic $C_D = b(C_L)$, which, for hypersonic velocities, may be assumed to be independent of the Mach number.

Also, we rewrote the equation such that we can use Eq. (5.103):

$$\frac{L}{mg} = \left(1 - \frac{V^2}{V_c^2}\right) \tag{5.115}$$

This result substituted in Eq. (5.114) gives us the second derivative, $\frac{\mathrm{d}V}{\mathrm{d}s}$, we were looking for:

$$\frac{\mathrm{d}V}{\mathrm{d}s} = -\frac{\frac{D}{L}g\left(1 - \frac{V^2}{V_c^2}\right)}{V} \tag{5.116}$$

With $\frac{\mathrm{d}\rho}{\mathrm{d}V}$ given by Eq. (5.112) and $\frac{\mathrm{d}V}{\mathrm{d}s}$ by (5.116), we can substitute these results into Eq. (5.111) to find

$$\bar{\gamma} \approx \sin\bar{\gamma} = -\frac{1}{\beta}\frac{2}{L/D}\frac{g}{V^2} \tag{5.117}$$

and with $V_c \approx \sqrt{R_e g}$ the final result is

$$\bar{\gamma} \approx \sin\bar{\gamma} = -\frac{1}{\beta R_e}\frac{2}{L/D}\frac{V_c^2}{V^2} \tag{5.118}$$

In this expression $\beta R_e = 890$, so that the descent angle $\bar{\gamma}$ will be small if the velocity is not too low. For $V/V_c = 0.10$ and $L/D = 2$ it follows that $\bar{\gamma} = -0.112$ rad $= -6.44°$. So, at higher velocities the previously used simplifications ($\cos\gamma = 1$ and $\frac{\mathrm{d}\gamma}{\mathrm{d}t} = 0$) will be valid.

5.4.1.2 Flight Range and Flight Duration

For both the flight range and the flight duration in the equilibrium flight, simple expressions can be derived, again with the assumption that the descent angle is so small that the weight term in Eq. (5.100) can be neglected:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \approx -D \tag{5.119}$$

Similarly, we found with $\frac{\mathrm{d}\gamma}{\mathrm{d}t} \approx 0$ the result of Eq. (5.103):

$$L = mg\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.120}$$

To get a relation between velocity and the range differential we rewrite $\frac{\mathrm{d}V}{\mathrm{d}t}$ as in Eq. (5.114) and with Eq. (5.120) we find Eq.(5.116) written in a slightly different form:

$$V\mathrm{d}V = -\frac{D}{L}g\left(1 - \frac{V^2}{V_c^2}\right)\mathrm{d}s \tag{5.121}$$

To solve this we need to integrate the above equation, and to facilitate this we use the substitution $x = \frac{V^2}{V_c^2}$. With

$$\mathrm{d}x = \frac{2V}{V_c^2}dV \Rightarrow dV = \frac{V_c^2}{2V}\mathrm{d}x$$

Equation (5.121) can thus be rewritten as

$$V\frac{V_c^2}{2V}\mathrm{d}x = \frac{V_c^2}{2}\mathrm{d}x = -\frac{D}{L}g\,(1-x)\,\mathrm{d}s \tag{5.122}$$

or

$$-\frac{1}{2}\frac{L}{D}\frac{V_c^2}{g}\frac{\mathrm{d}x}{1-x} = \mathrm{d}s \tag{5.123}$$

The equation to solve is

$$R_f = \int_0^{s_f}\mathrm{d}s = -\frac{1}{2}\frac{L}{D}\frac{V_c^2}{g}\int_{x_1}^{x_2}\frac{1}{1-x}\mathrm{d}x \tag{5.124}$$

and with the solution to the standard integral

$$\int_{x_1}^{x_2}\frac{1}{1-x}\mathrm{d}x = -\ln\left(\frac{1-x_2}{1-x_1}\right) \tag{5.125}$$

integration yields

$$R_f = \int \mathrm{d}s = +\frac{1}{2}\frac{L}{D}\frac{V_c^2}{g}\ln\left(\frac{1-x_2}{1-x_1}\right) \tag{5.126}$$

Or, with $x_1 = \frac{V_E^2}{V_c^2}$ and $x_2 = \frac{V_F^2}{V_c^2}$

$$R_f = \frac{1}{2}\frac{L}{D}\frac{V_c^2}{g}\ln\left(\frac{1-\frac{V_E^2}{V_c^2}}{1-\frac{V_E^2}{V_c^2}}\right) \tag{5.127}$$

Knowing that $V_c^2 \approx gR_e$ and realising that $V_F << V_E$, Eq. (5.127) simplifies to:

$$\frac{R_f}{R_e} = -\frac{1}{2}\frac{L}{D}\ln\left(1-\frac{V_E^2}{V_c^2}\right) \tag{5.128}$$

In Fig. 5.34 this relationship has been presented graphically. With, for example, $L/D = 2$ (a representative value for hypersonic velocities), it follows for $V_E/V_c = 0.8$ that R_f/R_e is about 1. The distance travelled is then roughly 6,400 km. This distance will increase with increasing L/D. When V_E approaches V_c, the ratio of the two approaches "one", which drives the natural logarithm towards minus infinity. This means that the range increases, and in the limit case approaches plus infinity. Physically this means that the vehicle keeps on orbiting the Earth, thereby increasing its "range" indefinitely.

Example 5.7 (*Glide range*) For a gliding re-entry vehicle, the requirement is that it has to cover a range of $R = 6{,}000$ km to reach its landing site. What should be the minimum lift-to-drag ratio (L/D) of the vehicle such that this can be achieved? Assume that the vehicle enters the atmosphere with a velocity equal to 80% of the circular velocity, and that the Earth's equatorial radius is $R_e = 6{,}378$ km.

Using Eq. (5.128), it follows for L/D:

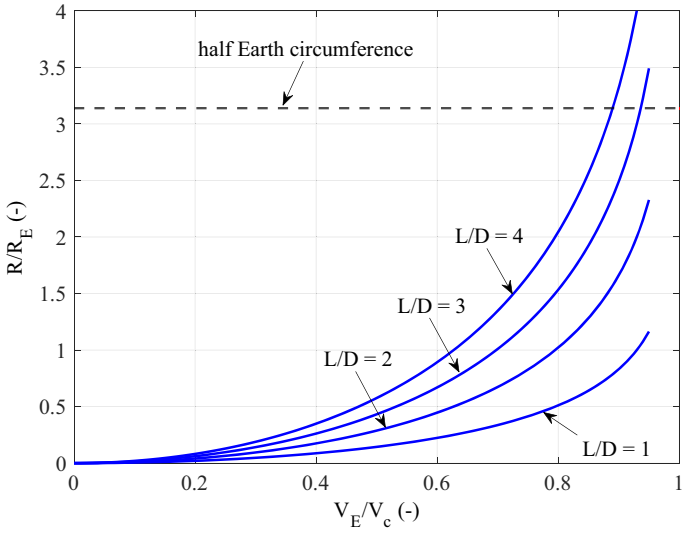


Fig. 5.34 Normalised flight range of a gliding vehicle, for values of the lift-to-drag ratio, L/D

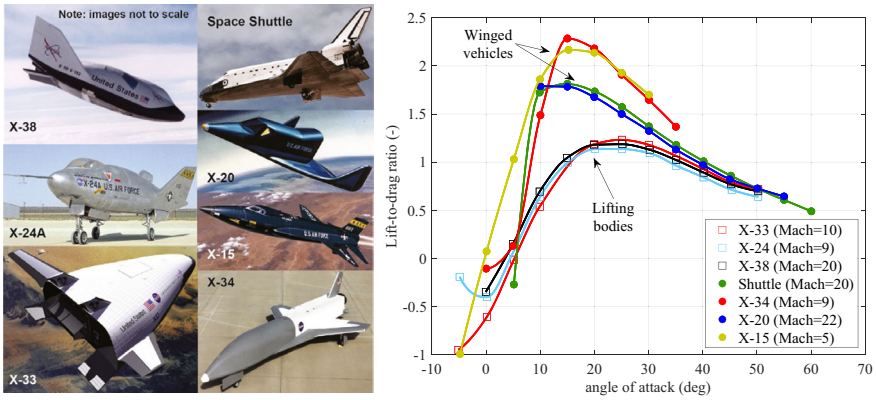


Fig. 5.35 Hypersonic L/D comparisons for seven lifting reentry configurations (Vehicle images: NASA; data taken from Whitmore and Dunba (2003))

$$\frac{L}{D} = -2\frac{\frac{R}{R_e}}{\ln\left(1 - \frac{V_E^2}{V_c^2}\right)} \tag{5.129}$$

Substituting $R = 6{,}000\,\text{km}$, $R_e = 6{,}378\,\text{km}$ and $\frac{V_E^2}{V_c^2} = 0.8^2 = 0.64$, it follows that $L/D = 1.84$

■

In terms of L/D, passenger aircraft can have values in between 20 and 25, but for entry vehicles the lift-to-drag ratio is a lot smaller, see Fig. 5.35. Even though the Space Shuttle may look like a plane, in terms of aerodynamic performance it is definitely not!

The flight duration can also be derived from Eq. (5.114), together with Eq. (5.120):

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{D}{L}g\frac{L}{mg} = -\frac{D}{L}g\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.130}$$

or

$$\mathrm{d}t = -\frac{\frac{L}{D}}{g\left(1 - \frac{V^2}{V_c^2}\right)}\mathrm{d}V \tag{5.131}$$

The total flight duration is obtained by integrating Eq. (5.126). However, to avoid complex expressions we apply the following substitution

$$x = \frac{V}{V_c}$$

$$\mathrm{d}x = \frac{\mathrm{d}V}{V_c} \Rightarrow \mathrm{d}V = V_c \mathrm{d}x = V_c \mathrm{d}\left(\frac{V}{V_c}\right)$$

This substituted into Eq. (5.131) yields:

$$\mathrm{d}t = -\frac{\frac{L}{D}}{g\left(1-x^2\right)} V_c \mathrm{d}x = -\frac{L}{D}\frac{V_c}{g}\frac{\mathrm{d}x}{1-x^2} \tag{5.132}$$

The integrand part on the right-hand side leads to a standard integral with a known solution, i.e.,

$$\int \frac{\mathrm{d}x}{1-x^2} = \frac{1}{2}\int \frac{\mathrm{d}x}{1+x} + \frac{1}{2}\int \frac{\mathrm{d}x}{1-x} = \frac{1}{2}\ln\frac{1+x}{1-x} \tag{5.133}$$

So, with boundaries $x_1 = \frac{V_E}{V_c}$ and $x_2 = \frac{V_F}{V_c}$, Eq. (5.132) gives us with Eq. (5.133)

$$t_{flight} = \int \mathrm{d}t = -\frac{1}{2}\frac{V_c}{g}\frac{L}{D}\left(\ln\frac{1+x_2}{1-x_2} - \ln\frac{1+x_1}{1-x_1}\right) \Rightarrow \tag{5.134}$$

$$t_{flight} = -\frac{1}{2}\frac{V_c}{g}\frac{L}{D}\ln\left(\frac{1+x_2}{1-x_2}\cdot\frac{1-x_1}{1+x_1}\right) \Rightarrow \tag{5.135}$$

$$t_{flight} = -\frac{1}{2}\frac{V_c}{g}\frac{L}{D}\ln\left(\frac{1+V_F/V_c}{1-V_F/V_c}\cdot\frac{1-V_E/V_c}{1+V_E/V_c}\right) \tag{5.136}$$

If we realise now that towards the end of the flight $V_F << V_c$ we can introduce $\frac{V_F}{V_c} \approx 0$ in the above equation, which finally gives us for the flight duration

$$t_{flight} = \frac{1}{2}\frac{V_c}{g}\frac{L}{D}\ln\left(\frac{1+V_E/V_c}{1-V_E/V_c}\right) \tag{5.137}$$

where we have exchanged numerator and denominator to get rid of the leading minus sign. Note that if one wants to establish the flight duration of only part of the flight, then the condition $V_F << V_c$ may no longer hold.

With $V_c = 7.92\,\mathrm{km/s}$, $V_E/V_c = 0.8$ and $L/D = 2$, it follows from Eq. (5.137) that the flight duration is $t_{flight} = 29.6\,\mathrm{min}$. If it further assumed that for the considered vehicle the lift parameter is $W/SC_L = 6{,}000\,\mathrm{N/m^2}$, it then follows for an isothermal atmosphere ($H_s = 7{,}200\,\mathrm{m}$) from Eq. (5.104) that in the equilibrium flight path the initial altitude is $h_E = 67\,\mathrm{km}$.

5.4.2 Thermo-Mechanical Loads

5.4.2.1 Deceleration

To find an expression for the deceleration in the equilibrium flight we begin again with $m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma$, as we also did for the ballistic flight. Of course, we could ignore the weight term due to the small γ, but keeping it will allow us to see the impact on, for instance, the maximum deceleration. So, in this case the acceleration a is defined as

$$\frac{a}{g} = \frac{1}{g}\frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{D}{mg} - \sin\gamma \tag{5.138}$$

With the previous results on equilibrium-glide condition and equilibrium angle, Eqs. (5.103) and (5.118),

$$L = mg\left(1 - \frac{V^2}{V_c^2}\right)$$

$$\bar{\gamma} \approx \sin\bar{\gamma} = -\frac{1}{\beta R_e}\frac{2}{L/D}\frac{V_c^2}{V^2}$$

we can write for the *deceleration* $\bar{a}$ (note the minus sign in the equation below):

$$\frac{\bar{a}}{g} = -\frac{1}{g}\frac{\mathrm{d}V}{\mathrm{d}t} = +\frac{D}{W} + \sin\gamma = +\frac{D}{L}\left[1 - \frac{V^2}{V_c^2} - \frac{2}{\beta R_e}\frac{1}{\frac{V^2}{V_c^2}}\right] \tag{5.139}$$

The term $-\frac{D}{L}\frac{2}{\beta R_e}\frac{1}{\frac{V^2}{V_c^2}}$ is the contribution of the weight term and will have a diminishing effect on the deceleration. The deceleration for a given L/D becomes a maximum at

$$\frac{\mathrm{d}\bar{a}}{\mathrm{d}\,(V/V_c)} = 0 \tag{5.140}$$

which gives us

$$\frac{V'}{V_c} = \left(\frac{2}{\beta R_e}\right)^{1/4} = 0.218 \tag{5.141}$$

This velocity ratio of about 22% means that the maximum deceleration takes place quite late in the trajectory. Note: if the weight term would have been ignored the corresponding velocity ratio would have been $\frac{V'}{V_c} = 0$. Further, substituting the found velocity ratio into Eq.(5.139) gives us the expression for the maximum deceleration, i.e.,

$$\frac{\bar{a}_{max}}{g} = +\frac{D}{L}\left(1 - 2\sqrt{\frac{2}{\beta R_e}}\right) \cong 0.905\frac{D}{L} \tag{5.142}$$

With $L/D = 2$ this means that $\frac{\bar{a}_{max}}{g} = 0.45$. Without the weight term this would be $\frac{\bar{a}_{max}}{g} = \frac{D}{L}$, or $\frac{\bar{a}_{max}}{g} = 0.5$, so including the weight term decreases the maximum deceleration by 10%. Even then, the decelerations during the gliding flight are small, which was intuitively already obvious, in light of the reasonably long time during which the velocity change is achieved.

The density at which the maximum deceleration takes place follows from Eq. (5.104),

$$\rho = 2\frac{W/S}{C_L}\left(\frac{1}{V^2} - \frac{1}{V_c^2}\right)$$

so with $V'/V_c = 0.218$ and $V_c = 7.92\,\text{km/s}$:

$$\rho' = 0.639\frac{W/S}{C_L}10^{-6} \tag{5.143}$$

With $W/SC_L = 6{,}000\,\text{N/m}^2$ and an exponential atmosphere this correlates to an altitude of $h' = 40\,\text{km}$.

To finalise the discussion about the influence of the weight term, we first summarise the main findings. Without the weight term,

$$\frac{\bar{a}}{g} = +\frac{D}{L}\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.144}$$

The maximum deceleration is found at the end of the equilibrium flight, at sea level. As $V_F << V_c$, it follows from Eq. (5.106) that $\frac{\bar{a}_{max}}{g} = \frac{D}{L}$ at

$$V_{min} = \sqrt{\frac{2}{\rho_0}\frac{W/S}{C_L}} \tag{5.145}$$

For the aforementioned numerical example, one then finds $\frac{\bar{a}_{max}}{g} = 0.5$ at $V_{min} \approx 100\,\text{m/s}$ (360 km/hr).

From the expressions above, notably Eqs. (5.139) and (5.142), the relationship between $\frac{a}{a_{max}} = \frac{\mathrm{d}V/\mathrm{d}t}{(\mathrm{d}V/\mathrm{d}t)_{max}}$ and $\frac{V}{V_c}$ follows directly, and is shown in Fig. 5.36. The influence of the weight component along the trajectory on the resulting decelerations is clearly visible.

5.4.2.2 Stagnation Heatflux

To discuss the heat flux in or close to the stagnation point, we begin again with Chapman's equation:

$$q_c = c_1\frac{1}{R_N^n}\rho^{1-n}V^m \tag{5.146}$$

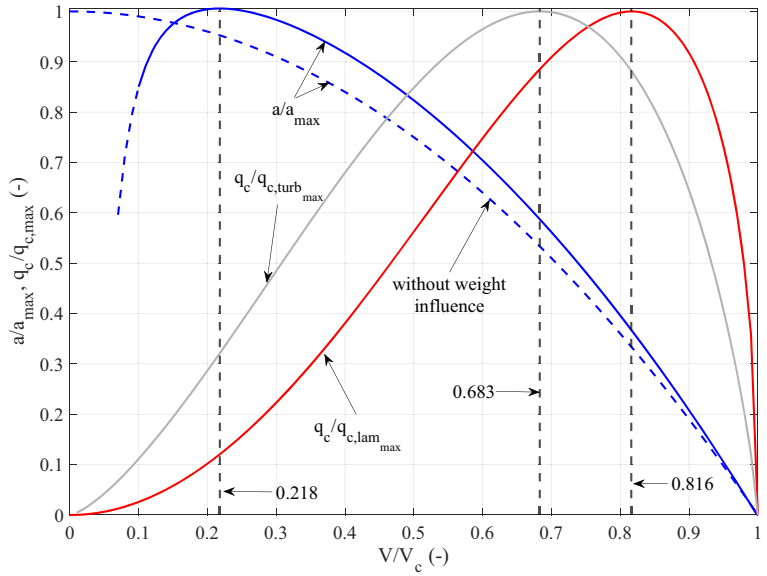


Fig. 5.36 Generalised curves for the glide trajectory

wherein $n = 0.5$ for a laminar boundary layer and $n = 0.2$ for a turbulent boundary layer, and m is an empirical constant commonly taken as $m = 3$ (for Earth). The constants c_1, n and m can be used to adjust to empirical data from different numerical models and/or for different planets, see also the discussion about the Huygens stagnation-flux model in Sect. 5.3.4.

We will now derive expressions for the velocity and altitude at which the maximum stagnation heat flux, $q_{c,max}$, occurs. We begin by substituting Eq. (5.104),

$$\rho = 2\frac{W/S}{C_L}\left(\frac{1}{V^2} - \frac{1}{V_c^2}\right) \tag{5.147}$$

into Eq. (5.146), yielding

$$q_c = c_1 \frac{1}{R_N^n}\left[2\frac{W/S}{C_L}\left(\frac{1}{V^2} - \frac{1}{V_c^2}\right)\right]^{1-n} V^m \tag{5.148}$$

or

$$q_c = c_1 \frac{1}{R_N^n}\left(\frac{2}{V_c^2}\frac{W/S}{C_L}\right)^{1-n} V_c^m \left(\frac{V_c^2}{V^2} - 1\right)^{1-n}\left(\frac{V}{V_c}\right)^m \tag{5.149}$$

For a given (constant) nose radius, lift parameter and circular velocity, and substituting $x = \frac{V}{V_c}$ for ease of use we rewrite the above equation as

$$q_c = c_2 \left(\frac{1}{x^2} - 1 \right)^{1-n} x^m \tag{5.150}$$

with $c_2 = c_1 \frac{1}{R_N^n} \left(\frac{2}{V_c^2} \frac{W/S}{C_L} \right)^{1-n} V_c^m$.

The maximum stagnation heat flux is obtained from $\frac{dq_c}{dx} = 0$, yielding

$$\frac{dq_c}{dx} = c_2(1-n) \left(\frac{1}{x^2} - 1 \right)^{-n} \frac{-2}{x^3} x^m + c_2 \left(\frac{1}{x^2} - 1 \right)^{1-n} m x^{m-1} = 0 \Rightarrow$$

$$-2(1-n) \left(\frac{1}{x^2} - 1 \right)^{-n} x^{m-3} + m \left(\frac{1}{x^2} - 1 \right)^{1-n} x^{m-1} = 0 \Rightarrow$$

$$2(1-n) \frac{1}{x^2} = m \left(\frac{1}{x^2} - 1 \right) = 0 \Rightarrow$$

$$2(1-n) = m(1-x^2) \Rightarrow$$

$$x^2 = 1 - \frac{2(1-n)}{m} \Rightarrow$$

$$x = \frac{V''}{V_c} = \sqrt{\frac{m - 2(1-n)}{m}} \tag{5.151}$$

With Eq. (5.147) for the density ρ, it can be shown that the maximum value for q_c occurs at

$$\rho'' = 4 \frac{W/S}{C_L} \frac{1}{V_c^2} \left(\frac{1-n}{m + 2(n-1)} \right) \tag{5.152}$$

With the commonly used value for m for Earth entry studies, $m = 3$, Eqs. (5.151) and (5.152) change into

$$\frac{V''}{V_c} = \sqrt{\frac{1+2n}{3}} \tag{5.153}$$

$$\rho'' = 4 \frac{W/S}{C_L} \frac{1}{V_c^2} \left(\frac{1-n}{1+2n} \right) \tag{5.154}$$

As mentioned before, in these expressions n determines the type of boundary layer. For a laminar one ($n = 0.5$) it follows:

$$\frac{V''}{V_c} = \sqrt{\frac{2}{3}} = 0.816 \tag{5.155}$$

$$\rho'' = \frac{W/S}{C_L}\frac{1}{V_c^2} \tag{5.156}$$

The velocity of 81.6% of V_c is well away from the corresponding value for maximum mechanical load (22.8%). From a structural point of view, this large separation is good for the vehicle, because a high temperature weakens the vehicle structure. Now, the vehicle has the chance to cool down before it encounters the peak *g-load*, which will then be easier to endure.

For the isothermal atmosphere, the maximum heat flux q_c occurs at an altitude of $h'' = 68.1\,\text{km}$ for $W/SC_L = 6{,}000\,\text{N/m}^2$. Similarly, for a turbulent boundary layer ($n = 0.2$) we get

$$\frac{V''}{V_c} = \sqrt{\frac{7}{15}} = 0.683 \tag{5.157}$$

$$\rho'' = \frac{16}{7}\frac{W/S}{C_L}\frac{1}{V_c^2} \tag{5.158}$$

resulting in an altitude of $h'' = 62.1\,\text{km}$.[6]

An expression for $q_{c,max}$ (laminar flow) follows by substituting Eqs. (5.155) and (5.156) into Eq. (5.146):

$$q_{c,max} = \frac{c_1}{\sqrt{R_N}}\sqrt{\frac{W/S}{C_L}}\sqrt{\frac{1}{V_c^2}}\sqrt{\frac{2}{3}}^3 V_c^3 = \frac{c_1^*}{\sqrt{R_N}}\sqrt{\frac{W/S}{C_L}} \tag{5.159}$$

The dependency of $q_{c,max}$ on the remaining variables is clear. For larger nose radii the maximum heat flux will go down. This is the main reason why hypersonic vehicles have blunt noses: sharp noses will get very hot and may need active cooling to keep the temperatures below material limits. A similar effect can be obtained by lowering the weight or increasing the surface area exposed to the flow. A lighter and "wider" vehicle will experience a smaller maximum thermal loading. And last but not least, by increasing the lift coefficient, C_L, $q_{c,max}$ will decrease as well. Effectively, the vehicle should fly in a $C_{L,max}$ configuration, which means for hypersonic flow that the angle of attack should be as large as possible. For this reason, the Space Shuttle came down with an angle of attack in excess of 40°, and basically this holds for all gliding vehicles.

With the equations found earlier, Eqs. (5.153) and (5.154), it also follows that

$$\frac{q_c}{q_{c,max}} = \left(\frac{\rho}{\rho''}\right)^{1-n}\left(\frac{V}{V''}\right)^3 \Rightarrow \tag{5.160}$$

[6] For a scale height of $H_s = 6{,}950\,\text{m}$—as we used before for some of the examples on ballistic entry—the corresponding altitudes are $h'' = 65.7\,\text{km}$ and $h'' = 60.1\,\text{km}$, respectively.

$$= \left[\frac{1+2n}{2(1-n)} V_c^2 \left(\frac{1}{V^2} - \frac{1}{V_c^2}\right)\right]^{1-n} \left(\frac{3}{1+2n}\right)^{\frac{3}{2}} \left(\frac{V}{V_c}\right)^3 \Rightarrow$$

$$= \left\{\frac{1+2n}{2(1-n)} \left(\frac{V_c}{V}\right)^2 \left[1 - \left(\frac{V}{V_c}\right)^2\right]\right\}^{1-n} \left(\frac{3}{1+2n}\right)^{\frac{3}{2}} \left(\frac{V}{V_c}\right)^3$$

and finally

$$\frac{q_c}{q_{c,max}} = \left(\frac{3}{1+2n}\right)^{\frac{3}{2}} \left(\frac{V}{V_c}\right)^{1+2n} \left\{\frac{1+2n}{2(1-n)} \left[1 - \left(\frac{V}{V_c}\right)^2\right]\right\}^{1-n} \tag{5.161}$$

For $n = 0.5$, the above expression reduces to

$$\frac{q_c}{q_{c,max}} = \frac{3}{2}\sqrt{2} \left(\frac{V}{V_c}\right)^2 \sqrt{1 - \left(\frac{V}{V_c}\right)^2} \tag{5.162}$$

and for $n = 0.2$, which cannot be written out so nicely:

$$\frac{q_c}{q_{c,max}} = \left(\frac{15}{7}\right)^{\frac{3}{2}} \left(\frac{V}{V_c}\right)^{1.4} \left\{\frac{7}{8} \left[1 - \left(\frac{V}{V_c}\right)^2\right]\right\}^{0.8} \tag{5.163}$$

Equations (5.162) and (5.163) have been used to plot the heat-flux profiles as a function of $\frac{V}{V_c}$, see Fig. 5.36.

To derive expressions for the heat supply to the entire vehicle, we follow the same approach as for the ballistic trajectory, beginning with Eq. (5.69):

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{C_F}{4} \rho V^3 S_W \tag{5.164}$$

With Eq. (5.147) substituted, it is found that $\mathrm{d}Q/\mathrm{d}t$ is maximal for

$$\frac{V'''}{V_c} = \frac{1}{\sqrt{3}} = 0.577 \tag{5.165}$$

and

$$\rho''' = 4 \frac{W/S}{C_L} \frac{1}{V_c^2} \tag{5.166}$$

Additionally, it holds that:

$$\frac{\frac{\mathrm{d}Q}{\mathrm{d}t}}{\left(\frac{\mathrm{d}Q}{\mathrm{d}t}\right)_{max}} = \frac{3}{2}\sqrt{3} \left(1 - \frac{V^2}{V_c^2}\right) \frac{V}{V_c} \tag{5.167}$$

The density ρ''' is four times as large as ρ''; this implies that the altitude at which the maximum value for $\mathrm{d}Q/\mathrm{d}t$ in the exponential atmosphere occurs at roughly 10 km beneath that for maximum q_c in the stagnation point (laminar boundary layer).

The total heat that is supplied to the vehicle during the entire flight follows from integrating $\mathrm{d}Q/\mathrm{d}t$:

$$Q = \int_{V_E}^{V_F} \frac{\frac{\mathrm{d}Q}{\mathrm{d}t}}{\frac{\mathrm{d}V}{\mathrm{d}t}} \mathrm{d}V \tag{5.168}$$

This integral can be solved quite easily if the weight influence on $\frac{\mathrm{d}V}{\mathrm{d}t}$ is neglected (i.e., the $\sin\gamma$ term):

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{Dg}{W} = -\frac{C_D \frac{1}{2}\rho V^2 S g}{W} \tag{5.169}$$

With Eq. (5.164) integration yields an expression for Q, similarly what we did for the ballistic trajectory (conform Eqs. (5.82) through (5.85)):

$$Q = -\frac{C_F}{2}\frac{W S_W}{C_D S g}\int_{V_E}^{V_F} V \mathrm{d}V = -\frac{C_F}{4}\frac{W S_W}{C_D S g}\left(V_F^2 - V_E^2\right) \tag{5.170}$$

and with the kinetic energy at entry $\mathcal{E}_k = \frac{1}{2}mV_E^2 = \frac{1}{2}\frac{W}{g}V_E^2$ this gives us

$$\frac{Q}{\mathcal{E}_k} = \frac{Q}{\frac{1}{2}\frac{W}{g}V_E^2} = \frac{1}{2}\frac{C_F S_W}{C_D S}\left[1 - \left(\frac{V_F}{V_E}\right)^2\right] \tag{5.171}$$

At the end of the equilibrium flight it holds that $V_F << V_E$, so that:

$$\frac{Q}{\mathcal{E}_k} = \frac{1}{2}\frac{C_F S_W}{C_D S}$$

In this case, for a given initial velocity, V_E, the heat supplied to the vehicle (Q) decreases as the total drag surface area increases in relation to the frictional drag surface area. This would once again require a blunt shape for the vehicle. However, this has also a negative impact on the glide number, so the ability of converting kinetic energy into flight range is compromised. A blunt shape is used for the nose and the leading edges of wings to limit the heat flow in the stagnation point(s).

In general, $\frac{C_F S_W}{C_D S}$ will be higher for a glide vehicle than for a ballistic vehicle. This does not necessarily have to lead to a higher thermal load of the vehicle. Even though as a result of the longer flight time of the glide vehicle, high skin temperatures will be reached during the flight, but the gliding vehicle can also radiate more heat as compared to ballistic vehicles.

5.4.3 *Entry Corridor*

The enclosure of all entry and descent trajectories that fulfil the constraints and guarantee a safe flight until landing is the so-called entry corridor, a three-dimensional spatial tube (Fig. 5.37). An important feature of supercircular entries—to a lesser extent this is also true for near-circular entries—is that of the guidance accuracy acquired to survive entry. If the guidance error results in undershooting an intended trajectory too much, the vehicle will enter the atmosphere at an excessively steep flight-path angle. In that case the deceleration will be too much for either the vehicle or the ability to execute an evasive manoeuvre. If, on the other hand, the result of poor guidance is overshooting the intended trajectory, the atmospheric pull may not be strong enough to capture the vehicle, and it may disappear forever into space or enter a highly elliptic orbit, prolonging dangerously the lapse time until the next return.

For some planets the entry corridor is very narrow, which puts a lot of emphasis on the robustness of the guidance system, of course, only when propulsive manoeuvres can be executed prior to entry, or when the vehicle has sufficient aerodynamic lift. In the latter case, the corridors of certain planets can be broadened significantly. This is not true, however, for the corridors of Mars and Titan (Chapman 1960). Chapman also showed that for non-lifting vehicles having fixed aerodynamic coefficients, the corridor width is independent of vehicle weight and dimensions. But, for both lifting and non-lifting vehicles, all corridor widths decrease rapidly as the entry velocity is increased.

As an example, in Fig. 5.38 the entry corridor of Apollo-11 is shown. It was defined by flight-path angle limits as a function of entry velocity (Graves and Harpold 1972), with the entry interface being defined at $h_E = 121.92\,\text{km}$ (i.e., 400,000 ft). The overshoot boundaries were based on the 4,630 and 2,380 km entry ranges (2,500 and

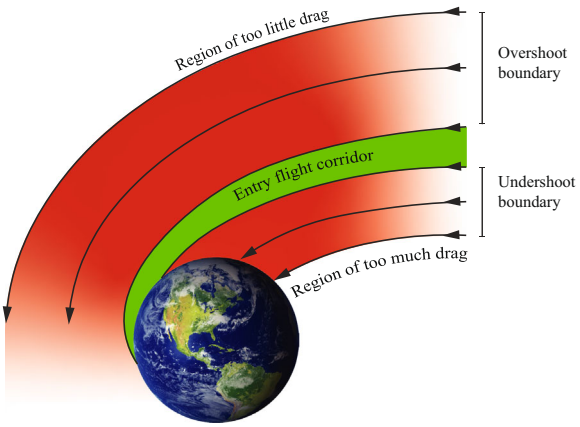


Fig. 5.37 The entry corridor, with overshoot and undershoot boundaries

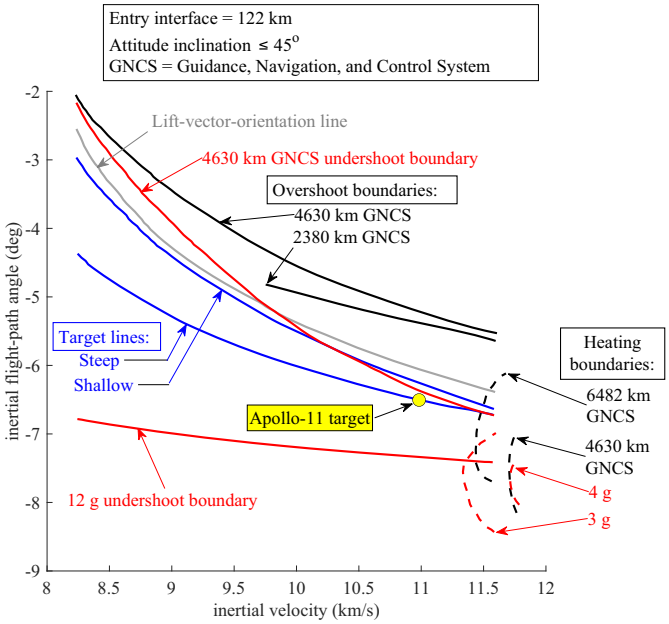


Fig. 5.38 Apollo-11 entry corridor. The original range boundaries are given as 1,285, 2,500 and 3,500 nautical miles. Data taken from Graves and Harpold (1972)

1,285 nautical miles, respectively), and represented the shallowest flight-path angles for which these minimum ranges could be achieved. The two indicated undershoot boundaries were based on a maximum *g-load* of 12 g, and a maximum range of 2,500 nmi. The 12-g constraint represented the steepest flight-path angle. The aerodynamic heating boundaries were controlled to a constant-load factor (3 g and 4 g, respectively), as well as for two different maximum-range constraints. They were based on a maximum temperature of 315° C at the bondline between the heat shield and the heat-shield substructure.

If we would only look at the global picture, e.g., the maximum thermo-mechanical loads on the vehicle, it is relatively straight-forward to construct the entry corridor, provided we know the entry conditions—we will show this later in this section. In case the entry conditions are part of the mission-design process, the problem becomes slightly more complicated, but also for this case Chapman (1960) has provided an extensive analysis of the problem at hand, an analysis summarised by others as well, see for instance the books by Loh (1968) and Vinh et al. (1980). It becomes even more complicated, though, when also sub-system constraints should be included in the analysis.

That constructing the entry corridor is not always easy is shown with the following example of a propelled, air-breathing launch system. Although not completely representative for an unpowered re-entry vehicle, it will get the idea across that there is more to it than just system-level performance aspects.

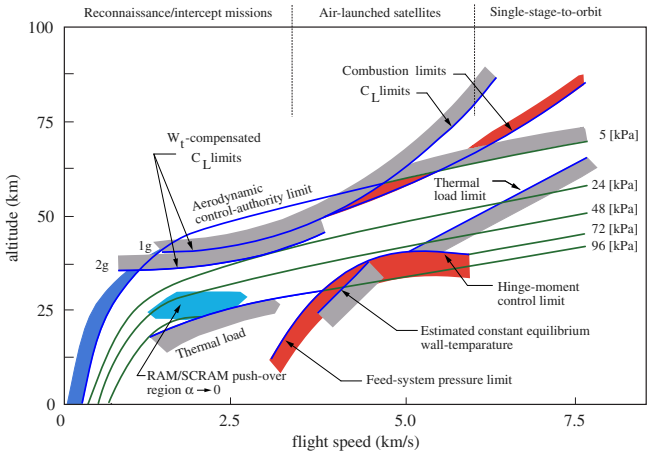


Fig. 5.39 The attainable performance envelope of an air-breathing space plane, adapted from McRuer (1991)

The interactions of airframe, propulsion system and controls may define many regions where the vehicle cannot be flown (McRuer 1991). Therefore, the attainable performance envelope of space planes cannot be defined solely by point-mass performance or considerations related to the operating point of the propulsion system. As an example, Fig. 5.39 shows a possible excluded region in flight corridors and operational areas, bounded by, for instance, thermal and structural limits. This is a region with low angles of attack in the RAM/SCRAM area that may be excluded because of bow and other shock impingements on the vertical control surfaces, in combination with directional divergencies and reduced directional control effectiveness and control power. Other impacts may be due to manoeuvring needs (e.g., a 2-g turn), which can cause a further reduction of the admissible flight corridor.

For initial re-entry trajectory design, usually three conditions are taken into account, i.e., the equilibrium-flight condition, the maximum allowable heat flux in the stagnation point and the maximum total deceleration, or *g-load*. In the following, it will be explained how each of these three will translate into a boundary in the *h*-*V* diagram and together will form the entry corridor.

The equilibrium-glide condition, derived from $\frac{\mathrm{d}\gamma}{\mathrm{d}t}$, should be seen as the theoretical ceiling under which no skipping flight will occur, the limit of which is $\gamma = 0$:

$$L = mg\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.172}$$

To transform this condition into a trajectory or path constraint in the *h*-*V* diagram, we proceed as follows: for a given value of the velocity, V_{eq}, we will reformulate Eq. (5.172) such that we can calculate the corresponding altitude, h_{eq}. We begin with

$$C_L \frac{1}{2} \rho_{eq} V_{eq}^2 S = mg \left(1 - \frac{V_{eq}^2}{V_c^2} \right) \tag{5.173}$$

which can be rewritten as

$$\rho_{eq} = 2 \frac{W/S}{C_L} \left(\frac{1}{V_{eq}^2} - \frac{1}{V_c^2} \right) \tag{5.174}$$

which is, of course, conform the definition of the equilibrium trajectories, Eq. (5.104). The altitude can be obtained from any atmosphere model, but we will stick to the exponential atmosphere for the ease of use. In that case

$$h_{eq} = -H_s \ln \frac{\rho_{eq}}{\rho_0} \tag{5.175}$$

It is noted that to calculate ρ_{eq} we need to know C_L. We can either assume a constant coefficient or use the more realistic $C_L = C_L(\alpha, M)$. To calculate the Mach number, we need to know the speed of sound, which is constant for the exponential atmosphere, but is otherwise a function of temperature, and thus, altitude. So, for any other atmosphere model where the speed of sound is not constant, an iterative process is needed to calculate h_{eq}. Using the exponential atmosphere may be less accurate, but it is an easier way to calculate this trajectory constraint.

The stagnation heat flux was defined earlier as, Eq. (5.146):

$$q_c = c_1 \frac{1}{R_N^n} \rho^{1-n} V^m \tag{5.176}$$

The original definition includes a few more constants that are now combined in the constant c_1. As already mentioned in Example 5.6, the alternative formulation for Chapman's equation reads

$$q_c = c^* \frac{1}{R_N^n} \left(\frac{\rho}{\rho_0} \right)^{1-n} \left(\frac{V}{V_c} \right)^m \tag{5.177}$$

with $c^* = 1.1097 \cdot 10^8 \sqrt{\text{m}}$, $\rho_0 = 1.225\,\text{kg/m}^3$, and $V_c = 7{,}905.4\,\text{m/s}$; as before, for Earth entry $n = 0.5$ (laminar flow) and $m = 3$. The reason to formulate it in this way is to facilitate studying the influence of each of these parameters, and to be consistent with other trajectory constraints, such as the equilibrium-glide condition that also includes V_c.

Suppose now that the maximum *allowable* heat flux is given by $q_{c,max}$. As before, for a given $V_{q,c}$, Eq. (5.177) gives us

$$q_{c,max} = c^* \frac{1}{R_N^n} \left(\frac{\rho_{q,c}}{\rho_0} \right)^{1-n} \left(\frac{V_{q,c}}{V_c} \right)^m \tag{5.178}$$

or

$$\rho_{q,c} = \rho_0 \left[R_N^n \frac{q_{c,max}}{c^*} \left(\frac{V_c}{V_{q,c}} \right)^m \right]^{\frac{1}{1-n}} \tag{5.179}$$

The corresponding altitude, $h_{q,c}$, follows again from Eq. (5.175).

The last of the three constraints determining the corridor, the *g-load* or n_g, is defined by the acceleration due to the *total* aerodynamic force, so the combined effect of drag and lift, and normalised by the gravitational acceleration at sea level:

$$n_g = \frac{\sqrt{D^2 + L^2}}{m g_0} \tag{5.180}$$

As before, let us define the maximum *allowable g-load* by $n_{g,max}$ and for a given V_g we derive an expression for h_g. Equation (5.180) is rewritten as

$$n_{g,max} = \frac{1}{2} \rho_g V_g^2 S \frac{\sqrt{C_D^2 + C_L^2}}{m g_0} \tag{5.181}$$

or

$$\rho_g = 2 n_{g,max} \frac{m g_0}{V_g^2 S \sqrt{C_D^2 + C_L^2}} \tag{5.182}$$

with h_g obtained with Eq. (5.175). Both C_D and C_L are functions of α and M, although the Mach effect is relatively small in the hypersonic regime.

An example of how to construct the entry corridor for a lifting entry vehicle is discussed next.

Example 5.8 (*HORUS entry corridor*) HORUS is a winged re-entry vehicle similar to the Space Shuttle, and will be used throughout this text for different case studies (e.g., in Chap. 7 for navigation- and control-system design, notably Sects. 7.3.4 and 7.5.7). More information about this reference vehicle can be found in Appendix B.

To minimise the maximum heat flux an entry vehicle should enter the atmosphere in a maximum-lift configuration, which means at maximum angle of attack. For that reason it is logical that the equilibrium flight condition is evaluated at the same α_{max}, which for HORUS is $\alpha_{max} = 40°$. This maximum lift gives the largest corridor, but we cannot assume a constant angle of attack for the complete flight. We will therefore check the influence of lowering the angle of attack to $\alpha = 20°$, which is a value close to the maximum lift-to-drag ratio configuration. In this configuration, the maximum cross- and downrange can be obtained. To evaluate Eq. (5.174) we need C_L, and for HORUS the tabulated values for C_L as a function of Mach number and angle of attack are found in Table B.8.

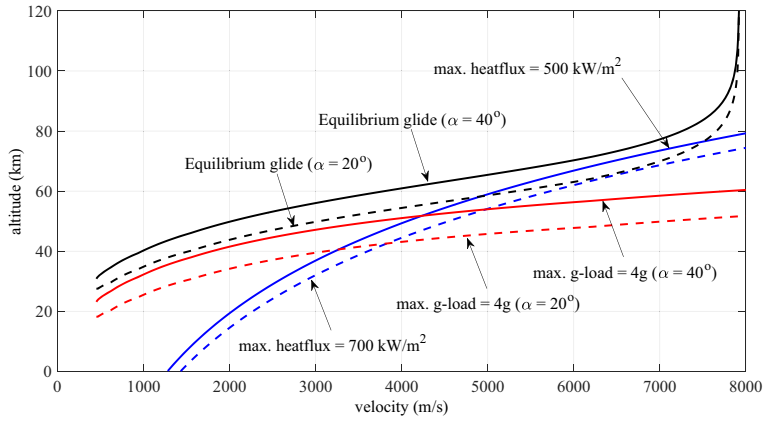


Fig. 5.40 The entry corridor of HORUS

The maximum heat flux is set at $q_{c,max} = 500\,\text{kW/m}^2$, which gives a wall temperature (radiation equilibrium) of $T_{wall} \approx 1{,}800\,\text{K}$. To study the effect of a relaxation of this constraint, also a value of $q_{c,max} = 700\,\text{kW/m}^2$ is chosen, equivalent to $T_{wall} \approx 1{,}950\,\text{K}$. Note that it is usually the other way round, notably that the maximum flux values are derived from the material properties of, for instance, the vehicle nose that can withstand a certain maximum temperature.

Finally, also two different values for the maximum *g-load* are studied. The commonality between the two is that they are both $n_g = 4\,\text{g}$, but the first one is evaluated for an angle of attack of $\alpha = 40°$ (the entry or maximum-lift configuration), whereas the second one is evaluated for $\alpha = 20°$, a value close to the maximum L/D configuration for the second half of the flight. As before, the reader is referred to Tables B.7 and B.8 for tabulated data on C_D and C_L as a function of M and α.

The resulting entry corridor for HORUS is shown in Fig. 5.40, where the solid lines indicate the nominal corridor, and the dashed lines show the effect of changing the constraints. At its smallest the corridor is only 3 km wide. This means that very tight trajectory control is required to either avoid violating the heat-flux constraint, or by entering a skipping flight when the equilibrium-glide condition is crossed. Relaxing the heat-flux constraint up to $700\,\text{kW/m}^2$ will add another 6 km to the corridor width. However, lowering the angle of attack to $\alpha = 20°$ will drastically lower the equilibrium-glide condition. This is of course not an issue at the beginning of the trajectory, because to minimise the maximum heat flux one should fly at the largest possible angle of attack. However, when the peak heat flux has passed the angle of attack should be lowered to increase the cross- and downrange capability. This will increase the possibility of reaching positive flight-path angles if nothing else is done.

Fortunately, there is another guidance mode that can take care of that. In principle, skipping flight is driven by the vertical component of the lift force. When the lift force is tilted away from the vertical plane, a process known as banking, means that skip

can be controlled. This will not endanger violating the heat-flux constraint, because the *magnitude* of the lift force is still the same (and defined by the angle of attack). So, in short, with α one can control the heat flux, and with the bank angle σ (or in combination with α) one can control skipping flight.

The last of the constraints is the maximum *g-load* that kicks in only much later. At $V \approx 4{,}200$ m/s the heat-flux constraint crosses the *g-load* constraint, which at that moment becomes the driving constraint. Relaxing the heat-flux constraint shifts the intersection of the two constraints to the right, which means that because of the steeper trajectory the maximum allowable deceleration is reached sooner. Lowering the angle of attack to $\alpha = 20°$ will relax the *g-load* constraint, because the aerodynamic forces will be lower and thus the total deceleration. Once beyond the maximum heat-flux constraint, lowering the angle of attack will keep the corridor width more or less constant in terms of n_g and equilibrium glide. However, as mentioned before, the bank angle can be effectively used to stay away from the equilibrium glide condition.

■

5.4.4 Footprint

To lower the decelerations and thermal loads on the vehicle, instead of following a ballistic flight-path a gliding entry is the better alternative. The capability to generate lift is in this context the keyword, so the vehicle should be able to fly at a non-zero angle of attack and remain stable. The solution can be either passive, to have a fixed offset in the location of the centre of mass, like Apollo, or active: the vehicle is trimmed to a certain angle of attack by means of control surfaces, like the Space Shuttle. The amount of lift that can be generated is not only dependent on the angle of attack, but also on the shape of the vehicle. To be more precise, it is not just the lift that determines the performance of the vehicle, it is the combined effect of lift and drag, in the form of the lift-to-drag ratio. The more slender the shape, the higher the L/D.

The larger the L/D the larger the range the vehicle can fly: for given lifting capabilities that extend the flight time and thus the range, the energy dissipating effect of the drag is less, so more kinetic energy remains. It can easily be imagined that for maximum (down)range, L/D should be maximal. But, if the lifting capabilities of the vehicle are too large, it enters a skipping flight (see Sect. 5.5). To avoid that from happening the lift vector should be "tilted" out of the vertical plane by a process called banking. This is in principle a three-dimensional problem, because the out-of-plane component of the lift will induce a lateral motion, associated with the crossrange of the vehicle.

If we see the induced crossrange, i.e., the deviation from the nominal groundtrack, as a mission-design parameter, then we can use it to reach a designated landing area that is "away" from the nominal path. In Fig. 5.41 this concept is illustrated.

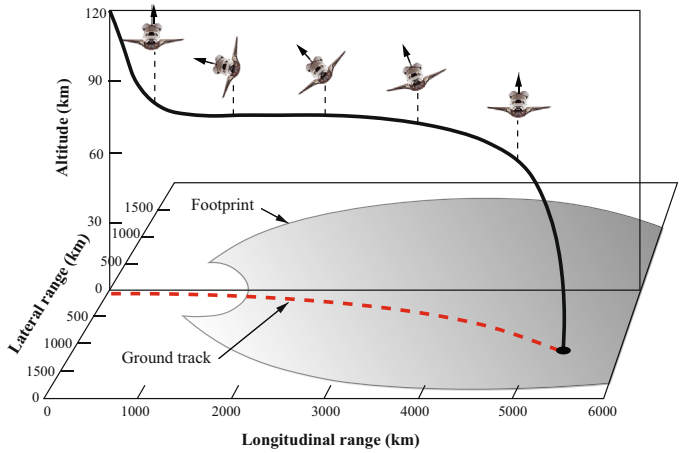


Fig. 5.41 The influence of banking on the landing location of a gliding vehicle

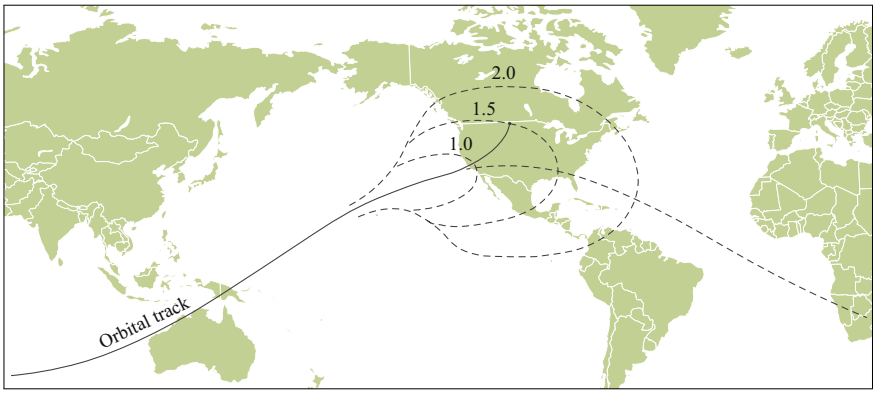


Fig. 5.42 The influence of the lift-to-drag ratio on the footprint of a gliding vehicle

Maximum lateral ranges are obtained by applying large bank angles early in the trajectory, followed by a reduced one once the the desired heading is achieved. A mapping of the maximum achievable longitudinal and lateral range boundaries is referred to as a footprint, which represents the region where recovery can be effected, or where the landing strip should be. Figure 5.42 shows the possible landing areas for lifting vehicles as a function of L/D. Vehicles with an L/D of around two can be landed anywhere in an area extending thousands of kilometres, both forward and laterally, from a given entry point.

How do we get this footprint? Simply said, if we would specify the angle-of-attack profile, and systematically vary the bank angle, but keeping it constant for a single trajectory, then we would get an enclosure of landing points that is the

footprint. But what happens when we change the α profile? We know that moving towards the maximum L/D configuration we will increase the range and thus the size of the footprint. And what about the thermo-mechanical constraints? What if we would specify a too low α at the beginning of flight, and the heat-flux constraint is exceeded? So, for the safe operation of an entry vehicle the question to answer is: how can we predict, from any initial position and velocity, the set of feasible landing sites without violating any of the constraints posed on the vehicle or the crew? This landing footprint will then provide critical information for mission planning in both nominal and abort situations.

So, to repeat: how do we get this footprint? Based on the brief discussion above, the footprint is formed by the end points of a significant number of optimal trajectories, and solving a family of optimal control problems is quite the effort. Features that make this task difficult, but cannot be neglected, are the constraints on heating, acceleration and dynamic pressure, and the Coriolis terms in the entry dynamics. The Coriolis terms, i.e., the effect of the Earth's rotation, introduces a significant asymmetry in the landing footprint. Vinh et al. (1981) discuss a method for producing optimal landing footprints in an idealised setting with no vehicle constraints or Coriolis effects. Continuing this work, Ngo and Doman (2002) developed a method for generating near-optimal landing footprints, but also did not account for Coriolis effects or vehicle constraints. However, Saraf et al. (2004) developed an elegant method to generate landing footprints in near real-time, which could be implemented in an on-board flight-management system. This method does include path constraints, accounts for Coriolis effects and constructs angle-of-attack profiles that are near optimal for long-range glide. This method was used by Mooij et al. (2007) to generate the footprint of HORUS, see Fig. 5.43.

For an understanding of the underlying principle and relations of the footprint, we cannot use the formulation of, for instance, Saraf et al. (2004). As before, we will derive an analytical approximation. From Chap. 4, Eq. (4.112) gives the rate of change of the heading angle, χ:

$$\begin{aligned} V\cos\gamma\dot{\chi} = &-\frac{L\sin\sigma}{m} + 2\omega_{cb}V(\sin\delta\cos\gamma - \cos\delta\sin\gamma\cos\chi) + \\ &+\frac{V^2}{R}\cos^2\gamma\tan\delta\sin\chi + \omega_{cb}^2 R\cos\delta\sin\delta\sin\chi \end{aligned} \tag{5.183}$$

For small flight-path angle and a flight along the equator ($\delta = 0°$), the above equation reduces to

$$mV\frac{\mathrm{d}\chi}{\mathrm{d}t} = -L\sin\sigma \tag{5.184}$$

The lateral motion is solely driven by the lateral component of the lift, and no longer by the internal dynamics of the system.

The same dynamics drive the longitudinal and lateral flight-range, R_{f_x} and R_{f_y}, respectively. In Fig. 5.44, the range geometry for an arbitrary trajectory is shown. By banking the vehicle it starts moving away from the initial flight path, indicated

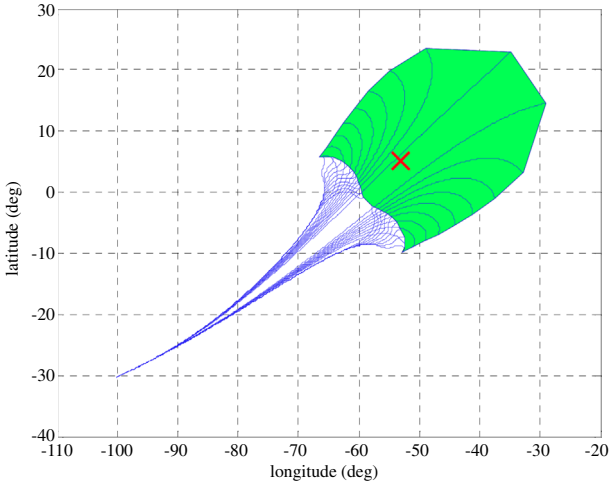


Fig. 5.43 Footprint-analysis result for HORUS. 'X' marks the Heading Alignment Cylinder, a targeting point for terminal-area flight (Mooij et al. 2007)

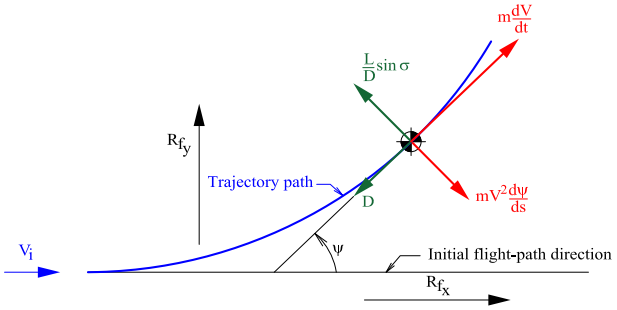


Fig. 5.44 Coordinate system for the simplified flight-range analysis

by the angle ψ. The two components of the range are defined with respect to the original flight path, i.e., R_{f_x} in along-path direction, and R_{f_y} perpendicular to this. Note that if the original flight path would be along the equator (due east), ψ would be the complementary angle to the flight heading, χ, which is defined with respect to the north.

So, in line with Eq. (5.184), but with opposite sign as the heading is the complimentary angle, we write for the rate of change of ψ:

$$mV\frac{\mathrm{d}\psi}{\mathrm{d}t} = L\sin\sigma \tag{5.185}$$

Introducing the flight range along the current trajectory, s, as we did in Eq. (5.107), we can split the derivative, i.e.,

$$mV\frac{\mathrm{d}\psi}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = mV^2\frac{\mathrm{d}\psi}{\mathrm{d}s} = L\sin\sigma \tag{5.186}$$

Reformulating Eqs. (5.100) and (5.101) with range as independent variable, and using the vertical lift component, $L\cos\sigma$, rather than the total lift, L, gives:

$$mV\frac{\mathrm{d}V}{\mathrm{d}s} = -D - mg\sin\gamma \approx -D \tag{5.187}$$

$$mV^2\frac{\mathrm{d}\gamma}{\mathrm{d}s} = L\cos\sigma - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \approx L\cos\sigma - mg\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.188}$$

where we have incorporated the assumptions of $D >> mg\sin\gamma$ and $|\gamma| << 1$. Combining Eqs. (5.186) and (5.187) yields

$$\mathrm{d}\psi = -\frac{L}{D}\sin\sigma\frac{\mathrm{d}V}{V} \tag{5.189}$$

For a constant $\frac{L}{D}\sin\sigma$, to solve the above equation it is formulated as

$$\int_{\psi_0}^{\psi} d\psi = -\frac{L}{D}\sin\sigma\int_{V_0}^{V}\frac{\mathrm{d}V}{V} \Rightarrow \tag{5.190}$$

$$\psi - \psi_i = -\frac{L}{D}\sin\sigma\int_{V_i}^{V}\frac{\mathrm{d}V}{V} \tag{5.191}$$

with ψ_i and V_i being the heading and velocity at the beginning of the lateral manoeuvre. Introducing the dimensionless velocity $\bar{V} = \frac{V}{V_c}$ for ease of writing, the solution to Eq. (5.191) is written as:

$$\psi - \psi_i = -\frac{L}{D}\sin\sigma\ln\frac{\bar{V}}{\bar{V}_i} = \frac{L}{D}\sin\sigma\ln\frac{\bar{V}_i}{\bar{V}} \tag{5.192}$$

For simplicity, ψ_i can be taken as zero, as initial condition for $\bar{V} = \bar{V}_i$.

From Fig. 5.44 the incremental longitudinal and lateral displacements are

$$\mathrm{d}R_{f_x} = \mathrm{d}s\cos\psi \tag{5.193}$$

$$\mathrm{d}R_{f_y} = \mathrm{d}s\sin\psi \tag{5.194}$$

To obtain the total displacements we need to integrate Eqs. (5.193) and (5.194), using Eq. (5.192). The latter equation depends, of course, on the flown trajectory.

One of the trajectories of interest is the equilibrium glide trajectory. Beginning with the equilibrium condition from Eq. (5.188), we write:

$$L\cos\sigma = mg\left(1-\frac{V^2}{V_c^2}\right) = \frac{L}{D}D\cos\sigma = -mV\frac{L}{D}\frac{\mathrm{d}V}{\mathrm{d}s}\cos\sigma \tag{5.195}$$

where we substituted Eq. (5.187). This can be rewritten as

$$\frac{L}{D}\cos\sigma\frac{V}{\frac{V^2}{V_c^2}-1}\mathrm{d}V = g\mathrm{d}s \tag{5.196}$$

and, with $\bar{V} = \frac{V}{V_c}$:

$$\frac{L}{D}\cos\sigma\frac{\bar{V}}{\bar{V}^2-1}\mathrm{d}\bar{V} = \frac{g}{V_c^2}\mathrm{d}s = \frac{\mathrm{d}s}{R} \approx \frac{\mathrm{d}s}{R_e} \tag{5.197}$$

Combining Eq. (5.197) with Eqs. (5.192) and (5.194) leads to

$$\frac{\mathrm{d}R_{f_y}}{R_e} = \frac{L}{D}\cos\sigma\frac{\bar{V}d\bar{V}}{\bar{V}^2-1}\sin\left(\frac{L}{D}\sin\sigma\ln\frac{\bar{V}_i}{\bar{V}}\right) \tag{5.198}$$

If the lateral deflection angle, ψ, is small ($\sin\psi \approx \psi$) and the bank manoeuvre is initiated at circular velocity ($\bar{V}_i = 1$), then Eq. (5.198) simplifies to

$$\frac{\mathrm{d}R_{f_y}}{R_e} = \left(\frac{L}{D}\right)^2\cos\sigma\sin\sigma\frac{\bar{V}\ln\bar{V}d\bar{V}}{1-\bar{V}^2} \tag{5.199}$$

which integrates to (for the limits $\bar{V}_i = 1$ to $\bar{V} = 0$)[7]

$$\frac{R_{f_y}}{R_e} = \frac{\pi^2}{24}\left(\frac{L}{D}\right)^2\cos\sigma\sin\sigma \tag{5.200}$$

[7] The integral

$$\int\frac{x\ln x}{1-x^2}\mathrm{d}x$$

has the solution

$$\frac{1}{2}\left[Li_2(1-x) - Li_2(-x) - \ln(x)\ln(x+1)\right]$$

Here, $Li_2(1-x)$ and $Li_2(-x)$ are polylogarithms (also known as Jonquière's function), a so-called *special function* defined as $Li_n(z) = \sum_{k=1}^{\infty}\frac{z^k}{k^n}$. Numerical values are $Li_2(-1) = -\frac{\pi^2}{12}$, $Li_2(0) = 0$ and $Li_2(1) = \frac{\pi^2}{6}$. So, the solution to the integral evaluated on the interval [1,0] yields $\frac{\pi^2}{24}$. More information on how to solve such an integral, as well as polylogarithms, can be found on http://integrals.wolfram.com.

This equation gives a maximum lateral range at $\sigma = 45°$.

However, in general the lateral deflection is not restricted to small angles, nor is the manoeuvre limited to be initiated at circular velocity. That means the integration of Eq. (5.198) is not so straightforward. Slye (1960) applies a Taylor-series expansion to evaluate the sine and cosine functions in the expressions for R_{f_x} and R_{f_y}. Only the first three terms of the series have to be retained for an accuracy of 1% up to angles of 90°. Let the first three terms of the Taylor-series expansion of $\sin x$ and $\cos x$ be given by

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots \tag{5.201}$$

$$\cos x \approx 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots \tag{5.202}$$

Further, we define $\xi = \frac{L}{D} \sin\sigma \ln \bar{V}$. The sine-term of Eq. (5.198) evaluates as follows:

$$\sin\left(\frac{L}{D} \sin\sigma \ln \frac{\bar{V}_i}{\bar{V}}\right) \Rightarrow$$

$$\sin\left[\frac{L}{D} \sin\sigma \left(\ln \bar{V}_i - \ln \bar{V}\right)\right] \Rightarrow$$

$$\sin\left(\xi_i - \xi\right)$$

which can be written as:

$$\sin\left(\xi_i - \xi\right) = \sin\xi_i \cos\xi - \sin\xi \cos\xi_i$$

The Taylor-series expansions substituted yields

$$\left(\xi_i - \frac{\xi_i^3}{3!} + \frac{\xi_i^5}{5!}\right)\left(1 - \frac{\xi^2}{2!} + \frac{\xi^4}{4!}\right) - \left(\xi - \frac{\xi^3}{3!} + \frac{\xi^5}{5!}\right)\left(1 - \frac{\xi_i^2}{2!} + \frac{\xi_i^4}{4!}\right)$$

and evaluating the multiplications while retaining only terms up to order 5 results in:

$$\xi_i - \xi_i \frac{\xi^2}{2!} + \xi_i \frac{\xi^4}{4!} - \frac{\xi_i^3}{3!} + \frac{\xi_i^3}{3!}\frac{\xi^2}{2!} + \frac{\xi_i^5}{5!}$$

for the first pair, and

$$-\xi + \xi\frac{\xi_i^2}{2!} - \xi\frac{\xi_i^4}{4!} + \frac{\xi^3}{3!} - \frac{\xi^3}{3!}\frac{\xi_i^2}{2!} - \frac{\xi^5}{5!}$$

for the second one. After combining all terms, Eq. (5.198) is then written as

$$\frac{\mathrm{d}R_{f_y}}{R_e} = \left(\frac{L}{D}\right)\cos\sigma\frac{\bar{V}d\bar{V}}{\bar{V}^2-1}\left[\left(\xi_i - \frac{\xi_i^3}{3!} + \frac{\xi_i^5}{5!}\right) + \left(-1 + \frac{\xi_i^2}{2!} - \frac{\xi_i^4}{4!}\right)\xi + \right.$$
$$\left. + \left(-\frac{\xi_i}{2!} + 2\frac{\xi_i^3}{4!}\right)\xi^2 + \left(\frac{1}{3!} - 2\frac{\xi_i^2}{5!}\right)\xi^3 + \frac{\xi_i}{4!}\xi^4 - \frac{1}{5!}\xi^5\right] \tag{5.203}$$

If we also define

$$\Phi_n = \int_{\bar{V}_i}^{\bar{V}} \frac{\eta\,(\ln\eta)^n}{1-\eta^2}d\eta \tag{5.204}$$

then

$$\frac{R_{f_y}}{R_e} = \frac{L}{D}\cos\sigma\left[\left(-\xi_i + \frac{\xi_i^3}{3!} - \frac{\xi_i^5}{5!}\right)\Phi_0 + \left(1 - \frac{\xi_i^2}{2!} + \frac{\xi_i^4}{4!}\right)\frac{L}{D}\sin\sigma\Phi_1 + \right.$$
$$+ \left(\frac{\xi_i}{2!} - 2\frac{\xi_i^3}{4!}\right)\left(\frac{L}{D}\right)^2\sin^2\sigma\Phi_2 + \left(-\frac{1}{3!} + 2\frac{\xi_i^2}{4!}\right)\left(\frac{L}{D}\right)^3\sin^3\sigma\Phi_3 +$$
$$\left. - \frac{\xi_i}{4!}\left(\frac{L}{D}\right)^4\sin^4\sigma\Phi_4 + \frac{1}{5!}\left(\frac{L}{D}\right)^5\sin^5\sigma\Phi_5\right] \tag{5.205}$$

For $\xi_i = 0$ (equivalent to $\bar{V}_i = 1$, i.e., circular entry velocity), this equation reduces to

$$\frac{R_{f_y}}{R_e} = \frac{L}{D}\cos\sigma\left[\frac{L}{D}\sin\sigma\Phi_1 - \frac{1}{3!}\left(\frac{L}{D}\right)^3\sin^3\sigma\Phi_3 + \frac{1}{5!}\left(\frac{L}{D}\right)^5\sin^5\sigma\Phi_5\right] \tag{5.206}$$

The expression for longitudinal range, derived from Eq. (5.193) along the same lines as Eq. (5.203), is:

$$\frac{R_{f_x}}{R_e} = \frac{L}{D}\cos\sigma\left[\left(-1 + \frac{\xi_i^2}{2!} - \frac{\xi_i^4}{4!}\right)\Phi_0 + \left(-\xi_i + \frac{\xi_i^3}{3!}\right)\frac{L}{D}\sin\sigma\Phi_1 + \right.$$
$$+ \left(\frac{1}{2!} - 6\frac{\xi_i^2}{4!}\right)\left(\frac{L}{D}\right)^2\sin^2\sigma\Phi_2 + \frac{\xi_i}{3!}\left(\frac{L}{D}\right)^3\sin^3\sigma\Phi_3 + \tag{5.207}$$
$$\left. - \frac{1}{4!}\left(\frac{L}{D}\right)^4\sin^4\sigma\Phi_4\right]$$

which, for $\xi_i = 0$, reduces to:

$$\frac{R_{f_x}}{R_e} = \frac{L}{D}\cos\sigma\left[-\Phi_0 + \frac{1}{2!}\left(\frac{L}{D}\right)^2\sin^2\sigma\Phi_2 - \frac{1}{4!}\left(\frac{L}{D}\right)^4\sin^4\sigma\Phi_4\right] \tag{5.208}$$

Integral equation (5.204) should be evaluated numerically. For $n = 0$, though, the integration is explicit and yields:

$$\Phi_0 = \int_{\bar{V}_i}^{\bar{V}} \frac{\eta}{1-\eta^2} d\eta = \left[-\frac{1}{2} \ln\left(1-\eta^2\right) \right]_{\bar{V}_i}^{\bar{V}} = \frac{1}{2} \ln \frac{1-\bar{V}_i^{\,2}}{1-\bar{V}^2} \tag{5.209}$$

With this result and by putting $\sin\sigma$ to zero, i.e., no out-of-plane motion, Eq. (5.207) gives the equilibrium-glide result for longitudinal range:

$$\frac{R_{f_x}}{R_e} = \frac{1}{2} \frac{L}{D} \ln \frac{1-\bar{V}^2}{1-\bar{V}_i^{\,2}} \tag{5.210}$$

Realising that $\bar{V} = \frac{V}{V_c}$, defining $V_E = V_i$ and $V_F = V$, and determining the range for a final velocity $V_F << V_E$, Eq. (5.210) simplifies to:

$$\frac{R_{f_x}}{R_e} = -\frac{1}{2} \frac{L}{D} \ln\left(1 - \frac{V_E^2}{V_c^2}\right) \tag{5.211}$$

which is exactly the one we derived earlier, see Eq. (5.128).

The above simple analytical methods appear to be adequate for studying small entry-angle trajectories for lift-to-drag ratios up to about 1.5. Numerical studies done by Slye (1960) indicated that neglecting the Earth's rotation in the calculations may introduce errors of as much as 10% in the resultant lateral range. Important conclusions of his were that the velocity at which the banked manoeuvre is initiated bears importantly on the resultant lateral displacement. Delay in initiation of the manoeuvre to a velocity of 0.8 V_c may result in a loss of lateral-range potential of 50%. The effect of γ_E on lateral range was found to be small in comparison with its strong effect on longitudinal range.

Example 5.9 (*Lateral and Longitudinal Range*) In this example we will study the effect of bank angle on lateral range, and will do so for both a non-rotating and a rotating Earth. In Fig. 5.45 this effect of (constant) bank angle on the lateral range is shown. The analytical curve in the figure has been calculated with Eq. (5.205) for $L/D = 1$, $V_E = V_c$ and $V_F = 0.1\ V_c$. The numerical solution has been obtained for an initial flight-path angle of $\gamma_E = -2°$ at an altitude of $h_E = 120\,\text{km}$.[8] The vehicle mass was (quite arbitrarily) set at $m = 1{,}000\,\text{kg}$, with a hypersonic drag coefficient of $C_D = 1.45$ ($S_{ref} = 12\,\text{m}^2$). It is noted, though, that the lateral range is quite insensitive to variation of $\frac{C_D S_{ref}}{m}$ (Slye 1960). In all cases, entry was parallel to the equator, in a flight from west to east ($\chi = 90°$).

[8] As mentioned before, for the maximum lateral range Slye (1960) found only a marginal dependency on γ_E; we selected $\gamma_E = -2°$, because values closer to 0° would yield large skips as we will show later.

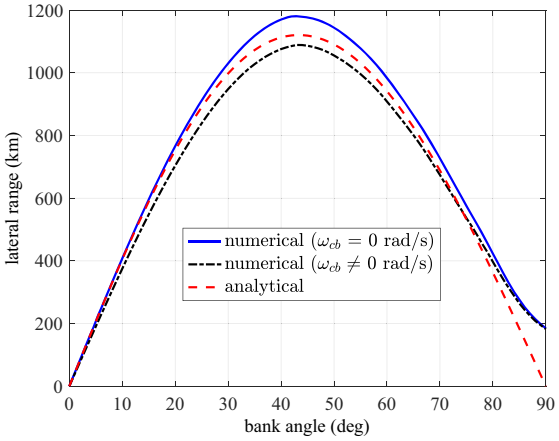


Fig. 5.45 Analytical lateral range as a function of bank angle ($L/D = 1$)

The results show that there is a good agreement between the analytical and numerical results, with a maximum difference of about 70 km (non-rotating Earth) for bank angles smaller than $\sigma \approx 85°$. This corresponds with the results of Slye (1960). The maximum lateral range occurs for a bank angle of $\sigma \approx 43°$. For bank angles close to $\sigma = 90°$ a relatively large deviation between analytical and numerical range can be seen ($\Delta R_{f_y} = 200$ km). As the analytical solution has been derived for an equilibrium glide, it is required to have some vertical lift component. When $\sigma = 90°$ the vertical lift is zero, so the analytical approximation no longer holds.

In case of a rotating Earth ($\omega_{cb} = 7.292115 \cdot 10^{-5}$ rad/s), the maximum (numerical) lateral range reduces from 1,181 km to 1,089 km, a difference of about 8%. The difference with the analytical solution has reduced, but be aware that this difference is also depending on the heading angle at entry. The corresponding curve has been generated with $V_E = 7{,}366.9$ m/s, which is equal to V_c minus the relative velocity with respect to the rotating Earth ($= \omega_{cb}(R_e + h_E)$).

To show the limitations of the analytical expression for longitudinal range we have done 4 (non-linear) simulations with different flight-path angles at entry, i.e., $\gamma_E = -2°$, $\gamma_E = -3°$, $\gamma_E = -4°$ and $\gamma_E = -5°$ (constant bank angle $\sigma = 0°$). The altitude profile as a function of flight time is shown in Fig. 5.46, and it is obvious that depending on the value of γ_E more or less skip will occur, greatly affecting the longitudinal range (see Table 5.6). Since in Eq. (5.207) there is no dependency on γ_E this means that the analytical expression will provide a constant longitudinal range ($R_{f,x} = 12{,}460$ km). This means that Eq. (5.207) is of little practical use.

In conclusion, Eq. (5.205) provides a reasonable first estimate of the maximum lateral range. It is noted, though, that this expression does not take any thermo-mechanical constraints into account, which may conflict with the constant bank-angle flight. To fly in the allowable entry corridor, bank-angle modulation is usually required (see Sect. 5.4.3). Therefore, for an accurate representation of the reachable footprint, one should resort to numerical methods as discussed by, e.g., Saraf et al. (2004). ■

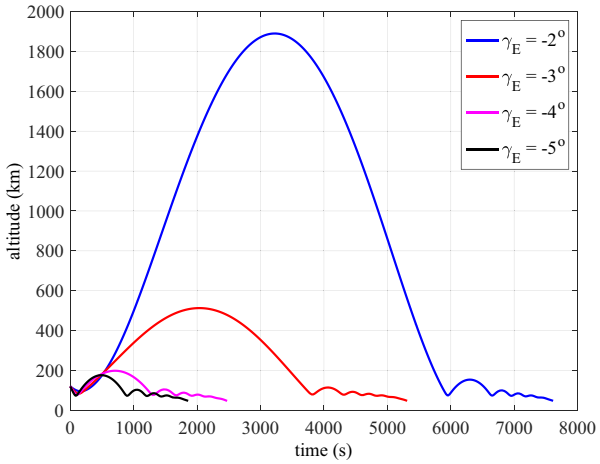


Fig. 5.46 Altitude profiles for different entry angles ($L/D = 1$)

Table 5.6 Longitudinal distance for different entry angles ($L/D = 1$)

γ_E (°)	$R_{f,x}$ (km)
−2	43,151
−3	33,302
−4	14,131
−5	9,798

To conclude the discussion on the footprint, in the Intermezzo below we will briefly summarise the equations for cross- and downrange that can be used for a non-linear, three-dimensional entry analysis.

*** Intermezzo 5.1 (Three-dimensional Crossrange and Downrange)**

The general geometry of the cross- and downrange in relation to the original and actual trajectory is shown in Fig. 5.47. In this figure, all depicted arcs are segments of so-called *great circles*, which are the largest possible circles on the surface of a (reference) sphere. The length of the arcs can be expressed in terms of radians, and the distance in meters between any two points on the sphere is obtained by simply multiplying the arc length in radians with the radius of the sphere.

The derivation of the equations for the downrange δ_d and the cross range τ_d of a Point $\mathbf{P}(\delta, \tau)$, given the initial Point $\mathbf{P_0}(\delta_\mathbf{0}, \tau_\mathbf{0})$ and the initial heading angle χ_0, involves the derivation of some auxiliary arcs and spherical angles, which are all depicted in Fig. 5.47. It is explicitly noted that the following equations are only valid for longitude and latitude values in the intervals $-180° \leq \tau < 180°$ and $-90° \leq \delta \leq 90°$, respectively. Other range definitions might require additional expressions to ensure that the correct spherical quadrants are used.

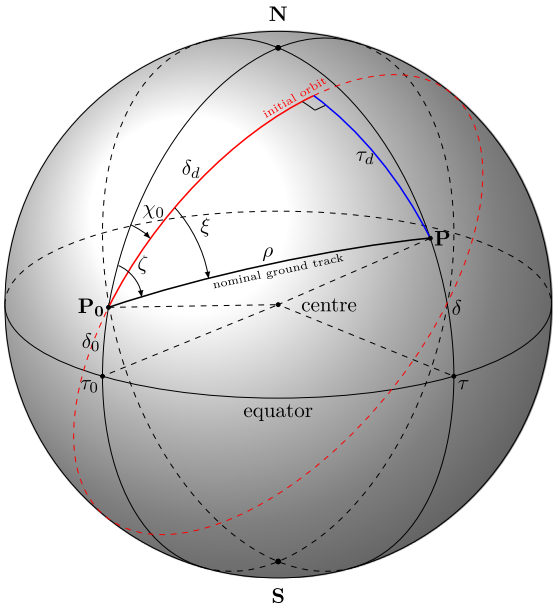


Fig. 5.47 Geometry for crossrange and downrange

The great circle arc ρ connecting $\mathbf{P_0}$ and $\mathbf{P}$, equal to the distance between the points on the surface of the sphere, is the angle between the vectors $\mathbf{p_0}$ and $\mathbf{p}$ pointing to $\mathbf{P_0}(\delta_0, \tau_0)$ and $\mathbf{P}(\delta, \tau)$, respectively. For the sake of the derivation, we assume the sphere to be a unit sphere, i.e., with radius 1.[9] The Cartesian equivalent of the spherical positions are

$$\mathbf{p_0} = \begin{pmatrix} \cos\delta_0 \cos\tau_0 \\ \cos\delta_0 \sin\tau_0 \\ \sin\delta_0 \end{pmatrix} \tag{5.212}$$

$$\mathbf{p} = \begin{pmatrix} \cos\delta \cos\tau \\ \cos\delta \sin\tau \\ \sin\delta \end{pmatrix} \tag{5.213}$$

The angle ρ between the unit vectors $\mathbf{p_0}$ and $\mathbf{p}$ follows from the dot product:

$$\begin{aligned} \cos\rho = \mathbf{p_0} \cdot \mathbf{p} &= \cos\delta_0 \cos\tau_0 \cos\delta \cos\tau + \cos\delta_0 \sin\tau_0 \cos\delta \sin\tau + \sin\delta_0 \sin\delta \\ &= \cos\delta_0 \cos\delta \left[\cos\tau_0 \cos\tau + \sin\tau_0 \sin\tau\right] + \sin\delta_0 \sin\delta \\ &= \cos\delta_0 \cos\delta \cos(\tau - \tau_0) + \sin\delta_0 \sin\delta \end{aligned} \tag{5.214}$$

[9] A unit sphere will lead to the same angular expressions as a sphere with a radius different from 1. However, in terms of distances along the surface of the sphere the metric distance on a unit sphere will be the same as the angular distance. For a non-unit sphere the scale factor between the two will be the planetary radius.

The angle ζ between the meridian passing through $\mathbf{P_0}$ and ρ can be found with the Law of Cosines for Sides, with ρ and $\frac{\pi}{2} - \delta$ being the known side lengths:

$$\cos\left(\frac{\pi}{2} - \delta\right) = \cos\rho\cos\left(\frac{\pi}{2} - \delta_0\right) + \sin\rho\sin\left(\frac{\pi}{2} - \delta_0\right)\cos\zeta \tag{5.215}$$

This gives, after some rearrangements,

$$\cos\zeta = \frac{\sin\delta - \cos\rho\sin\delta_0}{\sin\rho\cos\delta_0}. \tag{5.216}$$

The angle ξ between the initial great circle and the great circle passing through $\mathbf{P_0}$ and $\mathbf{P}$ is then

$$\xi = \zeta - \chi_0. \tag{5.217}$$

The cross range τ_d, the downrange δ_d and the auxiliary arc ρ form a spherical triangle with a right angle between τ_d and δ_d. Using the relations for right spherical triangles (see Appendix C, Sect. C.1), the final expressions for the cross- and downrange of $\mathbf{P}$, given $\mathbf{P_0}$ and χ_0 are:

$$\tau_d = \arcsin\left(\sin\rho\sin\xi\right) \tag{5.218}$$

$$\delta_d = \arccos\left(\frac{\cos\rho}{\cos\delta_d}\right) \tag{5.219}$$

Equations (5.218) and (5.219) correspond with the expressions given by Cavallo et al. (1996). Depending on the choice of the initial conditions and $\mathbf{P}$, the cross- and downrange can be used for navigational purposes: if $\mathbf{P_0}$ and χ_0 are fixed at a point of time during the flight, τ_d and δ_d represent the traverse and lateral deviations from the initial trajectory; if $\mathbf{P}$ is chosen to be the target point, τ_d and δ_d represent the traverse and lateral distance between the target and the lander at position $\mathbf{P_0}$, flying in the direction χ_0.

■

5.4.5 *Case Study: Horus Entry*

In this case study we will address the typical gliding trajectory of a winged re-entry vehicle. As we did in Example 5.8, we have selected the HORUS entry vehicle, since this vehicle has been fully documented in publicly available literature (MBB 1988). As mentioned earlier, in Appendix B a summary of vehicle and mission characteristics can be found.

The mission of HORUS starts in Earth orbit. A deorbit-burn manoeuvre targets the vehicle towards the atmospheric entry interface, here selected to be $h_E = 122\,\text{km}$.

The actual re-entry starts at this interface with the following initial conditions: V_E = 7435.5 m/s and $\gamma_E = -1.43°$. In the original study, the European launch site in Kourou (French Guiana) was going to be extended with facilities to allow for landing of space planes. Also for this case study we will assume the vehicle is targeted for a runway in Kourou (the initial flight heading is χ_E = 70.75°). Some 80 km from the landing site guidance switches to the Terminal Area Energy Management (TAEM) logic. This latter part of the mission will be discussed in Chap. 8. At all times HORUS is in a trimmed condition, i.e., there is pitch equilibrium, and any change in aerodynamic properties due to the relevant control-surface deflections is taken into account in the simulation.

Before we get to the actual mission of HORUS, we do a brief analysis on the sensitivity of the trajectory to the initial flight-path angle, γ_E. The nominal angle, $\gamma_E = -1.43°$, will be perturbed with $\pm 0.1°$ and $\pm 0.2°$ and we will look at the effect on the *h-V* diagram, the stagnation heat flux and the profile of the flight-path angle.[10] The simulations will be done with the simplified model of Sect. 5.2, but with the Coriolis acceleration included to have a more accurate representation of the non-linear model. The atmospheric density and speed of sound are based on the exponential atmosphere, which will lead to an underestimation of the heat flux. Stop condition for the simulations is a final altitude of h_F = 60 km; we will also inspect the corresponding flight time and velocity.

In Figs. 5.48, 5.49 and 5.50 the results are shown. In the *h-V* diagram it shows that a steeper entry will force the vehicle deeper into the atmosphere. What also becomes obvious is the need for bank-angle control. The simulations were executed with a constant $\alpha_E = 40°$ and $\sigma_E = 0°$. Without modulating the bank angle, i.e., controlling the vertical component of the lift, the vehicle enters into a skipping flight. Of course, because the entry velocity is sub-orbital, it will not skip out of the atmosphere. By diving deeper into the atmosphere the vehicle experiences a larger drag and thus its loss in kinetic energy will be larger, showing as a lower velocity at corresponding altitudes.

A steeper entry will lead to larger peak heat fluxes, clearly seen in Fig. 5.49. Even a difference of only $\Delta\gamma_E = 0.1°$ will already increase $q_{c,max}$ with 8%. But there is more to it than just an increased peak value: as can be seen, the frequency of the oscillations increases for steeper entries. That means that also the gradient of the flux increases, leading to much stronger transient effects on the structure. The thermal-protection system will be faced with a more rapid heating-up and cooling-down effect, which will lead to larger thermal stresses in the structure. The local steepness of the trajectory, as well as the oscillation frequency, is supported by Fig. 5.50 that shows the flight-path angle profile.

Finally, in Table 5.7 the flight time and final velocity are listed. Steeper entries will result in shorter flight times, as the energy dissipation is more rapid. Overall, due

[10] With an entry velocity of V_E = 7435.5 m/s, the vertical velocity component at $\gamma_E = -1.43°$ and $\gamma_E = -1.53°$ is $V_v = V_E \sin\gamma_E = -185.6$ m/s and -198.5 m/s, respectively. This means that to achieve 0.1° accuracy in flight-path angle, the vertical velocity has to be managed with 12.9 m/s (= 0.17%) accuracy.

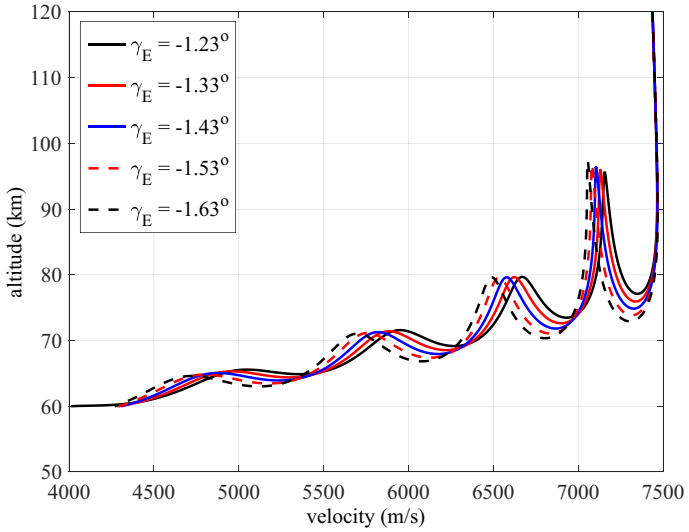


Fig. 5.48 Altitude-velocity diagram for varying entry angles ($h_F = 60\,\text{km}$)

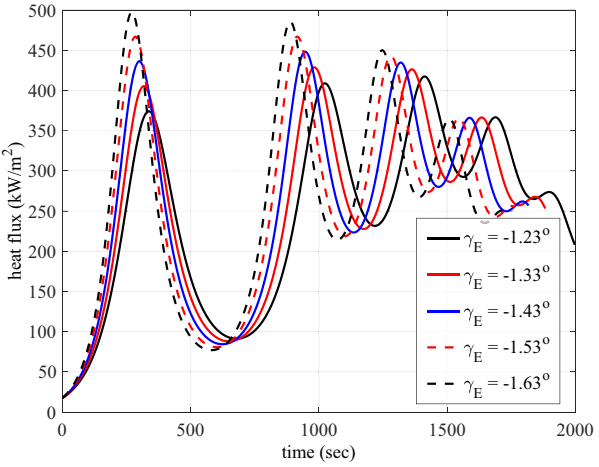


Fig. 5.49 Stagnation heat flux as a function of time for varying entry angles ($h_F = 60\,\text{km}$)

to the skipping motion, the flight time is larger than for a regular gliding flight. Due to the oscillations, there is no clear conclusion about V_F to be drawn, as it depends at what moment the vehicle crosses the 60-km altitude marker.

The nominal mission discussed by MBB (1988) was the result of a trajectory-optimisation activity and provides us with the commanded angle of attack and absolute bank angle. The absolute bank angle determines the magnitude of the vertical lift, and in an actual guidance system the so-called lateral guidance algorithm determines

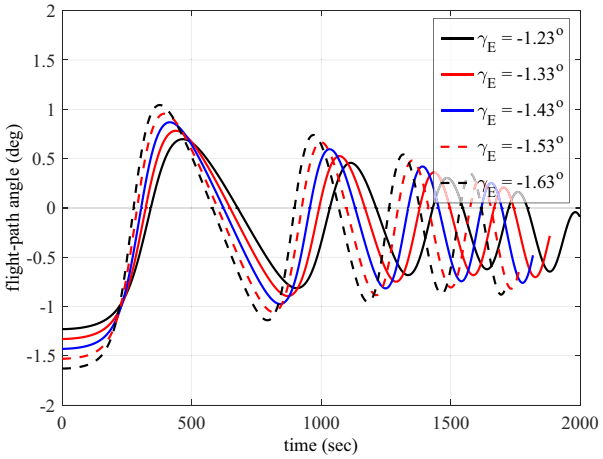


Fig. 5.50 Flight-path angle as a function of time for varying entry angles ($h_F = 60\,\text{km}$)

Table 5.7 Flight time and end velocity at $h = 60\,\text{km}$ for varying entry angles

γ_E [°]	t_F [s]	V_F [m/s]
−1.23	1,999	4,012
−1.33	1,883	4,285
−1.43	1,818	4,304
−1.53	1,764	4,292
−1.63	1,717	4,269

the sign of the bank angle based on the final target. This lateral guidance is discussed in Chap. 7, Sect. 7.6. In Fig. 5.51 the nominal guidance commands are shown. The initial angle of attack is set at α_E 40° (maximum-lift configuration) to minimise the maximum heat flux—in this study set at $q_c = 500\,\text{kW/m}^2$. The moment the heat-flux peak is reached and skipping flight may occur, bank-angle control is initiated (around $t \approx 260\,\text{s}$). An almost horizontal lift ($\sigma \approx 80°$) is required to prevent skipping flight. The bank angle is modulated throughout flight to remain inside the entry corridor.

Once HORUS is clear from the thermal constraints the angle of attack is lowered towards the maximum L/D configuration to extend the cross- and downrange, see also Fig. 5.52 for the change of L/D with time. It is clear that the maximum-lift configuration, necessary to lower the maximum heat flux as much as possiblem has an L/D of about one. After the peak heating has occurred the guidance system commands HORUS to a lower angle of attack, gradually increasing the L/D and thus the cross- and downrange capabilities. The maximum L/D is more than two at lower Mach numbers.

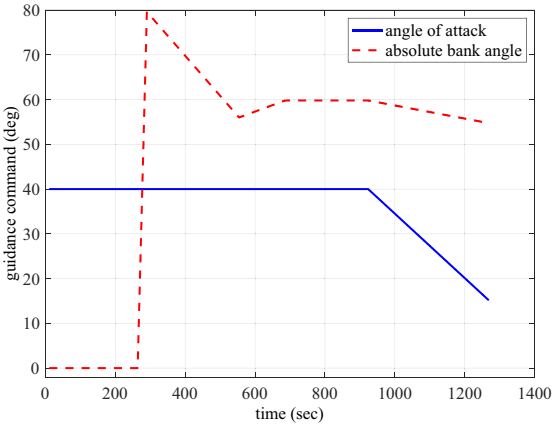


Fig. 5.51 HORUS' nominal guidance commands

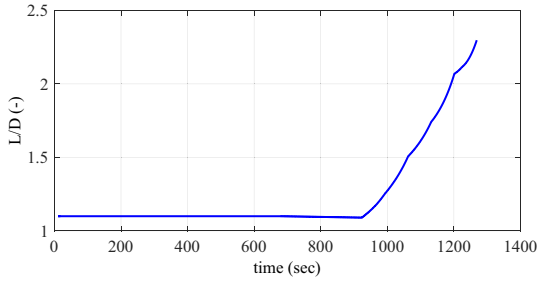


Fig. 5.52 HORUS' lift-to-drag ratio as a function of time

Figure 5.53 shows HORUS's trajectory in the entry corridor. This corridor has been established using the nominal angle-of-attack profile of Fig. 5.51, which means that C_D and C_L in the equilibrium-glide and maximum *g-load* constraints, Eqs. (5.174) and (5.182), have been calculated for this varying α. Around $V \approx 6{,}500\,\text{m/s}$ ($t \approx 260\,\text{s}$) the flight-path angle is approximately zero, and the large bank angle of $\sigma \approx 80°$ forces the vehicle towards the middle of the entry corridor. Due to the dynamics involved this shows as a low-frequency oscillation. For the remaining part of the flight the trajectory is well within the entry corridor, guaranteeing a safe nominal flight.

For the sake of completeness, in Fig. 5.54 we have again shown the h-V diagram and the guidance commands, but now with the locations of the bank reversals included as well as the actual bank angle rather than the absolute value. Out of the four reversals, three are executed at relative close distance to the landing site, because the closer the vehicle gets the faster the heading error will grow and initiate a reversal. Again, this will be discussed in more detail in Sect. 7.6.

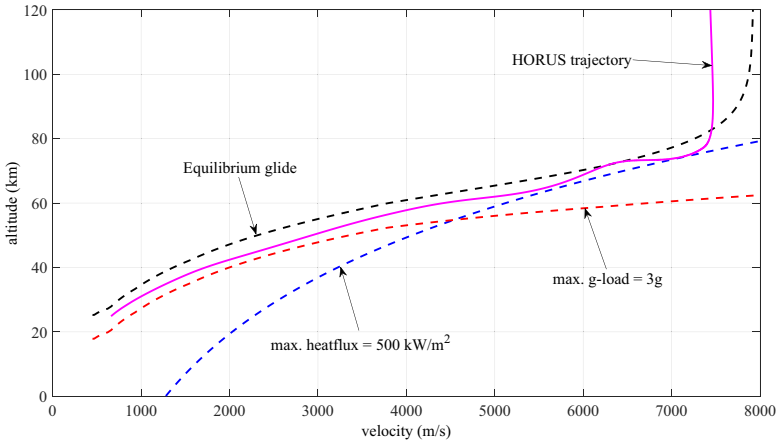


Fig. 5.53 HORUS's flight through the entry corridor

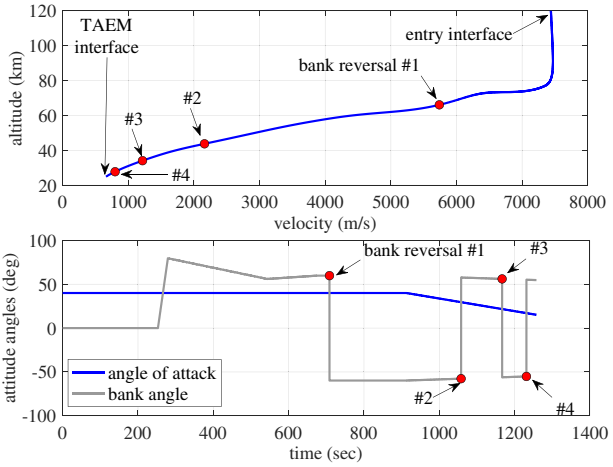


Fig. 5.54 Nominal control history (top) and altitude-velocity profile (bottom)

In Fig. 5.55 a three-dimensional rendering of the entry and hypersonic descent is shown. The final condition at about 80 km from Kourou is $h_F \approx 25$ km at a Mach number of $M \approx 2.5$.

To conclude this case study, in Fig. 5.56 the thermo-mechanical loads have been plotted. The peak heat flux stays just under the constraint value of $q_c = 500$ kW/m^2. The maximum occurring *g-load*, close to the end of the hypersonic descent, does not exceed 2 g. It is obvious that the maximum deceleration occurs well after the (first) peak in stagnation heat flux. This observation is somewhat obscured, because the guidance system employs a form of heat-flux tracking (i.e., staying close to the

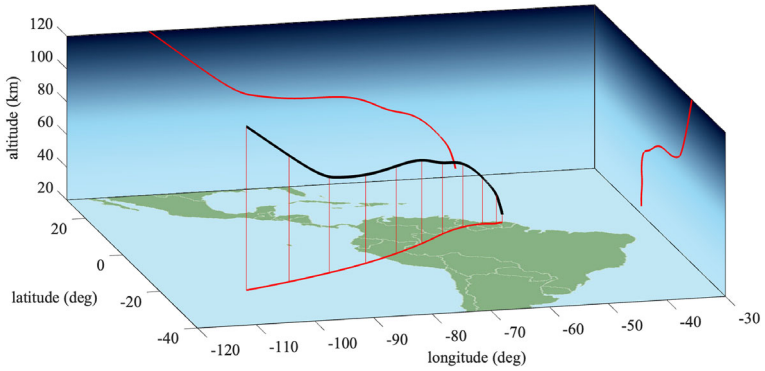


Fig. 5.55 HORUS's trajectory in three-dimensional space, with projections on the three planes

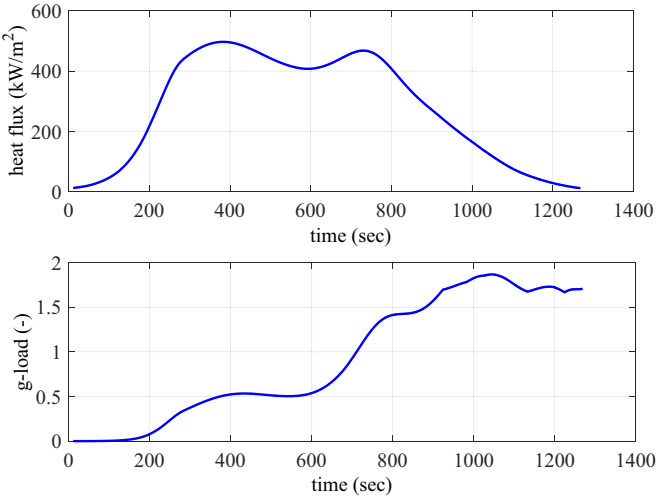


Fig. 5.56 HORUS's thermo-mechanical loads as a function of time

maximum heat-flux constraint to lower the integrated heat load). This shows as a second peak in the heat flux much later in the trajectory.

5.5 Skipping Entry

5.5.1 Introduction

The third and last entry "technique" that we will discuss is the skip entry, which involves one or more successive "skips" off the atmosphere, like a pebble thrown flat on the water. This technique has several purposes and advantages:

- In case of an Earth approach at super-circular speeds, i.e., parabolic in a return flight from the Moon or hyperbolic in case of an interplanetary approach, a skipping flight helps, before final entry, to dissipate the huge amount of heat that is usually generated on faster descents.
- As the skip trajectory can be thought of as the successive out-of-the-atmosphere portions of Kepler's planetary ellipse, each portion jointed to the next by a skipping phase during which the vehicle enters the atmosphere, negotiates a turn, and is then ejected from the atmosphere. This type of manoeuvre can thus be used to achieve a change in the orbital plane.
- The range modulation made possible by skip entry allows a spacecraft to reach a wider landing area, or to reach a designated landing point from a wider range of possible entry times, which is especially important in abort situations.

Skip entry requires a very precise guidance, because while in the atmosphere the vehicle should at all times remain within the entry corridor already discussed in Sect. 5.4.3. An overshoot trajectory at super-circular speed, in particular one larger than escape velocity, will result in a trajectory into deep space without any chance to return to Earth. Even at velocities lower than the escape velocity, being exposed to the Van Allen belts longer than strictly necessary can be hazardous to the astronauts on-board manned spacecraft. And an undershoot trajectory, on the other hand, will possibly lead to destruction of the vehicle, because of too large thermo-mechanical loads. No matter how important the guidance is, in this section we will only focus on the flight-mechanics aspects of the atmospheric portion of the flight.

To induce skipping flight is not very difficult. As we saw in the previous Sect. 5.4, gliding vehicles can enter into skipping flight when the vertical lift acting on the vehicle is larger than the corresponding value of the equilibrium-glide condition. For orbital entry velocities this usually results in an upward flight inside the atmosphere. However, the definition of skipping flight that we will adhere to is one where the vehicle is required to leave the atmosphere. Normally this can occur for an entry velocity larger than the local circular velocity V_c (in combination with "some" positive lift), or an abundant positive lift, also for $V < V_c$.

Skipping flight was first introduced and studied by Eugen Sänger in the 1930s (for more details on Sänger, see also Chap. 1). He was convinced that the way into space was by using a rocket plane rather than a rocket, and in 1933 he wrote a book called *Raketenflugtechnik* (or *Rocket Flight Technique*), which was published in Austria and originally served as his doctoral thesis. Although not many people believed in the concept of a rocket propelled, liquid-fuelled vehicle (called the *Silbervogel*

or *Silverbird*), it caught the attention of senior officials within the German High Command as a possibility to bomb the main land United States. They invited Sänger to open a research institute in Trauen for the specific purpose of developing a manned spaceplane (Rose 2008). When war was on its way, this project was reviewed and became known as the Sänger Amerika Stratospheric or Antipodal Bomber.

Towards the late 1930s major design changes were introduced by Sänger and Irene Bredt (his assistant and later wife), and the *Silverbird Bomber* became a design of a rocket-powered sub-orbital bomber aircraft. In World War II the Germans showed interest in the design of the *Silverbird Bomber*. It was going to be launched from a horizontal slide. and it would bring the bomber to a speed of 500 m/s in 11 s over 3 km. The *Silverbird* would subsequently climb to about 140 km altitude after which it would make a series of atmosphere skips and hereafter a phugoid motion with a large amplitude. The atmospheric skips and phugoid motion would substantially increase the travelled distance, such that the *Silverbird* was able to reach the United States from Germany, drop its bomb and continue onwards to its landing site. Some of the proposed missions would allow the spaceplane to return to its German base, other possible landing sites were at secure islands in the Pacific Ocean, probably occupied by the Japanese.

Design work continued until early 1942, when scale-size experimental engines were beginning to be tested, and work had started on building the first full-size rocket engine. However, due to the exorbitant cost of sustaining the war the project was cancelled for budgetary reasons. The flight profile of the *Silverbird Bomber* and the bomber itself can be seen in Figs. 5.57 and 5.58a, respectively.

In the United States during the 1950s, as the X-15 was taking shape, a parallel set of initiatives sought to define a follow-on hypersonic program that could actually achieve orbit. The plans were inspired in large measure by the *Silverbird*, which, in fact, greatly influenced postwar Soviet, American, and European thinking about hypersonics and long-range "antipodal" boost-skip-glide vehicles. In 1952 the Bell Aircraft Corporation proposed the BoMi, a long-range rocket-bomber designs for the United States Airforce. Initially intended to fly 3,500 mi, Bell officials gained funding from the Air Force's Wright Air Development Center (WADC) to study longer-range 4,000-mile and 6,000-mile systems under the umbrella of the top-secret Air Force project MX-2276 (Jenkins 2001; Rose 2008).

Different military programs emerged, such as RoBo and HYWARDS, in one way all variations on the *Silverbird* design. In 1957 they were consolidated in the Dyna-Soar (Fig. 5.58b), a concatenation of *Dynamic Ascent* and *Soaring Flight*. The Dyna-Soar, with the official government designation of X-20, went from manned space bomber, to a reconnaissance platform to a high-speed test vehicle. Even though based on the *Silverbird* vehicle and mission concept, the X-20 would not actually perform a skip entry, but would be boosted into space after which it would basically execute a gliding flight. Although it never reached operational status (it was cancelled in December 1963, before the first prototype was built), its technological legacy had a major impact on future re-entry vehicles, amongst others the Space Shuttle (the windtunnel programme alone covered 1,800 hr of subsonic testing, 3,700 hr at supersonic and 8,500 hr at hypersonic speeds). The X-20 was way ahead of its

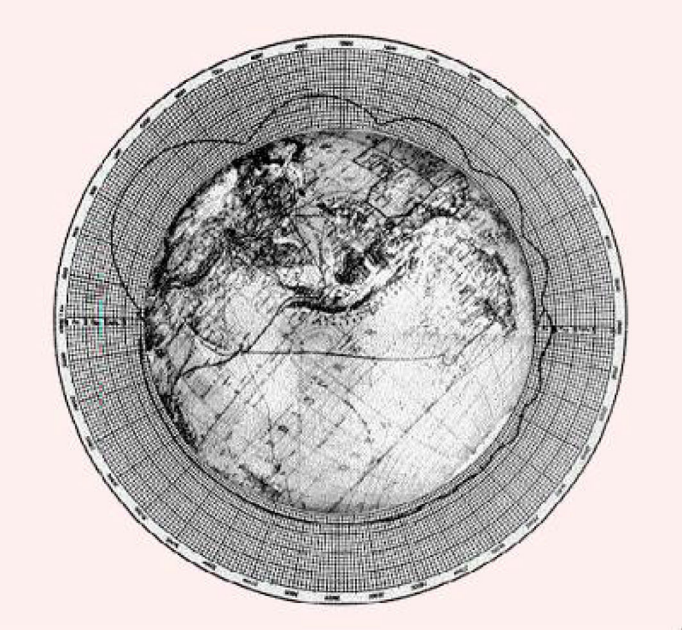
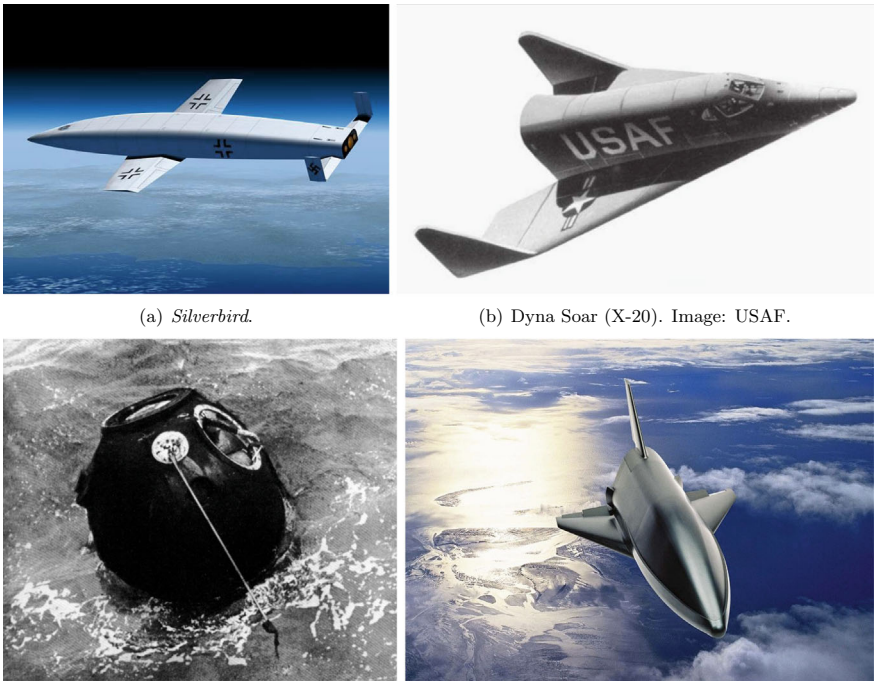


Fig. 5.57 The flight profile of the *Silverbird Bomber*. Adapted from Sänger and Bredt (1944)

time, and some of the developed solutions, e.g., the thermal protection system of a combination of the high-temperature alloy René 41, the silica-fibre insulation called Dyna-Quartz and skin panels of D-36 Columbium, would probably have served the Space-Shuttle better than the chosen solution of ceramic tiles (Sunday and London III 1990).

Unlike the American lunar program, which used common hardware for lunar landing and lunar orbital missions, the Soviet Union chose to develop a separate spacecraft with its own booster for these two missions (Harvey 2007). The Zond spacecraft, also known as Soyuz 7K-L1, was designed to sling a single cosmonaut around the Moon and back in a non-orbital loop. Despite the problem-ridden mission around the Moon, the unmanned Zond-5, with on-board amongst others two small turtles, managed to get into an inbound entry trajectory on September 21, 1968. The plan was to recover Zond on Soviet territory after a skip trajectory. However, due to a gyro failure a skip entry was no longer possible, and the vehicle entered into a ballistic trajectory directing it to a landing in the Indian Ocean (Fig. 5.58c). Despite the high temperatures of up to 13,000° C and a high *g-load* of 16 g, Zond-5 was recovered safely, and even the two turtles had survived.

On November 10, 1968, Zond-6 was launched. After successfully rounding the Moon it re-entered over the Indian Ocean and dived to 45 km. To generate more lift, the guidance system commanded to tilt the heat shield at 90° to the flight path and Zond-6 was thus pushed back into space. This was the first successful skipping



(a) *Silverbird*. (b) Dyna Soar (X-20). Image: USAF.

(c) Zond-5 after landing (Harvey, 2007). (d) Sub-orbital Hopper. Image: ESA.

Fig. 5.58 Examples of skip-entry vehicles

flight ever, in the process reducing its velocity from 11 km/s down to 7.6 km/s. After several minutes it began its second re-entry. Unfortunately, several things went wrong at that stage, amongst others the parachute system did not initiate, so in the end the vehicle crashed into hard soil. Zond-7, however, launched on August 7, 1969, had a successful mission including the entry one week later. It entered the atmosphere over the Indian Ocean in a skip trajectory, and after the second entry it achieved a soft landing in a region south of present-day Kustanai. And, to conclude, Zond-8 (launched October 20, 1970) used a new variant of the double-dip re-entry. Six days after launch, it came in over the north pole, bounced off the atmosphere while being tracked by Soviet radar stations as it soared south over the Soviet Union. It then made a final precision re-entry followed by splashdown at the recovery point in the Indian Ocean.

To put the achievement of a successful skip entry in perspective, the Apollo Command Module, when returning from the moon, was capable of a one-skip entry. However, this capability was apparently never used in-flight, as the maximum flight range achieved by a manned Apollo mission was "only" 3,400 km (Orloff 2000). The Apollo-derived Orion spacecraft crew module, part of the cancelled Constellation Programme to bring the Americans back to the Moon, was planned to be capable of skip entry, to allow for targeting the landing site from a greater variety of abort

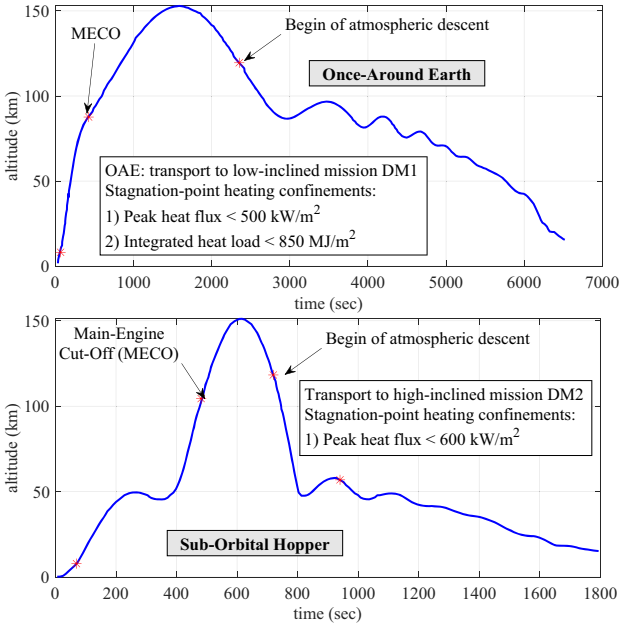


Fig. 5.59 Mission profiles of FESTIP's Concept 15: Once-Around Earth (top) and Sub-Oribital Hopper (bottom). Data taken from FESTIP (1998)

trajectories (Hoelscher 2007). So, to date, the Zond spacecraft have been the only ones to do this complex entry!

The Future European Space Transportation Investigation Programme (FESTIP) was an ESA program that aimed to define the (reusable) follow-up of Ariane-5. In 1998, out of 16 concepts the Sub-Orbital Hopper (Fig. 5.58d) was chosen to be the most cost-effective and most efficient concept of a reusable space transportation system. However, after the first glide test of the scale model Phoenix in May 2004, the project was eventually cancelled. The Hopper came with two mission variations, i.e., the actual Sub-Orbital Hopper (SOH) and the so-called Once-Around Earth (OAE), although they both had in common that they were to be launched horizontally on a magnetic skid sled running for 4 km in Kourou, French Guiana (note the similarity with the *Silverbird* design). After payload delivery above the atmosphere, the SOH would immediately return to Earth by performing a (minor) skipping entry. The duration was less than half an hour from the horizontal take-off until return at a designated (downrange) landing site—see Fig. 5.59. The OAE, on the other hand, would glide once around Earth back to the launch site and land there. It is noted that the SOH follows a much steeper entry path, which causes large thermal gradients and poses a design challenge for the thermal-protection system.

From the above, brief description it has become clear that many vehicles, which are claimed to have followed a skip trajectory, actually have not done that. At most they flew a trajectory that will periodically show an increasing altitude with time,

as we also saw in Sect. 5.4.5. Actually leaving the atmosphere again with a lower velocity as they had at first entry and cooling down by radiating heat, is something that has only been achieved by a handful of Russian capsules, notably Zond-5 to Zond-8. That the skipping entry is an important entry mechanism, both for reducing the thermo-mechanical loads and increasing the flight range is clear, so we will discuss several aspects of this entry type in the next section.

5.5.2 *Flight Mechanics*

For velocities V larger than the circular velocity, V_c, the equilibrium flight can only be executed with a negative lift L. This case is not treated further here, but will be discussed in the case study of this section (Sect. 5.5.5). Another option of deceleration in the atmosphere, applicable for entry velocities both greater and smaller than the circular velocity, is the skip trajectory. The vehicle dives into the atmosphere and is deflected back out of the atmosphere. As a result of the acquired deceleration, the vehicle exits the atmosphere with a velocity lower than the entry velocity. Outside the atmosphere, the vehicle describes a ballistic trajectory until it re-enters the atmosphere and so forth. The deceleration is thus established in a pendulous motion, respectively in and outside the (assumed bounded) atmosphere (Fig. 5.1). The skip trajectory is feasible not only for winged vehicles, but for any vehicle that can generate sufficient (positive) lift.

Starting point for the discussion about several characteristics of skipping flight is again given by the set of simplified equations derived in Sect. 5.2:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{5.220}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.221}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma \tag{5.222}$$

A simple approximation of the motion in the sip trajectory can be obtained in the case where the aerodynamic forces L and D are assumed to be large in relation to the mass-terms in the equations of motion (5.220) and (5.221). In that case, with $D >> W$ and $L >> W$, it follows that

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D \tag{5.223}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L \tag{5.224}$$

Dividing the above equations directly yields:

$$\frac{1}{V}\frac{\mathrm{d}V}{\mathrm{d}\gamma} = -\frac{D}{L} \tag{5.225}$$

Assuming L/D to be constant, we find for initial conditions V_E and γ_E

$$\int_{V_E}^{V} \frac{1}{V^*} dV^* = -\frac{D}{L}\int_{\gamma_E}^{\gamma} d\gamma^* \Rightarrow$$

$$\ln\frac{V}{V_E} = -\frac{D}{L}(\gamma - \gamma_E)$$

so

$$\frac{V}{V_E} = e^{-\frac{\gamma-\gamma_E}{L/D}} \tag{5.226}$$

According to Eq. (5.224) $\frac{\mathrm{d}\gamma}{\mathrm{d}t} > 0$ and therefore γ becomes less negative, or $\gamma > \gamma_E$. With $\Delta\gamma = \gamma - \gamma_E > 0$ substituted into Eq. (5.226) it follows that $V < V_E$ and the vehicle is thus decelerated.

Rewriting Eq. (5.224),

$$V\frac{\mathrm{d}\gamma}{\mathrm{d}t} = V\frac{\mathrm{d}\gamma}{\mathrm{d}p}\frac{\mathrm{d}p}{\mathrm{d}h}\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{L}{m} \tag{5.227}$$

gives us the starting point to look for the course of the altitude during the skip trajectory. From the hydrostatic equation, Eq. (5.26), we also have

$$\frac{\mathrm{d}p}{\mathrm{d}h} = -\rho g \tag{5.228}$$

and the kinematic equation gives

$$\frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma \tag{5.229}$$

These two derivatives substituted into Eq. (5.227) yields

$$V\frac{\mathrm{d}\gamma}{\mathrm{d}p} = -\frac{L}{m}\frac{1}{\rho g V\sin\gamma} \tag{5.230}$$

or, with $L = C_L\frac{1}{2}\rho V^2 S$:

$$V\frac{\mathrm{d}\gamma}{\mathrm{d}p} = -\frac{C_L S V}{2W\sin\gamma} \Rightarrow$$

$$\sin\gamma \mathrm{d}\gamma = -\frac{1}{2}\frac{1}{\frac{W/S}{C_L}}\mathrm{d}p \tag{5.231}$$

For a constant value of $\frac{W/S}{C_L}$, integration gives:

$$\cos\gamma - \cos\gamma_E = \frac{1}{2\frac{W/S}{C_L}}(p - p_E) \tag{5.232}$$

No assumptions with respect to the atmosphere have been made as yet. For an assumed exponential atmosphere, it follows with

$$\frac{p}{\rho} = RT = \frac{g}{\beta} \tag{5.233}$$

that

$$\cos\gamma - \cos\gamma_E = \frac{g}{2\beta}\frac{1}{\frac{W/S}{C_L}}(\rho - \rho_E) \tag{5.234}$$

It is generally assumed that at the beginning of the skip trajectory at high altitude (i.e., the atmospheric interface) $p_E = 0$ and $\rho_E = 0$. Thus

$$\cos\gamma - \cos\gamma_E = \frac{g}{2\beta}\frac{1}{\frac{W/S}{C_L}}\rho \tag{5.235}$$

According to Eqs. (5.232) and (5.235) the change of the flight-path angle with pressure or density, thus altitude, is independent of the entry velocity V_E.

At the lowest point p of the skip we know that $\gamma_p = 0$. Substituted into Eq. (5.235) this leads to

$$1 - \cos\gamma_E = \frac{g}{2\beta}\frac{C_L}{W/S}\rho_p = \frac{g}{2\beta}\frac{C_L}{W/S}\rho_0 e^{-h_p\beta} \tag{5.236}$$

With

$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha = 1 - 2\sin^2\alpha \tag{5.237}$$

Equation (5.236) is rewritten as

$$\rho_p = \rho_0 e^{-h_p\beta} = \frac{4\beta}{g}\frac{W/S}{C_L}\sin^2\frac{\gamma_E}{2} \tag{5.238}$$

or

$$h_p = -\frac{1}{\beta}\ln\left(\frac{4\beta}{g\rho_0}\frac{W/S}{C_L}\sin^2\frac{\gamma_E}{2}\right) \tag{5.239}$$

For the velocity, with $\gamma_p = 0$ substituted into Eq. (5.226), it easily follows that:

$$\frac{V_p}{V_E} = e^{\frac{\gamma_E}{L/D}} \tag{5.240}$$

To prevent the vehicle from hitting the ground, the condition $\rho_p < \rho_0$ should be satisfied, or, with Eq. (5.238):

$$\frac{W/S}{C_L} < \frac{g\rho_0}{4\beta} \frac{1}{\sin^2 \frac{\gamma_E}{2}} \cong \frac{21,600}{\sin^2 \frac{\gamma_E}{2}} \frac{\text{N}}{\text{m}^2}$$

which will usually be met in practical situations.

For $\rho_p > \rho > \rho_E$, Eq. (5.234) is satisfied by $+\gamma$ and $-\gamma$ in the ascending and descending segments of the trajectory, respectively. Last but not least, when we look at the end condition for the skip, we know that $\rho_F = \rho_E = 0$. This means

$$\cos\gamma_F - \cos\gamma_E = \frac{g}{2\beta} \frac{C_L}{W/S} (\rho_F - \rho_E) = 0 \tag{5.241}$$

thus $\gamma_F = -\gamma_E$ (as the vehicle is moving upwards). The atmosphere is thus left under a climb angle, in magnitude equal to the descent angle on atmospheric entry. For V_F it holds:

$$\frac{V_F}{V_E} = e^{\frac{2\gamma_E}{L/D}} \tag{5.242}$$

From the characteristics of the (ellipse-shaped) ballistic trajectory that follows on leaving the atmosphere we find that for the subsequent skip trajectory again: $\gamma_E = -\gamma_F$ and $V_E = V_F$. The second skip trajectory will thus have a lower entry velocity, but the same entry angle as the first skip trajectory. It now follows from Eq. (5.239) that the altitude of the lowest point is identical for both skip trajectories (as it is independent of V_E). In successive skip trajectories the altitude above the surface of the Earth will thus not decrease. Of course, once the velocity has decreased below a certain limit, skipping flight can no longer be sustained, and the vehicle will follow a glide path towards the surface.

Generalised, for each subsequent skip it holds that $\gamma_{E,i+1} = -\gamma_{F,i}$ and $V_{E,i+1} = V_{F,i}$. For n skip trajectories it is thus valid that

$$\frac{(V_F)_n}{(V_E)_1} = e^{n\frac{2\gamma_E}{L/D}} = \left\{\frac{(V_F)_1}{(V_E)_1}\right\}^n \tag{5.243}$$

Example 5.10 (*Skip characteristics*) For a vehicle with $L/D = 2$ entering the atmosphere with $\gamma_E = -10°$ it follows from Eq. (5.242) that $V_F = 0.840\ V_E$, which denotes an important reduction in velocity. If it is assumed that for the vehicle $W/SC_L = 6{,}000\,\text{N/m}^2$, then with Eq. (5.238) $\rho_p = 2.582 \cdot 10^{-3}\,\text{kg/m}^3$ and thus with Eq. (5.239) $h_p = 44.37\,\text{km}$. For this lowest point, $V_p = 0.916\ V_E$ (Eq. (5.240)).

In Table 5.8 some characteristic values for skipping flights with different entry angles are listed. Also the variation with L/D is shown. To point out some of the obvious conclusions: (i) the velocity decrease remains the same when both γ_E and L/D are increased with the same factor, (ii) the altitude for the lowest point in the trajectory is independent of L/D, and (iii) the lower the L/D the larger the velocity reduction. Summarised, in Figure 5.60 the influence of γ_E and L/D on the exit velocity has been visualised. For the range of the variation, the effect of γ_E is almost linear, whereas the L/D variation shows much more as a parabola. However, it is noted that in both cases the effect is, of course, a logarithmic one (Fig. 5.60).

For a (parabolic) return flight from the Moon, $V_E = 11.2$ km/s. A single skip may not significantly reduce the velocity, unless the trajectory is very steep. But in that case, the loads during the first skip may be too large. Thus, it is better to do more skips and reduce the velocity more gradually. Doing three successive skips (so $n = 3$) the velocity at the end of the three skips is $V_F = 6.638$ km/s $= 0.84\ V_c$. This means that for the actual entry the thermo-mechanical loads will be drastically reduced as compared to an entry with $V_E = 11.2$ km/s.

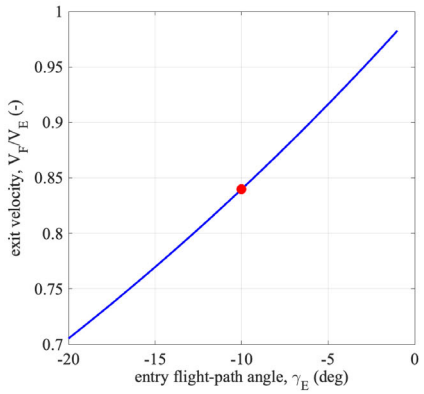
Note: Zond-6 reduced its velocity in a single skip from 11 km/s down to 7.6 km/s, a reduction of 31%. Assuming an L/D of about 0.3, similar to that of Apollo, this means that Zond-6 would have entered with $\gamma_E \approx -3.2°$. ■

5.5.2.1 The Horizontal Distance Travelled

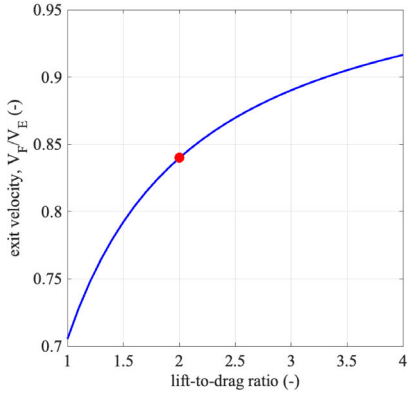
For the distance travelled in horizontal sense from entry up to a specific point, it follows with $mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L$:

Table 5.8 Characteristic values for the single-skip re-entry trajectory ($W/SC_L = 6{,}000\,\mathrm{N/m^2}$, $H_s = 7{,}200$ m)

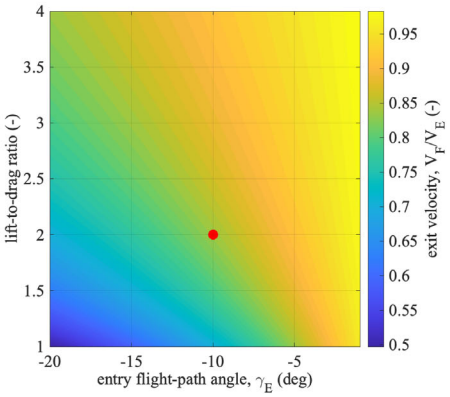
γ_E (°)	$\frac{L}{D}$ (–)	$\frac{V_p}{V_E}$ (–)	h_p (km)	$\frac{V_F}{V_E}$ (–)
−7.5	0.5	0.77	48.5	0.59
−15.0	0.5	0.59	38.6	0.35
−30.0	0.5	0.35	28.7	0.12
−7.5	1.0	0.88	48.5	0.77
−15.0	1.0	0.77	38.6	0.59
−30.0	1.0	0.59	28.7	0.35
−7.5	2.0	0.94	48.5	0.88
−15.0	2.0	0.88	38.6	0.77
−30.0	2.0	0.77	28.7	0.59



(a) $L/D = 2$



(b) $\gamma_E = -10°$



(c) Combined variation of L/D and γ_E

Fig. 5.60 Effect of entry flight-path angle and lif-to-drag ratio on the exit velocity. The red dot indicates the exit velocity for $L/D = 2$ and $\gamma_E = -10°$

$$X = \int_{t_E}^{t} V \cos\gamma \mathrm{d}t = \int_{\gamma_E}^{\gamma} \frac{V \cos\gamma}{\mathrm{d}\gamma/\mathrm{d}t}\mathrm{d}\gamma = \int_{\gamma_E}^{\gamma} \frac{V^2 \cos\gamma}{L/m}\mathrm{d}\gamma \Rightarrow \tag{5.244}$$

$$X = \int_{\gamma_E}^{\gamma} 2\frac{W}{SC_L g\rho}\cos\gamma \mathrm{d}\gamma \tag{5.245}$$

From Eq. (5.234) we can write for ρ

$$\rho = \rho_E + (\cos\gamma - \cos\gamma_E)\,\frac{2\beta}{g}\,\frac{W/S}{C_L} \tag{5.246}$$

and substituted into Eq. (5.245) yields:

$$X = \frac{1}{\beta}\int_{\gamma_E}^{\gamma} \frac{1}{\left(\cos\gamma - \cos\gamma_E + \frac{\rho_E g}{2\beta}\frac{C_L}{W/S}\right)} \cos\gamma \mathrm{d}\gamma \tag{5.247}$$

Note that for reasons that will become clear shortly, we cannot put $\rho_E = 0$.

To solve X as a function of γ, we introduce an effective entry angle, γ'_E in Eq. (5.247), as defined by:

$$\cos\gamma'_E = \cos\gamma_E - \frac{\rho_E g}{2\beta}\frac{C_L}{W/S} \tag{5.248}$$

which gives us to solve:

$$X = \frac{1}{\beta}\int_{\gamma_E}^{\gamma} \frac{\cos\gamma \mathrm{d}\gamma}{\cos\gamma - \cos\gamma'_E} \tag{5.249}$$

Solution of the above integral is a rather tedious process, details of which can be found in the Intermezzo below. Here, we will just highlight the major steps. For constant $\frac{W/S}{C_L}$, Eq. (5.249) can be rewritten as:

$$X = \frac{1}{\beta}\left(\gamma - \gamma_E + \cos\gamma'_E \int_{\gamma_E}^{\gamma} \frac{\mathrm{d}\gamma}{\cos\gamma - \cos\gamma'_E}\right) \tag{5.250}$$

The integral in this expression can be solved by introducing a new variable $t = \tan\left(\frac{\gamma}{2}\right)$, for which $\mathrm{d}\gamma = 2\frac{\mathrm{d}t}{1+t^2}$. The obtained result is:

$$X = \frac{1}{\beta}\left[\gamma - \gamma_E + \cot\gamma'_E \ln\left(\frac{\tan\frac{\gamma'_E}{2} + \tan\frac{\gamma}{2}}{\tan\frac{\gamma'_E}{2} - \tan\frac{\gamma}{2}} \cdot \frac{\tan\frac{\gamma'_E}{2} - \tan\frac{\gamma_E}{2}}{\tan\frac{\gamma'_E}{2} + \tan\frac{\gamma_E}{2}}\right)\right] \tag{5.251}$$

Since Eq. (5.234) relates the altitude with the flight-path angle, the trajectory is thus determined by Eq. (5.251): $X = X(h)$. For the distance travelled between the beginning and the end of a single-skip trajectory, it follows with $\gamma = \gamma_F = -\gamma_E$:

$$X_f = -\frac{1}{\beta}\left[2\gamma_E + 2\cot\gamma'_E \ln\frac{\tan\frac{\gamma'_E}{2} + \tan\frac{\gamma_E}{2}}{\tan\frac{\gamma'_E}{2} - \tan\frac{\gamma_E}{2}}\right] \tag{5.252}$$

From this expression, it appears that in the derivation above one may *not* set $\rho_E = 0$, because in that case $\gamma'_E = \gamma_E$ and Eq. (5.252) would yield an infinitely large value for X_f. Equation (5.252) can be simplified further by deriving an approximating relation between $\tan\left(\frac{\gamma'_E}{2}\right)$ and $\tan\left(\frac{\gamma_E}{2}\right)$ from Eq. (5.248). To do so, the latter equation

is rewritten with Eq. (5.238) as:

$$\cos\gamma_E' = \cos\gamma_E - \frac{\rho_E}{\rho_p}\left(1-\cos\gamma_E\right) \tag{5.253}$$

With this expression it follows that

$$\tan\frac{\gamma_E'}{2} = \sqrt{\frac{1-\cos\gamma_E'}{1+\cos\gamma_E'}} = \sqrt{\frac{1-\cos\gamma_E+\frac{\rho_E}{\rho_p}\left(1-\cos\gamma_E\right)}{1+\cos\gamma_E-\frac{\rho_E}{\rho_p}\left(1-\cos\gamma_E\right)}} \Rightarrow \tag{5.254}$$

$$\tan\frac{\gamma_E'}{2} = \sqrt{\frac{1+\frac{\rho_E}{\rho_p}}{\frac{1+\cos\gamma_E}{1-\cos\gamma_E}-\frac{\rho_E}{\rho_p}}} = \tan\frac{\gamma_E}{2}\sqrt{\frac{1+\frac{\rho_E}{\rho_p}}{1-\frac{\rho_E}{\rho_p}\left(\tan\frac{\gamma_E}{2}\right)^2}} \tag{5.255}$$

With $\frac{\rho_E}{\rho_p} << 1$ we can approximate for small and medium entry angles γ_E

$$\sqrt{\frac{1+\frac{\rho_E}{\rho_p}}{1-\frac{\rho_E}{\rho_p}\left(\tan\frac{\gamma_E}{2}\right)^2}} \approx \sqrt{1+\frac{\rho_E}{\rho_p}} \approx 1+\frac{1}{2}\frac{\rho_E}{\rho_p} \tag{5.256}$$

with the last term on the right-hand side being the first-order Taylor term. Thus,

$$\tan\frac{\gamma_E'}{2} = \tan\frac{\gamma_E}{2}\left(1+\frac{1}{2}\frac{\rho_E}{\rho_p}\right) \tag{5.257}$$

Substitution of Eq. (5.257) into Eq. (5.252) yields, with the approximation $\cot\gamma_E' \approx \cot\gamma_E$:

$$X_f = -\frac{1}{\beta}\left[2\gamma_E + 2\cot\gamma_E \ln\left(1+4\frac{\rho_p}{\rho_E}\right)\right] \tag{5.258}$$

Obviously, the term $4\frac{\rho_p}{\rho_E} >> 1$, so

$$\ln\left(1+4\frac{\rho_p}{\rho_E}\right) \approx \ln 4\frac{\rho_p}{\rho_E} = -\beta\left(h_p-h_E\right) + \ln 4 \approx -\beta\left(h_p-h_E\right) \tag{5.259}$$

with which Eq. (5.258) becomes:

$$X_f = -2\left(h_E-h_p\right)\cot\gamma_E + \frac{2}{\beta}\left(-\gamma_E+\ln 2\right) \tag{5.260}$$

The second term on the right-hand side indicates the increase in X_f with respect to a rectilinear descending and ascending flight, towards and from the lowest point of the trajectory, respectively.

*** Intermezzo 5.2 (Horizontal distance during skip)**
Equation (5.249) poses an integral that at first sight is not so straightforward to solve. To be complete, we will present the solution here, although for a general understanding of the derivation of the skip distance, the path to this solution is not really essential.

Starting point for the derivation, Eq. (5.249), is repeated here for convenience:

$$X = \frac{1}{\beta}\int_{\gamma_E}^{\gamma} \frac{\cos\gamma \mathrm{d}\gamma}{\cos\gamma - \cos\gamma'_E} \Rightarrow \tag{5.261}$$

$$X = \frac{1}{\beta}\int_{\gamma_E}^{\gamma} \frac{\cos\gamma - \cos\gamma'_E + \cos\gamma'_E}{\cos\gamma - \cos\gamma'_E}\mathrm{d}\gamma \Rightarrow \tag{5.262}$$

$$X = \frac{1}{\beta}\int_{\gamma_E}^{\gamma} \left(1 + \frac{\cos\gamma'_E}{\cos\gamma - \cos\gamma'_E}\right)\mathrm{d}\gamma \Rightarrow \tag{5.263}$$

$$X = \frac{1}{\beta}\left(\gamma - \gamma_E + \cos\gamma'_E \int_{\gamma_E}^{\gamma} \frac{\mathrm{d}\gamma}{\cos\gamma - \cos\gamma'_E}\right) \Rightarrow \tag{5.264}$$

To solve

$$\int_{\gamma_E}^{\gamma} \frac{\mathrm{d}\gamma}{\cos\gamma - \cos\gamma'_E} \tag{5.265}$$

we introduce a new variable $t = \tan\left(\frac{\gamma}{2}\right)$, for which $\mathrm{d}\gamma = 2\frac{\mathrm{d}t}{1+t^2}$. From the definition of t it follows that $t^2 = \frac{\sin^2\left(\frac{\gamma}{2}\right)}{\cos^2\left(\frac{\gamma}{2}\right)}$, which we can use to find the equivalent for $\cos\gamma$:

$$\cos\gamma = \cos\left(\frac{\gamma}{2} + \frac{\gamma}{2}\right) = \cos^2\left(\frac{\gamma}{2}\right) - \sin^2\left(\frac{\gamma}{2}\right) \Rightarrow$$

$$\cos\gamma = \cos^2\left(\frac{\gamma}{2}\right)\left[1 - t^2\right] = \cos^2\left(\frac{\gamma}{2}\right)\frac{1 - t^2}{\cos^2\left(\frac{\gamma}{2}\right) + \sin^2\left(\frac{\gamma}{2}\right)} \Rightarrow$$

$$\cos\gamma = \frac{1 - t^2}{1 + t^2} \tag{5.266}$$

Using these definitions we rewrite Eq. (5.265) as

$$\int_{\gamma_E}^{\gamma} \ldots \equiv \int_{t_E}^{t} \ldots = \int_{t_E}^{t} \frac{2}{1+t^2} \frac{1}{\frac{1-t^2}{1+t^2} - \frac{1-t_E'^2}{1+t_E'^2}} \mathrm{d}t \Rightarrow$$

$$\int_{t_E}^{t} \ldots = \int_{t_E}^{t} 2 \frac{1+t_E'^2}{(1-t^2)(1+t_E'^2) - (1-t_E'^2)(1+t^2)} \mathrm{d}t \Rightarrow$$

$$\int_{t_E}^{t} \ldots = \int_{t_E}^{t} 2 \frac{1+t_E'^2}{-2t^2+2t_E'^2} \mathrm{d}t = (1+t_E'^2) \int_{t_E}^{t} \frac{\mathrm{d}t}{t_E'^2 - t^2} \Rightarrow$$

$$\int_{t_E}^{t} \ldots = \left(\frac{1}{t_E'} + t_E'\right) \int_{\frac{t_E}{t_E'}}^{\frac{t}{t_E'}} \frac{\mathrm{d}\left(\frac{t}{t_E'}\right)}{1 - \left(\frac{t}{t_E'}\right)^2} \tag{5.267}$$

The solution to the standard integral $\int \frac{\mathrm{d}x}{1-x^2}$ has already been given in Eq. (5.133), i.e.,

$$\int \frac{\mathrm{d}x}{1-x^2} = \frac{1}{2} \int \frac{\mathrm{d}x}{1+x} + \frac{1}{2} \int \frac{\mathrm{d}x}{1-x} = \frac{1}{2} \ln \frac{1+x}{1-x}$$

which leads us to

$$\int_{t_E}^{t} \ldots = \frac{1}{2} \left(\frac{1}{t_E'} + t_E'\right) \left[\ln \left(\frac{1 + \frac{t}{t_E'}}{1 - \frac{t}{t_E'}} \right) - \ln \left(\frac{1 + \frac{t_E}{t_E'}}{1 - \frac{t_E}{t_E'}} \right) \right] \Rightarrow$$

$$\int_{t_E}^{t} \ldots = \frac{1}{2} \left(\frac{1}{t_E'} + t_E'\right) \ln \left(\frac{t_E' + t}{t_E' - t} \cdot \frac{t_E' - t_E}{t_E' + t_E} \right) \Rightarrow$$

$$\int_{t_E}^{t} \ldots = \frac{1}{2} \left(\frac{1}{t_E'} + t_E'\right) \ln \left(\frac{t_E' + t}{t_E' - t} \cdot \frac{t_E' - t_E}{t_E' + t_E} \right) \Rightarrow$$

$$\int_{t_E}^{t} \ldots = \frac{1+t_E'^2}{2t_E'} \ln \left(\frac{t_E' + t}{t_E' - t} \cdot \frac{t_E' - t_E}{t_E' + t_E} \right) = \frac{1}{\sin \gamma_E'} \ln \left(\frac{t_E' + t}{t_E' - t} \cdot \frac{t_E' - t_E}{t_E' + t_E} \right) \tag{5.268}$$

We have used here, in line with Eq. (5.266), the expression

$$\sin \gamma = \frac{2t}{1+t^2} \tag{5.269}$$

Table 5.9 Travelled distance for the single-skip re-entry trajectory (H_s = 7,200 m)

γ_E (°)	h_p (–)	X_{f_1} (km)	X_{f_2} (km)	X_f (km)
−7.5	48.5	782	12	794
−15.0	38.6	459	14	473
−30.0	28.7	247	18	265

With the solution of integral Eq. (5.265), given by Eq. (5.268), and substituting back $t = \tan\frac{\gamma}{2}$, $t'_E = \tan\frac{\gamma'_E}{2}$ and $t_E = \tan\frac{\gamma_E}{2}$, we finally get for Eq. (5.264):

$$X = \frac{1}{\beta}\left[\gamma - \gamma_E + \cot\gamma'_E \ln\left(\frac{\tan\frac{\gamma'_E}{2} + \tan\frac{\gamma}{2}}{\tan\frac{\gamma'_E}{2} - \tan\frac{\gamma}{2}} . \frac{\tan\frac{\gamma'_E}{2} - \tan\frac{\gamma_E}{2}}{\tan\frac{\gamma'_E}{2} + \tan\frac{\gamma_E}{2}}\right)\right] \tag{5.270}$$

which is identical to Eq. (5.251) and was our goal to proof.

Example 5.11 (*Skip distance*) For the same vehicle as before, i.e., L/D = 2 and W/SC_L = 6,000 N/m^2, we can now look at the distance travelled during a single-skip trajectory. Here, we will distinguish between the first and second term of Eq. (5.260). We assume an entry altitude of h_E = 100 km. In Table 5.9 the results are shown. It is clear that the horizontal distance travelled in the skip trajectory decreases with increasing entry angle, while the additional distance as a result of trajectory curving (second term) is relatively small. Furthermore, the distances X_f are in this case so small that the curvature of the Earth has no role of significance at all.

■

5.5.2.2 Ballistic Phase

For a flight outside the atmosphere the motion is purely gravitational. That means that the horizontal velocity is constant, and the vertical velocity reduces through g only. With a horizontal and vertical velocity of $V_h = V_F \cos\gamma_F$ and $V_v = V_F \sin\gamma_F$ we find at the top of the ballistic flight that $V_{v,top} = 0 = V_F \sin\gamma_F - gt_{top}$, so

$$t_{top} = \frac{V_F \sin\gamma_F}{g} \tag{5.271}$$

The altitude difference covered is $\Delta h = V_F \sin\gamma_F t_{top} - \frac{1}{2}gt^2_{top}$. Further, $\gamma = 0$ and $V = V_h$, and with t_{top} given the horizontal distance flown is

$$X_{top} = V_F \cos\gamma_F t_{top} \tag{5.272}$$

Of course, this distance will be twice as much when the vehicle enters the atmosphere again, i.e., including the distance covered during the descending leg towards the atmosphere.

One thing should be noted, though. We are actually not free to specify the altitude when the atmosphere is entered. From Eq. (5.234),

$$\cos\gamma - \cos\gamma_E = \frac{g}{2\beta}\frac{1}{\frac{W/S}{C_L}}(\rho - \rho_E)$$

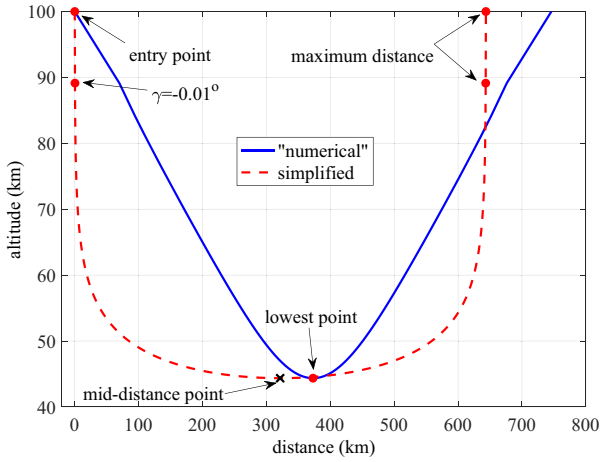
at $h = h_E$, we only confirm that 0 = 0. And for $\rho = \rho_E = 0$ (the assumption that we made earlier), the exponential atmosphere gives infinity for the corresponding altitude. A practical solution would be to specify a small offset in γ, e.g., 0.01°, in which case the above equation would give us the corresponding density for which we could start the analysis. With $\rho_0 = 1.225\,\text{kg/m}^3$ and $H_s = 7{,}200\,\text{m}$, this results in $h = 89.1\,\text{km}$.

However, if we would have specified $h_E = 100\,\text{km}$, for $V_E = 8\,\text{km/s}$ and $\gamma_E = -10°$, the altitude difference of 10 km would be covered in $\Delta t = \frac{\Delta h}{V_E \sin\gamma_E} \approx 8\,\text{s}$, during which $\Delta X_f = V_E \cos\gamma_E \approx 62\,\text{km}$ is flown. In this simple analysis we have assumed that the velocity and flight-path angle remain constant during this part of the flight, because of the absence of aerodynamic forces in the upper layer of the atmosphere. In conclusion, one should be aware of the possible error that is made using the simplified analysis discussed in this section, although it does give a good first indication what is happening during skipping flight.

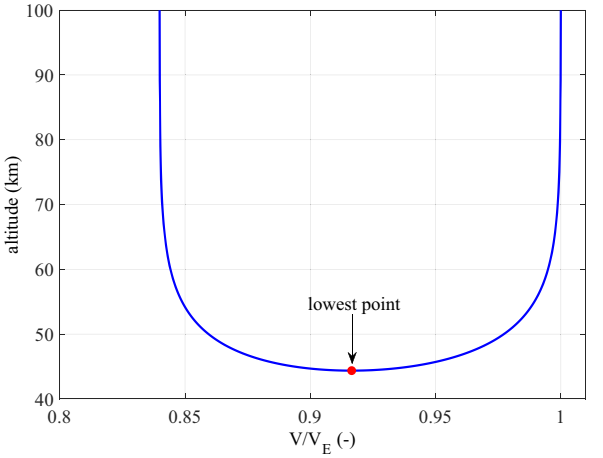
Example 5.12 (*Skip geometry*) In this example we will look at the geometry of the skipping flight by combining the atmospheric and ballistic flight phases. From Example 5.10 we found for a vehicle with $L/D = 2$ and $W/SC_L = 6{,}000\,\text{N/m}^2$ that the characteristic points for γ_E = -10° were: h_p = 44.37 km, V_p = 0.916 V_E and V_F = 0.840 V_E. In Fig. 5.61a the change of altitude as a function of flight range has been plotted. The two curves have been obtained in different ways. For the (blue) solid line, density (and thus altitude) has been calculated according to Eq. (5.235) by varying γ between γ_E (= -10°) and γ_F (= +10°) with steps of $\Delta\gamma = 0.01°$. The range has been calculated with Eq. (5.251), again for varying γ.

For the (red) dashed line, we calculated the altitude in the same way as discussed before, but used Eq. (5.260) to calculate the final distance, X_f. Then, this distance was divided into equidistant intervals. A few observations can be made. The simplified, dashed solution obviously has the lowest point, h_p, at mid distance. From the "numerical" solution, it is clear that there is an asymmetry in the skip. Due to the simplifications, there is a discrepancy in the final distance in the order of 100 km (10–15%). Overall, the "numerical" solution shows a more intuitive feel of being correct.

In Fig. 5.61b the altitude-velocity diagram has been shown, with the velocity calculated with Eq. (5.226) for a given variation of γ. Initially, the velocity does not decrease as the retarding drag force is small. At lower altitudes, the deceleration is severe, which we also saw earlier for ballistic and gliding flight. With $V_p = 0.916\,V_E$



(a) Altitude versus range



(b) Altitude-velocity diagram

Fig. 5.61 Skipping flight with single leg ($\rho_0 = 1.225\,\text{kg/m}^3$, and $H_s = 7{,}200\,\text{m}$)

and $V_F = 0.840\ V_E$, the asymmetry is only small, but present nonetheless. Inspecting the change of flight-path angle with velocity (not plotted here) shows a close-to-linear variation, with only a marginal increase with decreasing velocity. This (concave) effect becomes stronger for steeper entries, as the push of the atmosphere becomes larger.

Finally, in Fig. 5.62 the ballistic flight phase after leaving the atmosphere is shown. The motion is driven by gravity alone, and because the kinetic and potential energy exchange is conservative, the trajectory is symmetric. The velocity when "hitting" the atmosphere again is the same as the one at the beginning of the ballistic flight,

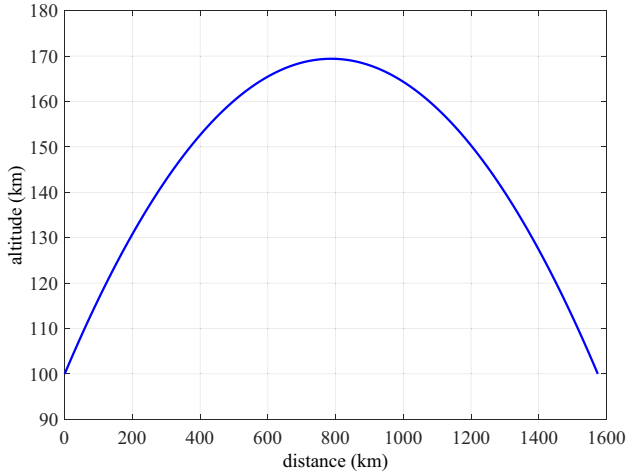


Fig. 5.62 Ballistic flight after leaving the atmosphere for skipping flight

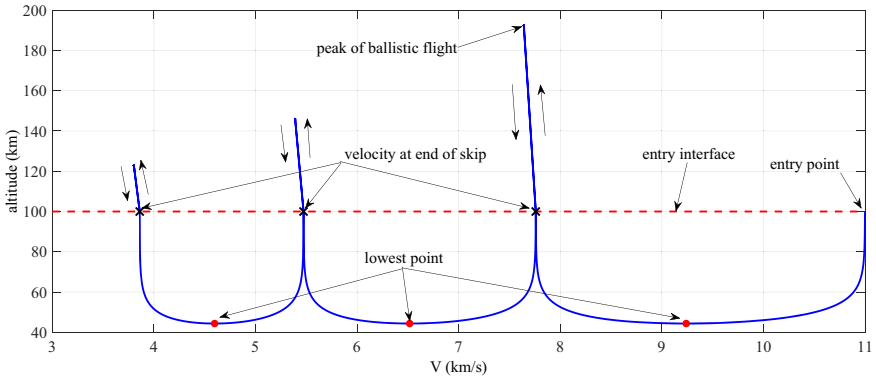


Fig. 5.63 Skipping flight with 3 legs, altitude-velocity diagram ($\rho_0 = 1.225\,\text{kg/m}^3$, and $H_s = 7{,}200\,\text{m}$)

only the flight-path angle has changed sign. By comparing Figs. 5.61a and 5.62 is is obvious that the horizontal distance covered during the ballistic phase is more than twice compared to the one in the atmospheric phase (1574 versus 746 km). This clearly shows the importance of the ballistic phase for range control.

To conclude this example we will study the effects of multiple skips on the entry of an Apollo-like vehicle returning from the Moon. For this example we assume a lift-to-drag ratio of $L/D = 1$, for an otherwise similar lift parameter, $W/SC_L = 6{,}000\,\text{N/m}^2$. The entry velocity is $V_E = 11\,\text{km/s}$ at $h_E = 100\,\text{km}$. The entry angle is the same as before, i.e., $\gamma_E = -10°$. In Fig. 5.63 the altitude-velocity diagram is shown. The previously indicated asymmetry is visible again; note that for each of the three skips the lowest point is at the same altitude (Eq. (5.239)), so independent of the

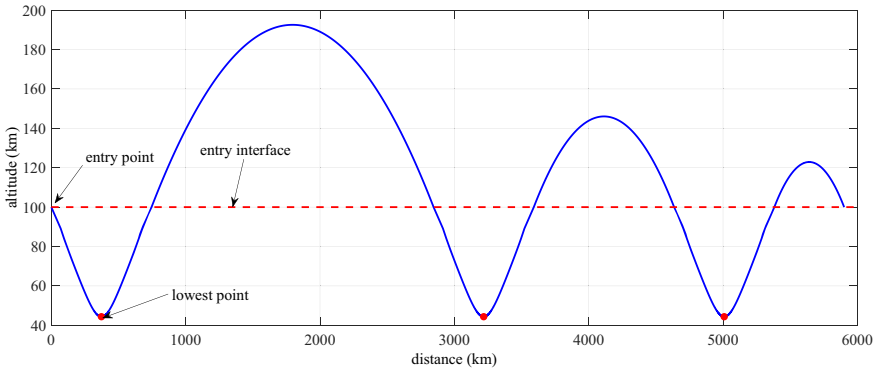


Fig. 5.64 Skipping flight with 3 legs, range versus altitude ($\rho_0 = 1.225\,\text{kg/m}^3$, and $H_s = 7{,}200\,\text{m}$)

velocity at the beginning of the skip. The ballistic flight phase shows as a single line, first as a decrease of velocity with increasing altitude and then an opposite motion with increasing velocity and decreasing altitude. Due to the decreasing velocity at the beginning of the next skip, the velocity change becomes smaller with each skip. Also the maximum altitude during the ballistic flight becomes smaller. The final velocity at the end of the third skip is $V_F = 3.86\,\text{km/s}$, a mere 35% of the original entry velocity.

In Fig. 5.64 the change of altitude with horizontal range is shown. After three skips the total range is close to 6,000 km. From this figure it is clear that the range in the atmospheric phase is the same for each of the three skips. Looking at Eq. (5.260) this is obvious, as only h_p could affect this, but given the assumptions made the lowest point is constant for each skip. V_E does not appear in this expression, the range is solely driven by γ_E. ■

Example 5.13 (*Numerical and Analytical Skip Parameters*) During the derivations of the equations related to the trajectory shape and the horizontal distance covered, several assumptions were made. In combination with the typical larger velocity and/or steeper flight path at entry, one might expect that the above derived first-order approximations may differ from the results of a full non-linear simulation. The current example will address this, as well as a verification of the earlier presented results. For the first simulation, we assume an Apollo-like capsule returning from the Moon, with $L/D = 1.0$. At entry ($h_E = 120\,\text{km}$), the entry conditions are $V_E = 11{,}000\,\text{m/s}$ and $\gamma_E = -10°$. The mass of the vehicle is taken as 5,000 kg.

Settings for the numerical simulation are a rotating, spherical Earth with a central gravity field, and a tabulated atmosphere (Cospar International Reference Atmosphere, 1986). To enforce skipping flight, the bank angle is set to zero (full lift up); note that for the actual Apollo entry a bank angle larger than 90° had to be flown (negative lift) to guarantee entry (see also the case study in Sect. 5.5.5, where, amongst others, the sensitivity to bank-angle variations is studied). The trim angle of attack

Table 5.10 Simulation cases for skip analysis

Case #	m (kg)	S_{ref} m^2	$\frac{W}{C_L S_{ref}}\Big\vert_{L/D=1}$ (N/m^2)	$\frac{W}{C_L S_{ref}}$ $(L/D = 2)$ (N/m^2)
1	5,000	12	2,722	1,361
2	2,500	12	1,361	680
3	1,000	12	544	272
4	500	12	272	136
5	500	10	327	163
6	500	8	408	204
7	500	6	544	272

is set to $\alpha_{trim} = -23°$. Last but not least, the aerodynamic coefficients are assumed to be constant, where $C_D = 1.5$, and C_L is computed from the given L/D, i.e., $C_L = 1.5$ (the reference area is based on a spherical heat shield with diameter $d = 3.9\,\text{m}$, and thus $S_{ref} = 12\,\text{m}^2$).

Basis for the comparison is a number of key parameters, i.e., the conditions at the lowest point of the skip, h_p and V_p, and those at the end of the skip, given by V_F and γ_F. For the analytical model, we evaluate these parameters with Eqs. (5.239), (5.240), and (5.242), and the notion that $\gamma_F = -\gamma_E$. Underlying these first-order approximations is the use of an exponential atmosphere model. In this example, we will use $H_s = 7{,}050\,\text{m}$ and $\rho_0 = 1.225\,\text{kg/m}^3$. In Sect. 5.6.2, a more in-depth verification is performed, and besides the first-order approximation also the so-called second-order approximation will be analysed.

The run with the above mentioned settings shows quite some discrepancies between the analytical and the numerical results. Therefore, as a first analysis of the possible reasons where these differences come from, we will define seven cases, listed in Table 5.10, where a few aspects of the problem are varied. These seven cases will subsequently be executed for $L/D = 1$ and $\gamma_E = -10°$ (where we will execute the numerical simulation for a rotating and non-rotating Earth), $L/D = 2$ and $\gamma_E = -10°$, and $L/D = 2$ and $\gamma_E = -15°$.

In Fig. 5.65 the results of the first batch are shown. A few observations can be made. From theory, we know that V_F is only depending on L/D, which is constant, so also V_F is constant. The numerical simulation shows a significant discrepancy, and also an influence of the vehicle mass (Cases #1 through #4). Since the mass has been ignored in developing the analytical solutions, this effect is then obviously not present. However, it does not fully explain the difference, but as we will see later (Figs. 5.66 and 5.67), the difference gets smaller with steeper flight paths and larger L/D. So, it may be concluded that for the first-order approximation to be a valid approach, the flight path should be steep, the L/D larger than 2, and the vehicle should not be too heavy (Fig. 5.65).

The analytical flight-path angle at exit, γ_F, is the same (but with opposite sign) as γ_E. Also in this case we see that this does not hold for the numerical cases. As the

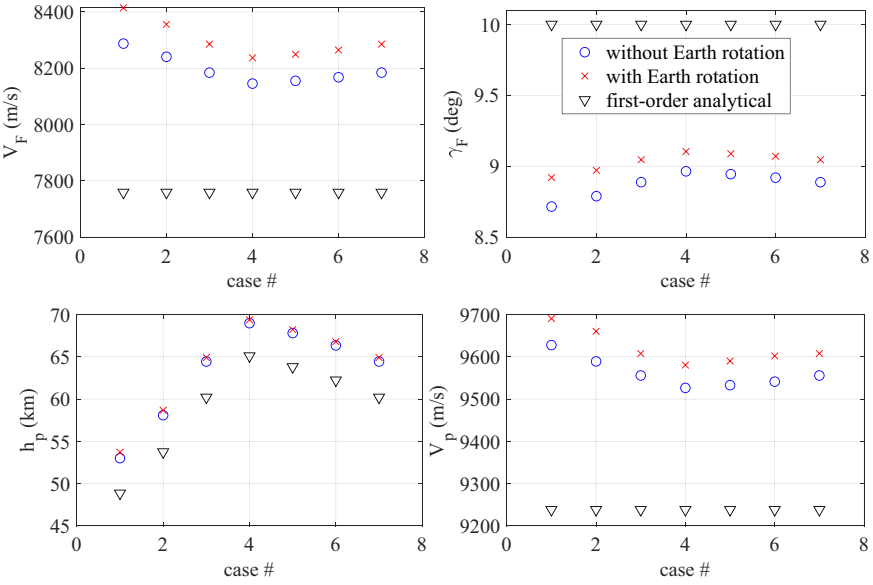


Fig. 5.65 Comparison for seven skip cases (Table 5.10), with $L/D = 1$ and $\gamma_E = -10°$

vertical force equilibrium is affected by the weight of the vehicle, the remaining lift is comparably smaller and can thus not "push" the vehicle enough in the ascending leg, hence resulting in a smaller exit angle. In the descending leg, on the other hand, when the dynamic pressure is larger due to the still larger velocity, this push is sufficiently large to prevent the vehicle from going lower into the atmosphere, hence reaching the lowest point in its trajectory at a slightly higher altitude. The corresponding velocity is also larger, mainly due to the additional acceleration due to gravity.

Including the rotation of the Earth acts as an additional push on the vehicle, due to the centrifugal effect. Even though the altitude of the lowest point is not so much affected, the velocity profile is. The centrifugal force reduces the effective deceleration, resulting in an overall larger velocity (at any point in the trajectory). It also increases the local curvature of the trajectory, and gives a slightly larger exit angle that is thus closer to the analytical value.

Skip Cases #5–#7 are a variation of Case #4, in the sense that the mass is kept constant at $m = 500\,\text{kg}$ (and, of all cases, minimising the effect of ignoring the mass on the analytical solution), but with an adjusted reference area, S_{ref}. The idea behind this that a (much) lower weight would possibly mean a (much) smaller vehicle. To study that effect, S_{ref} is reduced (from $12\,\text{m}^2$) to 10, 8, and $6\,\text{m}^2$. For the numerical simulation this means that the drag area is reduced, whereas for the analytical solution this will have an effect on the lift parameter that appears in the expression for h_p. The lowest altitude obtained from the numerical simulation is similarly affected, so the difference remains the same. However, this is not true for the velocity profile. The reduction in S_{ref} and thus also the drag area yields a larger velocity at corresponding

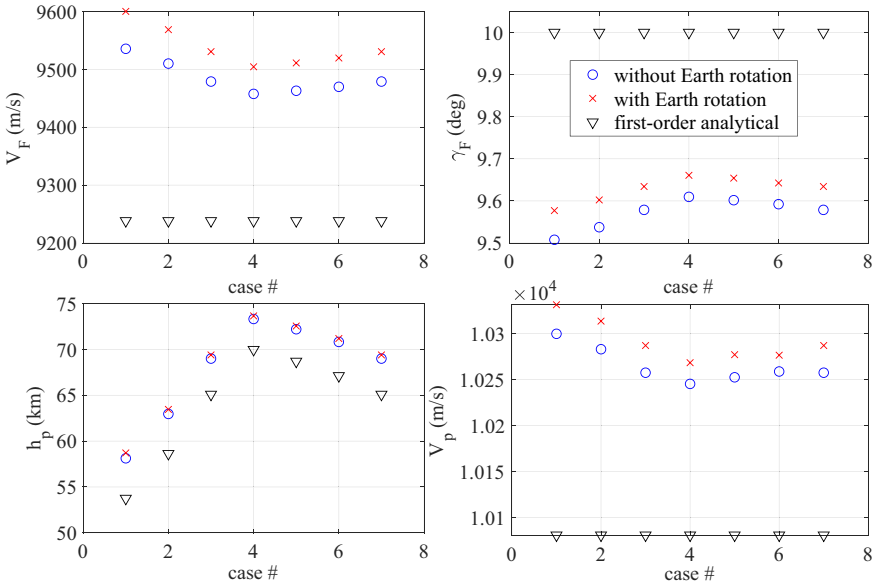


Fig. 5.66 Comparison for seven skip cases (Table 5.10), with $L/D = 2$ and $\gamma_E = -10°$

altitudes. Therefore, the difference in V_p and V_F increases with smaller reference area (Figs. 5.66 and 5.67).

In Figs. 5.66 and 5.67 the results for (L/D = 2, $\gamma_E = -10°$) and (L/D = 2, $\gamma_E = -15°$) are shown, confirming our earlier statement that increasing L/D and reducing γ_E, i.e., making it more negative, will reduce the difference between the two approaches. Also a reduction of vehicle mass will bridge the gap, as we have mentioned earlier. To support these conclusions, Tables 5.11, 5.12, 5.13 and 5.14 list the errors made when using the first-order approximations for all cases. The largest errors are those for γ_F and h_p, for the low L/D and shallower-entry case (errors up to 15%). However, for these single-point performance indicators the remaining approximations are not too bad. It is fair to say, though that one may have to consider the absolute error as well, if this approximation has to be used in, for instance, a guidance-system design.

Unfortunately we cannot draw definite conclusions from just this analysis. There is also the actual trajectory to consider, i.e., the relation between V, γ, and h, as well as maximum thermo-mechanical loads. These aspects will come up for discussion during a more elaborate verification process in Sect. 5.5.4. There, the skip entry will be analysed for Earth, Mars and Venus aerocapture missions, in anticipation of the discussion about aeromanoeuvring in Chap. 11. ■

Table 5.11 Error in V_F, analytical with respect to numerical

ase	$L/D = 1, \gamma_E = -10°$		$L/D = 2, \gamma_E = -10°$		$L/D = 2, \gamma_E = -15°$	
#	Rotating	Non-rotating	Rotating	Non-rotating	Rotating	Non-rotating
1	−6.4	−7.8	−3.1	−3.8	−1.9	−2.4
2	−5.8	−7.1	−2.9	−3.5	−1.8	−2.2
3	−5.2	−6.4	−2.5	−3.1	−1.6	−2.0
4	−4.7	−5.8	−2.3	−2.8	−1.5	−1.8
5	−4.9	−5.9	−2.4	−2.9	−1.5	−1.9
6	−5.0	−6.1	−2.4	−3.0	−1.5	−1.9
7	−5.2	−6.4	−2.5	−3.1	−1.6	−2.0

Table 5.12 Error in γ_F, analytical with respect to numerical

Case #	$L/D = 1, \gamma_E = -10°$		$L/D = 2, \gamma_E = -10°$		$L/D = 2, \gamma_E = -15°$	
	Rotating	Non-rotating	Rotating	Non-rotating	Rotating	Non-rotating
1	14.7	12.1	5.2	4.4	4.8	4.3
2	13.8	11.5	4.8	4.1	4.5	4.0
3	12.5	10.5	4.4	3.8	4.0	3.6
4	11.6	9.8	4.1	3.5	3.7	3.3
5	11.8	10.0	4.1	3.6	3.8	3.4
6	12.1	10.3	4.3	3.7	3.9	3.5
7	12.5	10.5	4.4	3.8	4.0	3.6

Table 5.13 Error in h_p, analytical with respect to numerical

Case #	$L/D = 1, \gamma_E = -10°$		$L/D = 2, \gamma_E = -10°$		$L/D = 2, \gamma_E = -15°$	
	Rotating	Non-rotating	Rotating	Non-rotating	Rotating	Non-rotating
1	−7.8	−9.0	−7.5	−8.5	−4.3	−4.8
2	−7.5	−8.4	−6.9	−7.6	−4.8	−5.2
3	−6.6	−7.2	−5.7	−6.2	−4.7	−5.0
4	−5.6	−6.2	−4.6	−5.0	−4.2	−4.4
5	−5.9	−6.5	−4.9	−5.3	−4.4	−4.6
6	−6.2	−6.8	−5.2	−5.7	−4.5	−4.8
7	−6.6	−7.2	−5.7	−6.2	−4.7	−5.0

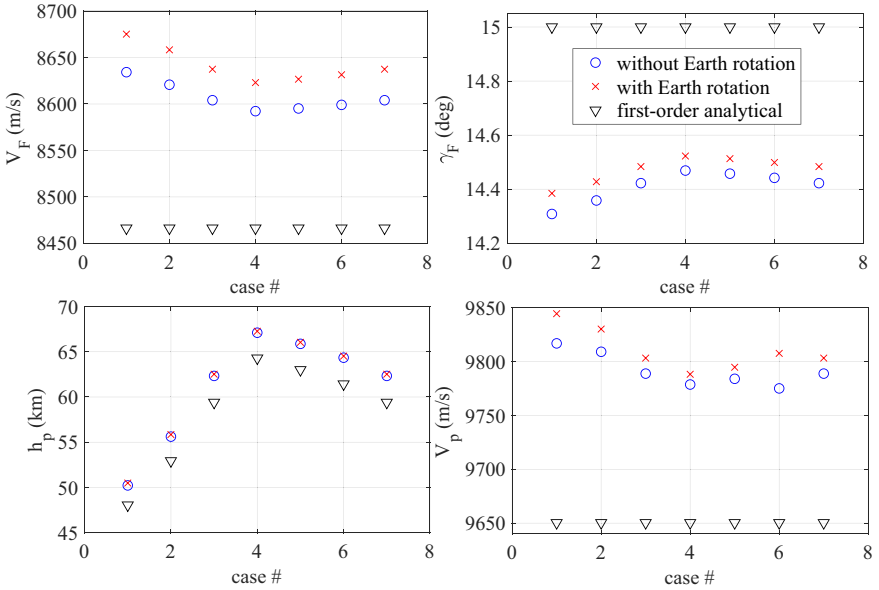


Fig. 5.67 Comparison for seven skip cases (Table 5.10), with $L/D = 2$ and $\gamma_E = -15°$

Table 5.14 Error in V_p, analytical with respect to numerical

Case #	$L/D = 1, \gamma_E = -10°$		$L/D = 2, \gamma_E = -10°$		$L/D = 2, \gamma_E = -15°$	
	Rotating	Non-rotating	Rotating	Non-rotating	Rotating	Non-rotating
1	−4.0	−4.7	−2.1	−2.4	−1.7	−2.0
2	−3.7	−4.4	−2.0	−2.3	−1.6	−1.8
3	−3.3	−3.8	−1.7	−2.0	−1.4	−1.6
4	−3.0	−3.6	−1.6	−1.8	−1.3	−1.4
5	−3.1	−3.7	−1.7	−1.9	−1.4	−1.5
6	−3.2	−3.8	−1.7	−1.9	−1.3	−1.6
7	−3.3	−3.8	−1.7	−2.0	−1.4	−1.6

5.5.3 Thermo-Mechanical Loads

5.5.3.1 Deceleration

For the ballistic and gliding trajectory we only looked at the deceleration in flight direction, even though for gliding flight there is also a component normal to the flight path due to the lift. For skipping flight, the lift is usually much larger and, therefore, should be included in the analysis. It will increase the loads on the vehicle significantly. First, we consider the deceleration in trajectory direction. If this deceleration is, as before, defined as positive, Eq. (5.223) gives us:

$$\frac{a_V}{g} = -\frac{\mathrm{d}V/\mathrm{d}t}{g} = \frac{D}{W} = \frac{C_D \frac{1}{2}\rho V^2 S}{W} \tag{5.273}$$

Second, the acceleration perpendicular to the trajectory follows from Eq. (5.224):

$$\frac{a_n}{g} = \frac{L}{mg} = \frac{L}{D}\frac{D}{mg} = \frac{L}{D}\frac{a_V}{g} \tag{5.274}$$

The total resulting acceleration that the vehicle is subjected to is found to be:

$$\frac{a_{res}}{g} = \sqrt{\left(\frac{a_V}{g}\right)^2 + \left(\frac{a_n}{g}\right)^2} = \frac{a_V}{g}\sqrt{1 + \left(\frac{L}{D}\right)^2} \tag{5.275}$$

This means that to find an expression for maximum deceleration, we just need to focus on a_V. So, from Eq. (5.273) for a constant C_D, the deceleration is directly proportional to the dynamic pressure, $\bar{q} = \frac{1}{2}\rho V^2$. Earlier we found for the velocity and density in the skip trajectory, Eqs. (5.226) and (5.235):

$$V = V_E e^{-\frac{(\gamma - \gamma_E)}{L/D}} \tag{5.276}$$

and

$$\rho = \frac{2\beta}{g}\frac{W/S}{C_L}(\cos\gamma - \cos\gamma_E) \tag{5.277}$$

where again we have assumed $\rho_E = 0$. Substitution into Eq. (5.273) gives:

$$\frac{a_V}{g} = \frac{\beta V_E^2}{g}\frac{D}{L}(\cos\gamma - \cos\gamma_E)\, e^{-\frac{2(\gamma - \gamma_E)}{L/D}} \tag{5.278}$$

For given entry conditions (V_E and γ_E) the deceleration is solely a function of γ and L/D. By taking the derivative of Eq. (5.278) with respect to γ and setting it equal to zero, the condition for maximum deceleration can be found:

$$\frac{\mathrm{d}\left(\frac{a_V}{g}\right)}{\mathrm{d}\gamma} = \frac{\beta V_E^2}{g}\frac{D}{L}\left[-\sin\gamma - \frac{2}{L/D}(\cos\gamma - \cos\gamma_E)\right] e^{-\frac{2(\gamma - \gamma_E)}{L/D}} = 0 \Rightarrow$$

$$\sin\gamma = -\frac{2}{L/D}(\cos\gamma - \cos\gamma_E) \tag{5.279}$$

From flight-mechanical aspects we know that $-\gamma_E < \gamma < +\gamma_E$ and thus $\cos\gamma > \cos\gamma_E$, therefore $\sin\gamma$ will be negative. This means that the maximum deceleration will occur in the descending segment of the skip trajectory, so before the lowest point is reached.

To solve the goniometric relationship Eq. (5.279) we will re-introduce the "half" flight-path angle, $\frac{\gamma}{2}$, and express $\cos\gamma$ and $\sin\gamma$ accordingly. With

$$\sin\gamma = 2\sin\left(\frac{\gamma}{2}\right)\cos\left(\frac{\gamma}{2}\right) = \frac{2\tan\left(\frac{\gamma}{2}\right)}{1+\tan^2\left(\frac{\gamma}{2}\right)} \tag{5.280}$$

$$\cos\gamma = \cos^2\left(\frac{\gamma}{2}\right) - \sin^2\left(\frac{\gamma}{2}\right) = \frac{1-\tan^2\left(\frac{\gamma}{2}\right)}{1+\tan^2\left(\frac{\gamma}{2}\right)} \tag{5.281}$$

Equation (5.279) is rewritten as

$$\frac{2\tan\left(\frac{\gamma}{2}\right)}{1+\tan^2\left(\frac{\gamma}{2}\right)} = -\frac{2}{L/D}\left(\frac{1-\tan^2\left(\frac{\gamma}{2}\right)}{1+\tan^2\left(\frac{\gamma}{2}\right)} - \cos\gamma_E\right) \Rightarrow$$

$$\frac{L}{D}\tan\left(\frac{\gamma}{2}\right) + \left[1-\tan^2\left(\frac{\gamma}{2}\right)\right] - \cos\gamma_E\left(1+\tan^2\left(\frac{\gamma}{2}\right)\right) = 0$$

which results in the following quadratic equation:

$$(1+\cos\gamma_E)\tan^2\left(\frac{\gamma}{2}\right) - \frac{L}{D}\tan\left(\frac{\gamma}{2}\right) - (1-\cos\gamma_E) = 0 \tag{5.282}$$

Since it must hold that $\gamma < 0$ (descent), only the negative root of this equation is relevant[11]:

$$\tan\left(\frac{\gamma}{2}\right) = \frac{\frac{L}{D} - \sqrt{\left(\frac{L}{D}\right)^2 + 4\sin^2\gamma_E}}{2\left(1+\cos\gamma_E\right)} \Rightarrow$$

$$\tan\left(\frac{\gamma}{2}\right) = \frac{\frac{L}{D}}{2\left(1+\cos\gamma_E\right)}\left(1 - \sqrt{1+\frac{4\sin^2\gamma_E}{\left(\frac{L}{D}\right)^2}}\right) \tag{5.283}$$

so

$$\gamma_{a,max} = 2\arctan\left[\frac{\frac{L}{D}}{2\left(1+\cos\gamma_E\right)}\left(1 - \sqrt{1+\frac{4\sin^2\gamma_E}{\left(\frac{L}{D}\right)^2}}\right)\right] \tag{5.284}$$

Once the flight-path angle for maximum deceleration is known, the resulting acceleration a_{max} can be found from Eqs. (5.275) and (5.278). It is noted that these equations hold for the condition that there is sufficient lift (and drag) to sustain a skipping flight. That means that for smaller $\frac{L}{D}$ these equations will not hold, with of course a singularity for $\frac{L}{D} = 0$, the condition for ballistic flight.

[11] The square-root term is always larger than $\frac{L}{D}$, so a subtraction of this root from $\frac{L}{D}$ will yield a negative $\tan\left(\frac{\gamma}{2}\right)$ and thus a negative γ.

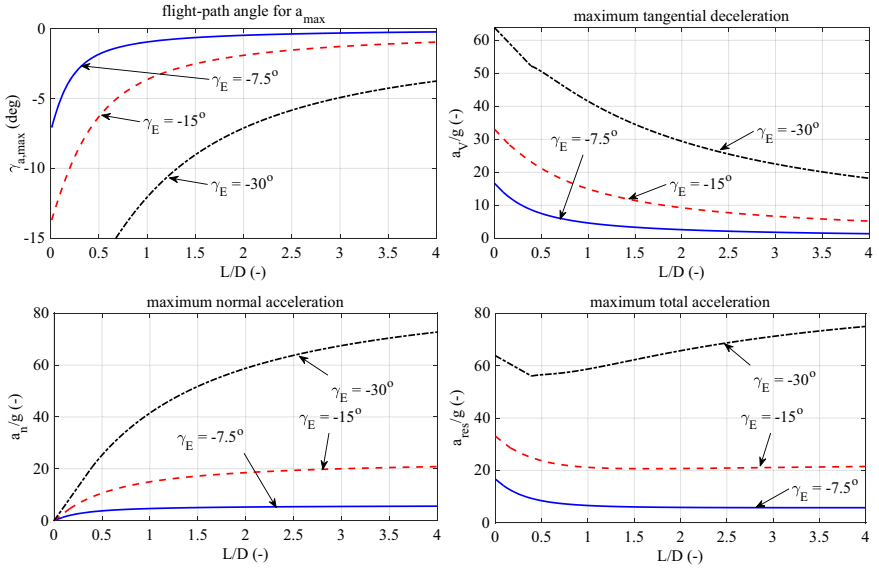


Fig. 5.68 Acceleration parameters as function of $\frac{L}{D}$ and entry angle, γ_E ($H_s = 7{,}200$ m)

In Fig. 5.68 the mechanical-load related parameters have been plotted. To start with, the flight-path angle for which the maximum deceleration occurs (Eq. (5.284)) is shown in the top-left plot, as a function of lift-to-drag ratio and three values for the flight-path angle at entry. The $\frac{L}{D}$ ranges from 0 to 4, to get the full picture, but as stated, for lower values of $\frac{L}{D}$ the assumptions of D and L being much larger than W do not hold. Also, three values for γ_E have been analysed, notably $\gamma_E = -7.5°$, $-15°$ and $-30°$. In all cases $\gamma_{a,max}$ does not exceed γ_E in absolute sense, indeed showing that a_{max} occurs during the descending leg. With increasing $\frac{L}{D}$ the deceleration peaks later, i.e., closer to the lowest point: for $\gamma_E = -7.5°$ and $\frac{L}{D} = 4$, it is almost *at* the lowest point.

Substituting $\gamma_{a,max}$ into Eq. (5.278) produces the curves of the top-right plot of Fig. 5.68. It is noted that Eq. (5.47) has been used to calculate a_{max} for $\frac{L}{D} = 0$, i.e., the ballistic value ($V_E = 7{,}000$ m/s). A few of the points for values of $\frac{L}{D}$ close to zero had to be removed from the curve (especially for $\gamma_E = -30°$), because they were affected by the singularity at $\frac{L}{D} = 0$ and showed an erratic behaviour. The maximum tangential deceleration decreases with increasing $\frac{L}{D}$, a phenomenon that was already shown for gliding flight. Of course, a steeper trajectory will lead to an increase in peak deceleration. It is clear that even for not too steep entries accelerations about 20–30 times the acceleration of gravity are possible during the skipping phase. This has a rather negative effect on the vehicle structure, and, for manned vehicles, on the physiology of the astronauts. Careful mission planning is advised, as well as a robust guidance system to account for uncertainties in the mission design.

A further observation is that while the tangential acceleration decreases the normal acceleration increases monotonically with increasing $\frac{L}{D}$. Hence, the total acceleration has a minimum with respect to $\frac{L}{D}$; for instance, this minimum occurs for $\frac{L}{D} \approx 0.4$ if γ_E $-30°$ and $\frac{L}{D} \approx 1.7$ (although difficult to see) for $\gamma_E = -15°$. Note that for smaller entry angles the maximum total acceleration has become independent of $\frac{L}{D}$.

In Table 5.15 a number of numerical values is provided for the three cases of a skip trajectory, with flight-path entry angles of $\gamma_E = -7.5°$, $-15°$, and $-30°$ for an entry velocity of $V_E = 7$ km/s ($\frac{L}{D} = 2$). Once more it is confirmed that the maximum deceleration in flight direction and acceleration perpendicular to the trajectory, on the curved path, increase with increasing flight-path angle γ_E and can become quite large. To limit the loads on the vehicle and possibly its passengers, only small entry angles are allowed for high velocities. The loss of velocity in the skip trajectory is in that case less than for larger angles, but has to be accepted. One must also realise that a_{max}, as well as the accompanying a_n and a_{res}, increase quadratically with the entry velocity V_E, see Eq. (5.278).

For small entry angles the following will be true:

$$\frac{4\sin^2\gamma_E}{\left(\frac{L}{D}\right)^2} << 1 \tag{5.285}$$

and with the approximation $\sqrt{1+\varepsilon} = 1 + \frac{\varepsilon}{2}$, Eq. (5.283) can be transformed into:

$$\tan\left(\frac{\gamma}{2}\right) = -\frac{\frac{L}{D}}{2\left(1+\cos\gamma_E\right)}\frac{2\sin^2\gamma_E}{\left(\frac{L}{D}\right)^2} \Rightarrow$$

$$\tan\left(\frac{\gamma}{2}\right) = -\frac{\frac{L}{D}}{\left(1+\cos^2\left(\frac{\gamma_E}{2}\right)-\sin^2\left(\frac{\gamma_E}{2}\right)\right)}\frac{4\sin^2\left(\frac{\gamma_E}{2}\right)\cos^2\left(\frac{\gamma_E}{2}\right)}{\left(\frac{L}{D}\right)^2} \Rightarrow$$

$$\tan\left(\frac{\gamma}{2}\right) = -\frac{2\sin^2\left(\frac{\gamma_E}{2}\right)}{L/D} \tag{5.286}$$

Since we assumed small entry angles it thus follows:

$$\gamma \cong -\frac{\gamma_E^2}{L/D} \tag{5.287}$$

These expressions show that the maximum deceleration for small entry angles will occur close to the lowest point of the trajectory (and, as stated before, in the descending leg, where $\gamma < 0$). Table 5.15 also contains the trajectory parameters for the lowest point of the trajectory. It appears that for small γ_E ($|\gamma_E| < 15°$) the maximum deceleration is well approximated by the deceleration in the lowest point.

Table 5.15 Example of trajectory parameters in the skip trajectory for several entry angles γ_E

	$V_E = 7\,\text{km/s}$, $\frac{L}{D} = 2$, $\frac{W/S}{C_L} = 6{,}000\text{N/m}^2$ exponential atmosphere ($H_s = 7{,}200\,\text{m}$)			
	Parameter	$\gamma_E = -7.5°$	$\gamma_E = -15°$	$\gamma_E = -30°$
Point in trajectory with maximum deceleration	γ (°)	−0.49	−1.92	−7.24
	$\frac{\rho}{\rho_0}$	0.00118	0.00465	0.01748
	h (km)	48.5	38.7	29.1
	$\frac{V}{V_E}$	0.94	0.89	0.82
	$\frac{\bar{a}_{max}}{g}$	2.6	9.3	29.4
	$\frac{a_{n,max}}{g}$	5.2	18.5	58.8
	$\frac{a_{res,max}}{g}$	5.8	20.7	65.7
Point in trajectory with minimum altitude	γ (°)	0	0	0
	$\frac{\rho_p}{\rho_0}$	0.00119	0.00473	0.01858
	h_p (km)	48.5	38.6	28.7
	$\frac{V}{V_E}$	0.94	0.88	0.77
	$\frac{\bar{a}_{max}}{g}$	2.6	9.1	27.5
Point in trajectory where vehicle exits atmosphere	$\frac{V_F}{V_E}$	0.88	0.77	0.59
	V_F (km/s)	6.14	5.39	4.15

For this deceleration it holds true according to Eq. (5.278):

$$\frac{a_V}{g} = \frac{\beta V_E^2}{g}\frac{D}{L}\left(1 - \cos\gamma_E\right) e^{-\frac{2\gamma_E}{L/D}} \tag{5.288}$$

5.5.3.2 Heat Load

For the heat supply per unit of time in the stagnation point we begin again with:

$$q_c = c_1 \frac{1}{R_N^n} \rho^{1-n} V^m \tag{5.289}$$

For the heat supplied to the entire vehicle, likewise:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{C_F}{4}\rho V^3 S_W \tag{5.290}$$

To find the point where these heat flows are maximal for a given vehicle in the skip trajectory, the maximum of $f(h, V) = \rho^{1-n} V^m$ needs to be examined, whereby $n = 0$ and $m = 3$ for $\frac{\mathrm{d}Q}{\mathrm{d}t}$ on Earth. From Eqs. (5.226) and (5.235) we have:

$$f^*(\gamma) = c_2 \left(\cos\gamma - \cos\gamma_E\right)^{1-n} e^{\frac{-m(\gamma-\gamma_E)}{L/D}} \tag{5.291}$$

where we assume a constant nose radius, included now in the constant $c_2 = c_1 \left(\frac{2\beta}{g}\frac{W/S}{C_L}\right)^{1-n} \frac{1}{R_N^n}$. With the condition $\frac{\partial f^*}{\partial \gamma} = 0$ it follows that

$$\sin\gamma = -\frac{m}{\left(\frac{L}{D}\right)(1-n)}\left(\cos\gamma - \cos\gamma_E\right) \tag{5.292}$$

As before, introducing $\tan\left(\frac{\gamma}{2}\right)$, see Eqs. (5.280)–(5.281),

$$\frac{2\tan\left(\frac{\gamma}{2}\right)}{1+\tan^2\left(\frac{\gamma}{2}\right)} = -\frac{m}{(L/D)(1-n)}\left(\frac{1-\tan^2\left(\frac{\gamma}{2}\right)}{1+\tan^2\left(\frac{\gamma}{2}\right)} - \cos\gamma_E\right) \tag{5.293}$$

leads to the quadratic equation:

$$(1+\cos\gamma_E)\tan^2\left(\frac{\gamma}{2}\right) - \frac{2}{m}\frac{L}{D}(1-n)\tan\left(\frac{\gamma}{2}\right) - (1-\cos\gamma_E) = 0 \tag{5.294}$$

This equation has as solution, with the negative root taken to guarantee a negative γ:

$$\tan\left(\frac{\gamma}{2}\right) = \frac{\frac{2}{m}\frac{L}{D}(1-n) - \sqrt{\frac{4}{m^2}\left(\frac{L}{D}\right)^2(1-n)^2 + 4\sin^2\gamma_E}}{2(1+\cos\gamma_E)} \Rightarrow$$

$$\tan\left(\frac{\gamma}{2}\right) = \frac{\frac{2}{m}\frac{L}{D}(1-n)\left(1 - \sqrt{1 + \frac{4\sin^2\gamma_E}{\frac{4}{m^2}\left(\frac{L}{D}\right)^2(1-n)^2}}\right)}{2(1+\cos\gamma_E)} \Rightarrow$$

or

$$\tan\left(\frac{\gamma}{2}\right) = \left(\frac{1-n}{m}\right)\left(\frac{1}{1+\cos\gamma_E}\right)\frac{L}{D}\left(1 - \sqrt{1 + \frac{m^2\sin^2\gamma_E}{\left(\frac{L}{D}\right)^2(1-n)^2}}\right) \tag{5.295}$$

For small angles γ_E the approximation $\sqrt{1+\varepsilon} = 1 + \frac{\varepsilon}{2}$ can be introduced for the root of this equation, yielding in a similar fashion as for the maximum deceleration:

Table 5.16 Characteristic points in the skip trajectory for heating ($m = 3$) and deceleration $V_E = 7\,\text{km/s}$, $\gamma_E = -15°$, $\frac{L}{D} = 2$, $\frac{W/S}{C_L} = 6{,}000\,\text{N/m}^2$ exponential atmosphere ($H_s = 7{,}200\,\text{m}$)

		"Exact"		Approximation	
		γ (°)	$\frac{V}{V_E}$	γ (°)	$\frac{V}{V_E}$
$q_{c,max}$	$n = 0.5$ (lam)	−5.17	0.918	−5.89	0.924
	$n = 0.2$ (turb)	−3.47	0.904	−3.68	0.906
$\left(\frac{dQ}{dt}\right)_{max}$	$n = 0$	−2.83	0.899	−2.95	0.900
$\frac{\bar{a}_{max}}{g}$	–	−1.92	0.892	−1.96	0.892
h_p	–	0	0.877	0	0.877

$$\tan\left(\frac{\gamma}{2}\right) = \frac{1}{2}\left(\frac{1}{1+\cos\gamma_E}\right)\left(\frac{m}{1-n}\right)\frac{\sin^2\gamma_E}{\left(\frac{L}{D}\right)}$$

With

$$\frac{\sin^2\gamma_E}{1+\cos\gamma_E} = \frac{(1-\cos\gamma_E)(1+\cos\gamma_E)}{1+\cos\gamma_E} = 1-\cos\gamma_E = 2\sin^2\left(\frac{\gamma_E}{2}\right)$$

we get

$$\tan\left(\frac{\gamma}{2}\right) = \left(\frac{m}{1-n}\right)\frac{\sin^2\left(\frac{\gamma_E}{2}\right)}{\frac{L}{D}} \tag{5.296}$$

and for the small angles we consider

$$\gamma \cong \frac{1}{2}\left(\frac{m}{1-n}\right)\frac{\gamma_E^2}{L/D} \tag{5.297}$$

Using the data in Table 5.15 and $\gamma_E = -15°$, γ and V/V_E have been listed in Table 5.16 for various heat-flow conditions. For comparison, the corresponding results for the maximum deceleration and the lowest point are also included. Apart from the values of γ and corresponding V/V_E according to the "exact" Eq. (5.295), the values obtained with the approximating Eq. (5.297) have been given too. The differences between both cases are still small for $\gamma_E = -15°$, but with larger γ_E the discrepancies will increase.

It is clear from Table 5.16 that, in the descending segment of the skip trajectory, the successively occurring points are: maximum q_c for the laminar boundary layer, maximum q_c for the turbulent boundary layer, maximum $\frac{dQ}{dt}$ and maximum deceleration. This sequence is in agreement with the conclusions for the ballistic re-entry trajectory and is independent of the characteristics of the vehicles themselves.

As the entry velocity is reduced for each successive skip trajectory, the largest heat flows and the greatest deceleration will occur in the first dive into the atmosphere. For small entry angles γ_E the lowest point in the trajectory is found to be a good approximation of the point with maximum heat flows.

The heat supply during the entire skip trajectory follows again with the expressions

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{C_F}{4}\rho V^3 S_W \tag{5.298}$$

and

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{D}{m} = -\frac{C_D}{m}\frac{1}{2}\rho V^2 S \tag{5.299}$$

substituted in

$$Q = \int_{V_E}^{V_F} \frac{\mathrm{d}Q}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}V}\mathrm{d}V \tag{5.300}$$

From the above expressions the same expression as derived for the ballistic and glide trajectories follows:

$$\frac{Q}{\mathcal{E}_k} = \frac{1}{2}\frac{C_F}{S_W}C_D S\left[1 - \left(\frac{V_F}{V_E}\right)^2\right] \tag{5.301}$$

It can easily be shown that this expression is valid not only for one skip trajectory through the atmosphere, but also for a successive number of skip trajectories, whereby V_E is the entry velocity of the first dive into the atmosphere and V_F is the end velocity after the last skip trajectory. It is assumed hereby that for the $(n+1)^{th}$ skip trajectory: $V_E^{n+1} = V_F^n$ (with a ballistic flight outside of the atmosphere between two successive skip trajectories).

For a given entry kinetic energy $\mathcal{E}_k$, the part that is supplied to the vehicle as heat, for a desired value of $\frac{V_F}{V_E}$, is dependent on the ratio $\frac{C_F S_W}{C_D S}$, the value of which is largely determined by the applied shape of the vehicle. If the vehicle has gained a high temperature as a result of the first skip trajectory, a portion of the heat can be disposed off through radiation during the ballistic flight outside of the atmosphere.

It was already mentioned that an atmospheric re-entry to Earth is nowadays not yet realised by means of the skip trajectories described above, which is primarily due to the high thermal load for larger entry angles. These larger entry angles are necessary to obtain the ever so important reduction in velocity.

Table 5.17 Relevant parameters for Apollo-derived aerocapture vehicles

Mission	Diameter, d (m)	Aerodynamic reference area, S_{ref} (m/s^2)	Nose radius, R_N (m)	Mass, m (kg)
CRM	3.9116	12.017	4.694	5,500
SRM	0.75	0.441	0.90	38.77

Table 5.18 Initial states for aerocapture mission at three planets

Central Body	Altitude, h_E (km)	Velocity, V_E (m/s)	Flight-path angle, γ_E (°)
Earth	125	11,500	−7.4
Mars	125	8,200	−11.5
Venus	250	11,250	−6.1

5.5.4 *Numerical Verification of an Aerocapture Mission*

In Example 5.13, the effect of γ_E, L/D and vehicle mass was addressed, by comparing the first-order approximation with the results of a numerical simulation. In this section, we will go a little more in-depth for two particular aerocapture missions, i.e., a heavy vehicle for a manned mission, and a light-weight capsule for a sample-return mission. These two missions will be discussed in more detail in Chap. 11, Sect. 11.2, where the influence of selected target planet on the geometry of the entry corridors is determined.

The current verification will be done for three different planets, i.e., Venus, Mars, and Earth, such that the effect of different atmospheres and gravity fields on the results is included in the discussion. For all cases, the first-order solutions will be compared with the results obtained with a corresponding numerical simulation. The current verification process will be continued in Sect. 5.6, where the benefits of using a second-order solution will be demonstrated.

The reference vehicles and aerocapture missions that are considered here have been defined by Engelsma and Mooij (2020), and deal with a manned Crew Return Mission (CRM) and a Sample Return Mission (SRM). The underlying configuration for both is the Apollo Command Module, where the CRM resembles the original Apollo the most, and the SRM is a scaled-down version with dimensions and mass properties derived from the Hayabusa and Stardust sample-return capsules. The relevant data for the current analysis are listed in Table 5.17. The aerodynamic data are those of Apollo, and can in principle be found in Appendix B. However, for the sake of this verification we assume constant aerodynamic coefficients, i.e., $C_D = 1.2786$ and $C_L = 0.5627$ (so $L/D = 0.44$). The initial conditions for the three target planets Venus, Earth, and Mars are provided in Table 5.18.

Beginning with the Earth case, the simulations will be setup as follows. For the first-order solutions we will use the equations of motion Eqs. (5.222) through (5.224), as they served as the basis for obtaining the analytical approximations. Simplifications applied are that the weight terms on the right-hand side (yielding an additional acceleration) are neglected, gravitational acceleration is constant, and the density follows an exponential profile. We will use $\rho_0 = 1.225\,\text{kg/m}^3$ and $H_s = 7{,}050\,\text{m}$ for the two constants. The equations of motion will be integrated using a fourth-order Runge-Kutta scheme, with a step size of 0.01 s. The numerical simulation is based on the full set of non-linear equations of motion derived for a rotating, spherical central body (Eqs. (4.110)–(4.115)), albeit for planar motion, which puts the bank angle at zero. The atmosphere is modelled using a tabulated Cospar International Reference Atmosphere, CIRA (release date 1986), whereas the gravitational acceleration follows a central field model with the flattening of the Earth included as a J_2 zonal harmonic coefficient. The equations are integrated with a variable step-size Runge-Kutta-Fehlberg 5(6) integration method.

To study the effects of certain environmental effects, we follow this up with a few others, i.e., one where we remove J_2 (but this had no visible effect), one where the Earth rotational rate is put to zero, one where the atmosphere model is replaced by the exponential atmosphere, and finally where all this simplifications are combined, and even the gravity is kept constant at its surface value. All results are plotted in Fig. 5.69. It is clear that there are significant differences between the analytical and the numerical trajectories, even though in Example 5.13 we saw that some of the characteristic points matched quite well.

A few more detailed observations: for the analytical solutions, the exit flight-path angle and velocity reach their final values way before the altitude comes back at its entry value (at $t \approx 100\,\text{s}$ versus 160 s). As a consequence, the altitude profile is no longer symmetric around its lowest point. Even though for the numerical solution it is not symmetric, the lowest point occurs later in the trajectory, giving a more natural feel to the profile. The velocity and flight-path angle histories show a large difference between the two solutions, much more than the altitude profile, and have therefore hardly any practical use. Of the different numerical solutions, simulating a non-rotating Earth has the largest effect. It may be surprising that using an exponential atmosphere has hardly any effect, compared with the supposedly more accurate CIRA model. However, it must be noted that using a scale height of 7,200 m drives the solution in the direction of the analytical one, even if that one would use the same H_s. This reconfirms the earlier conclusion (Example 5.5) that tuning the exponential model with actual atmosphere data is a must. Last but not least, when all simplifications are "switched on" simultaneously the results are virtually the same as those for the non-rotating case, clearly establishing the dominance of this simplification among those simulated. No interactions between the simplifications seem to be present.

As before, in Example 5.13, the main origin of the difference is the assumption that the weight term cannot be neglected, in particular with the combination of low L/D and the relatively shallow entry. Contrary to the earlier results, due to the less favourable L/D, γ_E, in combination with the large mass, the key parameters for the

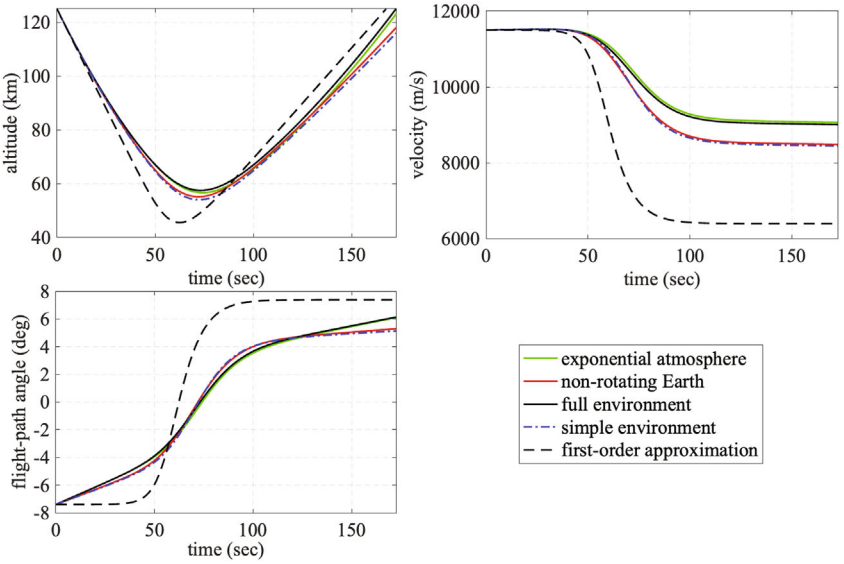


Fig. 5.69 Altitude, velocity, and flight-path angle profiles for various numerical simulations and the first-order analytical approximation

Table 5.19 Tabulation of various variables at three points of interest, taken both from the first-order analytical equations and the numerical simulation

Point of interest	Parameter	First order approx.	Simulation	Ratio [–]
Lowest altitude	Altitude (km)	45.5	57.5	0.8
Peak	Flight-path angle (°)	−2.0	−0.54	3.7
Acceleration	Load factor (g)	23.9	7.0	3.4
Atmospheric	Flight-path angle (°)	7.5	5.7	1.3
Exit	Velocity (m/s)	6,394	9,005	0.7

current example show larger differences, as shown in in Table 5.19. Thus, it can be concluded that the first-order approximation is not suitable for the Earth-CRM case.

To conclude the analysis, we will investigate the assumption of negligible mass effect compared the the aerodynamic effect. Using the data of the numerical simulation, we can determine the contribution of the gravitational and aerodynamic forces to $\dot{V}$ and $\dot{\gamma}$. Subsequently, the ratio of the corresponding components are calculated, defined as "aerodynamic-over-mass", such that a ratio larger than unity means that the aerodynamic acceleration dominates the gravitational one. Figure 5.70 shows the results for both the CRM and SRM.

It can be concluded that for the CRM vehicle flying through the lower parts of the atmosphere, where the aerodynamic forces are the largest, the mass-effect on the rate of change of the velocity can indeed be neglected. However, for the rate of change

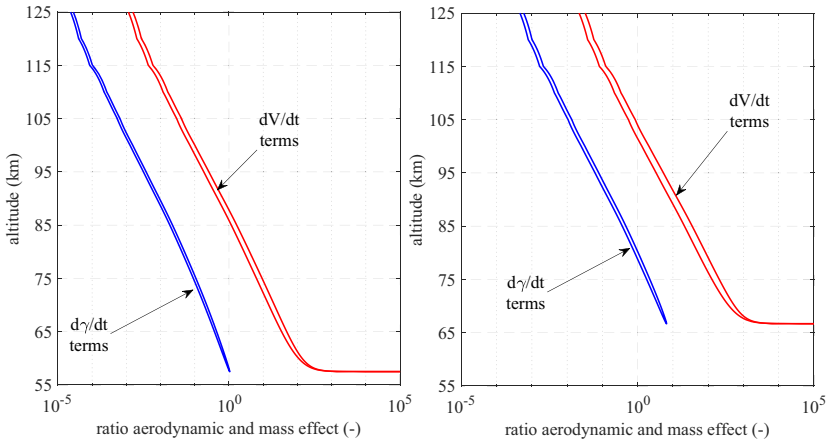


Fig. 5.70 Ratio between the aerodynamic and gravitational contributions in the Equations of Motion governing the rate of change of the velocity and flight-path angle

of the flight-path angle, both effects contribute roughly the same, and thus cannot be neglected. And, since steeper or shallower entries will affect the altitude (and thus density and dynamic pressure), there will also be a coupling with the velocity profile (see again Fig. 5.69). For the Earth-bound CRM it is thus confirmed that the assumption of ignoring the weight term is invalid.

For the Earth-SRM case a similar conclusion can be drawn. Since the aerodynamic reference area of the SRM vehicle about 14% of the CRM reference area, and the aerodynamic forces are proportional to this area, only 14% of the aerodynamic force is generated. But at the same time, the mass of the SRM vehicle is only 0.7% of that of the CRM vehicle. Therefore, the ratio between the aerodynamic and mass effect is increased by roughly one order of magnitude. Despite this, for a large part of the flight, the rates of change of the velocity and the flight-path angle are dominated by the mass-effect. Therefore, for the SRM the same conclusion holds.

Up to this point, we have evaluated the Earth aerocapture missions. In case the vehicles are bound for Venus and Mars, different conditions apply and we should repeat the analysis. For Mars, it was also found that selection of gravity-field model (central field versus constant) has a negligible effect on the trajectory. The error induced by the use of the exponential atmosphere model (Sect. 2.3.3) was smaller compared to Earth, but it is stressed once more that the proper selection of ρ_0 and H_s is important. The effect of assuming a stationary planet is comparable, but since Mars' rotational rate is smaller than that of Earth, the difference is slightly smaller. Finally, the aerodynamic-to-mass effect ratio is smaller than for the Earth cases, for the obvious reason that due to the lower density the aerodynamic forces are comparably smaller, even with the high velocity of aerocapture missions. Despite the lower gravity, it can be concluded that ignoring the weight terms induces larger errors for Mars.

Venus has a rotational rate that is about 240 times lower than Earth's, rotating once in about 5800 h. One can easily understand that the assumption of a non-rotating planet will have hardly any effect on the results. We also found that the error due to the exponential atmosphere model were significantly smaller than those for Earth, when using the custom fit for the middle altitudes discussed in Sect. 2.3.5 (namely $\rho_0 = 186{,}765\,\mathrm{kg/m^3}$ and $H_s = 4.5\,\mathrm{km}$. On Venus, due to the largely increased density, the aerodynamic forces are comparably larger for the same velocity. Because of the larger aerodynamic deceleration, the velocity is lowered quicker and the vehicle will thus reach the denser parts of the atmosphere with a lower velocity. This combined effect cause the ratio between the aerodynamic and mass terms to be slightly larger on Venus compared to Earth, but not yet large enough to justify neglecting the mass-effect for the CRM vehicle. For the SRM vehicle, on the other hand, where the ratio is already larger compared to CRM, the assumption to neglect the mass compared to the aerodynamics is applicable. From this analysis it can thus be concluded that for low vehicle masses in denser atmospheres, like the Venus-SRM case, the assumption of ignoring weight terms can be rightfully made.

5.5.5 *Case Study: Apollo Entry*

Even though Apollo never performed a skip out of the atmosphere, the entry guidance system of Apollo was designed to do just that. In Fig. 5.71 the original entry guidance phases have been indicated. The entry-guidance logic as developed by the MIT Instrumentation Laboratory had seven phases used to guide the Command Module (CM) to the desired landing point (Graves and Harpold 1972). Before the actual entry would commence, in the *Pre-entry Phase* the CM was maintained in a three-axis attitude-hold mode that oriented the vehicle to the aerodynamic trim attitude required for the hypersonic flight phase. The *Initial Roll and Constant-Drag Phase* was initiated at a *g-load* of 0.05 g, and the initial Lift-Vector Orientation was determined to ensure a safe entry and aerodynamic capture within the Earth's atmosphere.

During the *HUNTEST and Constant-Drag Phase* the basic characteristics of the remainder of the entry trajectory were determined, and any excessive energy was dissipated. The skip-out velocity and flight-path angle were predicted based on a predetermined value of the L/D in the vertical plane. During the *Down- and Up-Control Phases* the roll commands are calculated according to different guidance laws, taking the drag and range-rate errors as input, and using analytically established reference trajectories. Subsequently, the *Kepler Phase* was entered, beginning at a *g-load* smaller than 0.2 g. During this phase, the CM was placed in a three-axis hold mode. When the aerodynamic load factor had reached a sufficiently large value again, the CM entered the final guidance phase, notably the *Second-Entry Phase*. The roll commands were generated with a similar guidance law as during the previous phases, and would drive the range potential to the range-to-target until the drogue-parachute deployment altitude was reached.

The above mentioned guidance phases were designed for controlling the vertical lift-to-drag ratio, by rolling the CM around the velocity vector (i.e., banking), and thus creating a coupling between longitudinal and lateral motion. The lateral guidance logic would use the lateral-range error when it would exceed the predicted capability as a means to reverse the bank direction, a concept that was also used by the Space Shuttle, see Sect. 7.6.3.2 for more details.

To study the performance of the guidance system, the Apollo-10 post-flight entry data were compared with the pre-flight data (Graves and Harpold 1972), see also Fig. 5.72. Apollo-10 entered the atmosphere at May 26, 1969, with $V_E = 11.069$ km/s and $\gamma_E = -6.61°$. The range-to-go at entry was 2,395 km (= 1,293 nmi). The guidance computer commanded a lift-up orientation to minimise the first peak load (6.8 g). Then, a lift-down orientation was commanded to keep the *g-load* at a constant 4 g and to avoid an excessive skipping motion. The differences between actual and pre-calculated *g-load* were attributed to the gains in the guidance laws, and were updated for subsequent missions. All in all, the performance of the guidance and control system was deemed to be satisfactory.

Despite the capability of skip guidance, Graves and Harpold (1972) state about the skipping entry: *Positive aerodynamic control of the entry trajectory should be maintained throughout entry. Control to a long-range trajectory that requires a skip out from the atmosphere is a critical manoeuvre that requires precise execution. The consequences of an error in controlling (to) the skip out make this type of trajectory undesirable.*

Therefore, in the remainder of this case study we will look at several aspects of the trajectory sensitivity to the initial flight-path angle and the applied bank angle. To this end we use the model of Apollo (Appendix B). In summary, the Apollo Command

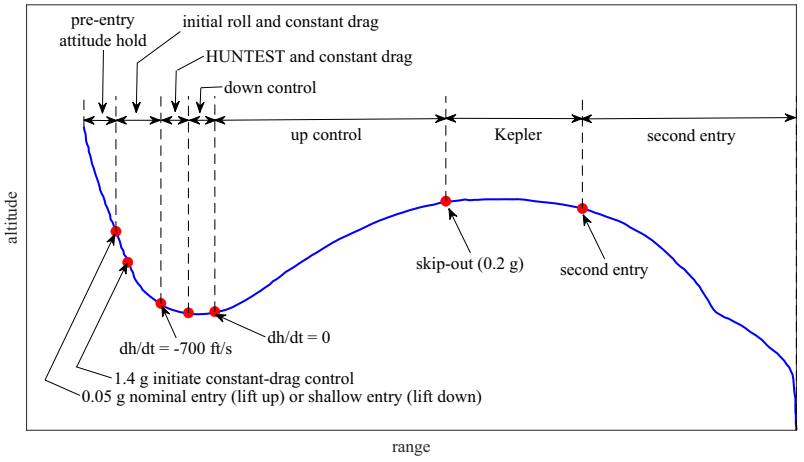


Fig. 5.71 Apollo entry-guidance phases. Data taken from Graves and Harpold (1972)

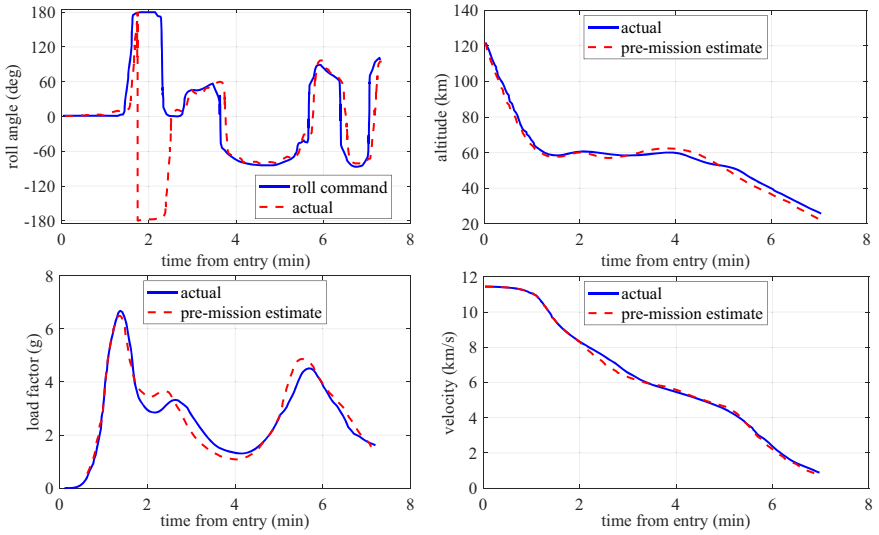


Fig. 5.72 Apollo-10 pre- and post-flight analysis time histories: roll angle and load factor (left), and velocity and altitude. Data taken from Graves and Harpold (1972)

Module, weighing m = 4,976 kg, is on a return flight from the Moon and has the following entry conditions at h_E = 220 km: V_E = 11 km/s and γ_E = -9.563°. Due to the large velocity the capsule should fly in a so-called lift-down configuration that allows for capturing by the atmosphere. This requires a bank angle larger than 90°.

In Fig. 5.73, the results of a so-called uncontrolled 6 degrees-of-freedom simulation are shown (that will also show the hypersonic trim attitude of $\alpha \approx$ -23°). The initial bank angle is σ = 110°, which remains constant for a long time and changes in a natural way to a more lift-up orientation of σ = 60°. Close towards the end of the flight the angle of attack exhibits some minor instability, even though it does not lead to a diverging attitude. The h-V diagram looks like a standard one, and the hypersonic descent shows as a straight line, as is common for low L/D vehicles.

The first sensitivity that we study is the one to a change in (constant) bank angle. We specifically say *constant*, as for this simulation we treat the vehicle as a mass point and only look at the translational motion. We take the same entry conditions as before, freeze the angle of attack at $\alpha_E = -23°$, and vary the bank angle for $\sigma \in$ [0,15°,30°,60°,90°,120°,150°,180°]. The resulting trajectories have been plotted in Fig. 5.74. It is clear that for bank angles from 0° to 60°, all corresponding with a lift-up orientation, the vehicle skips out of the atmosphere, with σ = 0° (maximumm positive lift) reaching the highest altitude (h_{max} = 33,000 km). For bank angles up to σ = 30° the velocity is still so large that a second skip occurs, albeit to a much lower extent. In the right plot of Fig. 5.74 we see the plots for bank angles $\sigma \geq$ 90°, all of which lead to capture in the atmosphere. For σ = 90° there is still a marginal altitude

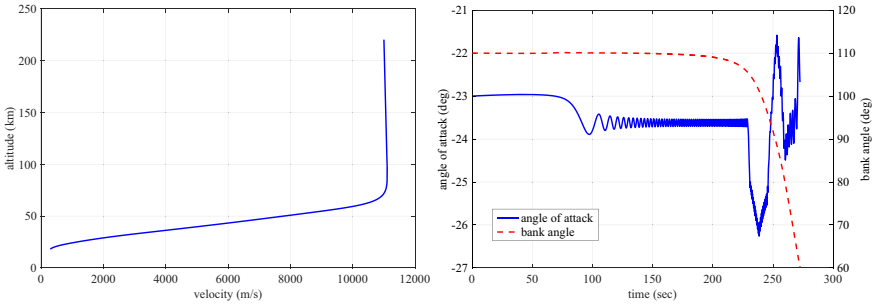


Fig. 5.73 The Apollo altitude-velocity profile (left) and the attitude angles (right), obtained from an uncontrolled entry simulation

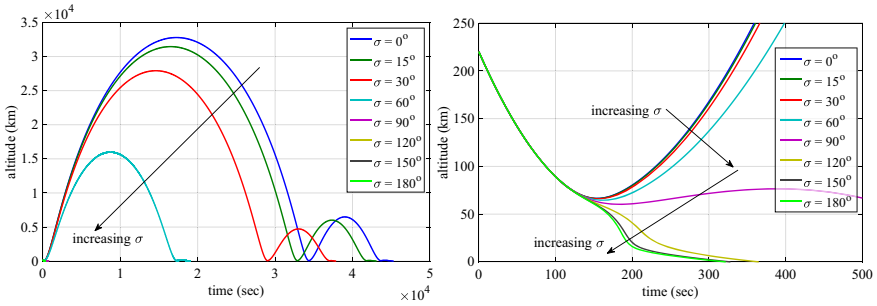


Fig. 5.74 Altitude profiles for bank-angle variations between $\sigma = 0°$ and $\sigma = 180°$ ($V_E = 11$ km/s and $\gamma_E = -9.563°$)

increase due to dynamic effects. Larger bank angles (including the nominal $\sigma = 110°$) will lead to "normal" entry trajectories.

The next sensitivity that we study is one to the initial flight-path angle. Similar entry conditions as before (with σ_E fixed to 110°) we vary $\gamma_E \in [-9.563°, -9.463°, -9.423°, -9.403°, -9.393°, -9.383°, -9.373°, -9.363°, -9.353°]$. We would like to point out the small variation in γ_E, i.e., $\Delta\gamma = 0.21°$ and mention that for $V_E = 11$ km/s a difference of $\Delta\gamma = 0.1°$ means a vertical velocity difference of only $\Delta V = 20$ m/s ($V_{vertical} = -V_E \sin\gamma_E$). This shows the accuracy with which the entry interface has to be met.

In Fig. 5.75 the results are shown. In the left plot one can see the trajectories that do not skip. The limit value for γ_E is $-9.393°$ ($\Delta\gamma = 0.17°$ with respect to the nominal value). In the right plot, the trajectory for this limit value has been plotted as well, but is difficult to spot, because of the skipping trajectories. The shallower the trajectories, the sooner the vehicle will bounce off the atmosphere. The last value of γ_E, i.e., -9.353° ($\equiv \Delta\gamma = 0.21°$), results in a maximum skip altitude of 9,000 km and a significant increase in range.

In the next Fig. 5.76 the mechanical load is plotted versus the altitude, where for reasons of clarity we have constrained the altitude plot range to 100 km. For

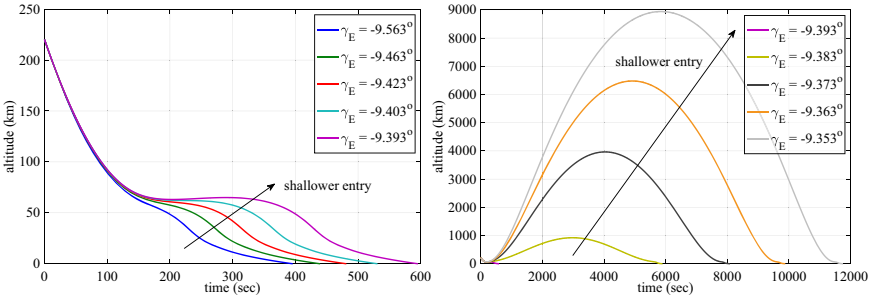


Fig. 5.75 Altitude profiles for flight-path-angle variations between $\gamma_E = -9.563°$ and $-9.353°$ ($V_E = 11$ km/s and $\sigma_E = 110°$)

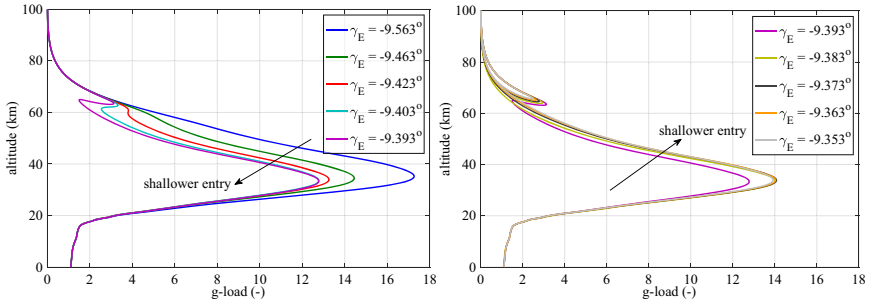


Fig. 5.76 G-load profiles for flight-path-angle variations between $\gamma_E = -9.563°$ and $-9.353°$ ($V_E = 11$ km/s and $\sigma_E = 110°$)

the nominal entry conditions the maximum *g-load* peaks at 17 g. For the shallower entries the peak decreases to value just under 13 g. It is noted that for entries starting at $\gamma_E = -9.423°$ a first peak starts appearing when—due to the "shallowness" of the trajectory—the atmosphere already starts pushing the vehicle, resulting in a more curved trajectory and hence the small deceleration peak. The effect is clearly seen for $\gamma_E = -9.393°$. The curves on the right side show that the maximum deceleration occurs during the second, proper entry, and is slightly larger than the limit (= no-skip) profile of $\gamma_E = -9.393°$, because at the beginning of second entry the path is slightly steeper. For each of the skipping entries the maximum deceleration is the same (i.e., 14 g).

Figure 5.77 shows the profiles for the stagnation-point heat flux. The non-skipping entries have maximum values ranging from 1,800 kW/m^2 ($\gamma_E = -9.563°$) down to 1,400 kW/m^2 ($\gamma_E = -9.393°$), each shallower angle seeing the peak happening at a slightly higher altitude. For the shallower entries in the plot on the right-hand side, the maximum heat flux occurs during the first entry, contrary to the one for the maximum *g-load*. Peak values up to 1,400 kW/m^2, with the smaller values for the shallower entries. During the second entry, still a significant heat flux is experienced,

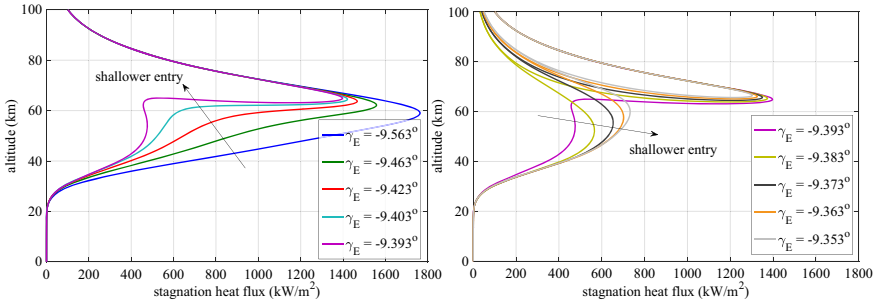


Fig. 5.77 Heat-flux profiles for flight-path-angle variations between $\gamma_E = -9.563°$ and $-9.353°$ ($V_E = 11$ km/s and $\sigma_E = 110°$)

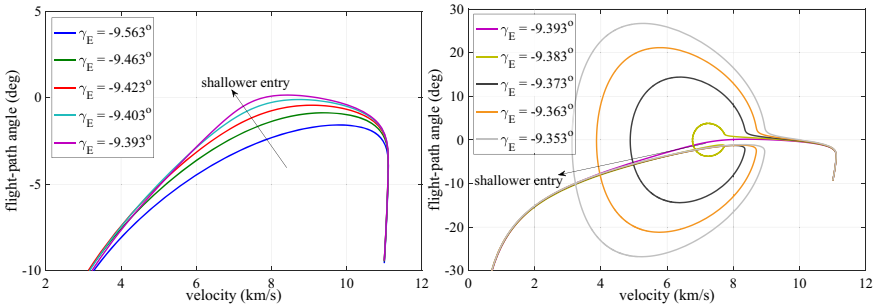


Fig. 5.78 Flight-path angle as a function of velocity for flight-path-angle variations between $\gamma_E = -9.563°$ and $-9.353°$ ($V_E = 11$ km/s and $\sigma_E = 110°$)

with values up to 700 kW/m^2. In this case the larger values happen for the shallower entries, as at the moment of second entry the velocity is larger.

This is very well shown in Fig. 5.78, where the flight-path angle is plotted as a function of the velocity. In the left plot, the "normal" trajectories are shown, with only the one flight-path angle profile exceeding the zero line. However, it is clear that in all cases the atmosphere is pushing the vehicle up, showing as a straight line up from the moment of entry, until the vehicle reaches a state close to horizontal flight. At lower velocities, all trajectories converge to the same profile, which is even better to be seen in the right plot.[12] The Keplerian phase outside the atmosphere—actually the used model extends to 1,000 km altitude, but the density is extremely low—shows as a circular profile, with $\gamma = 0°$ as the top of the trajectory. The shallower entries leave the atmosphere at larger velocities and thus also return into the atmosphere with a larger velocity.

[12] This is one of the advantages of using the velocity as independent variable, because it removes the time aspect from the trajectory. Despite the much longer flight times due to the skipping flight, the nature of the atmospheric flight at lower altitudes does not change, clear from the overlapping trajectories.

5.6 *Second-Order Approximation

Even though the first-order approximations discussed in this chapter give good insight in the behaviour of an entry vehicle in ballistic, gliding, and skipping flight, due to the assumptions made to obtain analytical solutions, certain discrepancies are introduced. It may be obvious that these discrepancies become more apparent when the relative influence of these assumptions is larger. In Sect. 5.2, Example 5.1, it was shown that ignoring the effect of the Earth's rotation, and thus ignoring the Coriolis and centrifugal acceleration, can lead to a large difference in skipping behaviour of the vehicle. Furthermore, Example 5.9 gave us insight in the limitations of analytically determining the longitudinal flight range due to the introduction of the equilibrium-glide condition and ignoring the weight compared to the drag force. A similar problem came forward in Example 5.13, where the limitations of the skip equations of motion were revealed when the vehicle weight could no longer be ignored.

As we have mentioned several times, the benefit of having analytical solutions that are fast to solve, is that they can be implemented in guidance systems to serve as input to mission planning and reference trajectory generation. Provided, of course, that they are accurate enough. Ignoring the effect of weight (in favour of the aerodynamic forces) has proven to be an assumption that is not always justified to make. Keeping the related terms in the starting equations will make for a much more complicated derivation, but will give much more accurate solutions. Several authors have derived these so-called second-order solutions, see, for instance, the works by Chapman (1959), Loh (1968), and Vinh et al. (1993).

To complete the discussion on planar motion, we will provide the reader with this model. The derivation will go a bit beyond the level (and requirements) of a standard textbook, but it may prove to be useful. Not only can it serve as a basis for guidance-system development, it could directly be used to accurately assess skipping flight, as well as aerocapture manoeuvres. The latter manoeuvres will be treated in significant detail in Chap. 11.

5.6.1 Equations of Motion

The derivation starts from the same set of equations as used for deriving the first-order solutions, i.e., those used for describing planar motion over a spherical, non-rotating body, Eqs. (5.8)–(5.10). They are restated here for convenience:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{5.302}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.303}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} = V \sin\gamma \tag{5.304}$$

All mentioned references have in common that the authors derive the equations using non-dimensional variables for both the independent and the dependent variables. In that way the resulting equations become more compact, although at the expense of readability and intuitive understanding. To remain in line with the quoted literature we will follow the same approach, however, in a more summarised way. The full derivation can easily be followed in the paper by Vinh et al. (1993).

The first step is to introduce a new dependent variable, the range angle θ. The range angle is a measure for the ground distance travelled by the vehicle, and is the negative of the auxiliary angle σ, as shown in Fig. 5.2. Following Eq. (5.5), the expression for the rate of change of the range angle is

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \frac{V}{R}\cos\gamma \tag{5.305}$$

Dividing Eqs. (5.302)–(5.304) by Eq. (5.305) changes the independent variable from t to θ:

$$m\frac{\mathrm{d}V}{\mathrm{d}\theta} = -\frac{DR}{V\cos\gamma} - mg\frac{R}{V}\tan\gamma \tag{5.306}$$

$$mV^2\frac{\mathrm{d}\gamma}{\mathrm{d}\theta} = \frac{LR}{\cos\gamma} - mgR\left(1 - \frac{V^2}{V_c^2}\right) \tag{5.307}$$

$$\frac{\mathrm{d}R}{\mathrm{d}\theta} = \frac{\tan\gamma}{R} \tag{5.308}$$

Next, non-dimensional variants for the radial position and velocity are introduced. The radial position is non-dimensionalised by converting it to a normalised inverted altitude or "penetration depth" below the radial position at entry, R_E:

$$\bar{h} = \frac{R - R_E}{R_E} \tag{5.309}$$

So, at the entry interface, the value of $\bar{h}$ is equal to zero and will decrease until it reaches a minimum at the lowest point in the trajectory. In case of a skipping flight, it will increase again until upon leaving the atmosphere $\bar{h} = 0$.

The choice for non-dimensionalising the velocity may be clear after inspecting Eq. (5.307) and is similar to the approach we have seen throughout this chapter:

$$u = \frac{V^2}{V_{c,E}^2} = \frac{V^2}{g_E R_E} \tag{5.310}$$

where $V_{c,E} = \sqrt{g_E R_E}$ is the (local) circular velocity at the entry interface.

To further "simplify" the equations of motion, three additional parameters are introduced, although from this point forward the equations become harder to understand in their compact form. As a consequence, the derivation is simply a mathematical process and we will re-establish the link with physics at the very end. Assume that lift and drag are related through the well-known parabolic drag polar, $C_D = C_{D_0} + KC_L^2$, with $C_{D,0}$ being the parasite drag coefficient and K being the induced-drag parameter.[13] The maximum lift-to-drag ratio is then defined as

$$E^* = \left(\frac{L}{D}\right)_{max} = \frac{1}{2\sqrt{KC_{D_0}}} \tag{5.311}$$

The second constant, B, specifies the entry altitude (through R_E and the corresponding density ρ_E) and physical characteristics of the vehicle (reference area S, mass m, and the aerodynamic constants $C_{D,0}$ and K):

$$B = \frac{\rho_E S R_E}{2m}\sqrt{\frac{C_{D_0}}{K}} \tag{5.312}$$

Because of the very low density at the entry interface, B has a small value as well.[14] The third and last constant is the normalised lift coefficient or lift-control parameter, λ, and allows for lift modulation:

$$\lambda = \frac{C_L}{\sqrt{C_{D_0}/K}} \tag{5.313}$$

It is equal to one when the vehicle is flying at maximum lift-to-drag ratio, zero for ballistic entry, and anything in between for modulated lift. While not mentioned in any of the sources, negative values for λ can be used to model lift-down flight, with negative one for a full lift-down flight, corresponding to a bank angle of 180°. With variations of this parameter, lift modulation due to continuous or discrete bank-angle control can be approximated. It is stressed that only the effect on the vertical lift force will be considered, as the starting equations are valid for planar flight. The induced lateral motion due to banking is not considered in this case.

Substituting the expressions for $\bar{h}$, u, E^*, B and λ into Eqs. (5.306)–(5.308) yields

$$\frac{\mathrm{d}u}{\mathrm{d}\theta} = -\frac{B(1+\bar{h})(1+\lambda^2)ue^{-\beta R_E \bar{h}}}{E^*\cos\gamma} - \frac{2}{1+\bar{h}}\tan\gamma \tag{5.314}$$

$$\frac{\mathrm{d}\gamma}{\mathrm{d}\theta} = \frac{B(1+\bar{h})\lambda e^{-\beta R_E \bar{h}}}{\cos\gamma} - \frac{1}{u(1+\bar{h})} + 1 \tag{5.315}$$

[13] Note that the symbol K is used conform the common notation; however, this symbol is not to be confused with the ballistic parameter, Eq. (5.23).

[14] Note that even though small ($\rho_E \approx 5.8 \cdot 10^{-7}$ kg/m^3 at 100 km altitude), the assumption that $\rho_E \approx 0$ will not be made, for reasons that will become clear later.

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}\theta} = (1 + \bar{h}) \tan\gamma \tag{5.316}$$

where it is noted that the above equations are fully equivalent to the original Eqs. (5.302)–(5.304), albeit with lift modulation included (and assuming a parabolic drag polar for the aerodynamics).

To allow for some simplifications at a later stage, we will change both independent variable θ and dependent variables for velocity, flight-path angle and "altitude" (u, γ, and $\bar{h}$) to τ, ν, ϕ, and y, respectively. To include information regarding the atmosphere in the independent variable, θ is scaled with $\sqrt{\beta R_E}$, such that the new independent variable is

$$\tau = \sqrt{\beta R_E}\theta \tag{5.317}$$

At the start of entry, $\tau = \theta = 0°$. The new dependent variable for velocity is

$$\nu = \frac{1}{\eta} \ln \frac{V_E^2}{V^2} \tag{5.318}$$

where V_E is the entry velocity and the constant η is defined as

$$\eta = \frac{B}{E^* \sqrt{\beta R_E}} = \frac{\rho_E S C_{D_0}}{m} \sqrt{\frac{R_E}{\beta}} \tag{5.319}$$

At atmospheric entry $\nu = 0$ and will become negative when the velocity decreases. The new dependent variable for the flight-path angle is

$$\phi = -\sqrt{\beta R_E} \sin\gamma \tag{5.320}$$

which is positive for negative γ. Finally, the new dependent variable for altitude is

$$y = e^{-\beta R_E \bar{h}} = \frac{\rho}{\rho_E} \tag{5.321}$$

where it has become clear that ρ_E cannot be zero. At atmospheric entry $y = 1$ after which it will increases until it reaches y_{max} at the lowest point in the trajectory, and for a skipping flight it decreases to $y = 1$ at atmospheric exit.

With the introduction of two more constants, i.e.,

$$k = \frac{2E^*}{\sqrt{\beta R_E} B} \qquad \alpha = \frac{g_E R_E}{V_E^2} \tag{5.322}$$

the final form of the non-dimensional equations of motion that still produce the exact same results as Eqs. (5.302)–(5.304) is obtained:

$$\frac{\mathrm{d}\nu}{\mathrm{d}\tau} = \frac{1+\bar{h}}{\cos\gamma}\left(1+\lambda^2\right)y - \frac{k\alpha\phi e^{\nu\eta}}{(1+\bar{h})\cos\gamma} \tag{5.323}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = -(1+\bar{h})B\lambda y + \frac{\cos\gamma}{1+\bar{h}}\alpha e^{\eta\nu} - \cos\gamma \tag{5.324}$$

$$\frac{\mathrm{d}y}{\mathrm{d}\tau} = \frac{1+\bar{h}}{\cos\gamma}y\phi \tag{5.325}$$

Before analytical expressions can be obtained from these rewritten equations of motion, however, two simplifications have to be introduced. The first of these simplifications is the small-angle assumption. By assuming the flight-path angle to be small, the geometric relations involving the *cosine* of the flight-path angle can be simplified to $\cos\gamma \approx 1$. The second simplification has to do with the non-dimensional variable $\bar{h}$. When entering the atmosphere of a planet, the altitude of the vehicle above the surface is negligible compared to the planetary radius. Mathematically, this means that $R - R_E \approx 0$, therefore $\bar{h} \approx 0$, and thus $1 + \bar{h} \approx 1$. Using these simplifications, the above exact equations of motion reduce to

$$\frac{\mathrm{d}\nu}{\mathrm{d}\tau} = \left(1+\lambda^2\right)y - k\alpha\phi e^{\eta\nu} \tag{5.326}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = -B\lambda y - (1-\alpha e^{\eta\nu}) \tag{5.327}$$

$$\frac{\mathrm{d}y}{\mathrm{d}\tau} = y\phi \tag{5.328}$$

These simplified non-dimensionalised equations of motion will be the starting point for the development of the analytical second-order expressions. However, before we get to that we will establish the first-order approximations. This is done by ignoring the variation in speed in the $e^{\eta\nu}$ term. By keeping the velocity equal to the entry velocity we obtain

$$e^{\eta\nu} = e^{\eta/\eta \ln\left(V_E^2/V_E^2\right)} = 1 \tag{5.329}$$

which corresponds with the first term of the Taylor series expansion of the term on the left-hand side. Substituting this results in the following reduced equations:

$$\frac{\mathrm{d}\nu}{\mathrm{d}\tau} = \left(1+\lambda^2\right) y - k\alpha\phi \tag{5.330}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = -B\lambda y - (1-\alpha) \tag{5.331}$$

$$\frac{\mathrm{d}y}{\mathrm{d}\tau} = y\phi \tag{5.332}$$

Dividing Eq. (5.331) by Eq. (5.332) results in

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}y} = \frac{\mathrm{d}\phi}{\mathrm{d}y} = \frac{-B\lambda - (1-\alpha)\frac{1}{y}}{\phi} \tag{5.333}$$

or

$$\phi \mathrm{d}\phi = \left[-B\lambda - (1-\alpha)\frac{1}{y}\right]\mathrm{d}y \tag{5.334}$$

Integrating the above equation from $y(0) = 1$ and $\phi(0) = \phi_E$, to an arbitrary point (y,ϕ) in the trajectory yields

$$\int_{\phi_E}^{\phi} \phi' \mathrm{d}\phi' = \int_{1}^{y} \left[-B\lambda - (1-\alpha)\frac{1}{y'}\right]\mathrm{d}y' \quad \Rightarrow \tag{5.335}$$

$$\frac{1}{2}(\phi^2 - \phi_E^2) = (-B\lambda y - (1-\alpha)\ln y) + B\lambda \quad \Rightarrow \tag{5.336}$$

$$\phi^2 = \phi_E^2 - 2B\lambda(y-1) - 2(1-\alpha)\ln y \tag{5.337}$$

Finally, introducing a new variable $x = \ln y$, using the definition of y in Eq. (5.321), so

$$x = \ln y = -\beta R_E \bar{h} = \frac{R_E - R}{R_E} \tag{5.338}$$

and substituting this in Eq. (5.337) yields

$$\phi^2 = \phi_E^2 - 2B\lambda\left(e^x - 1\right) - 2(1-\alpha)x \tag{5.339}$$

This new independent variable x is yet another measure for the altitude of the vehicle during entry. At the entry interface, $x = 0$, after which it increases to its maximum value x_p, although it remains relatively small. In skipping flight, after reaching the lowest point in the atmosphere it will decrease back to 0.

By replacing y in Eq. (5.332) by x, and using the definition for ϕ from Eq. (5.339), the following relation is obtained:

$$d\tau = \frac{dx}{\phi} = \frac{dx}{\pm\sqrt{\phi_E^2 - 2B\lambda\,(e^x - 1) - 2(1-\alpha)x}} \tag{5.340}$$

which can be used to get an expression relating x to τ. However, in the process of obtaining this expression, Vinh et al. (1993) approximate ϕ^2 with a second-order polynomial in x. The general form of this polynomial is

$$\phi_a^2 = a_1 x^2 + a_2 x + a_3 \tag{5.341}$$

such that the relation between x and τ takes the form

$$x = -\frac{a_2}{a_1} - \frac{\sqrt{a_2^2 - 4a_1a_3}}{2a_2}\sin\left(\sqrt{-a_1}\tau - \beta_\phi\right) \tag{5.342}$$

where β_ϕ is defined as

$$\sin\beta_\phi = \frac{a_2}{\sqrt{a_2^2 - 4a_1a_3}} \tag{5.343}$$

At this point Vinh et al. (1993) state that the three coefficients can be determined through any approximation scheme, although they give an analytical solution for the three coefficients as well, namely

$$a_1 = \frac{3}{x_p}\left[2(1-\alpha) - \frac{\phi_E^2 + 4B\lambda}{x_p} + \frac{4B\lambda\,(e^{x_p} - 1)}{x_p^2}\right] \tag{5.344}$$

$$a_2 = -6(1-\alpha) + \frac{2\left(\phi_E^2 + 6B\lambda\right)}{x_p} - \frac{12B\lambda\,(e^{x_p} - 1)}{x_p^2} \tag{5.345}$$

$$a_3 = \phi_E^2 \tag{5.346}$$

In these three equations, x_p is the value for x at the lowest point p in the trajectory and can be solved from Eq. (5.339) by putting $\phi^2 = 0$, which is the situation for the lowest point in the trajectory:

$$\phi_E^2 - 2B\lambda\left(e^{x_p} - 1\right) - 2(1-\alpha)x_p = 0 \tag{5.347}$$

This equation can only be solved numerically, e.g., by applying a root-finding scheme. Another way to solve it is by using an approximation. Since a first-order approximation for the altitude of the lowest point in the trajectory was already obtained in Sect. 5.5.2, Eq. (5.239), this approximation can be used to determine x_p, and thus a_1, a_2, and a_3 can be evaluated.

So far, a first-order approximation for the flight-path angle as a function of the altitude, Eq. (5.339), and for the altitude as a function of the range angle, Eq. (5.342),

has been obtained, the latter of which can easily be inverted to obtain the range angle as a function of altitude. The last remaining expression to derive is one for the velocity. This can be obtained by first inverting Eq. (5.331) to obtain an expression for y, inserting this into Eq. (5.330), followed by inserting the expression for ϕ as shown in Eq. (5.340), and lastly integrating the resulting equation from atmospheric entry to an arbitrary point in the trajectory. The reader is referred to the paper by Vinh et al. (1993) for more details; here, we simply state the end result, i.e., the first-order approximation for the velocity as a function of altitude and range angle:

$$\nu = \frac{\left(1+\lambda^2\right)}{B\lambda}(\phi_E - \phi) - k\alpha x - \frac{(1-\alpha)\left(1+\lambda^2\right)}{B\lambda}\tau \tag{5.348}$$

It should be noted that in this equation ϕ is required, and not ϕ^2. Equation (5.339) gives us ϕ^2, so the solution to this is $\pm\phi$ at best. However, with ϕ being the negative *sine* of the flight-path angle, and this operation retains the same sign as its argument for the range of angles considered valid for the flight-path angle ($\pm 90°$), the sign of ϕ is positive for the descending leg of the trajectory, and negative for the ascending leg.

With the first-order approximations developed, the accuracy of the solutions can be improved by taking the $e^{\eta\nu}$ term into account, again following the approach of Vinh et al. (1993). To facilitate finding analytical expressions, $e^{\eta\nu}$ is approximated by the first two terms of its Taylor series rather than just the first term, resulting in $e^{\eta\nu} \approx 1 + \eta\nu$. This yields the second-order approximations of the non-dimensionalised equations of motion:

$$\frac{\mathrm{d}\nu}{\mathrm{d}\tau} = \left(1+\lambda^2\right) y - k\alpha\phi - \eta k\alpha\phi\nu \tag{5.349}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = -B\lambda y - (1-\alpha) + \eta\alpha\nu \tag{5.350}$$

$$\frac{\mathrm{d}y}{\mathrm{d}\tau} = y\phi \tag{5.351}$$

Following the same procedure as before, Eq. (5.350) is divided by Eq. (5.351), and the independent variable y is replaced by x. Due to the inclusion of the second term of the Taylor series expansion, now ν appears in the differential equation for ϕ^2:

$$\frac{\mathrm{d}\phi^2}{\mathrm{d}x} = -2B\lambda e^x - 2(1-\alpha) + 2\eta\alpha\nu \tag{5.352}$$

Fortunately, Eq. (5.348) can be used for the "perturbation" on the right-hand side of Eq. (5.352) to obtain a relation that only depends on x and ϕ. Integrating this expression from atmospheric entry to an arbitrary point in the trajectory, an improved approximation for ϕ^2 can be obtained:

$$\phi^2 = c^2 - 2B\lambda\left(e^x - 1\right) - 2(1-\alpha)x + 2\eta\alpha I(x) \tag{5.353}$$

where $I = \int_0^x \nu dx'$, which, when worked out, results in

$$\begin{aligned} I = {} & \frac{\left(1+\lambda^2\right)}{B\lambda}\phi_E x - \frac{1}{2}k\alpha x^2 - \frac{(1-\alpha)\left(1+\lambda^2\right)}{B\lambda}\tau x + \\ & - \frac{\left(1+\lambda^2\right)}{B\lambda} I_1(x) + \frac{(1-\alpha)\left(1+\lambda^2\right)}{B\lambda} I_2(x) \end{aligned} \tag{5.354}$$

where $I_1 = \int_0^x \phi \mathrm{d}x' \approx \int_0^x \phi_a \mathrm{d}x'$, and $I_2 = \int_0^x \frac{x'\mathrm{d}x'}{\phi} \approx \int_0^x \frac{x'\mathrm{d}x'}{\phi_a}$, which gives, after integration:

$$I_1 = \frac{1}{2}\phi_a x - \frac{a_2}{4a_1}(\phi_E - \phi_a) + \frac{\left(4a_1a_3 - a_2^2\right)}{8a_1}\tau \tag{5.355}$$

$$I_2 = -\frac{1}{a_1}(\phi_E - \phi_a) - \frac{a_2}{2a_1}\tau \tag{5.356}$$

In these equations, the polynomial coefficients a_1, a_2, and a_3 have been derived before, Eqs. (5.344)–(5.346).

To obtain the second-order approximation for the non-dimensionalised velocity, the same procedure as before is followed. Similar to the derivation for the flight-path angle, the resulting equation is the same as the first-order equivalent with the addition of a term involving ν for which the first-order approximation is used. After integration, the final second-order analytical expression for the velocity as a function of altitude and range angle is

$$\nu = \frac{\left(1+\lambda^2\right)}{B\lambda}(\phi_E - \phi) - k\alpha x - \frac{(1-\alpha)\left(1+\lambda^2\right)}{B\lambda}\tau - \eta k\alpha I(x) + \frac{\eta\alpha\left(1+\lambda^2\right)}{B\lambda} J(x) \tag{5.357}$$

where I is the same as before, and $J(x) = \int_0^x \nu \mathrm{d}\tau$, which gives

$$J = \frac{\left(1+\lambda^2\right)}{B\lambda}\phi_E\tau - \frac{\left(1+\lambda^2\right)}{B\lambda}x - \frac{(1-\alpha)\left(1+\lambda^2\right)}{2B\lambda}\tau^2 - k\alpha I_2(x) \tag{5.358}$$

with I_2 given by Eq. (5.356).

An improved expression for the altitude as a function of the range angle is obtained by using the same three equations for the polynomial coefficients, Eqs. (5.344)–(5.346), but with a better approximation of x_p evaluated with the second-order approximation of the flight-path angle. A root-finding scheme will be required to solve x from $\phi^2(x) = 0$. Vinh et al. (1993) mention that for low L/D vehicles a

problem arises with the new approximation. The first-order approximation for x_p will be smaller than the more accurate second-order value, so during the root-finding scheme, values for x will be tested that are above x_p. As a reminder, a larger value of x indicates a lower point in the atmosphere. Therefore, stating that the value obtain for x_p from the first-order estimate is smaller than the more accurate second-order value means that the vehicle dives deeper into the atmosphere than expected from the first-order estimate. Whenever values for x are tested that are in excess of x_p, the normally positive $\phi_a^2(x)$ that is zero at x_p, will become negative for values in excess of x_p, and this leads to erroneous results. To remedy this, in case a value of x above x_p is tested, the value for x used in the $2\eta\alpha I(x)$ term should be held constant at x_p.

Using this updated value of x_p, a_1, a_2, and a_3 can be re-calculated and used to improve the second-order approximations. In theory, the process of computing an updated x_p, recalculating the polynomial coefficients, and updating the second-order approximations could be continued indefinitely. However, no significant increase in accuracy will be obtained after the first correction step.

The remaining tasks to complete before implementing these equations, is to make τ the only independent variable and to find expressions for the aerodynamic forces and heat load. The advantage, when these are based on the same independent variables and constants as the other equations, is that they can all be evaluated at once. Rewriting the found equations with τ as the independent variable is a rather straightforward process, the result of which is given below.

$$x(\tau) = -\frac{a_2}{2a_1} - \frac{\sqrt{a_2^2 - 4a_1a_3}}{2a_1} \sin\left(\sqrt{-a_1}\tau - \arcsin\frac{a_2}{\sqrt{a_2^2 - 4a_1a_3}}\right) \tag{5.359}$$

with a_1, a_2, and a_3 given by Eqs. (5.344)–(5.346).

$$\phi(\tau, \delta) = \delta\sqrt{\phi^2(\tau, \delta)} \tag{5.360}$$

with

$$\phi^2(\tau, \delta) = \phi_E^2 - 2B\lambda\left(e^{x(\tau)} - 1\right) - 2(1 - \alpha)x(\tau) + 2\eta\alpha I(\tau, \delta) \tag{5.361}$$

and where, for the sake of convenience, the short-hand variable $\delta = 2(1 - \alpha)$ has been introduced; α has been defined in Eq. (5.322). Further, B and η are defined by Eqs. (5.312) and (5.319). The auxiliary expressions required to evaluate ϕ are given by:

$$I(\tau,\delta) = \frac{\left(1+\lambda^2\right)}{B\lambda}\phi_E x(\tau) - \frac{1}{2}k\alpha x(\tau)^2 - \frac{(1-\alpha)\left(1+\lambda^2\right)}{B\lambda}\tau x(\tau) + \\ - \frac{\left(1+\lambda^2\right)}{B\lambda}I_1(\tau,\delta) + \frac{(1-\alpha)\left(1+\lambda^2\right)}{B\lambda}I_2(\tau,\delta) \tag{5.362}$$

$$I_1(\tau,\delta) = \frac{1}{2}\phi_a(\tau,\delta)x(\tau) - \frac{a_2}{4a_1}(\phi_E - \phi_a(\tau,\delta)) + \frac{\left(4a_1a_3 - a_2^2\right)}{8a_1}\tau \tag{5.363}$$

$$I_2(\tau,\delta) = -\frac{1}{a_1}(\phi_E - \phi_a(\tau,\delta)) - \frac{a_2}{2a_1}\tau \tag{5.364}$$

$$\phi_a^2(\tau) = a_1x(\tau)^2 + a_2x(\tau) + a_3 \tag{5.365}$$

$$\phi_a(\tau,\delta) = \delta\sqrt{\phi_a^2(\tau)} \tag{5.366}$$

with λ and k defined by Eqs. (5.313) and Eq. (5.322), respectively.

Finally,

$$\nu(\tau,\delta) = \frac{\left(1+\lambda^2\right)}{B\lambda}(\phi_E - \phi(\tau,\delta)) - k\alpha x(\tau) - \frac{(1-\alpha)\left(1+\lambda^2\right)}{B\lambda}\tau + \\ - \eta k\alpha I(\tau,\delta) + \frac{\eta\alpha\left(1+\lambda^2\right)}{B\lambda}J(\tau,\delta) \tag{5.367}$$

with

$$J(\tau,\delta) = \frac{\left(1+\lambda^2\right)}{B\lambda}\phi_E\tau - \frac{\left(1+\lambda^2\right)}{B\lambda}x(\tau) - \frac{(1-\alpha)\left(1+\lambda^2\right)}{2B\lambda}\tau^2 - k\alpha I_2(\tau,\delta) \tag{5.368}$$

Each of the above equations is only dependent on the known entry conditions, range-angle τ, and the auxiliary parameter δ. This parameter can take the value of 1 or -1, and specifies whether the flight is in the downwards or upwards leg, respectively. At atmospheric entry, x is zero and increases to its maximum value at the lowest point in the atmosphere, after which it decreases to zero again at atmospheric exit. With this known, the range of τ for which these equations can be evaluated can be determined by establishing values of τ for which $x(\tau) = 0$. The first root is known from the previously given definitions. At atmospheric entry with $\sqrt{-a_1}\tau = 0$, $x(\tau) = 0$. By considering the cyclic behaviour of the *sine* function, one set of roots can be expressed as $\sqrt{-a_1}\tau = 2\pi k$, where $k \in \mathbb{Z}$. The first root is part of this set. Similarly, a second set of roots can be expressed as $\sqrt{-a_1}\tau = (1+2k)\pi + 2\beta_\phi$, where $\beta_\phi = \arcsin\frac{a_2}{\sqrt{a_2^2-4a_1a_3}}$ and $k \in \mathbb{Z}$. Because only the first and second root are of interest (atmospheric entry and exit), only the smallest two roots described by

these two sets are of interest. For atmospheric entry and exit, these roots are $\tau = 0$ and $\tau = \pi + 2\arcsin \frac{a_2}{\sqrt{a_2^2 - 4a_1a_3}} \frac{1}{\sqrt{-a_1}}$.

When the above equations are evaluated from atmospheric entry to exit, the numerical results are not easily understood. Therefore, we conclude this section by stating the equations required to transform back to the relative spherical components, V, γ, and R.

$$V(\tau, \delta) = \sqrt{\frac{V_E^2}{e^{\eta\nu(\tau,\delta)}}} \tag{5.369}$$

$$\gamma(\tau, \delta) = \arcsin\left(\frac{\phi(\tau, \delta)}{-\sqrt{\beta R_E}}\right) \tag{5.370}$$

$$R(\tau) = R_E - \frac{x(\tau)}{\beta} \tag{5.371}$$

The used constants in these expresions are the velocity and radial position at entry, i.e., V_E and R_E, as well as the inverse of the atmospheric scale height, $\beta = \frac{1}{H_s}$.

5.6.2 *Verification of an Aerocapture Mission—Revisited*

To get an impression of how much more accurate the second-order solution is compared to the first-order solution, we will redo the verification of the aerocapture mission, as discussed in Sect. 5.5.4. The two vehicles, the Crew Return Mission (CRM) and the Sample Return Mission (SRM) have been defined in Table 5.17. The different parameters for an Earth, Venus, and Mars mission can be found in Table 5.18. We will follow a similar approach as before, i.e., we will integrate the equations of motion used to derive the second-order solution, based on exactly the same assumptions. This will guarantee that smooth solutions are obtained, and that these solutions exhibit the same limitations as the analytical model. From the "numerical" side, we will do two simulations, one with the complete model, i.e., for the Earth-based analysis, a rotating, spherical planet with the CIRA-86 atmosphere model and an inverse square gravity field, and one with a "simplified" model, i.e., a non-rotating planet with an exponential atmosphere. Focus of the discussion will be on the Earth CRM, but some general remarks will be made for the SRM, as well as the effect of executing the mission on Venus or Mars.

The output that we consider, besides the trajectory parameters (altitude, velocity and flight-path angle histories), include the aerodynamic load (drag and lift acceleration, expressed as $g - load$) and the thermal load, divided into a convective and a

radiative component, both evaluated for the stagnation point. As has been discussed in Sect. 3.6, for high entry velocities the radiative heat flux becomes the dominating component in the heat balance. The convective model is given by a simplified cold-wall model assuming a laminar boundary layer, i.e.,

$$q_c = c^* \sqrt{\rho} V^{3.15} \tag{5.372}$$

with $c^* = 5.16 \cdot 10^{-5}$, yielding a heat flux in W/m^2. The radiative heat flux emitted by the gas surrounding the vehicle, $q_{rad,gas}$, will be evaluated using the Tauber-Sutton model.

Starting with the trajectory, shown in Fig. 5.79, and comparing this with the results obtained earlier, it is evident that the second-order analytical expressions match the simplified numerical results almost exactly throughout the downwards leg of the flight. Only when the vehicle starts to climb again, the two solutions start to diverge. The use of the second-order analytical equations has improved the accuracy of the model significantly compared to the first-order analytical expressions. The analytical equations still overestimate the penetration depth into the atmosphere, and give a lower velocity at the atmospheric exit. However, the flight-path angle at atmospheric exit matches quite well. In Table 5.20 the values for the penetration depth and the values at atmospheric exit are tabulated for both numerical simulations and the analytical approximation.

The next Fig. 5.80 shows the aerodynamic load on the vehicle as a function of altitude, separate into a lift and a drag contribution. Also now, the close agreement between the approximate analytical and simplified numerical methods is obvious, but the results are not even that far off from the full numerical simulation: slightly shifted in altitude and a peak error that is about 25–30%. While the approximation

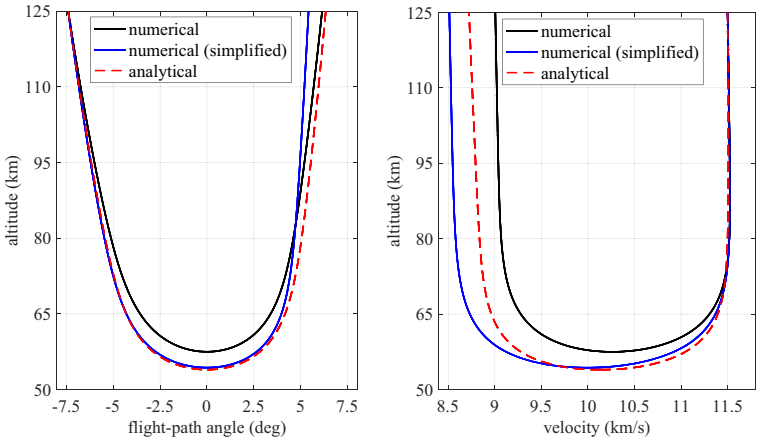


Fig. 5.79 Comparison between numerical and second-order analytical results for the trajectory, considering a rotating Earth, an eastward flight, and the CRM vehicle

Table 5.20 Tabulation of various variables at three points of interest, taken from the second-order analytical equations and both numerical simulations

Point of interest	Parameter	Unit	2nd Order approx.	Numerical (simplified)	Numerical
Lowest point	h_p	(km)	53.96	54.35	57.50
Peak	γ	(°)	−0.62	−0.66	−0.54
Acceleration	Lift	(g)	3.83	3.76	2.83
Load	Drag	(g)	8.70	8.54	6.43
Peak convective	γ	(°)	−1.63	−1.78	−5.84
Heat flux	$q_{c,max}$	(MW/m^2)	2.61	2.54	2.30
Peak radiative	γ_{q_c}	(°)	−2.21	−2.43	−5.86
Heat flux	$q_{rad,gas_{max}}$	(MW/m^2)	12.50	11.39	9.41
Peak total	$\gamma_{q_{rad}}$	(°)	−2.18	−2.41	−5.85
Heat flux	$q_{tot,max}$	(MW/m^2)	15.08	13.89	11.69
Atmospheric	γ_F	(°)	6.33	5.42	6.14
Exit	V_F	(km/s)	8.72	8.51	9.00

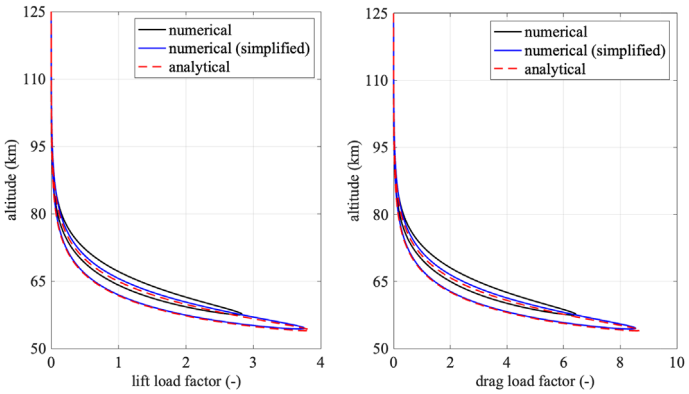


Fig. 5.80 Convective and radiative heat flux for the Earth Crew Return Mission

lacks accuracy in this situation, it is at least an overestimate and thus a conservative upper-bound of the peak acceleration experienced by the vehicle during the trajectory. Table 5.20 also includes the values for the peak aerodynamic load and the flight-path angles at which they occur, listed for all three solutions.

Finally, in Fig. 5.81, the aeroheating parameters have been plotted. Similar conclusions as for the aerodynamic loads can be drawn. The figure shows a close agreement between the analytical approximation and the simplified numerical simulation, while overestimating the full numerical one. In this case the overestimate is approximately 28%, which can also in this case serve as a conservative upper-bound of the peak-heat load. The shift in altitude—and flight-path angle—is more prominent than was the case for the aerodyanamic load, but if we look at the numerical results only, the actual

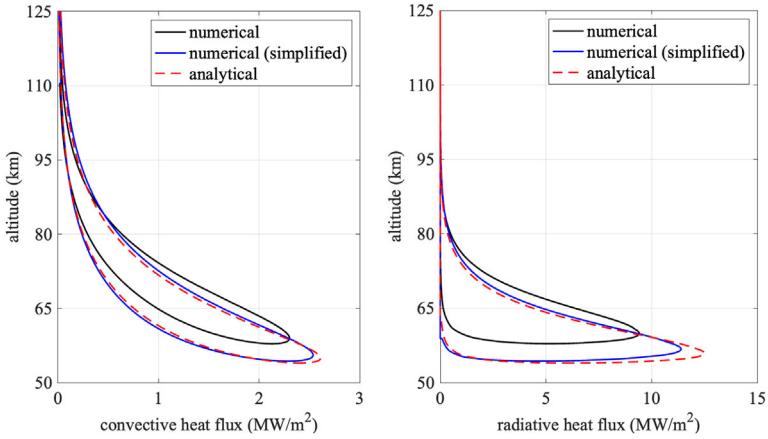


Fig. 5.81 Convective and radiative heat flux for the Earth Crew Return Mission

altitude for peak heating is still happening before the mechanical loading, which is a little bit more obscure for the analytical case. Peaking at a higher altitude has, of course, also an impact on the flight-path angle for which this occurs. In Table 5.20 the values for the peak heat load and the flight-path angles at which they occur have been summarised.

In conclusion, it can be stated that the second-order solution approximates the numerical results a lot better, because the major issue with the first-order solution, i.e., leaving out the weight term in the acceleration balance, has been solved. Not including the Earth's rotation has been identified to be the major contribution to the remaining discrepancies. Considering the fact that the analytical formulation yields very fast solutions compared to solving the differential equations of motion, the second-order solution could potentially be used as a predictor in on-board guidance systems for this type of mission. Of course, the robustness of such algorithm should be addressed by looking at different values for γ_E, V_E, L/D, and m. Also the fact that in the above simulations only eastward equatorial entries were performed should not be overlooked.

To get a first impression of applying the second-order solution in other conditions, a brief sensitivity analysis will be executed next. This analysis will include eastward and westward entries, a range of entry flight-path angles, the inclusion of Venus and Mars, and finally it will be done for both CRM and SRM. Preliminary analyses show that for a westward equatorial entry the analytical approximation loses its conservative nature. In that case the analytical approximations become slightly optimistic, underestimating the aerodynamic loads and heat flux by approximately 12% and 3%, respectively. However, it should be mentioned that a westward equatorial entry is the absolute worst-case scenario, as the rotational velocity of Earth is at its maximum. A second observation is that the second-order approximation seems to be getting better with steeper entries.

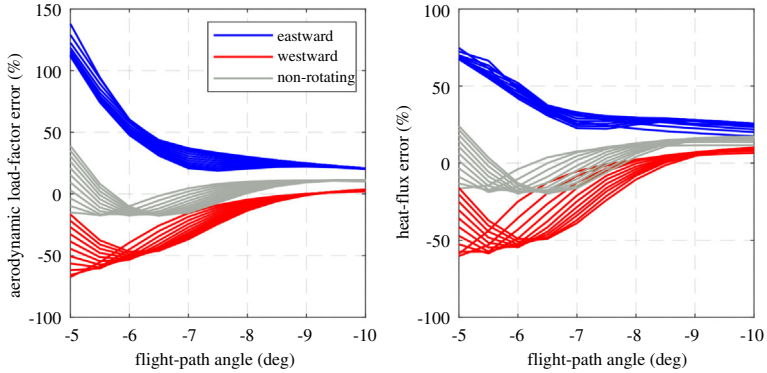


Fig. 5.82 Relative error between the second-order approximation and the numerical simulation (Earth, CRM)

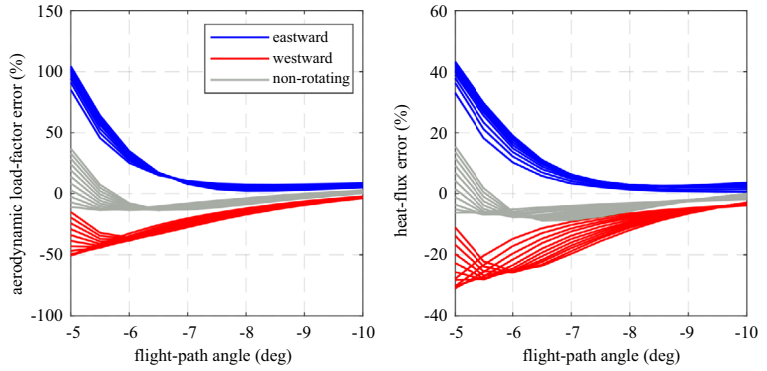


Fig. 5.83 Relative error between the second-order approximation and the numerical simulation (Earth, SRM)

The sensitivity analysis set up as follows. For each planet (Earth, Mars and Venus) and vehicle (CRM and SRM) the entry conditions are varied, i.e., for Earth $\gamma_E \in [-10, -5]^\circ$ (with steps of 0.5°) and $V_E \in [11{,}000, 16{,}000]$ m/s (with steps of 500 m/s). For Mars, the initial conditions change according to $\gamma_E \in [-13, -8]^\circ$ and $V_E \in [6{,}000, 10{,}000]$ m/s, whereas for Venus this is $\gamma_E \in [-11, -6.5]^\circ$ and $V_E \in [10{,}000, 14{,}000]$ m/s. Per set, three other variations are applied: the standard eastward entry, a westward one, and finally a simulation where the planet is non-rotating.

Figures 5.82 and 5.83 depict the error in the peak aerodynamic load factor and peak heat flux for the Earth CRM and SRM cases, respectively. A positive error means the analytical approximations overestimate the peak value, a negative error indicates an underestimate. Similarly, in Figs. 5.84 and 5.85, and 5.86 and 5.87 the corresponding errors for Mars (CRM and SRM), and Venus (CRM and SRM) are shown.

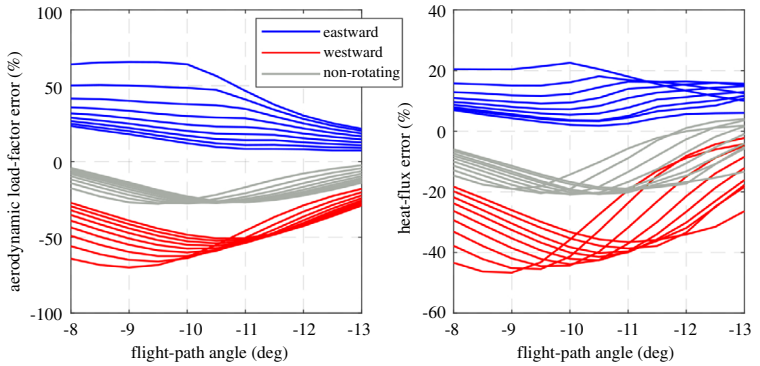


Fig. 5.84 Relative error between the second-order approximation and the numerical simulation (Mars, CRM)

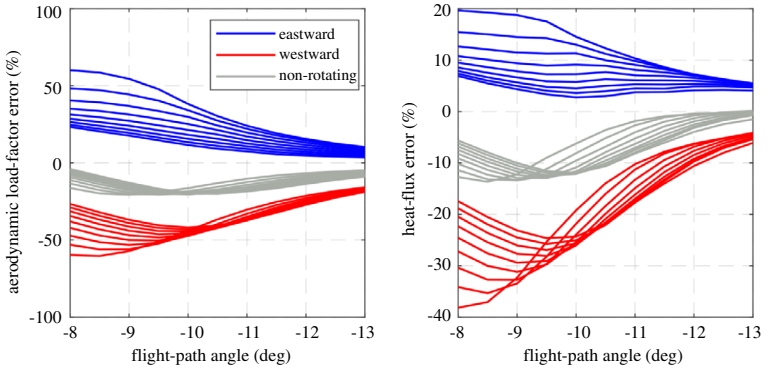


Fig. 5.85 Relative error between the second-order approximation and the numerical simulation (Mars, SRM)

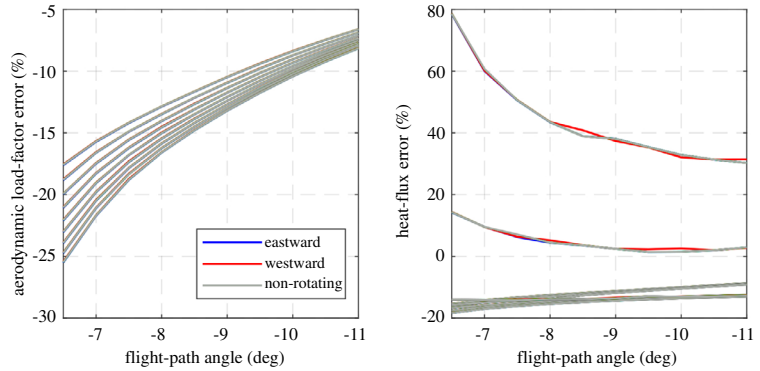


Fig. 5.86 Relative error between the second-order approximation and the numerical simulation (Venus, CRM)

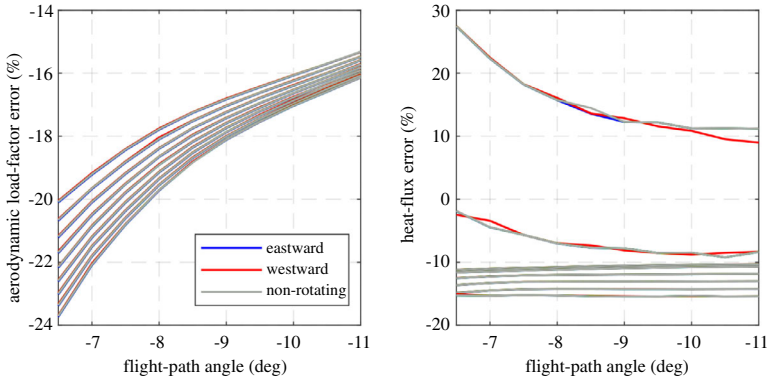


Fig. 5.87 Relative error between the second-order approximation and the numerical simulation (Venus, SRM)

From these figures several observations can be made. The first observation confirms what we found in the preliminary analyses, namely that when the entry becomes steeper, the second-order analytical approximations seems to be getting more accurate. This is most apparent in the two Earth cases. This is contrary to what was expected, however, as the second-order analytical approximations were developed using the small-angle assumption, which should induce a larger error as the flight-path angle increased. The reason for this phenomenon was found to be due to the errors in the exponential atmosphere model used, and not related to the small-angle approximation. The parameters of the exponential atmosphere model were selected for their accuracy around the expected point of deepest penetration, but this was only done once for the nominal entry conditions. For a shallow entry, the vehicle will not dive so deep into the atmosphere, which means that the vehicle does not reach the part of the atmosphere where the exponential model is the most accurate, resulting in larger errors. Additionally, as it appeared to be, in the upper atmosphere (90–125 km), the error made by the exponential atmosphere for Earth is greater than the error for either Venus or Mars.[15]

A second observation is that for the Venus case, the difference between westward or eastward flight or flight about a non-rotating planet is almost negligible. It is known that Venus has a much lower rotational rate than Earth and Mars. On top of that, it rotates in the opposite direction, but with the low rate this hardly matters. The third observation is that for the eastward flight on both Earth and Mars, the analytical approximation always overestimates the error. For a flight about a non-rotating planet, the error fluctuates between over- and underestimate, and the westward flight is

[15] For an implementation in an on-board guidance system the applied exponential atmosphere model should be tuned to the entry conditions, either during mission planning, or on-board using a simple estimator. However, for planets like Mars and Venus the density-versus-altitude profile exhibits a significant change in slope, which could require a dynamic update of the model parameters. In that way a better fit over the entire altitude range can be acquired, which improves the accuracy of the analytical approximations.

underestimated for all but the most steep entries. For Venus these statements are not true: the aerodynamic load factor is always underestimated, and the status of the heat-flux error depends on the velocity with which the entry occurred.

Problems

Judge whether the following statements are TRUE or FALSE (Problems 5.1 to 5.9).

5.1. Skipping entry has never been executed during a real-life mission.
5.2. A downwards lift force can only be obtained with a negative angle of attack.
5.3. While flying along the heat-flux constraint at zero bank angle, one cannot compensate for a larger heat flux.
5.4. For a gliding entry, the maximum *g-load* increases proportionally with the lift-to-drag ratio.
5.5. To escape from the sphere of influence of the Earth, it is sufficient to enter the atmosphere in a skipping flight with a velocity of larger than 11.2 km/s.
5.6. For any atmospheric entry, the peak mechanical load will always occur after the peak heat load.
5.7. During skipping flight, the normal acceleration is always of secondary importance.
5.8. Banking is the process of rolling around the X-axis of the body frame.
5.9. Given the nose radius, a shift in altitude-velocity profile will not affect the heat flux.

Equations of motion (Problem 5.10)

5.10. From memory, write down the three equations of planar motion.[16]

Ballistic entry (Problems 5.11 to 5.14).

Consider an MSL-like entry in the Martian atmosphere. The vehicle has the following characteristics: mass $m = 2{,}400\,\text{kg}$, drag coefficient $C_D = 1.45$, reference area $S = 16.6\,\text{m}^2$ (diameter $d = 4.6\,\text{m}$). The atmospheric density is assumed to follow an exponential profile with the following characteristics: $\rho_0 = 0.02\,\text{kg/m}^3$ (reference density), $H_s = 12{,}100\,\text{m}$ (scale height), $a = 240\,\text{m/s}$ (speed of sound). The (assumed constant) gravity on Mars is $g_0 = 3.69\,\text{m/s}^2$.

5.11. This problem is an exercise for doing standard derivations by yourself (try to do this from memory and using sound reasoning). Starting from the equations obtained under Problem 5.10, show that for a ballistic flight, the relation between velocity and altitude is given by

[16] Since these equations are at the heart of virtually all flight-mechanics problems in this book, it is recommended to spend some time on going through the derivation in Chap. 5 to make sure they can be reproduced from memory.

Table 5.21 Approximate planetary data

Parameter	Venus	Earth	Mars
R_e (10^3 m)	6,052	6,378	3,390
μ (10^{14} m^3/s^2)	3.2463	3.9832	0.4280
g_0 (m/s^2)	8.86	9.79	3.72
ρ_0 (kg/m^3)	64.79	1.225	0.0168
H_s (m)	15,900	7,050	8,800

$$\frac{V}{V_E} = \exp\left(\frac{1}{2}\frac{g_0\rho_0 e^{-\beta h}}{K\beta\sin\gamma_E}\right)$$

Clearly state any assumption that you make. Use can be made of the hydrostatic equation, $\mathrm{d}p = -\rho g \mathrm{d}h$, the ideal gas law, $p = \rho RT$, and the fact that the atmosphere follows an exponential density profile, $\rho = \rho_0 e^{-\beta h}$.

5.12. The vehicle enters the atmosphere with $V_E = 5.9$ km/s and $\gamma_E = -12.2°$. To decelerate the entry probe, a parachute system is deployed at Mach 2.2.

(a) What is the altitude, h, at which this happens?
(b) What is the corresponding dynamic pressure, $\bar{q}$?

5.13. Given a terminal velocity at retro-rocket firing of $V_f = 30$ m/s, what will be the required parachute drag area? State any assumption that you make.

5.14. A future Mars mission to prepare for establishing a Mars base aims at a vehicle mass of 4,000 kg. To lower the deployment loads on the parachute due to this mass increase, one could lower the dynamic pressure at deployment. One way to achieve this is to mount an inflatable aeroshell ($C_D = 1.35$) on the front of the vehicle. Note: the mass savings due to the inflatable have already been included in the vehicle mass.
For the altitude and dynamic pressure as found under Problem 5.12: what should be the diameter of this (circular) aeroshell for a reduction of this $\bar{q}$ of 30%?

Thermo-mechanical loads (Problems 5.15 to 5.26).

Consider the three target planets for atmospheric entry: Venus, Earth, and Mars, with relevant data listed in Table 5.21. For three types of entry capsules, with ballistic parameter of $K_m = 150$ kg/m^2 (MSL-like capsule), 100 kg/m^2, and 50 kg/m^2 (Stardust-like sample-return capsule), the following conditions at the entry interface are considered: $V_E = 0.9V_c$, and $\gamma_E = -7.5°$. Note that in Chap. 5 the ballistic parameter is *weight* referenced; the *mass* referenced values provided here are to indicate that they are related to the same entry capsule for each planet.

5.15. Calculate the circular velocity at the entry interface for the three planets: $h_E = 250$ km (Venus), $h_E = 100$ km (Earth), and $h_E = 80$ km (Mars).

5.16. Convert K_m to the weight-referenced ballistic parameters, K.

5.17. For the three planets and three capsules, calculate the nominal maximum deceleration, $\bar{a}_{max}$, and altitude at which this occurs, h'.
5.18. What can you conclude from the above results?
5.19. Any atmosphere model that is used is only an approximation of reality. Address the uncertainty in atmospheric density by assuming a 5% uncertainty on both ρ_0 and H_s. Vary these uncertainties according to a uniform distribution, and sample a sufficient number of times.[17] What is the effect on $\bar{a}_{max}$ and h'?
5.20. For the three planets and three capsules, calculate the nominal altitude at which the heat flux is a maximum, h''. Do this for both a laminar and a turbulent boundary layer. What do you observe?
5.21. Do a similar Monte Carlo analysis as before to study the effect on h''.
5.22. Assume that the heat flux is given by a cold-wall Chapman model. In Example 5.6, an expression for the maximum heat flux is derived, Eq. (5.81). This equation can be used to calculate $q_{c,max}$ for a capsule entering Earth's atmosphere, while experiencing a laminar boundary layer. For the three capsules (and Earth only), do a similar Monte Carlo analysis as before and calculate the variation in $q_{c,max}$. Discuss what you observe.
5.23. Generalise Eq. (5.81) for arbitrary n (type of boundary layer) and m (velocity exponent).
5.24. Consider the engineering expression Eq. (3.330) that relates the radiative heat flux coming from the surrounding gas (see also Sect. 3.6.3 for some more background), to the altitude-velocity profile:

$$q_{rad,gas} = C R_N \rho^{1.6} V^{8.5} \tag{5.373}$$

Derive expressions for velocity and altitude when $q_{rad,gas}$ reaches it peak value. Assume a ballistic flight.
5.25. What changes for a gliding flight? Derive supporting expressions, if need be.
5.26. For an Earth ballistic entry, evaluate the key thermo-mechanical load points, i.e., determine velocity ratio and corresponding altitude. Use $K = 100\,\mathrm{kg/m^2}$. Other data as before.

Ballistic, gliding, and skipping flight (Problems 5.27 to 5.30).

A science vessel of the Interstellar Federation has reached the planet Moijaussi. This obscure planet is rumoured to have vast amounts of the rare material Unobtanium that can double the efficiency of the Epstein drives. Before entering the atmosphere in a skipping/gliding flight, the captain decides to investigate the atmosphere with ballistic probes.
Given data: $g = g_0 = 15.4\,\mathrm{m/s^2} = \text{constant}$, $R_e = 8385.9\,\mathrm{km}$, $V_c = 11{,}498.8\,\mathrm{m/s}$.

[17] A sampling of uncertainties according to a particular distribution and evaluating the performance of an objective is also known as a Monte Carlo analysis. To answer this problem, use any software of your choice.

5.27. To establish the atmospheric-density profile, a spherical probe is released in the atmosphere that transmits data about friction and heat load back to the mothership. This sphere has a mass of 1 kg, a diameter of 0.2 m, and a drag coefficient of $C_D = 1.5$. Sensors in the probe measure the peak values for the thermo-mechanical loads and establish the altitude at which this occurs by radar altimetry. The entry conditions are given by $V_E = 9.3$ km/s and $\gamma_E = -9.5°$. For the peak deceleration it is found that $a_{max} = 169.4\,\text{m/s}^2$ at an altitude of $h' = 162.2$ km. Assuming an exponential density profile $\rho = \rho_0 e^{-h/H_s}$, what values can be derived for the scale height, H_s, and the density at the surface, ρ_0?

5.28. Inspecting the sensor data in more detail, it was found that the peak heat flux occurred 9743.4 m higher. Was the vehicle experiencing a laminar or turbulent flow? Assume a cold-wall stagnation flux model.

5.29. Having established the characteristics of the exponential atmosphere, a winged research vehicle is supposed to land 15,900 km in downrange direction. The following flight is performed: (1) a single skip in the atmosphere, (2) a ballistic flight phase outside the atmosphere, and (3) the final gliding flight towards the target. After the first skip, the vehicle leaves the atmosphere with velocity $V_F = 8{,}635.1$ m/s and $\gamma_F = +8.5°$. Calculate the total flight range during flight phase 2, $X_{h,2}$.

5.30. State/derive the expression for the flight range in the gliding portion of the flight, phase 3, $X_{h,3}$. Use this equation to determine the required lift-to-drag ration L/D to reach the target landing site. It is given that the downrange covered in flight phase 1 is 25% of that in flight phase 2: $X_{h,1} = 0.25 X_{h,2}$.

Skipping flight (Problems 5.31 to 5.35).

5.31. In a skipping flight, a vehicle with an L/D of 2 experiences an axial deceleration of 15 g. What will be the total mechanical load on the vehicle? State your assumptions.

5.32. Consider the motion of an arbitrary re-entry vehicle in the atmosphere of the Earth. Suppose that the re-entry vehicle returns in the atmosphere using a repeating skipping trajectory, entering the atmosphere twice. At first contact with the atmosphere, the entry angle $\gamma_E = -15°$ and the entry velocity $V_E = 8$ km/s. The lift-to-drag ratio of the vehicle is: $L/D = 2$ (constant). Compute the altitude of the lowest point of both parts of a dual skipping trajectory. For this purpose it is given that the "lift parameter" of the vehicle $(W/(SC_L))$ equals $0.2\,\text{N/cm}^2$, while the exponential-atmosphere parameters are $H_s = 7$ km and $\rho_0 = 1.225\,\text{kg/m}^3$. The constant gravitational acceleration $g_0 = 9.81\,\text{m/s}^2$.

5.33. For a vehicle with $L/D = 3$, how many skips are required to reduce the entry velocity by 25%?

5.34. Related to the previous problem, in case one want to achieve this (exact) reduction with a single skip, what should be the L/D?

5.35. What is the maximum deceleration that the vehicle experiences in case $L/D = 1.5$, 3, and 4.5? Discuss the results.

References

Allen HJ, Eggers AJ Jr (1958) A study of the motion and aerodynamic heating of missiles entering the Earth's atmosphere at high supersonic speeds, NACA Report 1381. Washington DC

Allen HJ, Eggers AJ, Jr, Neice SE (1958) A comparative analysis of the performance of long-range hypervelocity vehicle, NACA Report 1382, Washington, DC

Braun RD, Powell RW, Engelund WC, Gnoffo PA, Weilmuenster KJ, Mitcheltree RA (1995) Mars Pathfinder six-degree-of-freedom entry analysis. J Spacecraft and Rockets 32(6):993–1000

Cavallo A, Ferrara F (1996) Atmospheric re-entry control for low lift/drag vehicles. J Guid Control Dyn 19(1):47–53

Cawthorne A, Gardini B, Ongaro F, Pradier A, Vennemann D (2004) Enabling exploration - the technology development roadmap in ESA's Aurora programme, IAC-04-IAA.3.6.1.04. In: 55th international astronautical congress 2004 - Vancouver, Canada

Chapman DR (1959) An approximate analytical method for studying entry into planetary atmospheres, NASA Technical Report TR-R-11, Washington, DC

Chapman DR (1960) An analysis of the corridor and guidance requirements for supercircular entry into planetary atmospheres, NASA Technical Report TR-R-55, Washington, DC

Desai PN, Knocke PC (2004) Mars exploration rovers entry, descent, and landing trajectory analysis, AIAA 2004-5092. Astrodynamics specialist conference and exhibit, Providence, Rhode Island, 16–19 August 2004

Engelsma J, Mooij E (2020) Aerocapture mission analysis, AIAA-2020-1740. In: AIAA scitech forum, atmospheric flight mechanics conference, Orlando, FL, 6–10 January 2020

ESA (2001) Technologies for exploration: Aurora programme proposal: annex D. SP-1254

FESTIP Concept Team (1998) Slice-C progress meeting 8, FESTIP document SS-269, Ottobrunn, 17–18 February 1998

Graves CA, Harpold JC (1972) Apollo experience report - mission planning for Apollo entry, NASA TN D-6725, Washington, DC

Harvey B (2007) Soviet and Russian lunar exploration. Springer, Praxis Publishing, Chichester, UK

Hoelscher BR (2007) Orion entry, descent, & landing simulation, AIAA 2007-6428. In: AIAA guidance, navigation and control conference and exhibit, Hilton Head, SC, 20–23 August 2007

Jenkins DR (2001) Space Shuttle: the history of the national space transportation system - the first 100 missions, 3rd edn. Walsworth Publishing Company, Marceline, Missouri

Liever PA, Habchi SD, Burnell SI, Lingard JS (2003) Computational fluid dynamics prediction of the Beagle 2 aerodynamic database. J Spacecraft and Rockets 40(5):632–638

Loh WHT (1968) Re-entry and planetary entry physics and technology I. Springer, Berlin

Loh WHT (1969) Re-entry and planetary entry physics and technology II. Springer, Berlin

Mandell DS (1962) Maneuvering performance of lifting re-entry vehicles. Am Rocket Soc (ARS) J 32(3):346–354

MBB Space Communication and Propulsion Systems Division (1988) Study on re-entry guidance and control, ESA report reference: ESA CR (P) 2652

McRuer D (1991) Design and modeling issues for integrated airframe/propulsion control of hypersonic flight vehicles. In: Proceedings of the American control conference, Boston, Massachusetts, pp 729–735

Miele A (1962) Flight mechanics, vol 1. Addison-Wesley Pub, Co

Mooij E (1991) ESA Huygens probe entry and descent analysis. ESTEC working paper EWP 1634, ESTEC, Noordwijk

Mooij E (1992) ESA Huygens probe entry and descent analysis. ESTEC working paper EWP 1679, ESTEC, Noordwijk

Mooij E, Huot J-P, Ortega G (2006) Entry trajectory simulation using ESA Mars climate database version 4.1. AIAA 2006-6023. In: AIAA/AAS astrodynamics specialist conference and exhibit, Keystone, Colorado, 21–24 August 2006

Mooij E, Mease KD, Benito J (2007) Robust re-entry guidance and control system design and analysis, AIAA-07-6779. In: AIAA guidance, navigation, and control conference and exhibit, Hilton head, SC, 20–23 August 2007

Moster GE, Martin JG (1999) Oscillatory trajectories applied to NASA's DF-7 configuration, AIAA-99-4931. In: AIAA 9th international space planes and hypersonic systems and technologies conference, Norfolk, VA, 1–5 November 1999

Ngo AD, Doman DB (2002) Footprint determination for reusable launch vehicles experiencing control effector failures, AIAA-2002-4775. In: Guidance, navigation, and control conference, Monterey, 5–8 August 2002

North American Aviation, Inc. (1965) Aerodynamic data manual for project Apollo, NASA-CR-82907, 1 January 1965

Orloff RW (2000) Apollo by the numbers: a statistical reference, NASA SP-2000-4029. NASA History Division, Washington, DC

Regan FJ, Anandakrishnan SM (1993) Dynamics of atmospheric re-entry. AIAA education series, American Institute of Aeronautics and Astronautics

Rose B (2008) Secret projects: military space technology. Midland Publishing

Sänger E (1933) Raketen-flugtechnik. R. Oldenbourg, Berlin

Sänger E, Bredt J (1944) A rocket drive for long range bombers, Technical Information Branch, US, Navy Bureau of Aeronautics, Translation CGD-32 (1955)

Saraf A, Leavitt JA, Mease KD, Ferch M (2004) Landing footprint computation for entry vehicles, AIAA 2004-4774. In: AIAA guidance, navigation, and control conference and exhibit, Providence, RI, 16–19 August 2004

Slye RE (1960) An analytical method for studying the lateral motion of atmosphere entry vehicles, NASA-TN D-325

Spencer DA, Braun RD (1996) Mars Pathfinder atmospheric entry: trajectory design and dispersion analysis. J Spacecraft and Rockets 33(5):670–676

Sunday TL, London JR III (1990) The X-20 space plane: past innovation, future vision. Acta Astronautica 21(4):223–242

Vinh NX, Busemann A, Culp RD (1980) Hypersonic and planetary entry flight mechanics. University of Michigan Press

Vinh NX (1981) Optimal trajectories in atmospheric flight. Elsevier Scientific Publishing Company, Amsterdam, The Netherlands

Vinh NX, Kin E, Greenwood DT (1993) Second-order analytical solutions for re-entry trajectories, AIAA-1993-3679. In: Flight simulation and technologies conference, Monterey, CA, 9–11 August 1993

Whitmore SA, Dunba BJ (2003) Orbital space plane, past, present, and future, AIAA-2003-2718. In: AIAA/ICAS international air and space symposium and exposition: the next 100 years, Dayton, Ohio, 14–17 July 2003

Wright MJ, Hollis BR, Bose D, Walpot L (2006) Post-flight aerothermal analysis of Huygens probe. In: 3rd international planetary probe workshop, Anavyssos, Attiki, Greece, published as ESA SP-607, 27 June–1 July 2005

Chapter 6
Characteristic Motion of Entry Vehicles



Compared to (conventional) aircraft entry vehicles have not only a completely different mission profile, but also significantly different characteristics. Vehicles like the Space Shuttle, for instance, have a geometrical design that is optimised for the hypersonic flight regime, whereas the cruise characteristics in the high subsonic regime determine the configuration of conventional aircraft to a large extent. The situation for entry vehicles deviates even further if the vehicle is not aircraft-like such as the Shuttle, but a capsule such as the Apollo Command Module. The aerodynamic characteristics cannot even be compared anymore and also the means of control is completely different: Apollo only had a reaction control system (RCS) to counteract disturbances and lacks the aerodynamic control surfaces of aircraft and winged re-entry vehicles.

For conventional aircraft flying at subsonic velocities we are used to simplify the equations of motion, such that we can use simplified expressions to derive the stability and control derivatives, and to study, for instance, the characteristic motion. Typical assumptions made concern the use of a flat Earth, a constant gravity, and a constant atmospheric density for the considered manoeuvres.

But what happens if the assumptions made to simplify the equations are no longer valid, for instance, when the velocity is so large that curvature effects of the Earth become important or even the fact that the Earth rotates. Also the fact that we have an atmosphere is not always the case, for instance, when a Space Shuttle returns from the International Space Station it begins its mission outside the atmosphere and only gradually the atmospheric density and pressure start increasing. On the other side of the spectrum are capsules entering the dense atmospheres of Titan and Venus, the latter of which has a surface density of almost 55 times that of Earth!

For an accurate analysis of the motion of entry vehicles in the atmosphere we can no longer use simplified equations, as the non-linearities are so strong that ignoring them would lead to an inaccurate prediction of their trajectory and corresponding flight loads. However, even in this case the perturbed motion around the vehicle's nominal trajectory will give insight into its stability and controllability characteristics.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_6

To achieve this the non-linear equations of motion are linearised and through analysis of eigenvalues and eigenvectors insight in the characteristic motion of entry vehicles can be obtained.

To understand the influence of mission and vehicle configuration on the natural stability of entry vehicles, in this chapter we will study three different entry configurations, i.e., a winged vehicle, a manned entry capsule, and a spinning planetary probe. To get a feeling for the influence of hypersonic velocities on the characteristic motion of aircraft-like vehicles we will use the same winged reference vehicle that we have used before, i.e., the HORUS-2B. As mentioned, for this vehicle a complete mass and aerodynamic database is available. Another vehicle that is quite well published in public literature is the Apollo entry capsule. This will be the second vehicle that we will study, more to understand the difference between winged entry vehicles and entry capsules rather than to compare capsules with aircraft. Finally, the motion of the Huygens planetary probe will be evaluated to address the effect of spin on its stability characteristics.

To give a complete overview of how to do such studies we will begin this chapter with a summary of the non-linear equations of motion, commonly used in the study of re-entry dynamics (Sect. 6.1). Before we dive into the linearisation process, we will introduce the concept of eigenvalues and eigenvectors and how these relate to the study of characteristic motion (Sect. 6.2). The linearisation process in Sect. 6.3 is aimed at reaching the starting point that is used in subsequent sections to get to a particular formulation, i.e., the state-space form for linear time-invariant systems. In Sect. 6.4 the state-space model and subsequent characteristic motion of the winged entry vehicle HORUS is discussed. Section 6.5 has the same structure, but now the discussion is focussed on entry capsules, i.e., Apollo and Huygens.

6.1 Non-linear Equations of Motion

In Appendix A the non-linear equations of motion have been derived for both an inertial (Earth-centred, index I) and rotating (Earth-centred Earth-fixed (ECEF), index R) reference frame. The next step would be to choose a set of state variables, e.g., Cartesian or spherical position and velocity components or a combination of the two. These equations can then form the basis of a linearisation process that, in the case of low-speed aircraft stability studies, will give us non-dimensional equations in state-space form, expressed in the body frame. Several steps are usually involved, notably simplifying the non-linear equations by assuming a non-rotating and flat Earth, and transforming the equations of translational motion to the body frame before linearisation. For the study of the characteristic motion (or *eigenmotion*) of high-velocity entry vehicles, the assumptions of a non-rotating and flat Earth cannot be used, because of the large errors that will be otherwise introduced.

However, if these aspects are not ignored, setting up the equations of translational motion in the body frame is a complex process. For this reason, we will use a different set of equations as the starting point for the linearisation as compared to aircraft

studies. Traditionally, in re-entry studies one either uses inertial Cartesian position and velocity elements, or a spherical position and velocity in a rotating frame, notably the R-frame. Because the latter definition makes it far easier to understand and interpret the results, it will be our basis for linearisation. However, on the downside, the full derivation is long and tedious. Therefore, we will only present the final set of equations here. For more details on the derivation the interested reader is referred to Vinh (1981). Our only assumptions that affect the equations are that the Earth is spherical (with a spherical gravity field) and rotates with a constant angular velocity ω_{cb}.

For the rotational motion we will use the Euler equations as derived in Appendix A. Here, we assume that the vehicle has a plane of mass symmetry ($X_B Y_B$-plane), which means that $I_{xy} = I_{yz} = 0$. In terms of kinematic attitude equations we will not use the classical Euler angles ϕ, θ, and ψ, but a different set, notably the aerodynamic angles angle of attack, angle of sideslip, and bank angle. These angles are in principle groundspeed based, but again, since there is no wind, they will be equal to their airspeed-based counterparts. Even now we will not derive the corresponding equations, because also this derivation is long and tedious.

In Chap. 4, Fig. 4.4 the state variables for position and velocity are shown. As discussed there, the position is given by the distance R, longitude τ and latitude δ, whereas the velocity is expressed by its modulus, the groundspeed V, and two direction angles, i.e., flight-path angle γ and heading χ (note that since we do not consider wind, no subscript is added to differentiate between air- and groundspeed). The attitude of a vehicle, or, in mathematical terms, the orientation of the body-fixed reference frame with respect to the trajectory reference frame, is expressed by the so-called aerodynamic angles, i.e., the angle of attack α, the angle of sideslip β and the bank angle σ, see also Fig. 4.6 in Sect. 4.2. These angles form an asymmetric Euler set. The angular rate of the body is here defined as the rotation of the body frame with respect to the inertial frame, expressed in components along the body axes. These components are called roll rate p, pitch rate q, and yaw rate r (see Fig. 4.7).

Summarised, the dynamic equations of translational motion are given by

$$\dot{V} = -\frac{D}{m} - g\sin\gamma + \omega_{cb}^2 R\cos\delta(\sin\gamma\cos\delta - \cos\gamma\sin\delta\cos\chi) \tag{6.1}$$

$$\begin{aligned} V\dot{\gamma} = \frac{L\cos\sigma}{m} - g\cos\gamma + 2\omega_{cb}V\cos\delta\sin\chi + \frac{V^2}{R}\cos\gamma + \\ + \omega_{cb}^2 R\cos\delta(\cos\delta\cos\gamma + \sin\gamma\sin\delta\cos\chi) \end{aligned} \tag{6.2}$$

$$\begin{aligned} V\cos\gamma\dot{\chi} = \frac{L\sin\sigma}{m} + 2\omega_{cb}V(\sin\delta\cos\gamma - \cos\delta\sin\gamma\cos\chi) + \frac{V^2}{R}\cos^2\gamma\tan\delta\sin\chi + \\ + \omega_{cb}^2 R\cos\delta\sin\delta\sin\chi \end{aligned} \tag{6.3}$$

whereas the kinematic position equations are given by

$$\dot{R} = \dot{h} = V \sin\gamma \tag{6.4}$$

$$\dot{\tau} = \frac{V \sin\chi \cos\gamma}{R \cos\delta} \tag{6.5}$$

$$\dot{\delta} = \frac{V \cos\chi \cos\gamma}{R} \tag{6.6}$$

The Euler equations of rotational motion are defined by:

$$\dot{\boldsymbol{\omega}} = \mathbf{I}^{-1}\left(\tilde{\mathbf{M}}_{\text{cm}} - \boldsymbol{\omega} \times \mathbf{I}\boldsymbol{\omega}\right) \tag{6.7}$$

with

$\tilde{\mathbf{M}}_{\text{cm}} = \left(M_x, M_y, M_z\right)^T$ = Sum of external, Coriolis and relative moments about the c.o.m., expressed in components along the body axes

$\mathbf{I} = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{xy} & I_{yy} & -I_{yz} \\ -I_{xz} & -I_{yz} & I_{zz} \end{bmatrix}$ = Inertia tensor of the re-entry vehicle, referenced to the body frame

$\boldsymbol{\omega} = (p, q, r)^T$ = The rotation vector of the body frame with respect to the inertial frame, expressed in components along the body axes

Solving the above equations asks for the inversion of the inertia tensor. To explicitly write the solution as $\dot{p} = \ldots$ is an elaborate exercise. However, in literature these "analytical" expressions can be found for $\dot{p}, \dot{q}$, and $\dot{r}$ (Duke et al. 1988). Fortunately, in a number of cases Eq. (6.7) can be simplified such that it becomes relatively easy to formulate the required equations. For instance, for a (mass-symmetrical) vehicle, which is the case for our winged re-entry vehicle, two out of three products of inertia are equal to zero. When the plane of symmetry is the $X_B Z_B$-plane, $I_{xy} = I_{yz} = 0$. Equation (6.7) then simplifies to

$$\begin{bmatrix} I_{xx} & 0 & -I_{xz} \\ 0 & I_{yy} & 0 \\ -I_{xz} & 0 & I_{zz} \end{bmatrix} \begin{pmatrix} \dot{p} \\ \dot{q} \\ \dot{r} \end{pmatrix} = \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} - \begin{pmatrix} p \\ q \\ r \end{pmatrix} \times \begin{bmatrix} I_{xx} & 0 & -I_{xz} \\ 0 & I_{yy} & 0 \\ -I_{xz} & 0 & I_{zz} \end{bmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix} \tag{6.8}$$

Solving this system of equations for $\dot{p}, \dot{q}$ and $\dot{r}$ leads to

$$\dot{p} = \frac{I_{zz}}{I^*} M_x + \frac{I_{xz}}{I^*} M_z + \frac{\left(I_{xx} - I_{yy} + I_{zz}\right) I_{xz}}{I^*} pq + \frac{\left(I_{yy} - I_{zz}\right) I_{zz} - I_{xz}^2}{I^*} qr \tag{6.9}$$

$$\dot{q} = \frac{M_y}{I_{yy}} + \frac{I_{xz}}{I_{yy}}\left(r^2 - p^2\right) + \frac{I_{zz} - I_{xx}}{I_{yy}} pr \tag{6.10}$$

$$\dot{r} = \frac{I_{xz}}{I^*} M_x + \frac{I_{xx}}{I^*} M_z + \frac{\left(I_{xx} - I_{yy}\right) I_{xx} + I_{xz}^2}{I^*} pq + \frac{\left(-I_{xx} + I_{yy} - I_{zz}\right) I_{xz}}{I^*} qr \tag{6.11}$$

with $I^* = I_{xx} I_{zz} - I_{xz}^2$.

Finally, the kinematic attitude equations are given by (see also Intermezzo 4.1, Eqs. (4.137)–(4.139)):

$$\begin{aligned}\dot{\alpha}\cos\beta = & -p\cos\alpha\sin\beta + q\cos\beta - r\sin\alpha\sin\beta + \\ & + \sin\sigma\left[\dot{\chi}\cos\gamma - \dot{\delta}\sin\chi\sin\gamma + (\dot{\tau} + \omega_{cb})\left(\cos\delta\cos\chi\sin\gamma - \sin\delta\cos\gamma\right)\right] + \\ & - \cos\sigma\left[\dot{\gamma} - \dot{\delta}\cos\chi - (\dot{\tau} + \omega_{cb})\cos\delta\sin\chi\right]\end{aligned} \tag{6.12}$$

$$\begin{aligned}\dot{\beta} = & \; p\sin\alpha - r\cos\alpha + \\ & + \sin\sigma\left[\dot{\gamma} - \dot{\delta}\cos\chi - (\dot{\tau} + \omega_{cb})\cos\delta\sin\chi\right] + \\ & + \cos\sigma\left[\dot{\chi}\cos\gamma - \dot{\delta}\sin\chi\sin\gamma + (\dot{\tau} + \omega_{cb})\left(\cos\delta\cos\chi\sin\gamma - \sin\delta\cos\gamma\right)\right]\end{aligned} \tag{6.13}$$

$$\begin{aligned}\dot{\sigma} = & -p\cos\alpha\cos\beta - q\sin\beta - r\sin\alpha\cos\beta + \\ & + \dot{\alpha}\sin\beta - \dot{\chi}\sin\gamma - \dot{\delta}\sin\chi\cos\gamma + (\dot{\tau} + \omega_{cb})\left(\cos\delta\cos\chi\cos\gamma + \sin\delta\sin\gamma\right)\end{aligned} \tag{6.14}$$

In these equations, $\dot{\gamma}$, $\dot{\chi}$, $\dot{\delta}$, and $\dot{\tau}$ are given by Eqs. (6.2) through (6.6), and represent the effects of the rotation of the Earth and the rotation of the local horizontal plane. Note that the external moment components are expressed in the body-fixed reference frame.

The system of equations, Eqs. (6.1) through (6.14), is a time-varying system. Such a system cannot be easily studied in terms of natural motion, because this time-dependency will obscure the nature of the individual motion types. Also the non-linearities in the system make the behaviour of the system dependent on the magnitude of the input, so no crisp conclusions can be drawn from the results. Therefore, a common approach is to linearise the equations of motion and bring them in a form that can be used to analyse the natural motion.

However, the large variation of vehicle characteristics and operational conditions along the re-entry vehicle's trajectory will force us to use more than one model. Therefore, we will discretise the nominal trajectory into a number of time intervals. During each of the intervals, the state of the vehicle and other system parameters are assumed to be constant. For each of the time points, the system can then be investigated. Thus, by linearising the equations of motion (see Sect. 6.3) and discretising the trajectory, the overall system has been divided into a sequence of Linear Time-Invariant systems (LTI systems) of a form that will be presented in Sect. 6.4.1.

This so-called state-space form of the differential equations consists of a homogenous and a particular part, where

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{6.15}$$

is known as the homogeneous part of the state equation. It is this part that can be used to study the free response (also called open-loop behaviour, characteristic motion or eigenmotion) of the system. This is done through eigenvalue analysis, which will be discussed in the next section.

6.2 Eigenvalues and Eigenmotion

Eigenvalues and eigenvectors have originated in the field of linear algebra. They came about that after a linear transformation of a vector, the direction of the resultant vector is invariant under this transformation and only its length has changed. A vector, for which this property holds, is called an eigenvector, and the corresponding scaling factor is called the eigenvalue. As simple example will clarify this notion. Suppose we have two rubber bands, each with an arrow drawn on it (Fig. 6.1). Let us now pull the rubber bands in the direction of its centre line; the result is shown on the right side of the figure. The arrow on the top rubber band is stretched in the direction of elongation, but the arrow on the bottom rubber band has not only stretched, but also seems to have rotated. So, the former has maintained its direction, whereas the latter has changed. You can probably guess now which of the two is the eigenvector and which one is not.

In a mathematical formulation, the transformation can be written as follows. Let $\boldsymbol{\mu}$ be a vector, and $\mathbf{A}$ a (linear) transformation matrix. Then, in case the direction of $\boldsymbol{\mu}$ is invariant, the transformation is defined by:

$$\mathbf{A}\boldsymbol{\mu} = \lambda\boldsymbol{\mu} \tag{6.16}$$

In the above equation, $\boldsymbol{\mu}$ is called the eigenvector, and the corresponding scaling factor or multiplier, λ, is called the eigenvalue. It is said that $\boldsymbol{\mu}$ is the eigenvector of $\mathbf{A}$, and for an $n \times n$ matrix, there will be (at most) n different eigenvectors. At this place, we will not go into more detail about the mathematics involving eigenvalues and eigenvectors (i.e., the *eigenproblem*), but focus on one of the many applications of them, namely the one associated with determining the natural motion of mechanical systems. The interested reader should check out books on mathematics, specifically linear algebra, or linear systems theory for more information (Kreyszig 2011; Ogata 2004; Meirovitch 2001).

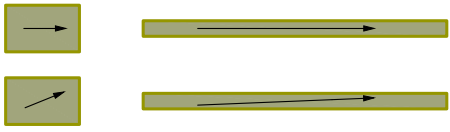


Fig. 6.1 Stretching of rubber bands

Equation (6.16) can be solved by

$$\mathbf{A}\boldsymbol{\mu} - \lambda\boldsymbol{\mu} = (\mathbf{A} - \lambda\mathbf{I})\,\boldsymbol{\mu} = 0 \tag{6.17}$$

The condition for a non-trivial solution is

$$|\mathbf{A} - \lambda\mathbf{I}\,| = 0 \tag{6.18}$$

Standard algorithms are available to calculate the eigenvalues and corresponding eigenvectors of a given matrix $\mathbf{A}$ (Press et al. 2007), which can also be found in software environments like Matlab®. The mathematical foundation of these algorithms is considered to be beyond the scope of the current text.

Example 6.1 (*Eigenvalues of a Matrix*) Let the matrix $\mathbf{A}$ be defined by

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \tag{6.19}$$

With Eq. (6.18) we can solve for the eigenvalues:

$$\det\left(\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = 0 \quad \Rightarrow \quad \begin{vmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = 0 \tag{6.20}$$

The characteristic equation to solve is thus

$$\lambda^2 - 5\lambda + 6 = 0 \tag{6.21}$$

which trivially yields $\lambda_1 = 3$ and $\lambda_2 = 2$. The corresponding eigenvectors follow easily, e.g., for λ_1:

$$\mathbf{A}\begin{pmatrix} \mu_{1,1} \\ \mu_{1,2} \end{pmatrix} - \lambda_1 \begin{pmatrix} \mu_{1,1} \\ \mu_{1,2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{6.22}$$

or

$$3\mu_{1,1} + \mu_{1,2} - 3\mu_{1,1} = 0 \quad \text{and} \quad 2\mu_{1,2} - 3\mu_{1,2} = 0 \tag{6.23}$$

which both give $\mu_{1,2} = 0$ with an arbitrary value for $\mu_{1,1}$. Let this value be 1, such that the first eigenvector is $\boldsymbol{\mu}_1 = \begin{pmatrix} \mu_{1,1} & \mu_{1,2} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 1 & 0 \end{pmatrix}^{\mathrm{T}}$. For the second eigenvalue we find

$$3\mu_{2,1} + \mu_{2,2} - 2\mu_{2,1} = \mu_{2,1} + \mu_{2,2} = 0 \quad \text{and} \quad 2\mu_{2,2} - 2\mu_{2,2} = 0 \tag{6.24}$$

Choosing $\mu_{2,2}$ to be 1, we get $\mu_{2,1} = -1$, such that $\boldsymbol{\mu}_2 = \begin{pmatrix}\mu_{2,1} & \mu_{2,2}\end{pmatrix}^{\mathrm{T}} = \begin{pmatrix}-1 & 1\end{pmatrix}^{\mathrm{T}}$ (or, when normalised, which is often done, $\boldsymbol{\mu}_2 = \begin{pmatrix}-\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2}\end{pmatrix}^{\mathrm{T}}$).

Consider now the following two-dimensional rotation matrix, $\mathbf{C}$:

$$\mathbf{C} = \begin{bmatrix}\cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha\end{bmatrix} \tag{6.25}$$

What we know of transformation matrices is that length is invariant (!) and that all vectors are rotated,[1] i.e., *changing direction*. Does this mean that such a matrix does not have any eigenvectors, as it is clearly not meeting with the earlier stated criterion? Let us check this with an example, using a rotation angle of $\alpha = 90°$, thus

$$\mathbf{C} = \begin{bmatrix}0 & 1 \\ -1 & 0\end{bmatrix} \tag{6.26}$$

The characteristic equation that follows is

$$\begin{vmatrix}\lambda & 1 \\ -1 & \lambda\end{vmatrix} = 0 \quad \Rightarrow \quad \lambda^2 + 1 = 0 \tag{6.27}$$

which clearly leads to complex eigenvalues: $\lambda_{1,2} = \pm i$. Repeating the exercise of above to derive the eigenvectors will give $\boldsymbol{\mu}_1 = \begin{pmatrix}1 & i\end{pmatrix}^{\mathrm{T}}$ and $\boldsymbol{\mu}_2 = \begin{pmatrix}1 & -i\end{pmatrix}^{\mathrm{T}}$ (or, as before, normalised: $\boldsymbol{\mu}_1 = \begin{pmatrix}\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2}i\end{pmatrix}^{\mathrm{T}}$ and $\boldsymbol{\mu}_2 = \begin{pmatrix}\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2}i\end{pmatrix}^{\mathrm{T}}$).

This last example holds a general truth, i.e., when the eigenvalues and/or eigenvectors of a matrix are complex, any input vector will exhibit some form of rotation, even if only small. This holds (at least) for all transformation matrices, as they all have complex eigenvalues. ■

Next, let us consider the motion of a single mass-spring system, of which the equation of motion is given by

$$m\ddot{x} + kx + c\dot{x} = 0 \tag{6.28}$$

where x is the position, m is the mass, k is the spring stiffness, and c is the damping coefficient. We will study the effect of different k and c in a few simple examples later. Converting the above equation into a set of first-order differential equations (and assuming $m = 1\,\mathrm{kg}$), we introduce the velocity as $v = \dot{x}$ and write the two equations in matrix form:

$$\begin{pmatrix}\dot{v} \\ \dot{x}\end{pmatrix} = \begin{bmatrix}-c & -k \\ 1 & 0\end{bmatrix}\begin{pmatrix}v \\ x\end{pmatrix} \quad \Rightarrow \quad \dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{6.29}$$

[1] Of course, the trivial vector $\mathbf{v} = \begin{pmatrix}0 & 0\end{pmatrix}$ is an exception: it is not rotated, but it also does not change length. So let us exclude the null vector from the discussion.

where $\mathbf{x} = \begin{pmatrix} v & x \end{pmatrix}^{\mathrm{T}}$ is used as state vector, and the coefficient matrix $\mathbf{A}$ is the so-called state or system matrix.

For the first analysis, an undamped system is considered, i.e., $c = 0$ Ns/m, and we take $k = 1$ N/m. These numbers substituted gives

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

which is effectively the rotation matrix, Eq. (6.25), with $\alpha = -90°$. We already know from before that the eigenvalues of rotation matrices are complex, and that such a transformation leads to the rotation of the input vector. For the current case, the two eigenvalues are actually the same, i.e., $\lambda_{1,2} = \pm i$; the direction of rotation does not matter, and both the regular and inverse rotation always give the same eigenvalues.[2]

How does having a pair of complex eigenvalues reflect on the motion of the mass? In the left plot of Fig. 6.3 position and velocity have been plotted against each other (starting from $\mathbf{x_0} = \begin{pmatrix} 0 & -5 \end{pmatrix}^{\mathrm{T}}$), giving a so-called phase-plane plot. As one can see, the relation between x and v shows as a circular (= rotational) motion, which is what we would expect from the eigenvalues. Since the system is conservative, i.e., $c = 0$, the motion is repeating itself, a continuous exchange of kinetic and potential energy. We call such motion *periodic*. In case we increase the spring stiffness, $k = 2$ N/m, the velocity amplitude increases, as do the eigenvalues: $\lambda_{1,2} = \pm\sqrt{2}i$. This is, in fact, also a general rule: the larger the complex part of an eigenvalue, the faster the rotation will be.

Example 6.2 (*Eigenvalues of Pendulum Motion*) A well-known example of a system in oscillation is the simple pendulum, originally examined by Galileo Galilei (1564–1642) in the beginning of the 17th century as a measure to keep time. The pendulum under consideration is a mass m suspended from a pivot so that it can swing freely. When hanging vertically down, it is in its equilibrium position. However, when displaced sideways from rest, it is subject to a restoring force due to gravity that will accelerate it back toward this equilibrium position. Due to its inertia it will overshoot and ideally it will keep oscillating due to the ever-changing restoring moment. The time for one complete cycle, a left swing and a right swing, is called the period. For an ideal, friction-less pendulum, the period only depends on the length of the pivot.

In this example we will set up the mathematical model, and use that to determine the eigenmotion of the pendulum. The general situation has been depicted in Fig. 6.2. For a frictionless pendulum, the swinging behaviour can be derived to be

$$\ddot{\theta} + \frac{g}{\ell} \sin\theta = 0 \tag{6.30}$$

[2] For a rotation of either +45° or –45°, the eigenvalues are in both cases the same, i.e., $\lambda_{1,2} = \frac{1}{2}\sqrt{2} \pm \frac{1}{2}\sqrt{2}i$. In this case, however, the eigenvalues have both a real and imaginary component.

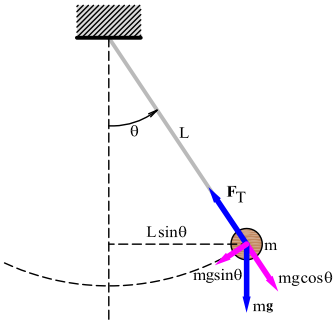


Fig. 6.2 The simple pendulum

For small angles, $\sin\theta \approx \theta$, so:

$$\ddot{\theta} + \frac{g}{\ell}\theta = 0 \tag{6.31}$$

This can be rewritten in the standard form of a second-order differential equation for undamped vibrational systems,

$$\ddot{x} + \omega_n^2 x = 0 \tag{6.32}$$

where ω_n is the natural frequency or eigenfrequency of the system. Comparing Eqs. (6.31) and (6.32) shows that $\omega_n = \sqrt{\frac{g}{\ell}}$. As the natural frequency defines the angular velocity in a corresponding circular motion, the period of a single "swing", P, can be obtained from $\omega_n = \frac{2\pi}{P} = \sqrt{\frac{g}{\ell}}$, so:

$$P = 2\pi\sqrt{\frac{\ell}{g}} \tag{6.33}$$

The well-known result that the period is independent from the mass of the pendulum is clearly shown here.

To calculate the eigenvalues of the system we will bring Eq. (6.31) into matrix form—also known as *state-space form*—by rewriting the second-order differential equation as two first-order equations:

$$\dot{\omega} = -\frac{g}{\ell}\theta \tag{6.34}$$

$$\dot{\theta} = \omega \tag{6.35}$$

so

$$\begin{pmatrix} \dot{\omega} \\ \dot{\theta} \end{pmatrix} = \begin{bmatrix} 0 & -\frac{g}{\ell} \\ 1 & 0 \end{bmatrix} \begin{pmatrix} \omega \\ \theta \end{pmatrix} \tag{6.36}$$

which is thus of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{6.37}$$

The eigenvalues of $\mathbf{A}$ can be obtained from

$$\left|\mathbf{A} - \lambda\mathbf{I}\right| = 0 \tag{6.38}$$

so

$$\begin{vmatrix} -\lambda & -\frac{g}{\ell} \\ 1 & -\lambda \end{vmatrix} = 0 \Rightarrow \lambda^2 + \frac{g}{\ell} = 0 \Rightarrow \lambda_{1,2} = \pm i\sqrt{\frac{g}{\ell}} \tag{6.39}$$

These eigenvalues form a complex conjugate pair, with the real part $\mathrm{Re}(\lambda) = 0$ and the imaginary part $\mathrm{Im}(\lambda) = \sqrt{\frac{g}{\ell}}$. A zero real part means that the oscillatory motion is undamped. ■

The general formulation for an (undamped) mass-spring system is given by Eq. (6.28), with $c = 0$:

$$m\ddot{x} + kx = 0 \tag{6.40}$$

From known solutions to ordinary differential equations we know that the solution is of the form $x = e^{pt}$, which, when substituted in Eq. (6.40) gives

$$mp^2e^{pt} + ke^{pt} = 0 \quad \Rightarrow \quad mp^2 + k = 0 \tag{6.41}$$

for this solution to hold for all t. Consequently,

$$p^2 = -\frac{k}{m} \quad \Rightarrow \quad p_{1,2} = \pm i\sqrt{\frac{k}{m}} = \pm i\omega \tag{6.42}$$

where $\omega = \sqrt{\frac{k}{m}}$ is the so-called *natural frequency* or *eigenfrequency* of the free oscillation. In the above we have not yet considered the constants of integration, following from the initial conditions (position and velocity). Considering those, the general solution is

$$x = c_1e^{+i\omega t} + c_2e^{-i\omega t} \tag{6.43}$$

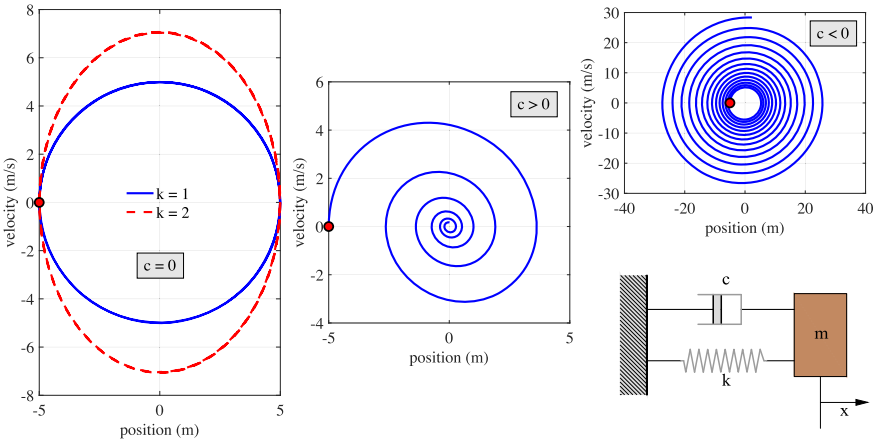


Fig. 6.3 Position-velocity phase plane for the motion of a single mass-spring system. Left: $c = 0$, middle: $c = –0.2$ Ns/m, and right: $c = 0.05$ Ns/m

with c_1 and c_2 being these constants. Of course, the position coordinate will not be a complex variable, as the physical coordinates are real numbers. Therefore, one can only conclude that c_1 and c_2 will have (complex) values to guarantee that.

Note that similarly the solution can be assumed to be

$$x = \gamma_1 \sin(\omega t) + \gamma_2 \cos(\omega t) \tag{6.44}$$

or, equivalently

$$x = \gamma_0 \sin(\omega t + \phi) \tag{6.45}$$

where $\gamma_0, \gamma_1, \gamma_2$ and ϕ are (real) constants of integration; ϕ is also known as the *phase angle*. The relation between the two solutions is evident from *Euler's Formula*:

$$e^{ix} = \cos x + i \sin x \tag{6.46}$$

Moving on, let us now consider a non-zero damping coefficient in Eq. (6.28), starting with $c = 0.2$ Ns/m. In the middle plot of Fig. 6.3 the effect is obvious, the motion has become damped, and the phase-plane curve spirals towards the centre, until all energy has dissipated. The eigenvalues of the system matrix are $\lambda_{1,2} = -0.1 \pm 0.995i$. It is interesting to see that the eigenvalues have now a (negative) real part; this is an important observation, because negative real parts are associated with damped (stable) motion. And, the more negative the real part is, the more rapid the decay will be. Finally, in the right figure the results for $c = –0.05$ Ns/m are shown, an unstable motion that spirals away from the centre. Intuitively we expect now a positive

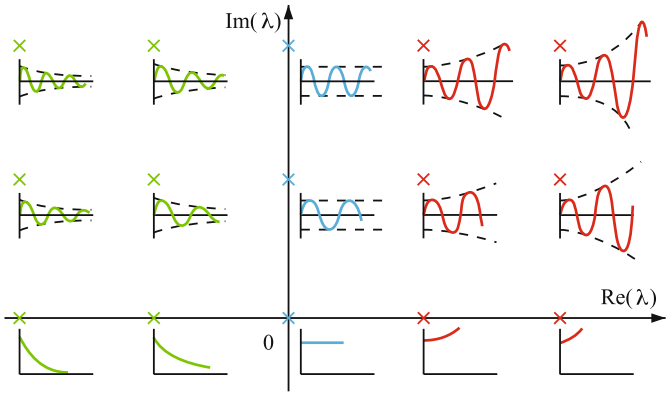


Fig. 6.4 Impulse responses for various eigenvalue locations in the complex plane, where the conjugate eigenvalues are omitted

real part for the eigenvalues, and this is indeed the case: $\lambda_{1,2} = 0.025 \pm 0.997i$. For this situation, the larger the real part will be, the stronger the divergence.

So, summarised, given the differential equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{6.47}$$

the eigenvalues λ of state matrix $\mathbf{A}$ and the corresponding eigenvectors $\boldsymbol{\mu}$ can be computed from the following equation:

$$\mathbf{A}\boldsymbol{\mu} = \lambda\boldsymbol{\mu} \tag{6.48}$$

Based on the eigenvalues, we can see whether a component of the motion is (un)stable and (a)periodic, see also Fig. 6.4. An eigenvalue can be real or complex. Complex eigenvalues appear in (conjugated) pairs and indicate a periodic eigenmotion, whereas real eigenvalues imply an aperiodic eigenmotion. The sign of the real part of the eigenvalue shows whether the eigenmotion will be converging (negative real part) or diverging (positive real part). When the real part is zero, the oscillations have a constant amplitude. A more detailed discussion can be found in, amongst others, DSouza (1988) and Kuo (1987).

To characterise the eigenmotion we compute some specific coefficients, i.e., the *period* P, when a pair of complex conjugate eigenvalues represents a periodic motion, defined as

$$P = \frac{2\pi}{\text{Im}(\lambda)} \tag{6.49}$$

An aperiodic motion is by definition not oscillatory, which also follows from the fact that the imaginary part of the eigenvalue is zero.

Next, we define the *halving time* $T_{\frac{1}{2}}$, indicating the time interval when the amplitude of the motion has become half its original value,

$$T_{\frac{1}{2}} = \frac{\ln \frac{1}{2}}{\mathrm{Re}(\lambda)} \tag{6.50}$$

However, when the real part of the eigenvalue is positive, the halving time becomes negative. In that case it is better to speak of the *doubling time* of the (diverging) eigenmotion:

$$T_2 = \frac{\ln 2}{\mathrm{Re}(\lambda)}$$

The *damping ratio* ζ for periodic eigenmotion (complex eigenvalues) can be computed with

$$\zeta = -\frac{\mathrm{Re}(\lambda)}{\sqrt{\mathrm{Re}(\lambda)^2 + \mathrm{Im}(\lambda)^2}} \tag{6.51}$$

In the case of damped eigenmotion, ζ is positive (a negative damping ratio is in principle an amplification ratio). For aperiodic motion, ζ is not defined.

Finally, the *natural frequency* ω_n for periodic eigenmotion is defined to be

$$\omega_n = \sqrt{\mathrm{Re}(\lambda)^2 + \mathrm{Im}(\lambda)^2} \tag{6.52}$$

The natural frequency is the theoretical frequency of the eigenmotion when the energy of the system is constant during that motion, which means that the amplitude is constant. Again, for aperiodic motion ω_n is not defined. It should be noted that the natural frequency is more a mathematical notion rather than a physical one.

Using the above definitions, we can write a relation between the period on the one hand, and the damping ratio and the natural frequency on the other:

$$P = \frac{2\pi}{\omega_n \sqrt{1 - \zeta^2}} \tag{6.53}$$

Example 6.3 (*The Oscillating Pendulum Revisited*) Returning to the oscillating pendulum of Example 6.2, with eigenvalues $\lambda = \pm\sqrt{\frac{g}{\ell}}$, we can now easily evaluate the period, P, and natural frequency, ω_n:

$$P = \frac{2\pi}{\mathrm{Im}(\lambda)} = \frac{2\pi}{\sqrt{\frac{g}{\ell}}} = 2\pi\sqrt{\frac{\ell}{g}} \tag{6.54}$$

$$\omega_n = \sqrt{\mathrm{Re}(\lambda)^2 + \mathrm{Im}(\lambda)^2} = \sqrt{\frac{g}{\ell}} \tag{6.55}$$

which is, of course, in line with the earlier results, Eq. (6.33).

The final aspect to be discussed is how the analysis will change if we add damping, because of friction. Let the friction coefficient be given by ν, and act as a multiplication factor for the (velocity-dependent) friction force. Equation (6.30) changes to

$$\ddot{\theta} + \frac{\nu}{m}\dot{\theta} + \frac{g}{\ell}\sin\theta = 0 \tag{6.56}$$

Again assuming small angular displacements, Eq. (6.36) changes to

$$\begin{pmatrix} \dot{\omega} \\ \dot{\theta} \end{pmatrix} = \begin{bmatrix} -\frac{\nu}{m} & -\frac{g}{\ell} \\ 1 & 0 \end{bmatrix} \begin{pmatrix} \omega \\ \theta \end{pmatrix} \tag{6.57}$$

such that the eigenvalues can be obtained from

$$\begin{aligned} \begin{vmatrix} -\lambda - \frac{\nu}{m} & -\frac{g}{\ell} \\ 1 & -\lambda \end{vmatrix} = 0 \Rightarrow \lambda^2 + \frac{\nu}{m}\lambda + \frac{g}{\ell} = 0 \\ \Rightarrow \lambda_{1,2} = -\frac{\nu}{2m} \pm \sqrt{\left(\frac{\nu}{2m}\right)^2 - \frac{g}{\ell}} \end{aligned} \tag{6.58}$$

It is obvious that now we have a different situation: for $\nu = 2m\sqrt{\frac{g}{\ell}}$, the eigenmotion is (critically) damped, i.e., $\zeta = 1$, and aperiodic—the eigenvalues are real ($\lambda_{1,2} = -\frac{\nu}{2m}$). A damped oscillation occurs when

$$\left(\frac{\nu}{2m}\right)^2 - \frac{g}{\ell} < 0 \Rightarrow \nu < 2m\sqrt{\frac{g}{\ell}} \tag{6.59}$$

The standard second-order equation for a system in a damped oscillation is:

$$\lambda^2 + 2\omega_n\zeta\lambda + \omega_n^2 = 0 \tag{6.60}$$

Equating this with the characteristic polynomial gives:

$$\lambda^2 + \frac{\nu}{m}\lambda + \frac{g}{\ell} = \lambda^2 + 2\omega_n\zeta\lambda + \omega_n^2 = 0 \tag{6.61}$$

So, for eigenfrequency and damping we find:

$$\omega_n = \sqrt{\frac{g}{\ell}} \qquad \zeta = \frac{\nu}{2m}\sqrt{\frac{\ell}{g}} \tag{6.62}$$

It is once more noted that the pendulum will only be in (damped) oscillation in case of complex eigenvalues. For three different values of ν, the resulting motion has

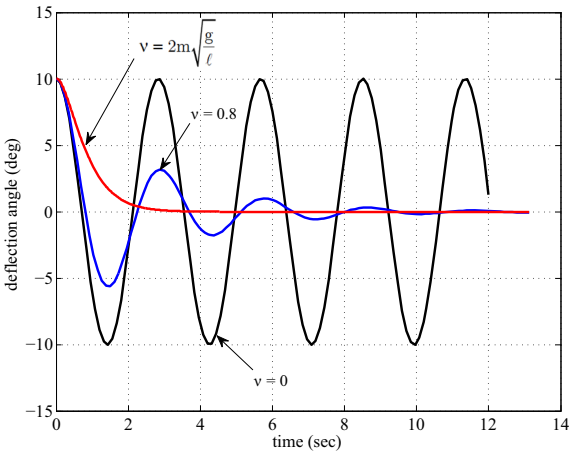


Fig. 6.5 Response curves for the simple pendulum, for three different values of friction coefficient ν

been depicted in Fig. 6.5, i.e., for the undamped case ($\nu = 0$), the critically damped, aperiodic motion ($\nu = 2m\sqrt{\frac{g}{\ell}}$), and the damped oscillation ($\nu = 0.8$). ■

So far, we have only discussed the stability characteristics of the eigenmotion, but we do not know yet in which of the state variables we can trace this motion. Studying the eigenvectors will help us to answer this question. To explain this in more detail, we need to explain the concept of *natural modes*, or *normal modes* when normalised. A normal mode of an oscillating system is a motion pattern, where all parts of the system move sinusoidally with the same frequency and with a fixed phase relation. This free motion takes place at a fixed frequency, which is the eigenfrequency of the system. This eigenfrequency is dependent on the characteristics of the system, e.g., its design, materials, and internal and external boundary conditions. For instance, in structural dynamics, a mode shape is the deformation that the component would show when vibrating at the natural frequency.

Let us have a closer look at natural modes by inspecting a system with two masses, connected to each other and supporting "walls" by springs, as indicated in Fig. 6.6. The equations of motion for such coupled mass-spring systems are available from any book on structural dynamics or vibration analysis, e.g., Donaldson (2006), Meirovitch (2001), but are also easily derived. They are given by:

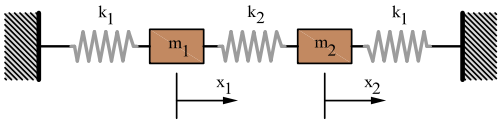


Fig. 6.6 Two-degree-of-freedom system: two masses connected by springs

$$\begin{aligned} m_1\ddot{x}_1 + (k_1+k_2)x_1 - k_2x_2 = 0 \quad &\Rightarrow \quad \ddot{x}_1 = -\frac{k_1+k_2}{m_1}x_1 + \frac{k_2}{m_1}x_2 \\ m_2\ddot{x}_2 + (k_2+k_1)x_2 - k_2x_1 = 0 \quad &\Rightarrow \quad \ddot{x}_2 = \frac{k_2}{m_2}x_1 - \frac{k_1+k_2}{m_2}x_2 \end{aligned} \tag{6.63}$$

or, in matrix form

$$\begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{pmatrix} = \begin{bmatrix} -\frac{k_1+k_2}{m_1} & \frac{k_2}{m_1} \\ \frac{k_2}{m_2} & -\frac{k_1+k_2}{m_2} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \Rightarrow \quad \ddot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{6.64}$$

with $\mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix}^{\mathrm{T}}$ and $\mathbf{A} = \begin{bmatrix} -\frac{k_1+k_2}{m_1} & \frac{k_2}{m_1} \\ \frac{k_2}{m_2} & -\frac{k_1+k_2}{m_2} \end{bmatrix}$.

If, as we did before for the single mass-spring system, we assume the form of the solution to be that of purely oscillatory motion, then:

$$\mathbf{x} = \boldsymbol{\mu} e^{i\omega t} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} e^{i\omega t} \tag{6.65}$$

Substituting this solution into Eq. (6.64) yields

$$\ddot{\mathbf{x}} = -\omega^2 \boldsymbol{\mu} e^{i\omega t} = -\omega^2 \mathbf{x} \tag{6.66}$$

and thus

$$\mathbf{A}\mathbf{x} = -\omega^2 \mathbf{x} = \lambda \mathbf{x} \tag{6.67}$$

with $\lambda = -\omega^2$. This is simply the formulation of an eigenvalue problem. Following the same procedure as before, the eigenvalues are easily calculated from

$$\begin{vmatrix} -\frac{k_1+k_2}{m_1} - \lambda & \frac{k_2}{m_1} \\ \frac{k_2}{m_2} & -\frac{k_1+k_2}{m_2} - \lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \begin{vmatrix} \alpha - \lambda & \gamma \\ \gamma & \beta - \lambda \end{vmatrix} = 0 \tag{6.68}$$

where we have introduced the coefficients α, β, and γ for convenience: $\alpha = -\frac{k_1+k_2}{m_1}$, $\beta = -\frac{k_1+k_2}{m_2}$, and $\gamma = \frac{k_2}{m_2}$. Expanding the determinant results in

$$\lambda^2 - (\alpha+\beta)\lambda + (\alpha\beta - \gamma^2) = 0 \tag{6.69}$$

such that

$$\lambda_{1,2} = -\omega_{1,2}^2 = \frac{(\alpha+\beta) \pm \sqrt{(\alpha+\beta)^2 - 4(\alpha\beta - \gamma^2)}}{2} \tag{6.70}$$

For the remainder of this discussion, let us now consider some numerical cases, which are easier to deal with and provide some more insight. Starting with $m_1 = m_2$

= 1 kg, and $k_1 = k_2 = 1$ N/m, this gives $\alpha = -2$, $\beta = -2$, and $\gamma = 1$. Thus, $\omega_1^2 = 3\,\text{rad}^2/\text{s}^2$ and $\omega_2^2 = 1\,\text{rad}^2/\text{s}^2$. Then, with $(\mathbf{A} + \omega_1^2\mathbf{I})\boldsymbol{\mu}_1 = 0$ the first eigenvector, $\boldsymbol{\mu}_1$, is found with

$$\left(\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}\right)\begin{pmatrix} \mu_{1,1} \\ \mu_{1,2} \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{pmatrix} \mu_{1,1} \\ \mu_{1,2} \end{pmatrix} = 0 \tag{6.71}$$

Solving the two equations results in $\boldsymbol{\mu}_1 = \begin{pmatrix} 1 & -1 \end{pmatrix}^{\mathrm{T}}$. Similarly, for the second eigenvector we find $\boldsymbol{\mu}_2 = \begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathrm{T}}$.

The two eigenvectors are the natural modes of the system, the meaning of which can be shown when we look at the general motion solution of the mass-spring system. With the assumed solution, Eq. (6.65), and realising that we have found *four* eigenfrequencies, the general solution for the mass-spring system Eq. (6.63) is given by:

$$\mathbf{x} = c_1\boldsymbol{\mu}_1 e^{i\omega_1 t} + c_2\boldsymbol{\mu}_1 e^{-i\omega_1 t} + c_3\boldsymbol{\mu}_2 e^{i\omega_2 t} + c_4\boldsymbol{\mu}_2 e^{-i\omega_2 t} \tag{6.72}$$

As before, for a real response the coefficients c_1 through c_4 need to be complex, and are, in fact, two complex conjugate pairs: c_1 and c_2, and c_3 and c_4. This means that if we rewrite Eq. (6.72) with the help of Euler's Formula, Eq. (6.46), half the terms will cancel out, leaving:

$$\mathbf{x}(t) = \gamma_1\boldsymbol{\mu}_1 \cos\omega_1 t + \gamma_3\boldsymbol{\mu}_1 \sin\omega_1 t + \gamma_2\boldsymbol{\mu}_2 \cos\omega_2 t + \gamma_4\boldsymbol{\mu}_2 \sin\omega_2 t \tag{6.73}$$

With the initial conditions at $t = 0$ s, we can now solve for the unknowns γ_1, γ_2, γ_3, and γ_4:

$$\mathbf{x}(0) = \gamma_1\boldsymbol{\mu}_1 + \gamma_2\boldsymbol{\mu}_2 \qquad \dot{\mathbf{x}}(0) = \omega_1\gamma_3\boldsymbol{\mu}_1 + \omega_2\gamma_4\boldsymbol{\mu}_2 \tag{6.74}$$

Suppose now that the velocity at $t = 0$ is zero, and that we shift the masses to an arbitrary (non-zero) position. With the non-zero eigenfrequencies, this yields $\gamma_3 = \gamma_4 = 0$, and thus the solution simplifies to $\mathbf{x}(t) = \gamma_1\boldsymbol{\mu}_1 \cos\omega_1 t + \gamma_2\boldsymbol{\mu}_2 \cos\omega_2 t$. The remaining unknowns, γ_1 and γ_2, follow from

$$\mathbf{x}(0) = \mathbf{M}\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \tag{6.75}$$

where we have concatenated the two eigenvectors, $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1 & \boldsymbol{\mu}_2 \end{bmatrix}$. The two coefficients can then be obtained from the inverse of Eq. (6.75), i.e.,

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \mathbf{M}^{-1}\mathbf{x}(0) \tag{6.76}$$

In Fig. 6.7, the motion of the two masses has been plotted for three different initial positions: $\mathbf{x}(0) = \boldsymbol{\mu}_1$ m, $\mathbf{x}(0) = \boldsymbol{\mu}_2$ m, and $\mathbf{x}(0) = \begin{pmatrix} 1 & 0 \end{pmatrix}$ m. From these results we

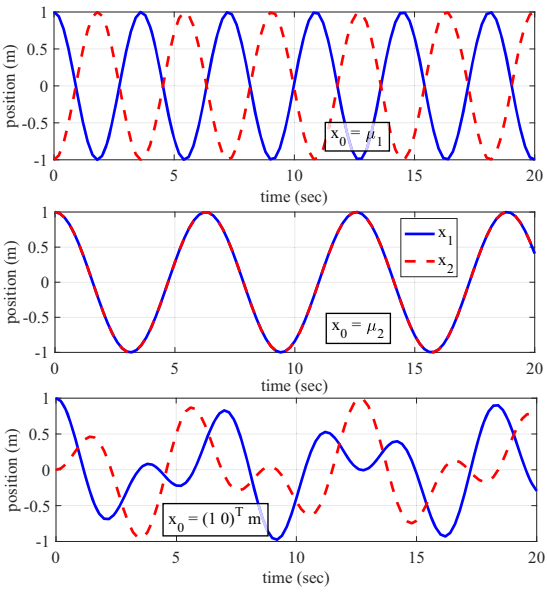


Fig. 6.7 Response of the two-degree-of-freedom mass-spring system for various initial conditions

see that when the initial condition is given by one of the eigenvectors, the system moves in its *natural mode*. The amplitude of m_1 and m_2 is given by the magnitude of the corresponding eigenvector element (related to x_1 and x_2), and the sign of the elements determines the direction of the motion: a positive and negative sign means that the masses move opposite to each other (out of phase), whereas the same sign means the motion is in sync (in phase) (the curves in the middle plot are overlapping). The situation where we have a different initial condition, the motion is the result of the combined natural motions, and it it becomes harder to analyse the system.[3] In the upcoming example we will dig a little deeper concerning the relation between system aspects and motion components.

Example 6.4 (*Second-Order Mass-Spring System*) Above, we have looked at a coupled mass-spring system, consisting of two masses. We identified the mode shapes by looking at the positions of the masses, and how they behaved. The analysis was based on the two second-order differential equations, Eq. (6.63). What changes if we transform this system to four first-order equations? In terms of state variables, they will consist of both velocity and position of each of the masses:

[3] Note that commonly the eigenvectors are normalised, which, of course, will not matter for the natural motion: only the amplitude is scaled, whereas the ratio between the components remains the same. Normalising $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ would give $\boldsymbol{\mu}_1 = \left(\frac{1}{2}\sqrt{2} \ -\frac{1}{2}\sqrt{2}\right)^{\mathrm{T}}$ and $\boldsymbol{\mu}_2 = \left(\frac{1}{2}\sqrt{2} \ \frac{1}{2}\sqrt{2}\right)^{\mathrm{T}}$. Using these vectors in Eq. (6.76) would simply scale down the amplitude by 0.707 in the top two figures of Fig. 6.7.

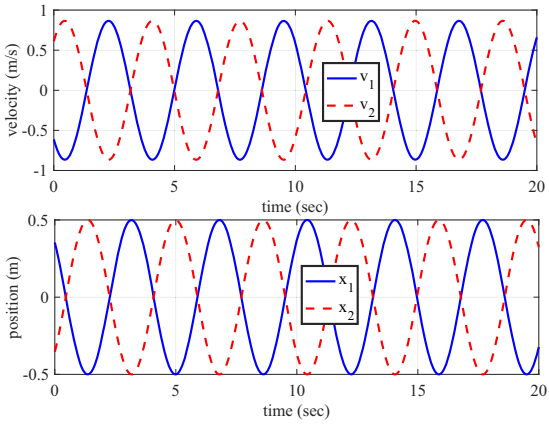


Fig. 6.8 Response of the two-degree-of-freedom mass-spring system (nominal model: $m_1 = m_2 = 1\,\text{kg}$)

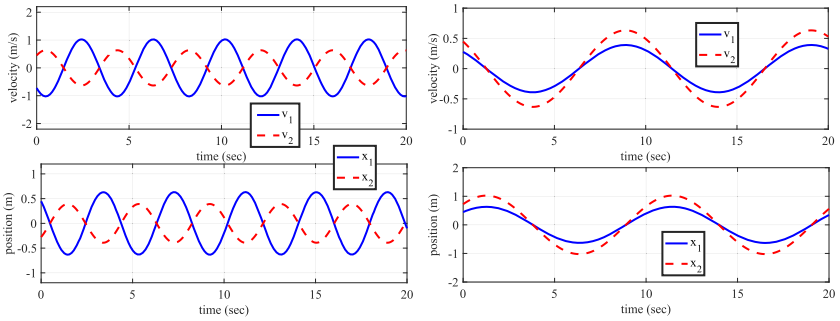
$$\begin{pmatrix} \dot{v}_1 \\ \dot{x}_1 \\ \dot{v}_2 \\ \dot{x}_2 \end{pmatrix} = \begin{bmatrix} 0 & -\frac{k_1+k_2}{m_1} & 0 & \frac{k_2}{m_1} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{k_2}{m_2} & 0 & -\frac{k_1+k_2}{m_2} \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} v_1 \\ x_1 \\ v_2 \\ x_2 \end{pmatrix} \tag{6.77}$$

which will resemble more the situation that we will encounter later: the state vector of the entry vehicles will include both position (attitude) and velocity (angular rate).

The earlier results showed that for $m_1 = m_2 = k_1 = k_2 = 1$ the corresponding (pairs) of natural frequencies are $\omega_1^2 = 3$ and $\omega_2^2 = 1$, which are equal to the negative of the two eigenvalues. In case we determine the eigenvalues for the coefficient matrix of Eq. (6.77), we find $\lambda_{1,2} = \pm\sqrt{3}i$ and $\lambda_{3,4} = \pm i$, which gives, of course, $\omega_1 = \sqrt{3}$ rad/s and $\omega_2 = 1$ rad/s (with Eq. (6.52)). The corresponding (normalised) eigenvectors are $\boldsymbol{\mu}_{1,2} = \begin{pmatrix} -0.6124 & \pm 0.3536i & 0.6124 & \mp 0.3536i \end{pmatrix}^{\mathrm{T}}$ and $\boldsymbol{\mu}_{3,4} = \begin{pmatrix} 0.5 & \mp 0.5i & 0.5 & \mp 0.5i \end{pmatrix}^{\mathrm{T}}$.

If we now simulate the system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, we do expect the same results as before. Indeed, with $\mathbf{x_0} = \boldsymbol{\mu}_1$ the results as shown in Fig. 6.8 are obtained.[4] In this case the maximum amplitude of the positions of m_1 and m_2 is 0.5 (rather than 1, see Fig. 6.7 top figure), because the eigenvectors have been normalised. What is important, though, is the ratio between the maximum amplitudes of velocity (0.866) and position (0.5), which is 1.7321. And this is exactly the ratio between the first and second element of $\boldsymbol{\mu}_1$, apart from the sign. This is an important observation: the order of the elements is the same as that of the state vector, and the absolute magnitude is an indication

[4] Defining $\mathbf{x_0} = \boldsymbol{\mu}_1$ is in this case slightly more involved than it appears. The magnitude is defined by the norm of the complex number, but as the elements are either real or complex, the corresponding sign of either the real or complex part is used. So, with $\boldsymbol{\mu}_1 = \begin{pmatrix} -0.6124 & 0.3536i & 0.6124 & -0.3536i \end{pmatrix}^{\mathrm{T}}$, it follows that $\mathbf{x_0} = \begin{pmatrix} -0.6124 & 0.3536 & 0.6124 & -0.3536 \end{pmatrix}^{\mathrm{T}}$ m.



(a) $\mathbf{x_0} = \boldsymbol{\mu}_1$ (b) $\mathbf{x_0} = \boldsymbol{\mu}_3$

Fig. 6.9 Response of the two-degree-of-freedom mass-spring system ($m_1 = 1$, $m_2 = 2$)

of the maximum amplitude of the natural motion. This means, that by inspecting the eigenvectors we can draw conclusions about the nature of the modes, i.e., which state variables are involved and how much "power" each state is contributing to the motion.

To verify this observation, we slightly change the model, i.e., m_2 is doubled compared to m_1. For this system, the eigenvalues of $\mathbf{A}$ are $\lambda_{1,2} = \pm 1.6180i$ and $\lambda_{3,4} = \pm 0.6180i$, with corresponding eigenfrequencies $\omega_1 = 1.618$ rad/s and $\omega_2 = 0.618$ rad/s. The eigenvectors are $\boldsymbol{\mu}_{1,2} = \begin{pmatrix} 0.7236 & \pm 0.4472i & 0.4472 & \mp 0.2764i \end{pmatrix}^{\mathrm{T}}$ and $\boldsymbol{\mu}_{3,4} = \begin{pmatrix} \pm 0.2764i & 0.4472 & \pm 0.4472i & 0.7236 \end{pmatrix}^{\mathrm{T}}$. Doing two simulations, with $\mathbf{x_0} = \boldsymbol{\mu}_1$ and $\mathbf{x_0} = \boldsymbol{\mu}_3$, the results as shown in Fig. 6.9 are obtained. The simulations show the normal-mode motions of the system, and the amplitudes of velocity and position components do indeed match the relative magnitude of the corresponding eigenvector elements. So it is confirmed that we can use the eigenvectors to identify the state variables involved in the eigenmotion. ■

Let us now formalise what we have learned from the above analysis. Suppose we compute the modulus of each of the (complex) components of an eigenvector $\boldsymbol{\mu}_i$, then we can easily spot the components which are involved. Furthermore, computing the argument of the complex component, we get an impression of the phase difference between the related components. In mathematical terms, the modulus, z, and argument, θ, of a complex number λ are defined to be (Fig. 6.10):

$$z = \sqrt{\mathrm{Re}(\lambda)^2 + \mathrm{Im}(\lambda)^2} \qquad \theta = \arctan\left[\frac{\mathrm{Im}(\lambda)}{\mathrm{Re}(\lambda)}\right] \tag{6.78}$$

Note that the argument of a real number is always 0 or π ($\lambda < 0$).

Each component of the eigenvector can in principle be plotted as a time vector using z and θ, which will give us a visual aid to inspect the eigenmotion in the sense that the relative magnitude and phase difference of the related components can be

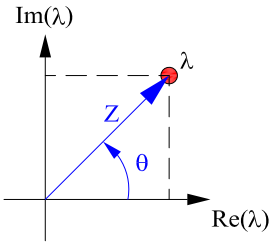


Fig. 6.10 The modulus z and argument θ of a complex number

distinguished, e.g., see Etkin (1972). In this chapter, we will not do that but restrict to a numerical inspection of the more important components of the state vector.

The last question to be answered is: can we reproduce the natural motion of the system, having only the eigenvalues and eigenvectors at our disposal? The answer is "yes", and with λ and $\boldsymbol{\mu}$, the eigenmotion $\mathbf{x}_{\lambda(t)}$ of the system follows from (see also the books by Kuo 1987; Meirovitch 2001):

$$\mathbf{x}_\lambda(t) = e^{\lambda t}\boldsymbol{\mu} \tag{6.79}$$

Since we are dealing with a linear system of n coupled equations, Eq. (6.79) gives only one solution. The general form is, because of the linearity of the system, written as

$$\mathbf{x}_\lambda(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t}\boldsymbol{\mu}_i \tag{6.80}$$

The constants c_i can be computed from specified initial conditions $\mathbf{x}(0)$, for instance, an initial perturbation in the nominal angle of attack:

$$\mathbf{x}(0) = \begin{bmatrix}\boldsymbol{\mu}_1 & \boldsymbol{\mu}_2 & \cdots & \boldsymbol{\mu}_\mathbf{n}\end{bmatrix}\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{bmatrix}\boldsymbol{\mu}_1 & \boldsymbol{\mu}_2 & \cdots & \boldsymbol{\mu}_\mathbf{n}\end{bmatrix}^{-1}\mathbf{x}(0) \tag{6.81}$$

Having covered all required theory about eigenmotion, we can now start preparing for analysing the characteristic motion of entry vehicles. In the upcoming section, the linearisation of the equations of motion will be discussed. This will enable us to set up the system matrix $\mathbf{A}$, which is at the heart of the analysis.

6.3 Linearisation of Equations of Motion

The characteristic motion, or the open-loop behaviour of a vehicle, can be described by the eigenvalues and eigenvectors of the equations of motion. However, the eigenvalues and eigenvectors can only be obtained when the time derivatives of the states are given as a linear combination of the states. In other words: the equations of motion have to be linearised. As a result, the motion given by this linearised system should be regarded as an equilibrium trajectory, so that the characteristic motions are small deviations from this nominal path. To simplify the process of linearisation, we will make two assumptions. In this way we can already remove some terms from the equations, which would otherwise disappear during the linearisation.

1. We will consider a non-rotating Earth ($\omega_{cb} = 0\,\text{rad/s}$), which is allowed since the rotation of the vehicle (and its local horizontal plane) is of a much higher frequency than the rotation of the Earth. As a result, the Coriolis and centripetal accelerations are zero as well,
2. The vehicle is assumed to be rotationally symmetric (with respect to mass) around the X_B-axis of the body-fixed reference frame, which means that all products of inertia are zero. In Intermezzo 6.1, we will revisit this assumption to prepare the equations for vehicles with a strong mass asymmetry, such as Apollo.

With the above assumptions, we can neglect the smaller terms. The resulting equations of motion become:

$$\dot{V} = -\frac{D}{m} - g\sin\gamma \tag{6.82}$$

$$\dot{\gamma} = \left(\frac{V}{R} - \frac{g}{V}\right)\cos\gamma + \frac{(L\cos\sigma - S\sin\sigma)}{mV} \tag{6.83}$$

$$\dot{\chi} = \frac{V}{R}\cos\gamma\tan\delta\sin\chi - \frac{(L\sin\sigma + S\cos\sigma)}{mV\cos\gamma} \tag{6.84}$$

$$\dot{R} = V\sin\gamma \tag{6.85}$$

$$\dot{\tau} = \frac{V\cos\gamma\sin\chi}{R\cos\delta} \tag{6.86}$$

$$\dot{\delta} = \frac{V}{R}\cos\gamma\cos\chi \tag{6.87}$$

$$\dot{p} = \frac{M_x}{I_{xx}} + \left(I_{yy} - I_{zz}\right)qr \tag{6.88}$$

$$\dot{q} = \frac{M_y}{I_{yy}} + (I_{zz} - I_{xx})\, pr \tag{6.89}$$

$$\dot{r} = \frac{M_z}{I_{zz}} + \left(I_{xx} - I_{yy}\right) pq \tag{6.90}$$

Since the kinematic attitude equations are very complex, we will introduce another simplification. We assume that the vehicle's attitude is only marginally influenced by an asymmetric translational motion, or in other words: we assume that the vehicle's trajectory is parallel to the equator ($\delta = 0°$ and $\chi = 90°$), and remains that way ($\dot{\delta} = 0°/\text{s}$). In that case Eqs. (6.12)–(6.14) reduce to:

$$\dot{\alpha} = q - (p\cos\alpha + r\sin\alpha)\tan\beta - \frac{L - mg\cos\gamma\cos\sigma}{mV\cos\beta} \tag{6.91}$$

$$\dot{\beta} = p\sin\alpha - r\cos\alpha - \frac{S + mg\cos\gamma\sin\sigma}{mV} \tag{6.92}$$

$$\dot{\sigma} = -\frac{p\cos\alpha + r\sin\alpha}{\cos\beta} - \frac{L - mg\cos\gamma\cos\sigma}{mV}\tan\beta + \frac{L\sin\sigma + S\cos\sigma}{mV}\tan\gamma \tag{6.93}$$

As we can see, the equations for χ, τ and δ are no longer coupled to the rest of the equations, so a 9-degrees-of-freedom linearised model for the state space formed by V, γ, R, p, q, r, α, β and σ can be derived.

The linearisation is done by developing a Taylor series to first order of the equations of motion (see for more details on Taylor series Sect. C.2 in Appendix C. A formal discussion on linearisation can be found in Sect. C.3). The outcome of the linearisation is a set of coupled linear differential equations, and can be written in matrix form:

$$\Delta\dot{\mathbf{x}} = \mathbf{A}\Delta\mathbf{x} + \mathbf{B}\Delta\mathbf{u} \tag{6.94}$$

where $\Delta\mathbf{x}$ is the n x 1 error state vector, $\Delta\mathbf{u}$ is the m x 1 noise input vector, and $\mathbf{A}$ and $\mathbf{B}$ are the $n \times n$ system or state matrix and $n \times m$ input or control matrix, respectively. $\mathbf{A}$ and $\mathbf{B}$ can directly be derived from the non-linear equations of motion with:

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{bmatrix} \tag{6.95}$$

evaluated at the nominal condition. In the above definitions, f_i are the equations of motion, $\dot{V}$, $\dot{\gamma}$, etc.

However, to provide more insight we will first explain the linearisation process in a slightly more elaborate way. We assume an equilibrium value of each of the state

variables (index 0), and we will look at small deviations from this equilibrium state. Note that the vehicle is not powered (no fuel consumption), which means that the mass properties are constant. So:

$$\begin{array}{ccc} V = V_0 + \Delta V & \gamma = \gamma_0 + \Delta\gamma & R = R_0 + \Delta R \\ p = p_0 + \Delta p & q = q_0 + \Delta q & r = r_0 + \Delta r \\ \alpha = \alpha_0 + \Delta\alpha & \beta = \beta_0 + \Delta\beta = \Delta\beta & \sigma = \sigma_0 + \Delta\sigma \end{array} \tag{6.96}$$

Note that by definition of the nominal state, this also results in nominal aerodynamic forces D_0, S_0 and L_0, as well as nominal aerodynamic moments $\mathcal{L}_0$, $\mathcal{M}_0$ and $\mathcal{N}_0$. Similarly, the small state deviations give force and moment deviations ΔD, ΔS, ΔL, $\Delta\mathcal{L}$, $\Delta\mathcal{M}$ and $\Delta\mathcal{N}$.

The nominal position (R_0), velocity (V_0 and γ_0) and corresponding control history (α_0, β_0 and σ_0) follow from the nominal trajectory (see Sect. 6.4.2), which leaves us with three unknowns p_0, q_0 and r_0. However, for controlled flight, since we want no perturbation of the nominal control (α_0 and σ_0) when the vehicle is flying, the equilibrium angular rates can be computed from the condition.[5]

$$\dot{\alpha} = \dot{\beta} = \dot{\sigma} = 0 \tag{6.97}$$

or, using Eqs. (6.91)–(6.93),

$$q_0 = \frac{L_0}{mV_0} - \frac{g_0}{V_0}\cos\gamma_0\cos\sigma_0 \tag{6.98}$$

$$p_0\sin\alpha_0 - r_0\cos\alpha_0 = \frac{g_0}{V_0}\cos\gamma_0\sin\sigma_0 \tag{6.99}$$

$$p_0\cos\alpha_0 + r_0\sin\alpha_0 = \frac{L_0}{mV_0}\tan\gamma_0\sin\sigma_0 \tag{6.100}$$

The expressions for p_0 and r_0 are easily solved when realising that in the matrix expression

$$\begin{bmatrix} \sin\alpha_0 & -\cos\alpha_0 \\ \cos\alpha_0 & \sin\alpha_0 \end{bmatrix} \begin{pmatrix} p_0 \\ r_0 \end{pmatrix} = \begin{pmatrix} \frac{g_0}{V_0}\cos\gamma_0\sin\sigma_0 \\ \frac{L_0}{mV_0}\tan\gamma_0\sin\sigma_0 \end{pmatrix} \tag{6.101}$$

the coefficient matrix is orthonormal, and thus

$$\begin{bmatrix} \sin\alpha_0 & -\cos\alpha_0 \\ \cos\alpha_0 & \sin\alpha_0 \end{bmatrix}^{-1} = \begin{bmatrix} \sin\alpha_0 & -\cos\alpha_0 \\ \cos\alpha_0 & \sin\alpha_0 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \sin\alpha_0 & \cos\alpha_0 \\ -\cos\alpha_0 & \sin\alpha_0 \end{bmatrix} \tag{6.102}$$

[5] In case the vehicle is uncontrolled, or put into a spin for stability reasons, the nominal angular rate can have different values. What this will be should be determined on a case-by-case basis.

such that p_0 and r_0 are easily isolated. Thus, p_0, q_0 and r_0 are found to be:

$$p_0 = c_1 \sin\alpha_0 + c_2 \cos\alpha_0 \tag{6.103}$$

$$q_0 = \frac{L_0}{mV_0} - \frac{g_0}{V_0}\cos\gamma_0 \cos\sigma_0 \tag{6.104}$$

$$r_0 = -c_1 \cos\alpha_0 + c_2 \sin\alpha_0 \tag{6.105}$$

with

$$c_1 = \frac{g_0}{V_0}\cos\gamma_0 \sin\sigma_0 \tag{6.106}$$

$$c_2 = \frac{L_0}{mV_0}\tan\gamma_0 \sin\sigma_0 \tag{6.107}$$

Note that since $\beta_0 = 0$, also $S_0 = 0$.

In case the model is to be used to analyse the stability during manoeuvring, where $\dot{\alpha}$, $\dot{\beta}$, and $\dot{\sigma}$ are not necessarily zero, the above expressions for p_0, q_0 and r_0 change accordingly. In that case

$$q_0 = \frac{L_0}{mV_0} - \frac{g_0}{V_0}\cos\gamma_0 \cos\sigma_0 + \dot{\alpha} \tag{6.108}$$

and to solve

$$\begin{bmatrix} \sin\alpha_0 & -\cos\alpha_0 \\ \cos\alpha_0 & \sin\alpha_0 \end{bmatrix} \begin{pmatrix} p_0 \\ r_0 \end{pmatrix} = \begin{pmatrix} \frac{g_0}{V_0}\cos\gamma_0 \sin\sigma_0 + \dot{\beta} \\ \frac{L_0}{mV_0}\tan\gamma_0 \sin\sigma_0 + \dot{\sigma} \end{pmatrix} \tag{6.109}$$

Even though the formulations for p_0 and r_0 do not change, Eqs. (6.103) and (6.105), the coefficients c_1 and c_2 do, i.e.,

$$c_1 = \frac{g_0}{V_0}\cos\gamma_0 \sin\sigma_0 + \dot{\beta} \tag{6.110}$$

$$c_2 = \frac{L_0}{mV_0}\tan\gamma_0 \sin\sigma_0 + \dot{\sigma} \tag{6.111}$$

Substituting Eq. (6.96) into Eqs. (6.82)–(6.93) gives us equations with nominal (e.g., V_0, L_0, etc.), first-order (e.g., ΔV_0, ΔL_0, etc.) and higher-order terms (e.g., $\Delta V \Delta\gamma$, $p_0 \Delta\alpha$, where the nominal rotational rates p_0, q_0 and r_0 can be treated as perturbations, because they are small). So if we neglect now the higher-order terms and subtract the nominal state (i.e., $\dot{V}_0 = \ldots$, etc.) we will get the linearised differential equations. With either of the two methods shown in Appendix C, Sect. C.3, the following nine first-order differential equations are obtained:

$$\Delta\dot{V} = -\frac{\Delta D}{m} + 2\frac{g_0}{R_0}\sin\gamma_0\Delta R - g_0\cos\gamma_0\Delta\gamma \tag{6.112}$$

$$\Delta\dot{\gamma} = \left(-\dot{\gamma}_0 + \frac{2V_0}{R_0}\cos\gamma_0\right)\frac{\Delta V}{V_0} + \left(\frac{2g_0}{R_0} - \frac{V_0^2}{R_0^2}\right)\frac{\cos\gamma_0}{V_0}\Delta R - \left(\frac{V_0^2}{R_0} - g_0\right)\frac{\sin\gamma_0}{V_0}\Delta\gamma + \\ - \frac{L_0}{mV_0}\sin\sigma_0\Delta\sigma + \frac{\cos\sigma_0}{mV_0}\Delta L - \frac{\sin\sigma_0}{mV_0}\Delta S \tag{6.113}$$

$$\Delta\dot{R} = \sin\gamma_0\Delta V + V_0\cos\gamma_0\Delta\gamma \tag{6.114}$$

$$\Delta\dot{p} = \frac{\Delta M_x}{I_{xx}} \tag{6.115}$$

$$\Delta\dot{q} = \frac{\Delta M_y}{I_{yy}} \tag{6.116}$$

$$\Delta\dot{r} = \frac{\Delta M_z}{I_{zz}} \tag{6.117}$$

$$\Delta\dot{\alpha} = \Delta q - \frac{1}{mV_0}\Delta L - \frac{g_0}{V_0}\cos\gamma_0\sin\sigma_0\Delta\sigma + \left(\frac{L_0}{mV_0^2} - \frac{g_0}{V_0^2}\cos\gamma_0\cos\sigma_0\right)\Delta V + \\ - \frac{g_0}{V_0}\sin\gamma_0\cos\sigma_0\Delta\gamma - \frac{2g_0}{R_0V_0}\cos\gamma_0\cos\sigma_0\Delta R \tag{6.118}$$

$$\Delta\dot{\beta} = \sin\alpha_0\Delta p - \cos\alpha_0\Delta r - \frac{\Delta S}{mV_0} - \frac{g_0}{V_0}\cos\gamma_0\cos\sigma_0\Delta\sigma + \frac{g_0}{V_0^2}\cos\gamma_0\sin\sigma_0\Delta V + \\ + \frac{2g_0}{R_0V_0}\cos\gamma_0\sin\sigma_0\Delta R + \frac{g_0}{V_0}\sin\gamma_0\sin\sigma_0\Delta\gamma \tag{6.119}$$

$$\Delta\dot{\sigma} = -\cos\alpha_0\Delta p - \sin\alpha_0\Delta r - \left(\frac{L_0}{mV_0} - \frac{g_0}{V_0}\cos\gamma_0\cos\sigma_0\right)\Delta\beta + \frac{L_0}{mV_0}\sin\sigma_0\Delta\gamma + \\ + \frac{\tan\gamma_0}{mV_0}\left(\sin\sigma_0\Delta L + \cos\sigma_0 L_0\Delta\sigma + \cos\sigma_0\Delta S - \frac{L_0}{V_0}\sin\sigma_0\Delta V\right) \quad (6.120)$$

In deriving the above equations, we have used the central-field definition of the gravitational acceleration, i.e., $g = \frac{\mu}{R^2}$, which gives us

$$g_0 + \Delta g = \frac{\mu}{(R_0 + \Delta R)^2} = \frac{\mu}{R_0^2}\frac{1}{1 + 2\frac{\Delta R}{R_0}} \approx \frac{\mu}{R_0^2}\left(1 - 2\frac{\Delta R}{R_0}\right)$$

or

$$\Delta g = -2g_0\frac{\Delta R}{R_0}$$

Now that we have linearised the equations of motion, we can write them in matrix form, also called the state-space form, see Eq. (6.94). This will be treated in the corresponding sections for both the winged entry vehicle and the entry capsules.

Intermezzo 6.1 (Rotational Dynamics for Vehicles with Mass Asymmetry)
In case a vehicle is not rotationally symmetric with respect to mass, for instance, when the centre of mass is shifted down or upwards in Z-direction to create a non-zero trim angle of attack, the assumption of I_{xz} may no longer hold. This is true for Apollo (Sect. 6.5.3), such that for the linearisation process we need to consider Eqs. (6.9) through (6.11) instead of Eqs. (6.88)–(6.90). Earlier, we also assumed that the rotational rates were small, and that products of a (nominal) rate with a perturbed rate could be treated as a second-order term. In this Intermezzo, though, we will not do that. This will lead to additional terms that might come in handy when studying the stability of spinning objects, e.g., launchers and missiles, or spinning entry probes, such as Huygens. It is noted, though, that in such a case typically only a single component of the angular-rate vector will be large.

Let us begin with Eq. (6.9):

$$\dot{p} = \frac{I_{zz}}{I^*}M_x + \frac{I_{xz}}{I^*}M_z + \frac{\left(I_{xx} - I_{yy} + I_{zz}\right)I_{xz}}{I^*}pq + \frac{\left(I_{yy} - I_{zz}\right)I_{zz} - I_{xz}^2}{I^*}qr \quad (6.121)$$

For constant inertia, the first two terms are expressed as before, Eqs. (6.115) and (6.117), where these terms may be a combination of aerodynamic and propulsive moments, either originating from the vehicle itself or the (controlled) actuators. They have to be developed on a case-by-case basis.

So, the related expressions are

$$\frac{I_{zz}}{I^*}\Delta M_x \quad \text{and} \quad \frac{I_{xz}}{I^*}\Delta M_z \quad (6.122)$$

The third term on the right-hand side linearises to two individual terms, i.e.,

$$I_{p_1} q_0 \Delta p \quad \text{and} \quad I_{p_1} p_0 \Delta q \tag{6.123}$$

with

$$I_{p_1} = \frac{(I_{xx} - I_{yy} + I_{zz}) I_{xz}}{I^*}$$

and, as has been defined earlier,

$$I^* = I_{xx} I_{zz} - I_{xz} I_{xz}$$

Similarly, linearisation of the last term on the right-hand side of Eq. (6.121) also yields two terms:

$$I_{p_2} r_0 \Delta q \quad \text{and} \quad I_{p_2} q_0 \Delta r \tag{6.124}$$

with

$$I_{p_2} = \frac{\left(I_{yy} - I_{zz}\right) I_{zz} - I_{xz}^2}{I^*}$$

In complete form:

$$\Delta \dot{p} = I_{p_1} q_0 \Delta p + (I_{p_1} p_0 + I_{p_2} r_0) \Delta q + I_{p_2} q_0 \Delta r + \frac{I_{zz}}{I^*} \Delta M_x + \frac{I_{xz}}{I^*} \Delta M_z \tag{6.125}$$

Inspecting this equations reveals several coupling terms. Apart from the inertia coupling through I_{p_1} and I_{p_2}, these (non-zero) constants multiply with the three angular-rate terms Δp, Δq, and Δr. Thus, all three axes link with $\Delta \dot{p}$, provided that the nominal rate components are non-zero. And, perturbations in the (control) moments about the Z-axis (ΔM_z) induce a perturbation on the roll axis, which becomes stronger with increasing I_{xz}.

The first term of Eq. (6.10),

$$\dot{q} = \frac{M_y}{I_{yy}} + \frac{I_{xz}}{I_{yy}} \left(r^2 - p^2\right) + \frac{I_{zz} - I_{xx}}{I_{yy}} pr \tag{6.126}$$

linearises to

$$\frac{\Delta M_y}{I_{yy}} \tag{6.127}$$

whereas the second term gives us two components:

$$2\frac{I_{xz}}{I_{yy}}\left(r_0\Delta r - p_0\Delta p\right) \tag{6.128}$$

Finally, linearising the third term on the right-hand side of Eq. (6.126) yields

$$\frac{I_{zz}-I_{xx}}{I_{yy}}\left(r_0\Delta p + p_0\Delta r\right) \tag{6.129}$$

such that the complete equation becomes:

$$\Delta\dot{q} = \left(I_{q_2}r_0 - 2I_{q_1}p_0\right)\Delta p + \left(2I_{q_1}r_0 + I_{q_2}p_0\right)\Delta r + \frac{\Delta M_y}{I_{yy}} \tag{6.130}$$

with

$$I_{q_1} = \frac{I_{xz}}{I_{yy}} \quad \text{and} \quad I_{q_2} = \frac{I_{zz}-I_{xx}}{I_{yy}}$$

Through the first two terms on the right-hand side of Eq. (6.130), longitudinal motion has become coupled with lateral motion.

Last, but not least, the third Eq. (6.11), linearises in the same manner as Eq. (6.121) and results in

$$\Delta\dot{r} = I_{r_1}q_0\Delta p + \left(I_{r_1}p_0 + I_{r_2}r_0\right)\Delta q + I_{r_2}q_0\Delta r + \frac{I_{xz}}{I^*}\Delta M_x + \frac{I_{xx}}{I^*}\Delta M_z \tag{6.131}$$

with

$$I_{r_1} = \frac{\left(I_{xx}-I_{yy}\right)I_{xx} - I_{xz}^2}{I^*}$$

$$I_{r_2} = \frac{\left(-I_{xx}+I_{yy}-I_{zz}\right)I_{xz}}{I^*}$$

The above set of equations, Eqs. (6.125), (6.130) and (6.131), includes all coupling terms, because of mass asymmetry and non-zero nominal angular rate. For spinning vehicles, typically only p_0 is sufficiently large to enforce this coupling. In that case, the equations reduce to:

$$\Delta\dot{p} = I_{p_1}p_0\Delta q + \frac{I_{zz}}{I^*}\Delta M_x + \frac{I_{xz}}{I^*}\Delta M_z \tag{6.132}$$

$$\Delta\dot{q} = -2I_{q_1}p_0\Delta p + I_{q_2}p_0\Delta r + \frac{\Delta M_y}{I_{yy}} \tag{6.133}$$

$$\Delta\dot{r} = I_{r_1} p_0 \Delta q + \frac{I_{xz}}{I^*}\Delta M_x + \frac{I_{xx}}{I^*}\Delta M_z \tag{6.134}$$

provided that $I_{xz} \neq 0$. In case the latter condition is not fulfilled, the auxiliary parameters change to $I^* = I_{xx} I_{zz}$, $I_{p_1} = I_{q_1} = I_{r_2} = 0$, $I_{p_2} = \frac{I_{yy}-I_{zz}}{I_{xx}}$, $I_{q_2} = \frac{I_{zz}-I_{xx}}{I_{yy}}$, and $I_{r_1} = \frac{I_{xx}-I_{yy}}{I_{zz}}$. In that case, the following set is obtained:

$$\Delta\dot{p} = \frac{\Delta M_x}{I_{xx}} \tag{6.135}$$

$$\Delta\dot{q} = \frac{I_{zz} - I_{xx}}{I_{yy}} p_0 \Delta r + \frac{\Delta M_y}{I_{yy}} \tag{6.136}$$

$$\Delta\dot{r} = \frac{I_{xx} - I_{yy}}{I_{zz}} p_0 \Delta q + \frac{\Delta M_z}{I_{zz}} \tag{6.137}$$

■

6.4 Eigenmotion of Winged Entry Vehicles

Before we can study the open-loop eigenmotion of the winged re-entry vehicle HORUS, the linearised equations obtained in the previous section has to be put into state-space form. This will be discussed in the next Sect. 6.4.1. As the state-space model is dependent on the trajectory that is flown, in Sect. 6.4.2 the HORUS nominal trajectory is introduced, and an overview of the parameters that are required to evaluate the system matrix **A** is provided. Then, Sect. 6.4.3 presents the characteristic motion of HORUS.

6.4.1 State-Space Model

When equations of motion are written in state-space form, they have the following form in matrix notation, as discussed before:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{6.138}$$

with $\mathbf{x}$ an $n \times 1$ state vector, $\mathbf{u}$ an $m \times 1$ control vector, and $\mathbf{A}$ and $\mathbf{B}$ the $n \times n$ state (or system) and $n \times m$ control coefficient matrices, respectively. Equation (6.138) is called the state equation; to complete the description of the state of the vehicle we also need a so-called output equation:

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} \tag{6.139}$$

In the above equation, $\mathbf{y}$ is the $k \times 1$ output vector, and $\mathbf{C}$ and $\mathbf{D}$ are the $k \times n$ output and $k \times m$ direct transmission matrices, respectively. For the time being, we

will restrict ourselves to the use of Eq. (6.138), and only in the next chapter we will revisit Eq. (6.139).

To write Eqs. (6.112)–(6.120) in state-space form, we must distinguish between state variables and control variables. The choice of state variables is obvious, if we look at the original equations of motion with their state variables. We write

$$\mathbf{x} = (\Delta V, \Delta\gamma, \Delta R, \Delta p, \Delta q, \Delta r, \Delta\alpha, \Delta\beta, \Delta\sigma)^T$$

The selection of control variables is less clear. The principle of control is obvious: by changing the magnitude and direction of external forces (and therefore in principle also the external moments), the trajectory of the vehicle can be changed. The guidance system makes sure that the vehicle will follow its nominal trajectory by adjusting the angle of attack and bank angle, the control variables of the guidance system. These control variables determine the size and direction of the aerodynamic force vector, the only controllable external force acting on the unpowered vehicle. (The other external force is of gravitational origin; this force is depending on the position of the vehicle and cannot be actively controlled.)

The attitude controller has to guarantee that the commanded attitude is obtained (and maintained) with a certain accuracy in a finite time, which means that eventually there should be moment equilibrium. Note that the actual angle of attack and bank angle are given by the kinematic equations, which means that they have defined values depending on the rotation of the vehicle and thus the size and direction of the external moments. So whereas the forces acting on the centre of mass (c.o.m.) of the vehicle are the guidance-control variables, the moments around the c.o.m. are the attitude-control variables.

Inspecting the equations of motion, we find in addition to the three force components ΔD, ΔS and ΔL, three moment components, i.e., ΔM_x, ΔM_y and ΔM_z. Part of the moment components is determined by the vehicle, depending on its actual attitude. However, the remaining part is given (and can be changed) by, for instance, aerodynamic control surfaces, reaction control thrusters, and momentum wheels, depending on the vehicle configuration and sub-systems. So, for the selection of the attitude-control effectors, we must have a closer look at the vehicle that is considered for the analysis. In this chapter, the focus will be on winged entry vehicles and on entry capsules. We have selected HORUS and the Apollo capsule to be representative entry vehicles, partially because a lot of (consistent) data is available. A description of both vehicles, including an overview of available (aerodynamic) data, can be found in Appendix B.

The setup of the state-space model in this section will be as general as possible such that it could also be used for other vehicles, with possibly only minor modifications. We will obviously focus on HORUS, and in terms of actuators it has more of them compared to an entry capsule, such as Apollo. This means that when using the model also for Apollo, we can easily adapt it to the capsule configuration, because in that case we can simply "remove" actuators from the model. HORUS-2B (MBB 1988) is a winged re-entry vehicle without a major propulsion system, but with five

aerodynamic control surfaces (two rudders, two wing flaps[6] and a body flap) and a number of reaction control thrusters. The body flap is only used for trim, and does not have to be considered as a control. The two wing flaps, or elevons, can be deflected symmetrically (elevator function) and asymmetrically (aileron function). The rudders are outward movable only, and will only be used one at a time.

It should be noted that by deflecting the control surfaces, an aerodynamic force is generated that gives a moment around the vehicle's c.o.m. depending on the moment arm, the distance to the c.o.m.. The deflection of a surface is in principle an analogous process, of course with a limited accuracy, which means that the generated moments are a continuous function of the deflection angle, as well as of flight and similarity parameters, such as the Mach number.

The reaction control system is only being used when the aerodynamic control surfaces are not sufficiently effective, e.g., in the upper layers of the atmosphere. The aerodynamic control surfaces are activated when their effectiveness is more than 10% of that of the corresponding thrusters and the thrusters are inhibited when their effectiveness is less than 10% of the corresponding aerodynamic surface.[7] Whereas the aerodynamic control moments are continuous functions, this is not the case with the reaction control moments. The principle of a thruster is that it is either ON or OFF. When it is on, it will generate a constant thrust force that will also result in a constant moment.[8] By switching the thrusters on and off repeatedly (so-called pulsing), the required moment can be approximated. Furthermore, in case there are more thrusters for generating moments about one particular axis, the magnitude of the moment can be varied by smartly combining the required thrusters. However, here we will assume continuous reaction control moments, for reasons of simplicity. Moreover, we will directly use the moments as control variables, instead of the thrust forces.

So, summarised we can write for the control vector $\mathbf{u}$:

$$\mathbf{u} = (\Delta\delta_e, \Delta\delta_a, \Delta\delta_r, \Delta M_{T,x}, \Delta M_{T,y}, \Delta M_{T,z})^T$$

with

δ_e = elevator deflection angle (rad)
δ_a = aileron deflection angle (rad)
δ_r = rudder deflection angle (rad)
$M_{T,x}$ = roll-thruster moment (Nm)

[6] To be more in line with other literature, we will not use the term *wing flaps*, but *elevons* since this control surface combines the elevator and aileron function.

[7] For the defined configuration of HORUS-2B, the maximal torque of the pitch thrusters is 10,400 Nm, the one of the roll thrusters is 1,600 Nm and the one of the yaw thrusters is 7,600 Nm. This results in activating the aerodynamic control surfaces at approximately 100 km and switching off the roll thrusters at about 75 km, while the yaw thrusters remain activated until the end of the descent.

[8] The thrust force is in principle depending on the atmospheric conditions, which change significantly during the flight. Since we do not have extended thruster models at our disposal, we will assume constant thrust forces. However, it is known from the experience with the Space Shuttle that the operation of the reaction control thrusters can be significantly influenced by atmospheric conditions.

$M_{T,y}$ = pitch-thruster moment (Nm)
$M_{T,z}$ = yaw-thruster moment (Nm)

These definitions of **x** and **u** result in the following state-space matrix equation:

$$\begin{pmatrix} \Delta\dot{V} \\ \Delta\dot{\gamma} \\ \Delta\dot{R} \\ \Delta\dot{p} \\ \Delta\dot{q} \\ \Delta\dot{r} \\ \Delta\dot{\alpha} \\ \Delta\dot{\beta} \\ \Delta\dot{\sigma} \end{pmatrix} = \begin{bmatrix} a_{VV} & a_{V\gamma} & a_{VR} & a_{Vp} & a_{Vq} & a_{Vr} & a_{V\alpha} & a_{V\beta} & a_{V\sigma} \\ a_{\gamma V} & a_{\gamma\gamma} & a_{\gamma R} & a_{\gamma p} & a_{\gamma q} & a_{\gamma r} & a_{\gamma\alpha} & a_{\gamma\beta} & a_{\gamma\sigma} \\ a_{RV} & a_{R\gamma} & a_{RR} & a_{Rp} & a_{Rq} & a_{Rr} & a_{R\alpha} & a_{R\beta} & a_{R\sigma} \\ a_{pV} & a_{p\gamma} & a_{pR} & a_{pp} & a_{pq} & a_{pr} & a_{p\alpha} & a_{p\beta} & a_{p\sigma} \\ a_{qV} & a_{q\gamma} & a_{qR} & a_{qp} & a_{qq} & a_{qr} & a_{q\alpha} & a_{q\beta} & a_{q\sigma} \\ a_{rV} & a_{r\gamma} & a_{rR} & a_{rp} & a_{rq} & a_{rr} & a_{r\alpha} & a_{r\beta} & a_{r\sigma} \\ a_{\alpha V} & a_{\alpha\gamma} & a_{\alpha R} & a_{\alpha p} & a_{\alpha q} & a_{\alpha r} & a_{\alpha\alpha} & a_{\alpha\beta} & a_{\alpha\sigma} \\ a_{\beta V} & a_{\beta\gamma} & a_{\beta R} & a_{\beta p} & a_{\beta q} & a_{\beta r} & a_{\beta\alpha} & a_{\beta\beta} & a_{\beta\sigma} \\ a_{\sigma V} & a_{\sigma\gamma} & a_{\sigma R} & a_{\sigma p} & a_{\sigma q} & a_{\sigma r} & a_{\sigma\alpha} & a_{\sigma\beta} & a_{\sigma\sigma} \end{bmatrix} \begin{pmatrix} \Delta V \\ \Delta\gamma \\ \Delta R \\ \Delta p \\ \Delta q \\ \Delta r \\ \Delta\alpha \\ \Delta\beta \\ \Delta\sigma \end{pmatrix} +$$

$$+ \begin{bmatrix} b_{Ve} & b_{Va} & b_{Vr} & b_{Vx} & b_{Vy} & b_{Vz} \\ b_{\gamma e} & b_{\gamma a} & b_{\gamma r} & b_{\gamma x} & b_{\gamma y} & b_{\gamma z} \\ b_{Re} & b_{Ra} & b_{Rr} & b_{Rx} & b_{Ry} & b_{Rz} \\ b_{pe} & b_{pa} & b_{pr} & b_{px} & b_{py} & b_{pz} \\ b_{qe} & b_{qa} & b_{qr} & b_{qx} & b_{qy} & b_{qz} \\ b_{re} & b_{ra} & b_{rr} & b_{rx} & b_{ry} & b_{rz} \\ b_{\alpha e} & b_{\alpha a} & b_{\alpha r} & b_{\alpha x} & b_{\alpha y} & b_{\alpha z} \\ b_{\beta e} & b_{\beta a} & b_{\beta r} & b_{\beta x} & b_{\beta y} & b_{\beta z} \\ b_{\sigma e} & b_{\sigma a} & b_{\sigma r} & b_{\sigma x} & b_{\sigma y} & b_{\sigma z} \end{bmatrix} \begin{pmatrix} \Delta\delta_e \\ \Delta\delta_a \\ \Delta\delta_r \\ \Delta M_{T,x} \\ \Delta M_{T,y} \\ \Delta M_{T,z} \end{pmatrix} \tag{6.140}$$

Before we derive the 135 matrix elements a_{ij} and b_{ik}, one more task remains to be done as we already indicated above: the evaluation of the force and moment variations in Eqs. (6.112)–(6.117), i.e., ΔD, ΔS, ΔL, ΔM_x, ΔM_y, ΔM_z, as a function of state and control variables. Starting with ΔD, we must first study the aerodynamic database of the HORUS-2B to find the dependency of drag D on flight parameters and deflection angles (MBB 1988). We find:

$$D = f(M, \alpha, h, \delta_e, \delta_r, \delta_b)$$

with

M = Mach number (-)
α = angle of attack (rad)
h = height (m)
δ_e = elevon deflection angle (rad)
δ_r = rudder deflection angle (rad)
δ_b = body-flap deflection angle (rad)

ΔD can thus be written as

$$\Delta D = \frac{\partial D}{\partial M}\Delta M + \frac{\partial D}{\partial \alpha}\Delta\alpha + \frac{\partial D}{\partial h}\Delta h + \frac{\partial D}{\partial \delta_e}\Delta\delta_e + \frac{\partial D}{\partial \delta_a}\Delta\delta_a + \frac{\partial D}{\partial \delta_r}\Delta\delta_r + \frac{\partial D}{\partial \delta_b}\Delta\delta_b \tag{6.141}$$

To evaluate the partial derivatives, we write D as

$$D = C_D \bar{q} S_{ref} \tag{6.142}$$

with

C_D = drag coefficient = f($M, \alpha, h, \delta_r, \delta_e, \delta_b$)
$\bar{q}$ = $\frac{1}{2}\rho V^2$ = dynamic pressure (N/m^2) = f(h, V)
S_{ref} = reference area (m^2)

The height dependency of the drag coefficient is only apparent for Mach numbers smaller than 1.5, and is small compared with the dependencies on Mach number and angle of attack. There are, however, two other dependencies on altitude, coming from the Mach number (as the speed of sound is a function of temperature and thus altitude), and the dynamic pressure through the atmospheric density. The body flap is only used for trim, so it has a nominal deflection angle only ($\Delta\delta_b = 0$). What remains are the contributions due to the elevators, ailerons and rudder. Since we are looking at corrective control only (the deflection angles are in principal in the order of a few degrees), and the contribution to the drag is small compared with the nominal drag, we will neglect these terms.[9] Thus, Eq. (6.141) simplifies to

$$\Delta D = \frac{\partial D}{\partial M}\Delta M + \frac{\partial D}{\partial \alpha}\Delta\alpha \tag{6.143}$$

However, in the end the linearised model has to be expressed in terms of the state variables ΔV and ΔR instead of ΔM, so we rewrite the above equation as

$$\Delta D = \frac{\partial D}{\partial V}\Delta V + \frac{\partial D}{\partial R}\Delta R + \frac{\partial D}{\partial \alpha}\Delta\alpha \tag{6.144}$$

Because the aerodynamic coefficients are dependent on Mach rather than V and R, we have to use the chain rule to evaluate the partial derivatives, i.e.,

$$\frac{\partial C_D}{\partial V} = \frac{\partial C_D}{\partial M}\frac{\partial M}{\partial V} \quad \text{and} \quad \frac{\partial C_D}{\partial R} = \frac{\partial C_D}{\partial M}\frac{\partial M}{\partial h}\frac{\mathrm{d}h}{\mathrm{d}R} \tag{6.145}$$

where $\frac{\mathrm{d}h}{\mathrm{d}R}$ has been added to link the altitude dependency to the actual state, R. Similarly, the altitude dependency of dynamic pressure in the term $\frac{\partial D}{\partial R}$ will be treated in the same way.

Evaluating the first term on the right-hand side of Eq. (6.144) yields

[9] Due to a similar reasoning the contribution of the elevons to the lift force, as compared to the nominal lift force, can be neglected as well.

$$\frac{\partial D}{\partial V} = \frac{\partial C_D}{\partial M}\frac{\partial M}{\partial V}\bar{q}S_{ref} + \frac{\partial \bar{q}}{\partial V}C_D S_{ref} \tag{6.146}$$

The partial derivative with of M with respect to V easily follows from its definition

$$M = \frac{V}{a} \quad \Rightarrow \quad \frac{\partial M}{\partial V} = \frac{1}{a} = \frac{M}{V} \tag{6.147}$$

with $a = f(h)$ being the speed of sound. However, since a is nowhere explicitly used in the equations of motion, $a = \frac{V}{M}$ has been substituted, as M and V are known variables.

For the dynamic-pressure derivative we get

$$\frac{\partial \bar{q}}{\partial V} = \frac{1}{2}\rho\frac{\mathrm{d}V^2}{\mathrm{d}V} = \rho V \tag{6.148}$$

such that Eq. (6.146) changes to:

$$\frac{\partial D}{\partial V} = \frac{\partial C_D}{\partial M}\frac{M}{V}\bar{q}S_{ref} + C_D S_{ref}\rho V = \frac{\partial C_D}{\partial M}\frac{M}{V}\bar{q}S_{ref} + 2\frac{D}{V} \tag{6.149}$$

or, with the definition of $D = C_D\frac{1}{2}\rho V^2 S_{ref}$:

$$\frac{\partial D}{\partial V} = \frac{\partial C_D}{\partial M}\frac{M}{V}\bar{q}S_{ref} + 2\frac{D}{V} \tag{6.150}$$

The next derivative to be evaluated is the one with respect to altitude, i.e.,

$$\frac{\partial D}{\partial R} = \frac{\partial C_D}{\partial M}\frac{\partial M}{\partial h}\frac{\mathrm{d}h}{\mathrm{d}R}\bar{q}S_{ref} + C_D\frac{\partial \bar{q}}{\partial h}\frac{\mathrm{d}h}{\mathrm{d}R}S_{ref} \tag{6.151}$$

with $\frac{\partial h}{\partial R}$ being equal to 1, as $R = R_e + h$ (for the spherical Earth that is typically considered for the linearised system). The altitude dependency of Mach number is derived as follows:

$$M = \frac{V}{a} \quad \Rightarrow \quad \frac{\partial M}{\partial h} = -\frac{V}{a^2}\frac{\mathrm{d}a}{\mathrm{d}h} = -\frac{M^2}{V}\frac{\mathrm{d}a}{\mathrm{d}h} \tag{6.152}$$

The formal way to deal with $\frac{\mathrm{d}a}{\mathrm{d}h}$ is as follows. From its definition, Eq. (2.83), $a = \sqrt{\gamma RT}$, where in principle all three parameters are dependent on the state of the atmosphere, i.e., being temperature and/or composition. The common parameter here is the altitude, but to make our life a bit easier, we will assume both γ and R (the ratio of specific heats and the gas constant, not the state variables) to be constant. In reality, they do change a bit, but this change is quite small and can thus be treated as a higher-order effect. So:

$$\frac{\mathrm{d}a}{\mathrm{d}h} = \frac{1}{2\sqrt{\gamma RT}} \frac{\mathrm{d}T}{\mathrm{d}h} = \frac{1}{2a} \frac{\mathrm{d}T}{\mathrm{d}h} \tag{6.153}$$

For tabulated atmospheres, $\frac{\partial T}{\partial h}$ can easily be derived numerically, although in some cases, e.g., the US76 atmosphere model, the temperature has been specified for the layers of the atmosphere with constant lapse rates (see Fig. 2.30, as well as Eq. (2.85) and the subsequent discussion). To avoid overcomplicating the linearisation process, we assume that the temperature derivative is obtained numerically. But then, for the same reason, the derivative of a with respect to h will also be directly available. Therefore, we will use the latter derivative in the subsequent state-space model, i.e., Eq. (6.152).

The dynamic-pressure dependency on altitude can be derived similarly, and also here we assume the availability of numerical values for $\frac{\mathrm{d}\rho}{\mathrm{d}h}$:

$$\frac{\partial \bar{q}}{\partial h} = \frac{1}{2} V^2 \frac{\mathrm{d}\rho}{\mathrm{d}h} \tag{6.154}$$

Substituting the obtained results back into Eq. (6.151) yields the final expression for $\frac{\partial D}{\partial R}$:

$$\frac{\partial D}{\partial R} = -\frac{\partial C_D}{\partial M} \frac{M^2}{V} \frac{\mathrm{d}a}{\mathrm{d}h} \bar{q} S_{ref} + C_D \frac{1}{2} V^2 \frac{\mathrm{d}\rho}{\mathrm{d}h} S_{ref} \tag{6.155}$$

It is noted that when the density follows an exponential profile, the corresponding derivative is given by $\frac{\partial \rho}{\partial h} = -\frac{\rho}{H_s}$, with H_s being the scale height. This expression has been used throughout Chap. 5.

Finally, the third derivative in Eq. (6.144) is simply given by

$$\frac{\partial D}{\partial \alpha} = \frac{\partial C_D}{\partial \alpha} \bar{q} S_{ref} \tag{6.156}$$

and Eqs. (6.150), (6.155), and (6.156) substituted into Eq. (6.144) results in the sought-for expression for ΔD:

$$\begin{aligned} \Delta D = & \left(\frac{\partial C_D}{\partial M} \frac{M_0}{V_0} \bar{q}_0 S_{ref} + 2 \frac{D_0}{V_0} \right) \Delta V + \\ & + \left(-\frac{\partial C_D}{\partial M} \frac{M_0^2}{V_0} \frac{\mathrm{d}a}{\mathrm{d}h} \bar{q}_0 S_{ref} + \frac{1}{2} C_{D,0} S_{ref} V_0^2 \frac{\mathrm{d}\rho}{\partial h} \right) \Delta R + \frac{\partial C_D}{\partial \alpha} \bar{q}_0 S_{ref} \Delta \alpha \end{aligned} \tag{6.157}$$

As we will see shortly, similar terms will also appear in the expressions for ΔL, $\Delta \mathcal{M}$, and $\Delta \mathcal{N}$. Depending on the relative magnitude of the terms, which is depending on

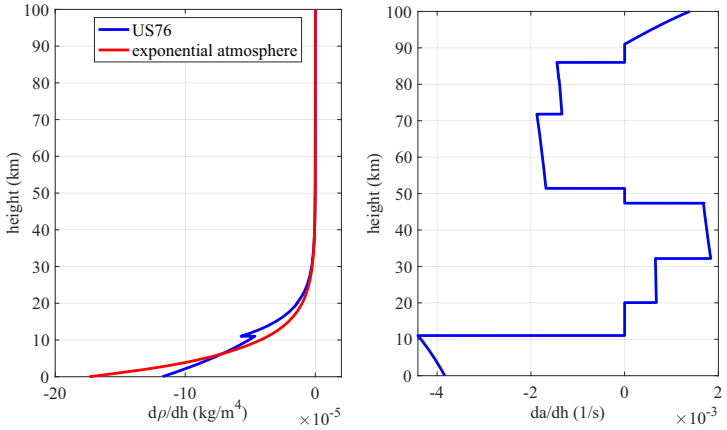


Fig. 6.11 Derivatives of density (left) and speed of sound with respect to altitude

the application, the ΔV and ΔR terms will induce a coupling between translational and rotational motion in the expressions for $\Delta\mathcal{M}$ and $\Delta\mathcal{N}$.

Example 6.5 (*First Derivatives of Atmospheric Properties*) In this example, the derivatives of speed of sound and density are evaluated for the United States 1976 atmosphere model (see Sect. 2.3.2.2 for a description of this atmosphere model), and it will be checked whether the corresponding terms are important to include in the linearised equations. For an altitude range of 0–100 km, the partial derivatives $\frac{\partial a}{\partial h}$ and $\frac{\partial \rho}{\partial h}$ have been plotted in Fig. 6.11. For comparison, the density derivative for the exponential atmosphere has been plotted as well. What is immediately obvious, is that the US76 derivative has a discrete jump at $h = 11$ km. This jump is due to the way the US76 atmosphere is defined: up to 11 km altitude, the temperature is linearly decreasing, whereas between 11 and 20 km the temperature is constant. As the density is derived from the ideal gas law, this shows as a (very) small bend in the density profile, but (obviously) the change in first derivative with respect to altitude is more pronounced. In the right figure, the speed-of-sound derivative is shown, which naturally follows a similar pattern as the temperature profile: whenever T is constant, $\frac{\mathrm{d}a}{\mathrm{d}h}$ will be zero.

From the figure several observations can be made. One might think that if $\frac{\mathrm{d}\rho}{\partial h}$ is important, it will only be so at lower altitudes, probably under 15 km. However, that will appear to be not true, since the derivative still has to be multiplied with V_0, which at higher altitudes is very large. The derivative $\frac{\mathrm{d}a}{\partial h}$ behaves in a different manner: it is significantly larger, and is not really decreasing with increasing altitude.

To evaluate the potential importance of these terms in a bit more detail, we will evaluate the ΔV and ΔR coefficient terms, i.e.,

Table 6.1 State-space coefficient values for different trajectory points

nr nr	h (km)	V (m/s)	M (–)	$V1$	$V2$	$R1$	$R2$
1	80.0	7423.8	26.0	−1.8	11.5	−0.07	−6.5
2	45.0	2325.8	7.1	−10.9	355.9	0.13	−56.3
3	26.6	748.0	2.5	−124.6	1028.0	0.20	−60.4

$$\overbrace{\frac{\partial C_D}{\partial M}\frac{M_0}{V_0}\bar{q}_0 S_{ref}}^{V1} + \overbrace{2\frac{D_0}{V_0}}^{V2}$$

and

$$\overbrace{-\frac{\partial C_D}{\partial M}\frac{M_0^2}{V_0}\frac{\mathrm{d}a}{\mathrm{d}h}\bar{q}_0 S_{ref}}^{R1} + \overbrace{\frac{1}{2}C_{D,0}S_{ref}V_0^2\frac{\mathrm{d}\rho}{\partial h}}^{R2}$$

and will do so for three flight conditions of HORUS, i.e., (1) $h = 80\,\mathrm{km}$ with $V_0 = 7{,}423.8\,\mathrm{m/s}$ and $M_0 = 26.0$, (2) $h = 45\,\mathrm{km}$ with $V_0 = 2{,}325.8\,\mathrm{m/s}$ and $M_0 = 7.1$, and (3) $h = 26.6\,\mathrm{km}$ with $V_0 = 748.0\,\mathrm{m/s}$ and $M_0 = 2.5$. The last point marks the end of the hypersonic descent. Note that these points have been taken from the nominal mission to be discussed in the next section.

The terms $V1$, $V2$, $R1$ and $R2$ can be found in Table 6.1. The terms $V1$ and $V2$ are in the same range, with $V1$ being between 3 and 16% of $V2$. Concerning $R1$ and $R2$, it is clear (and perhaps surprising) that $R1$ is much smaller than $R2$, at 80 km it is just over 1% at best. This means that for this vehicle and trajectory the speed of sound can be assumed to be constant in each trajectory point, but that the variation in density due to a perturbation in altitude should be included in the model. ■

The side force S is a function of M, α, β, δ_e, δ_a and δ_r. Again, we neglect the contribution due to the control surfaces as being second-order effects, which results in

$$\Delta S = \frac{\partial S}{\partial M}\Delta M + \frac{\partial S}{\partial \alpha}\Delta\alpha + \frac{\partial S}{\partial \beta}\Delta\beta \tag{6.158}$$

Evaluating all the terms in a similar way as above, we get for ΔS:

$$\Delta S = \frac{\partial C_S}{\partial \beta}\bar{q}_0 S_{ref}\Delta\beta \tag{6.159}$$

Note that $C_S = 0$ for $\beta = \beta_0 = 0°$ (independent from the other states), which is also true for the derivatives with respect to M and α evaluated at β_0. Therefore, also no altitude derivatives appear in the expression above.

For ΔL we find an expression very similar to the one for ΔL:

$$\Delta L = \left(\frac{\partial C_L}{\partial M}\frac{M_0}{V_0}\bar{q}_0 S_{ref} + 2\frac{L_0}{V_0}\right)\Delta V + \\ + \left(-\frac{\partial C_L}{\partial M}\frac{M_0^2}{V_0}\frac{\mathrm{d}a}{\mathrm{d}h}\bar{q}_0 S_{ref} + \frac{1}{2}C_{L,0}S_{ref}V_0^2\frac{\mathrm{d}\rho}{\partial h}\right)\Delta R + \frac{\partial C_L}{\partial \alpha}\bar{q}_0 S_{ref}\Delta\alpha \tag{6.160}$$

The variation of the moments consist of an aerodynamic and a propulsive component, i.e.,

$$\Delta M_x = \Delta\mathcal{L} + \Delta M_{T,x} \tag{6.161a}$$

$$\Delta M_y = \Delta\mathcal{M} + \Delta M_{T,y} \tag{6.161b}$$

$$\Delta M_z = \Delta\mathcal{N} + \Delta M_{T,z} \tag{6.161c}$$

Only the aerodynamic contributions have to be expanded to a state-variable and control-variable contribution, because the thruster moments are already control variables (and are assumed to be state independent).

We will apply the same way of derivation as for the aerodynamic forces, with one difference. In this case, the contributions of the control surfaces cannot be neglected, because of their relative magnitude. Also, if we did, it would reduce the means of control to reaction control only. After inspecting the aerodynamic database, we find that the ailerons contribute to both the roll and yaw moment, the elevators to the pitch moment and the rudder to the yaw moment. Furthermore, there are no dynamic damping terms included in the definition of the moments as is usual the case, i.e.,

$$C_{l_p} = C_{m_q} = C_{n_r} = 0$$

but only because they have not been computed. This does not mean that they are zero, as we know from a similar vehicle, the Space Shuttle (Trujillo 1986). However, if we want to take the damping terms into account in the derivation, so that the model does not have to be expanded once these terms become available, or when we want to apply the model to another vehicle, we will see that in this case the related terms will vanish. In the previous section it was stated that the nominal rotational rates are small, and can be treated in a similar manner as the variations ΔV, $\Delta\beta$, etc. This means that when we derive the small variation around a nominal damping term like

$$C_{l_p}\frac{p_0 b_{ref}}{2V_0}\bar{q}_0 S_{ref} b_{ref}$$

the result is a second-order term and for that reason can be neglected. Therefore, the damping terms do not have to be included in our derivation. However, when the rotational rates p_0, q_0 and r_0 are not small and can consequently not be neglected, the damping terms *will* appear in the model.

As a result, with the aerodynamic moments given by

$$\mathcal{L} = C_l \bar{q} S_{ref} b_{ref}$$

$$\mathcal{M} = C_m \bar{q} S_{ref} c_{ref}$$

$$\mathcal{N} = C_n \bar{q} S_{ref} b_{ref}$$

we get

$$\Delta\mathcal{L} = \frac{\partial C_l}{\partial \beta} \bar{q}_0 S_{ref} b_{ref} \Delta\beta + \frac{\partial C_l}{\partial \delta_a} \bar{q}_0 S_{ref} b_{ref} \Delta\delta_a \tag{6.162}$$

Because $C_l = 0$ for $\beta_0 = 0$, this means that, as was the case with the side force, the Mach-number derivative is zero, and also the density and speed-of-sound derivatives disappear.

$$\begin{aligned}\Delta\mathcal{M} = &\frac{\partial C_m}{\partial M} \frac{M_0}{V_0} \bar{q}_0 S_{ref} c_{ref} \Delta V - \frac{\partial C_m}{\partial M} \frac{M_0^2}{V_0} \frac{\mathrm{d}a}{\mathrm{d}h} \bar{q}_0 S_{ref} \Delta R + \\ &+ \frac{\partial C_m}{\partial \alpha} \bar{q}_0 S_{ref} c_{ref} \Delta\alpha + \frac{\partial C_m}{\partial \delta_e} \bar{q}_0 S_{ref} c_{ref} \Delta\delta_e\end{aligned} \tag{6.163}$$

Note that the nominal pitch coefficient $C_m = 0$ for trimmed flight, and it is assumed that trim is always guaranteed (see also Sect. 6.4.2, where the nominal trajectory will be discussed). Therefore, the terms $2\frac{\mathcal{M}_0}{V_0}\Delta V$ and $\frac{1}{2} C_{m,0} S_{ref} V_0^2 \frac{\mathrm{d}\rho}{\partial h} \Delta R$ do not appear in the above equation. We did retain the other ΔR term, as it is now the only coupling term with altitude, and it may be interesting to see its (possibly only small) effect.

It is noted that for the calculation of $\frac{\partial C_m}{\partial \alpha}$, the trim deflections $\delta_{b,trim}$ and $\delta_{e,trim}$ should be kept constant. For instance, when the derivative is obtained through central differencing,

$$\frac{\partial C_m}{\partial \alpha} = \frac{C_m(M_0, \alpha_0 + \delta\alpha, \delta_{b,trim}) - C_m(M_0, \alpha_0 - \delta\alpha, \delta_{b,trim})}{2\delta\alpha} \tag{6.164}$$

where it has been assumed that in this case trim is done with only the body flap. Thus, there is no new trim deflection evaluated for these "new" angles of attack, i.e., $\alpha_0 + \delta\alpha$ and $\alpha_0 - \delta\alpha$.

For the yaw moment, we get:

$$\Delta\mathcal{N} = \frac{\partial C_n}{\partial \beta}\bar{q}_0 S_{ref} b_{ref}\Delta\beta + \frac{\partial C_n}{\partial \delta_r}\bar{q}_0 S_{ref} b_{ref}\Delta\delta_r + \frac{\partial C_n}{\partial \delta_a}\bar{q}_0 S_{ref} b_{ref}\Delta\delta_a \tag{6.165}$$

Also now, $C_n = 0$ for $\beta_0 = 0$, as well as $\frac{\partial C_n}{\partial M} = 0$ and $\mathcal{N}_0 = 0$. Last but not least, the atmospheric-properties derivatives are cancelled out.

Finally, after substituting Eqs. (6.157)–(6.160) and (6.162)–(6.165) in Eqs. (6.112)–(6.120) and rearranging terms, we find expressions for each of the coefficients a_{ij} and b_{ik}. The obtained stability model, which has the desired form of Eq. (6.138)

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

or, fully written out,

$$\begin{pmatrix} \Delta\dot{V} \\ \Delta\dot{\gamma} \\ \Delta\dot{R} \\ \Delta\dot{p} \\ \Delta\dot{q} \\ \Delta\dot{r} \\ \Delta\dot{\alpha} \\ \Delta\dot{\beta} \\ \Delta\dot{\sigma} \end{pmatrix} = \begin{bmatrix} a_{VV} & a_{V\gamma} & a_{VR} & 0 & 0 & 0 & a_{V\alpha} & 0 & 0 \\ a_{\gamma V} & a_{\gamma\gamma} & a_{\gamma R} & 0 & 0 & 0 & a_{\gamma\alpha} & a_{\gamma\beta} & a_{\gamma\sigma} \\ a_{RV} & a_{R\gamma} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{p\beta} & 0 \\ a_{qV} & 0 & a_{qR} & 0 & 0 & 0 & a_{q\alpha} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{r\beta} & 0 \\ a_{\alpha V} & a_{\alpha\gamma} & a_{\alpha R} & 0 & a_{\alpha q} & 0 & a_{\alpha\alpha} & 0 & a_{\alpha\sigma} \\ a_{\beta V} & a_{\beta\gamma} & a_{\beta R} & a_{\beta p} & 0 & a_{\beta r} & 0 & a_{\beta\beta} & a_{\beta\sigma} \\ a_{\sigma V} & a_{\sigma\gamma} & a_{\sigma R} & a_{\sigma p} & 0 & a_{\sigma r} & a_{\sigma\alpha} & a_{\sigma\beta} & a_{\sigma\sigma} \end{bmatrix} \begin{pmatrix} \Delta V \\ \Delta\gamma \\ \Delta R \\ \Delta p \\ \Delta q \\ \Delta r \\ \Delta\alpha \\ \Delta\beta \\ \Delta\sigma \end{pmatrix} + \\ + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_{pa} & 0 & b_{px} & 0 & 0 \\ b_{qe} & 0 & 0 & 0 & b_{qy} & 0 \\ 0 & b_{ra} & b_{rr} & 0 & 0 & b_{rz} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta\delta_e \\ \Delta\delta_a \\ \Delta\delta_r \\ \Delta M_{T,x} \\ \Delta M_{T,y} \\ \Delta M_{T,z} \end{pmatrix} \tag{6.166}$$

can now be used to study the open-loop behaviour of the vehicle. This will be described in the next section. In Eq. (6.166), the elements of matrix $\mathbf{A}$, i.e., a_{ij}, are given by:

$$a_{VV} = -\frac{1}{mV_0}\left(M_0\frac{\partial C_D}{\partial M}\bar{q}_0 S_{ref} + 2D_0\right) \tag{6.167}$$

$$a_{V\gamma} = -g_0\cos\gamma_0 \tag{6.168}$$

$$a_{VR} = 2\frac{g_0}{R_0}\sin\gamma_0 - \frac{\partial C_D}{\partial M}\frac{M_0^2}{mV_0}\frac{\mathrm{d}a}{\mathrm{d}h}\bar{q}_0 S_{ref} + \frac{1}{2}C_{D,0}S_{ref}\frac{V_0^2}{m}\frac{\mathrm{d}\rho}{\partial h} \tag{6.169}$$

$$a_{V\alpha} = -\frac{1}{m}\frac{\partial C_D}{\partial \alpha}\bar{q}_0 S_{ref} \tag{6.170}$$

and, for the sake of convenience, the zero coefficients of the first row are:

$$a_{Vp} = a_{Vq} = a_{Vr} = a_{V\beta} = a_{V\sigma} = 0 \tag{6.171}$$

The coefficients of the other rows are:

$$a_{\gamma V} = \frac{1}{V_0}\left(-\dot{\gamma}_0 + \frac{2V_0}{R_0}\cos\gamma_0\right) + \frac{\cos\sigma_0}{mV_0^2}\left(M_0\frac{\partial C_L}{\partial M}\bar{q}_0 S_{ref} + 2L_0\right) \tag{6.172}$$

$$a_{\gamma\gamma} = -\left(\frac{V_0}{R_0} - \frac{g_0}{V_0}\right)\sin\gamma_0 \tag{6.173}$$

$$a_{\gamma R} = \left(\frac{2g_0}{V_0} - \frac{V_0}{R_0}\right)\frac{\cos\gamma_0}{R_0} + \frac{\cos\sigma_0}{mV_0}\left(\frac{1}{2}C_{L,0}S_{ref}V_0^2\frac{\mathrm{d}\rho}{\partial h} - \frac{\partial C_L}{\partial M}\frac{M_0^2}{V_0}\frac{\mathrm{d}a}{\mathrm{d}h}\bar{q}_0 S_{ref}\right) \tag{6.174}$$

$$a_{\gamma\alpha} = \frac{\cos\sigma_0}{mV_0}\frac{\partial C_L}{\partial \alpha}\bar{q}_0 S_{ref} \tag{6.175}$$

$$a_{\gamma\beta} = -\frac{\sin\sigma_0}{mV_0}\frac{\partial C_S}{\partial \beta}\bar{q}_0 S_{ref} \tag{6.176}$$

$$a_{\gamma\sigma} = -\frac{L_0}{mV_0}\sin\sigma_0 \tag{6.177}$$

$$a_{\gamma p} = a_{\gamma q} = a_{\gamma r} = 0 \tag{6.178}$$

$$a_{RV} = \sin\gamma_0 \tag{6.179}$$

$$a_{R\gamma} = V_0\cos\gamma_0 \tag{6.180}$$

$$a_{RR} = a_{Rp} = a_{Rq} = a_{Rr} = a_{R\alpha} = a_{R\beta} = a_{R\sigma} = 0 \tag{6.181}$$

$$a_{p\beta} = \frac{1}{I_{xx}}\frac{\partial C_l}{\partial \beta}\bar{q}_0 S_{ref} b_{ref} \tag{6.182}$$

$$a_{pV} = a_{p\gamma} = a_{pR} = a_{pp} = a_{pq} = a_{pr} = a_{p\alpha} = a_{p\sigma} = 0 \tag{6.183}$$

$$a_{qV} = \frac{1}{I_{yy}} \frac{M_0}{V_0} \frac{\partial C_m}{\partial M} \bar{q}_0 S_{ref} c_{ref} \tag{6.184}$$

$$a_{qR} = -\frac{1}{I_{yy}} \frac{M_0^2}{V_0} \frac{\partial C_m}{\partial M} \bar{q}_0 S_{ref} c_{ref} \frac{\mathrm{d}a}{\mathrm{d}h} \tag{6.185}$$

$$a_{q\alpha} = \frac{1}{I_{yy}} \frac{\partial C_m}{\partial \alpha} \bar{q}_0 S_{ref} c_{ref} \tag{6.186}$$

$$a_{q\gamma} = a_{qp} = a_{qq} = a_{qr} = a_{q\beta} = a_{q\sigma} = 0 \tag{6.187}$$

$$a_{r\beta} = \frac{1}{I_{zz}} \frac{\partial C_n}{\partial \beta} \bar{q}_0 S_{ref} b_{ref} \tag{6.188}$$

$$a_{rV} = a_{r\gamma} = a_{rR} = a_{rp} = a_{rq} = a_{rr} = a_{r\alpha} = a_{r\sigma} = 0 \tag{6.189}$$

$$a_{\alpha V} = -\frac{g_0}{V_0^2} \cos\gamma_0 \cos\sigma_0 - \frac{1}{m V_0^2} \left(M_0 \frac{\partial C_L}{\partial M} + C_L \right) \bar{q}_0 S_{ref} \tag{6.190}$$

$$a_{\alpha\gamma} = -\frac{g_0}{V_0} \sin\gamma_0 \cos\sigma_0 \tag{6.191}$$

$$a_{\alpha R} = -\frac{2g_0}{R_0 V_0} \cos\gamma_0 \cos\sigma_0 + \frac{1}{m V_0} \left(\frac{1}{2} C_{L,0} S_{ref} V_0^2 \frac{\mathrm{d}\rho}{\partial h} - \frac{\partial C_L}{\partial M} \frac{M_0^2}{V_0} \frac{\mathrm{d}a}{\mathrm{d}h} \bar{q}_0 S_{ref} \right) \tag{6.192}$$

$$a_{\alpha q} = 1 \tag{6.193}$$

$$a_{\alpha\alpha} = -\frac{1}{m V_0} \frac{\partial C_L}{\partial \alpha} \bar{q}_0 S_{ref} \tag{6.194}$$

$$a_{\alpha\sigma} = -\frac{g_0}{V_0} \cos\gamma_0 \sin\sigma_0 \tag{6.195}$$

$$a_{\alpha p} = a_{\alpha r} = a_{\alpha\beta} = 0 \tag{6.196}$$

$$a_{\beta V} = \frac{g_0}{V_0^2} \cos\gamma_0 \sin\sigma_0 \tag{6.197}$$

$$a_{\beta\gamma} = \frac{g_0}{V_0} \sin\gamma_0 \sin\sigma_0 \tag{6.198}$$

$$a_{\beta R} = \frac{2g_0}{R_0 V_0} \cos\gamma_0 \sin\sigma_0 \tag{6.199}$$

$$a_{\beta p} = \sin \alpha_0 \tag{6.200}$$

$$a_{\beta r} = -\cos \alpha_0 \tag{6.201}$$

$$a_{\beta\beta} = -\frac{1}{mV_0} \frac{\partial C_S}{\partial \beta} \bar{q}_0 S_{ref} \tag{6.202}$$

$$a_{\beta\sigma} = -\frac{g_0}{V_0} \cos \gamma_0 \cos \sigma_0 \tag{6.203}$$

$$a_{\beta q} = a_{\beta\alpha} = 0 \tag{6.204}$$

$$a_{\sigma V} = \frac{\tan \gamma_0 \sin \sigma_0}{mV_0^2} \left(M_0 \frac{\partial C_L}{\partial M} + C_L \right) \bar{q}_0 S_{ref} \tag{6.205}$$

$$a_{\sigma\gamma} = \frac{L_0}{mV_0} \sin \sigma_0 \tag{6.206}$$

$$a_{\sigma R} = \frac{\tan \gamma_0 \sin \sigma_0}{mV_0} \left(\frac{1}{2} C_{L,0} S_{ref} V_0^2 \frac{\mathrm{d}\rho}{\partial h} - \frac{\partial C_L}{\partial M} \frac{M_0^2}{V_0} \frac{\mathrm{d}a}{\mathrm{d}h} \bar{q}_0 S_{ref} \right) \tag{6.207}$$

$$a_{\sigma p} = -\cos \alpha_0 \tag{6.208}$$

$$a_{\sigma r} = -\sin \alpha_0 \tag{6.209}$$

$$a_{\sigma\alpha} = \frac{\tan \gamma_0 \sin \sigma_0}{mV_0} \frac{\partial C_L}{\partial \alpha} \bar{q}_0 S_{ref} \tag{6.210}$$

$$a_{\sigma\beta} = \frac{\tan \gamma_0 \cos \sigma_0}{mV_0} \frac{\partial C_S}{\partial \beta} \bar{q}_0 S_{ref} - \frac{L_0}{mV_0} + \frac{g_0}{V_0} \cos \gamma_0 \cos \sigma_0 \tag{6.211}$$

$$a_{\sigma\sigma} = \tan \gamma_0 \cos \sigma_0 \frac{L_0}{mV_0} \tag{6.212}$$

$$a_{\sigma q} = 0 \tag{6.213}$$

The elements of matrix $\mathbf{B}$, i.e., b_{ik}, are:

$$b_{Ve} = b_{Va} = b_{Vr} = b_{Vx} = b_{Vy} = b_{Vz} = 0 \tag{6.214}$$

$$b_{\gamma e} = b_{\gamma a} = b_{\gamma r} = b_{\gamma x} = b_{\gamma y} = b_{\gamma z} = 0 \tag{6.215}$$

$$b_{Re} = b_{Ra} = b_{Rr} = b_{Rx} = b_{Ry} = b_{Rz} = 0 \tag{6.216}$$

$$b_{pa} = \frac{1}{I_{xx}} \frac{\partial C_l}{\partial \delta_a} \bar{q}_0 S_{ref} b_{ref} \tag{6.217}$$

$$b_{px} = \frac{1}{I_{xx}} \tag{6.218}$$

$$b_{pe} = b_{pr} = b_{py} = b_{pz} = 0 \tag{6.219}$$

$$b_{qe} = \frac{1}{I_{yy}} \frac{\partial C_m}{\partial \delta_e} \bar{q}_0 S_{ref} c_{ref} \tag{6.220}$$

$$b_{qy} = \frac{1}{I_{yy}} \tag{6.221}$$

$$b_{qa} = b_{qr} = b_{qx} = b_{qz} = 0 \tag{6.222}$$

$$b_{ra} = \frac{1}{I_{zz}} \frac{\partial C_n}{\partial \delta_a} \bar{q}_0 S_{ref} b_{ref} \tag{6.223}$$

$$b_{rr} = \frac{1}{I_{zz}} \frac{\partial C_n}{\partial \delta_r} \bar{q}_0 S_{ref} b_{ref} \tag{6.224}$$

$$b_{rz} = \frac{1}{I_{zz}} \tag{6.225}$$

$$b_{re} = b_{rx} = b_{ry} = 0 \tag{6.226}$$

$$b_{\alpha e} = b_{\alpha a} = b_{\alpha r} = b_{\alpha x} = b_{\alpha y} = b_{\alpha z} = 0 \tag{6.227}$$

$$b_{\beta e} = b_{\beta a} = b_{\beta r} = b_{\beta x} = b_{\beta y} = b_{\beta z} = 0 \tag{6.228}$$

$$b_{\sigma e} = b_{\sigma a} = b_{\sigma r} = b_{\sigma x} = b_{\sigma y} = b_{\sigma z} = 0 \tag{6.229}$$

6.4.2 Nominal Mission

The analysis of the characteristic motion is centred around a nominal reference trajectory. Such a reference trajectory is usually the result of an optimisation process, taking mission requirements, performance criteria, and trajectory constraints into account. When neither a reference trajectory nor an optimisation program is at hand, one can generate a trajectory provided a control history is available. MBB (1988) discusses a nominal control history of α and σ as a function of time. With this control history as guidance output and applying ideal control, we can generate the corresponding re-entry trajectory using the vehicle model described in Appendix B

and the dynamics model to be discussed in the next section. The mission of HORUS starts in orbit with a deorbitation manoeuvre and successive coasting arc. The actual re-entry starts at the atmospheric entry interface (h = 122 km) with initial velocity conditions of V = 7435.5 m/s, $\gamma = -1.43°$ and $\chi = 70.75°$. The vehicle is assumed to be heading towards a runway in Kourou. Some 80 km from the landing site guidance switches to the Terminal Area Energy Management (TAEM) logic. Specifics about the nominal mission can be found in Appendix B, Sect. B.2.3. To facilitate the discussion below, in Fig. 6.12 the nominal control history is shown, as well as the altitude-velocity diagram. For a better understanding of the trajectory dynamics the moments of the four bank reversals have been indicated as well.

Studying the coefficients of **A** and **B**, we find that we need the following parameters of the nominal trajectory as a function of time:

$$\alpha_0, \sigma_0, V_0, \gamma_0, R_0, \bar{q}_0, M_0, g_0, D_0, L_0$$

The control history provides α_0 and σ_0 ($\beta_0 = 0°$), which is shown in the top plot of Fig. 6.12. The same holds for the altitude-velocity diagram shown in the bottom plot of the same figure, which provides V_0 and R_0. The other relevant parameters are given in Figs. 6.13 till 6.16. For a detailed discussion of the trajectory we refer to Mooij (1998). It should be noted that in the curves for the drag and the lift, the trim drag and lift are included. These (nominal) trim forces are too large to be ignored, as is usually done for conventional aircraft. In Fig. 6.16, finally, the nominal body-flap and elevator deflection for trim stability are shown. We see that only at the end of the flight the body-flap deflection is saturated, and the elevons are required to provide

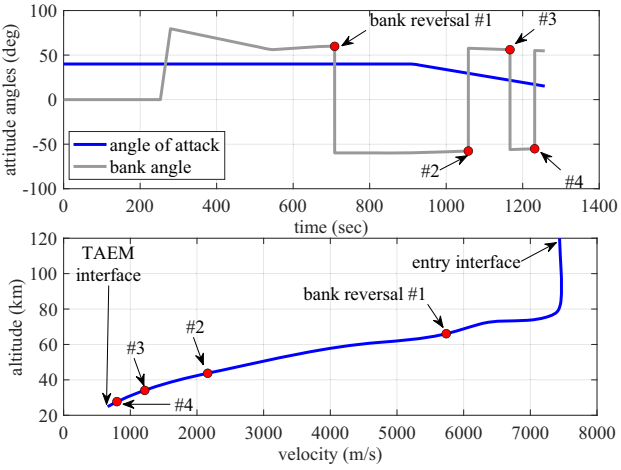


Fig. 6.12 Nominal control history (top) and altitude-velocity profile, with the time points of the four bank reversals indicated

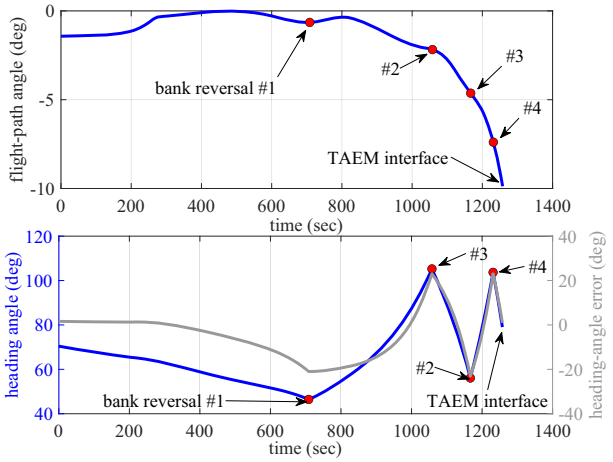


Fig. 6.13 Nominal flight-path angle (top) and heading error (bottom) as a function of time

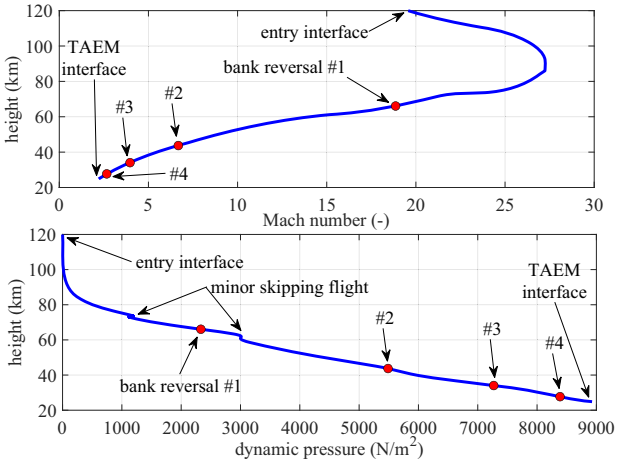


Fig. 6.14 Nominal Mach number (top) and dynamic pressure (bottom) versus altitude

the remaining trim moment. The deflections of the aerosurfaces take place with an infinite rate (zero time).

From coefficient $a_{\gamma V}$, we see that we need the value of $\dot{\gamma}_0$. This variable can be easily computed by substituting the related nominal values in Eq. (6.83),

$$\dot{\gamma}_0 = \left(\frac{V_0}{R_0} - \frac{g_0}{V_0}\right)\cos\gamma_0 + \frac{L_0 \cos\sigma_0}{m V_0} \tag{6.230}$$

Finally, we need the values of some aerodynamic coefficients and derivatives:

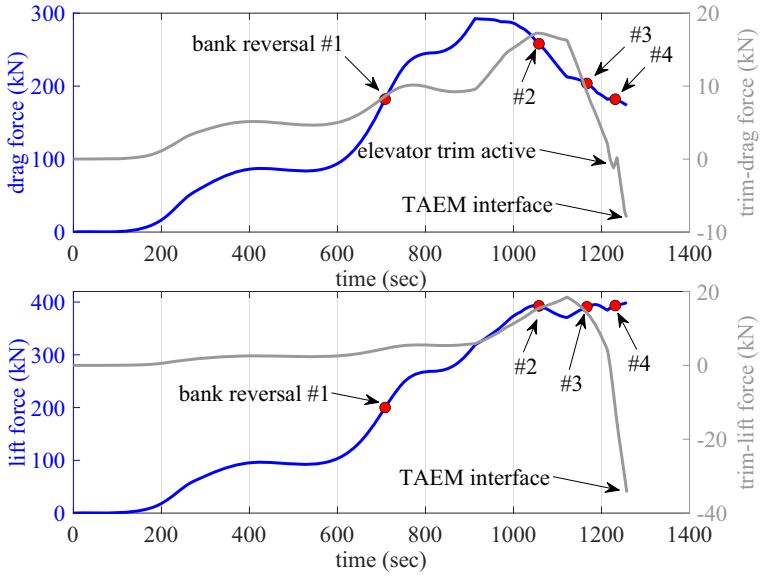


Fig. 6.15 Nominal aerodynamic forces, including trim forces

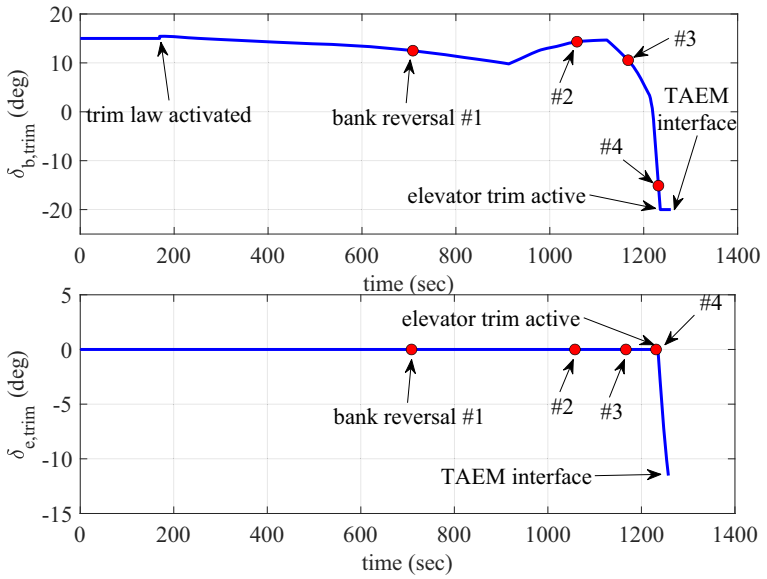


Fig. 6.16 Trim deflection for body flap (top) and elevator (bottom) as a function of time

- $\frac{\partial C_D}{\partial M}, \frac{\partial C_D}{\partial \alpha}, \frac{\partial C_S}{\partial \beta}, \frac{\partial C_L}{\partial M}, \frac{\partial C_L}{\partial \alpha}, \frac{\partial C_l}{\partial \beta}, \frac{\partial C_n}{\partial \beta}$ with contributions of the base vehicle only,

- $\frac{\partial C_m}{\partial M}$, $\frac{\partial C_m}{\partial \alpha}$ with contributions of the base vehicle, the body flap and the elevons,
- $\frac{\partial C_l}{\partial \delta_a}$, $\frac{\partial C_m}{\partial \delta_e}$, $\frac{\partial C_n}{\partial \delta_a}$ with contributions of the elevons only, and
- $\frac{\partial C_n}{\partial \delta_r}$ with contributions of the rudder only.

With M_0, α_0 and the trimmed deflection angles of body flap and elevator at hand, these derivatives can be computed off-line. The original data have been provided in the form of graphs (MBB 1988), and have been converted to numerical values in tabular form (Mooij 1995). Using these tables, linear interpolation is used to extract the information for a particular flight situation. A derivative follows from the the connecting line, when the flight situation is located in between two tabular values. At a boundary tabular value, either the left or right derivative is calculated, and for values outside the tabulated range, the values are constrained to their boundary values.

6.4.3 HORUS Characteristic Motion

Before we start the discussion on the eigenvalues and eigenmotions of HORUS, we must note that it is not our intention to give a complete analysis of the open-loop behaviour of HORUS. We will restrict ourselves to a more general discussion, following the brief introduction on the relation between eigenvalues and eigenmotion given in Sect. 6.2. To have a basis for comparison, the eigenmotions that characterise (conventional) subsonic aircraft will be summarised, because we expect at least partly similar modes (a component of the eigenmotion) for HORUS in certain speed regimes. These characteristic motions will be discussed for a number of time points. Finally, to illustrate the difficulty of analysing the characteristic motion from a non-linear simulation, this section is concluded with some results of a six-degrees-of-freedom open-loop re-entry simulation.

For a conventional subsonic aircraft in steady flight, five different modes can be distinguished, i.e., two longitudinal and three lateral modes. The longitudinal modes are called the short-period and the phugoid mode, whereas the lateral modes are called the lateral oscillation (or Dutch roll), the rolling convergence and the spiral mode. Furthermore, the longitudinal and lateral modes are decoupled, which means that a symmetric motion (e.g., a disturbance) will not affect the asymmetric motion and vice versa.

The *short-period oscillation* is a fast, periodic aircraft motion, which is usually well-damped. Because of the fast pitch rate, the aerodynamic forces acting on the vehicle are large. The flight path is nearly a straight line, and there is a negligible speed variation. The *phugoid* is a very slow oscillation that is poorly damped. The slowly oscillating speed results in small variations in the dynamic pressure and also the aerodynamic forces. Basically, the motion consists of translations in the aircraft plane of symmetry, while the rotation about the pitch axis is negligible. Since only

the height and velocity are changing, the phugoid can be interpreted as a continuous exchange of potential and kinetic energy.

The *lateral oscillation* resembles the short-period oscillation, in that sense that it is usually a fast, well-damped aircraft motion (sometimes moderately damped). The large aerodynamic forces are in principal due to a rotation about the top axis (yaw). There is also a strong roll rate, but this does not influence the lateral oscillation that much. The flight path and the velocity almost have constant values. The *roll convergence* is a strongly damped, aperiodic eigenmotion, during which the aircraft rotates purely about the X-axis (roll). The roll convergence is aperiodic, because the roll angle does not have any influence on the external moments and thus the motion. The *spiral mode*, finally, is seen to consist mainly of yawing at nearly zero sideslip with some roll. The aerodynamic forces are very small, resulting in a large time constant. The spiral mode can be either a stable or unstable, aperiodic eigenmotion.

We will now have a look at the eigenmotion of HORUS. As we saw in Sect. 6.4.1, the system matrix $\mathbf{A}$ is a 9×9 matrix, which means that there are nine eigenvalues and corresponding eigenvectors. Inspecting $\mathbf{A}$ will also tell us that there are some cross couplings between so-called symmetric (V, γ, R, q, α) and asymmetric variables (p, r, β, σ), indicating that the symmetric and asymmetric motion will not be completely decoupled. However, when these coupling terms are sufficiently small, decoupling will be possible. This will be confirmed later in this section.

In Fig. 6.17a through 6.17d the nine eigenvalues of HORUS along the nominal trajectory are plotted. The first two figures show the eigenvalues in the complex plane, whereas the latter two give the imaginary and real parts of the eigenvalues as a function of flight time. Since it is not easy to see which modes the curves represent, we will only draw some general conclusions from these figures.

Figure 6.17a shows a mixture of complex and real eigenvalues, indicating that we can expect both periodic and aperiodic eigenmotions. The maximum imaginary parts of about 1.5 imply quite short periods. The real parts are between –1 and 1, which means that the modes can be either converging or diverging. Looking at Fig. 6.17c, we see in principle two major periodic eigenmotions emerge, which are, as we will find out later, the lateral oscillation (the two outer curves) and the short-period oscillation (the two inner curves). The discontinuous jumps in the inner curves are due to the linearisation of the aerodynamic coefficients and the linear table interpolation: when we go from one table range to another, the aerodynamic derivatives sometimes change discontinuously. Figure 6.17d indicates that most of the time the eigenmotions are very lightly damped or undamped. Only towards the end of the flight strong divergencies and convergences appear.

To study the eigenmotion of HORUS in more detail, we have selected several time points for which we will compute the characteristic values (according to Eqs. (6.49) through (6.52)), and for which we will try to identify the modes by studying the eigenvectors. The selected time points are 1 (t = 0 s), 50 (t = 196 s), 100 (t = 396 s), 150 (t = 596 s), 200 (t = 796 s), 250 (t = 996 s), 300 (t = 1196 s) and 314 (t = 1252 s). The numerical results are presented in Table 6.2 through Table 6.9 at the end of this chapter.

Table 6.2 HORUS eigenvalues and corresponding characteristic values for time point 1 ($t = 0$ s)

	Short-period oscillation		Phugoid		Lateral oscillation		Aperiodic roll mode		Height mode
λ_i Re	$-7.3022 \cdot 10^{-8}$		$4.0251 \cdot 10^{-5}$		$8.6767 \cdot 10^{-4}$		$-1.7354 \cdot 10^{-3}$	$2.0734 \cdot 10^{-18}$	$-8.4019 \cdot 10^{-5}$
Im	$\pm 7.4154 \cdot 10^{-3}$		$\pm 1.2123 \cdot 10^{-3}$		± 0.0168		–	–	–
P (s)	847.3		5,183.0		373.8		∞	∞	∞
$T_{\frac{1}{2}}$ (s)	$9.492 \cdot 10^{6}$		–17,220.6		–798.9		399.4	$3.343 \cdot 10^{18}$	8,250.0
ζ (–)	$9.85 \cdot 10^{-6}$		–0.033		–0.052		–	–	–
ω_n (rad/s)	0.0074		0.0012		0.017		–	–	–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	θ (°)	z	z	z
ΔV	0.0128	$\mp$ 127.4	0.0013	± 178.8	$<10^{-4}$	± 37.9	$<10^{-4}$	$<10^{-4}$	0.0007
$\Delta\gamma$	$<10^{-4}$	± 91.6	$<10^{-4}$	± 89.6	$<10^{-4}$	$\mp$ 106.4	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
ΔR	**0.9997**	**0.0**	**1.0000**	**0.0**	$<10^{-4}$	± 168.3	$<10^{-4}$	$<10^{-4}$	**1.0000**
Δp	$<10^{-4}$	$\mp$ 160.9	$<10^{-4}$	± 176.6	**0.0166**	$\mp$ **96.1**	0.0021	0.0008	$<10^{-4}$
Δq	0.0002	$\mp$ 122.9	$<10^{-4}$	$\mp$ 12.4	$<10^{-4}$	$\mp$ 104.6	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δr	$<10^{-4}$	$\mp$ 148.1	$<10^{-4}$	$\mp$ 109.7	0.0012	$\mp$ 96.1	0.0002	0.0010	$<10^{-4}$
$\Delta\alpha$	**0.0226**	$\pm$**147.1**	$<10^{-4}$	$\mp$ 174.1	$<10^{-4}$	$\mp$ 68.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	± 121.9	$<10^{-4}$	± 158.4	**0.5883**	$\pm$ **171.0**	0.0078	$<10^{-4}$	$<10^{-4}$
$\Delta\sigma$	$<10^{-4}$	$\mp$ 23.4	$<10^{-4}$	$\pm$ 64.4	**0.8084**	**0.0**	**1.0000**	**1.0000**	$<10^{-4}$

Table 6.3 HORUS eigenvalues and corresponding characteristic values for time point 50 (t = 196 s)

	Short-period oscillation		Phugoid		Lateral oscillation		Aperiodic roll mode		Height mode
λ_i Re	$-2.3424 \cdot 10^{-5}$		$-1.1441 \cdot 10^{-4}$		$8.7234 \cdot 10^{-4}$		$-1.7687 \cdot 10^{-3}$	$8.8 \cdot 10^{-18}$	$7.9209 \cdot 10^{-5}$
Im	± 0.15831		$\pm 1.2559 \cdot 10^{-3}$		± 0.32569		–	–	–
P (s)	39.7		5,002.9		19.3		∞	∞	∞
$T_{\frac{1}{2}}$ (s)	$2.9591 \cdot 10^{4}$		6,058.3		–794.6		391.9	$7.9 \cdot 10^{18}$	–8,750.8
ζ (–)	$1.4796 \cdot 10^{-4}$		0.091		$-2.6785 \cdot 10^{-4}$		–	–	–
ω_n (rad/s)	0.1583		0.0013		0.326		–	–	–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	θ (°)	z	z	z
ΔV	**0.3038**	**∓92.0**	0.0013	∓174.3	$<10^{-4}$	∓115.9	$<10^{-4}$	$<10^{-4}$	0.0006
$\Delta\gamma$	$<10^{-4}$	±90.1	$<10^{-4}$	±96.4	$<10^{-4}$	±89.7	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
ΔR	**0.9522**	**0.0**	**1.0000**	**0.0**	$<10^{-4}$	∓88.6	$<10^{-4}$	$<10^{-4}$	**1.0000**
Δp	$<10^{-4}$	±0.4	$<10^{-4}$	±0.6	**0.3089**	**∓90.3**	0.0022	0.0008	$<10^{-4}$
Δq	0.0051	∓91.8	$<10^{-4}$	∓24.0	$<10^{-4}$	∓91.6	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δr	$<10^{-4}$	±0.3	$<10^{-4}$	±0.6	0.0224	∓90.3	0.0002	0.0010	$<10^{-4}$
$\Delta\alpha$	0.0320	±178.2	$<10^{-4}$	∓174.3	$<10^{-4}$	±169.4	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	∓90.1	$<10^{-4}$	∓130.2	**0.5568**	**±179.5**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\sigma$	$<10^{-4}$	±90.9	$<10^{-4}$	±141.3	**0.7707**	**0.0**	**1.0000**	**1.0000**	$<10^{-4}$

Table 6.4 HORUS eigenvalues and corresponding characteristic values for time point 100 (t = 396 s)

	Short-period oscillation		Phugoid		Lateral oscillation		Periodic roll mode		Height/Spiral mode
λ_i Re	$-2.0324 \cdot 10^{-4}$		$-9.9244 \cdot 10^{-4}$		$2.2249 \cdot 10^{-4}$		$1.0217 \cdot 10^{-4}$		$-1.1524 \cdot 10^{-15}$
Im	± 0.4292		$\pm 1.3585 \cdot 10^{-3}$		± 0.8844		$5.6080 \cdot 10^{-4}$		–
P (s)	14.6		4,625.1		7.1		11,203.9		∞
$T_{\frac{1}{2}}$ (s)	3,410.4		698.4		–3,115.4		6,784.2		$6.015 \cdot 10^{14}$
ζ (–)	$4.7359 \cdot 10^{-4}$		0.0590		$-2.5156 \cdot 10^{-4}$		–0.1792		–
ω_n (rad/s)	0.4292		0.0017		0.8844		0.0006		–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	θ (°)	z	θ (°)	z
ΔV	**0.9365**	**0.0**	0.0017	± 43.1	0.0076	∓ 179.8	0.0006	∓ 124.8	$<10^{-4}$
$\Delta\gamma$	$<10^{-4}$	∓ 177.9	$<10^{-4}$	∓ 53.7	0.0001	± 90.0	$<10^{-4}$	± 79.7	$<10^{-4}$
ΔR	**0.3485**	**$\pm$92.0**	**1.0000**	**180.0**	**0.8953**	**0.0**	**1.0000**	**0.0**	**1.0000**
Δp	$<10^{-4}$	∓ 179.8	$<10^{-4}$	∓ 34.0	**0.2943**	**$\mp$90.0**	$<10^{-4}$	∓ 153.4	$<10^{-4}$
Δq	0.0155	∓ 0.1	$<10^{-4}$	± 61.0	0.0001	∓ 179.9	$<10^{-4}$	∓ 45.1	$<10^{-4}$
Δr	$<10^{-4}$	∓ 179.8	$<10^{-4}$	∓ 34.0	0.0214	∓ 90.0	$<10^{-4}$	∓ 153.4	$<10^{-4}$
$\Delta\alpha$	0.0362	∓ 90.1	$<10^{-4}$	± 43.1	0.0005	± 90.1	$<10^{-4}$	∓ 124.8	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	± 90.2	$<10^{-4}$	∓ 87.9	**0.1954**	**180.0**	$<10^{-4}$	± 106.3	$<10^{-4}$
$\Delta\sigma$	$<10^{-4}$	∓ 78.3	$<10^{-4}$	± 27.1	**0.2704**	**0.0**	$<10^{-4}$	∓ 47.0	$<$**10^{-4}**

Table 6.5 HORUS eigenvalues and corresponding characteristic values for time point 150 (t = 596 s)

	Short-period oscillation		Phugoid		Lateral oscillation		Height/Spiral mode		
λ_i Re	$-2.5368 \cdot 10^{-4}$		$-1.0895 \cdot 10^{-3}$		$4.5831 \cdot 10^{-4}$		$2.8651 \cdot 10^{-4}$	$-5.6959 \cdot 10^{-4}$	$1.3032 \cdot 10^{-16}$
Im	± 0.4361		$\pm 1.5462 \cdot 10^{-3}$		± 0.9001		–	–	–
P (s)	14.4		4,063.6		7.0		∞	∞	∞
$T_{\frac{1}{2}}$ (s)	2,732.4		636.2		–1,512.4		–2,419.2	1,216.9	$-5.318 \cdot 10^{15}$
ζ (–)	$5.817 \cdot 10^{-4}$		0.0576		$-5.0916 \cdot 10^{-4}$		–	–	–
ω_n (rad/s)	0.4361		0.0019		0.9001		–	–	–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	θ (°)	z	z	z
ΔV	**0.8595**	**0.0**	0.0018	∓ 135.6	0.0081	∓ 179.8	0.0003	0.0011	$<10^{-4}$
$\Delta\gamma$	$<10^{-4}$	∓ 178.7	$<10^{-4}$	± 125.4	0.0001	± 90.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
ΔR	**0.5099**	**$\pm$91.3**	**1.0000**	**0.0**	**0.8799**	**0.0**	**1.0000**	**1.0000**	**1.0000**
Δp	$<10^{-4}$	∓ 179.7	$<10^{-4}$	± 141.6	**0.3170**	**$\mp$90.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δq	**0.0142**	**$\mp$0.1**	$<10^{-4}$	∓ 107.5	0.0001	∓ 179.8	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δr	$<10^{-4}$	∓ 179.7	$<10^{-4}$	± 141.6	0.0230	∓ 90.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\alpha$	**0.0326**	**$\mp$90.0**	$<10^{-4}$	∓ 135.6	0.0005	± 90.2	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	± 90.3	$<10^{-4}$	± 86.8	**0.2068**	**180.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\sigma$	$<10^{-4}$	∓ 73.6	$<10^{-4}$	∓ 155.3	**0.2862**	**0.0**	$\mathbf{<10^{-4}}$	$\mathbf{<10^{-4}}$	$\mathbf{<10^{-4}}$

Table 6.6 HORUS eigenvalues and corresponding characteristic values for time point 200 (t = 796 s)

	Short-period oscillation		Phugoid		lateral oscillation		Roll divergence		Height/Spiral mode
λ_i Re	$-9.4846 \cdot 10^{-4}$		$-2.7301 \cdot 10^{-3}$		$3.8175 \cdot 10^{-4}$		$-6.8063 \cdot 10^{-5}$		$2.0583 \cdot 10^{-16}$
Im	± 0.6924		$\pm 2.0061 \cdot 10^{-3}$		± 1.4315		$\pm 1.0843 \cdot 10^{-3}$		–
P (s)	30.8		3,132.0		4.4		5,794.7		∞
$T_{\frac{1}{2}}$ (s)	730.8		253.9		–1,815.7		10,183.9		$3.367 \cdot 10^{15}$
ζ (–)	$1.370 \cdot 10^{-3}$		0.806		$-2.667 \cdot 10^{-4}$		0.063		–
ω_n (rad/s)	0.6923		0.0034		1.4315		0.0011		–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	θ (°)	z	θ (°)	z
ΔV	**0.9419**	**0.0**	**0.0025**	**∓ 88.9**	0.0091	∓179.7	**0.0005**	**±78.5**	$<10^{-4}$
$\Delta\gamma$	$<10^{-4}$	∓179.0	$<10^{-4}$	±143.9	0.0002	±90.0	$<10^{-4}$	∓86.4	$<10^{-4}$
ΔR	**0.3347**	**±90.9**	**1.0000**	**0.0**	**0.8516**	**0.0**	**1.0000**	**180.0**	**1.0000**
Δp	$<10^{-4}$	±0.2	$<10^{-4}$	∓57.3	**0.4284**	**±90.1**	$<10^{-4}$	∓172.8	$<10^{-4}$
Δq	**0.0154**	**∓0.3**	$<10^{-4}$	∓73.5	0.0001	∓179.8	$<10^{-4}$	±145.9	$<10^{-4}$
Δr	$<10^{-4}$	±0.2	$<10^{-4}$	∓57.3	**0.0332**	**±90.1**	$<10^{-4}$	∓172.8	$<10^{-4}$
$\Delta\alpha$	**0.0222**	**∓90.2**	$<10^{-4}$	∓88.9	0.0004	±90.2	$<10^{-4}$	±78.5	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	∓89.7	$<10^{-4}$	∓93.6	**0.1746**	**±0.1**	$<10^{-4}$	±100.8	$<10^{-4}$
$\Delta\sigma$	$<10^{-4}$	±127.2	**$<10^{-4}$**	**∓59.2**	**0.2441**	**∓179.9**	**$<10^{-4}$**	**∓26.4**	$<10^{-4}$

Table 6.7 HORUS eigenvalues and corresponding characteristic values for time point 250 (t = 996 s)

	Short-period oscillation		Phugoid		Lateral oscillation		Roll divergence		Height/Spiral mode
λ_i Re	$-3.2450 \cdot 10^{-3}$		$-4.3861 \cdot 10^{-3}$		$1.6653 \cdot 10^{-3}$		$-1.9488 \cdot 10^{-3}$		$-3.9891 \cdot 10^{-15}$
Im	± 0.6924		$\pm 2.0061 \cdot 10^{-3}$		± 1.4315		$\pm 4.9957 \cdot 10^{-5}$		–
P (s)	9.1		2,192.8		4.2		125,772.4		∞
$T_{\frac{1}{2}}$ (s)	213.6		158.0		–416.2		355.7		$1.7376 \cdot 10^{14}$
ζ (–)	$4.678 \cdot 10^{-3}$		0.837		$-1.113 \cdot 10^{-3}$		1.000		–
ω_n (rad/s)	0.6936		0.0052		1.4960		0.0019		–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	θ (°)	z	θ (°)	z
ΔV	**0.8802**	**0.0**	**0.0029**	**∓62.6**	0.0130	∓179.3	**0.0008**	**±178.1**	$<10^{-4}$
$\Delta\gamma$	0.0001	∓179.1	$<10^{-4}$	±147.3	0.0005	±90.0	$<10^{-4}$	∓1.5	$<10^{-4}$
ΔR	**0.4742**	**±90.5**	**1.0000**	**0.0**	**0.9402**	**0.0**	**1.0000**	**180.0**	**1.0000**
Δp	$<10^{-4}$	±0.4	$<10^{-4}$	∓58.8	**0.2811**	**±90.0**	$<10^{-4}$	±178.5	$<10^{-4}$
Δq	**0.0112**	**∓0.7**	$<10^{-4}$	∓42.2	0.0001	∓179.5	$<10^{-4}$	±179.0	$<10^{-4}$
Δr	$<10^{-4}$	±0.4	$<10^{-4}$	∓58.8	**0.0335**	**±90.0**	$<10^{-4}$	±178.5	$<10^{-4}$
$\Delta\alpha$	**0.0161**	**∓90.3**	$<10^{-4}$	∓62.6	0.0004	±90.5	$<10^{-4}$	±178.1	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	∓89.3	$<10^{-4}$	∓91.9	**0.0890**	**0.0**	$<10^{-4}$	±177.0	$<10^{-4}$
$\Delta\sigma$	$<10^{-4}$	±157.5	**$<10^{-4}$**	**∓95.1**	**0.1670**	**180.0**	**$<10^{-4}$**	**±0.2**	**$<10^{-4}$**

Table 6.8 HORUS eigenvalues and corresponding characteristic values for time point 300 (t = 1196 s)

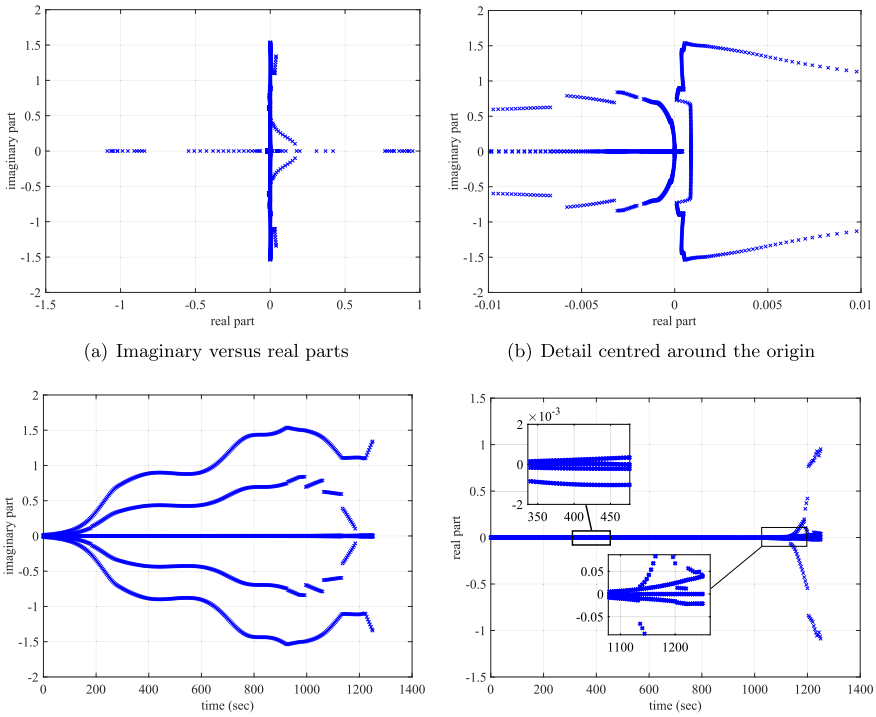
	Former short-period oscillation		Periodic pitch/Roll mode		Lateral oscillation		Pitch/Roll divergence	Height/Spiral mode	
λ_i Re	–0.5078	0.3695	$-1.5650 \cdot 10^{-2}$		$2.2908 \cdot 10^{-2}$		$8.1444 \cdot 10^{-2}$	$-2.8919 \cdot 10^{-4}$	$1.8542 \cdot 10^{-15}$
Im	–	–	$\pm 7.0349 \cdot 10^{-3}$		± 1.1129		–	–	–
P (s)	∞	∞	893.1		5.6		∞	∞	∞
$T_{\frac{1}{2}}$ (s)	1.4	–1.9	44.3		–30.3		–8.5	2,396.8	$3.7383 \cdot 10^{14}$
ζ (-)	–	–	0.912		–0.021		–	–	–
ω_n (rad/s)	–	–	0.0172		1.1131		–	–	–
$\boldsymbol{\mu}_i$	z	z	z	θ (°)	z	θ (°)	z	z	z
ΔV	**0.5578**	**0.4600**	**0.0018**	**∓19.5**	**0.0255**	**∓178.0**	**0.1206**	$<10^{-4}$	$<10^{-4}$
$\Delta\gamma$	0.0003	0.0003	$<10^{-4}$	∓24.2	0.0010	±88.9	0.0001	$<10^{-4}$	$<10^{-4}$
ΔR	**0.8300**	**0.8879**	**1.0000**	**180.0**	**0.9937**	**0.0**	**0.9927**	**1.0000**	**1.0000**
Δp	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	±121.4	**0.0802**	**±87.6**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δq	**0.0029**	**0.0016**	$<10^{-4}$	∓108.9	0.0001	±8.7	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δr	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	±121.4	**0.0135**	±87.6	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\alpha$	**0.0063**	**0.0039**	**$<10^{-4}$**	**±160.4**	0.0004	±90.8	**0.0002**	$<10^{-4}$	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	± 97.2	**0.0139**	**∓3.5**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\sigma$	$<10^{-4}$	0.0001	**$<10^{-4}$**	**±121.0**	**0.0718**	**±178.8**	**$<10^{-4}$**	**$<10^{-4}$**	**$<10^{-4}$**

Table 6.9 HORUS eigenvalues and corresponding characteristic values for time point 314 (t = 1252 s)

	Former short-period oscillation		Periodic pitch/Roll mode		Lateral oscillation		Pitch/Roll divergence	Height/Spiral mode	
λ_i Re	–1.0884	0.9518	–0.0213		0.0392		0.0400	$-1.0592 \cdot 10^{-4}$	$-2.7512 \cdot 10^{-15}$
Im	–	–	±0.0146		±1.3428		–	–	–
P (s)	∞	∞	430.4		4.7		∞	∞	∞
$T_{\frac{1}{2}}$ (s)	0.6	–0.7	32.6		–17.7		–17.3	6,544.1	$2.5195 \cdot 10^{14}$
ζ (-)	–	–	0.825		–0.029		–	–	–
ω_n (rad/s)	–	–	0.0258		1.3434		–	–	–
$\boldsymbol{\mu}_i$	z	z	z	θ (°)	z	θ (°)	z	z	z
ΔV	**0.7031**	**0.6634**	**0.0066**	**±162.3**	**0.0265**	**±1.6**	**0.0419**	$<10^{-4}$	$<10^{-4}$
$\Delta\gamma$	0.0009	0.0008	$<10^{-4}$	±146.2	0.0018	∓91.5	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
ΔR	**0.7106**	**0.7479**	**1.0000**	**0.0**	**0.9875**	**180.0**	**0.9991**	**1.0000**	**1.0000**
Δp	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	±89.5	**0.1244**	**±86.9**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δq	**0.0189**	**0.0164**	$<10^{-4}$	±51.1	0.0004	∓174.6	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δr	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	±89.5	**0.0060**	**±86.9**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\alpha$	**0.0190**	**0.0157**	$<$**10**$^{-4}$	**∓17.8**	0.0005	∓89.9	$<$**10**$^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	±55.1	**0.0222**	**∓4.8**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\sigma$	**0.0002**	**0.0002**	$<$**10**$^{-4}$	**±96.3**	**0.0900**	**±178.6**	$<$**10**$^{-4}$	$<10^{-4}$	$<$**10**$^{-4}$

Table 6.10 Apollo eigenvalues and corresponding characteristic values for time point $t = 0\,\mathrm{s}$

	"Longitudinal" oscillation		"Lateral" oscillation		Bank mode	Bank mode	Height oscillation		Height mode
λ_i Re	$-9.2099\cdot10^{-6}$		$-1.1715\cdot10^{-4}$		$-4.3878\cdot10^{-6}$	$1.3114\cdot10^{-4}$	$3.1900\cdot10^{-4}$		$-4.9954\cdot10^{-4}$
Im	$\pm7.0963\cdot10^{-3}$		$\pm8.1179\cdot10^{-3}$		–	–	$\pm1.4721\cdot10^{-3}$		–
P (s)	885.4		774.0		∞	∞	4,268.1		∞
$T_{\frac{1}{2}}$ (s)	75,261		5,916.5		157,970	5,285.5	2,172.8		1,387.6
ζ (–)	$1.2978\cdot10^{-3}$		$1.4430\cdot10^{-2}$		–	–	$-2.1178\cdot10^{-1}$		–
ω_n (rad/s)	$7.0964\cdot10^{-3}$		$8.1188\cdot10^{-3}$		–	–	$1.5063\cdot10^{-3}$		–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	z	z	θ (°)	z
ΔV	0.0013	95.4	0.0009	176.8	0.0019	0.0015	0.0009	166.6	0.0001
$\Delta\gamma$	$<10^{-4}$	155.4	$<10^{-4}$	92.2	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	83.6	$<10^{-4}$
ΔR	**0.2888**	**70.5**	**0.8325**	**0.0**	**0.4694**	**0.7547**	**1.0000**	**0.0**	**1.0000**
Δp	$<10^{-4}$	–170.0	0.0002	–81.5	0.0007	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
Δq	0.0050	–89.9	0.0011	–3.3	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
Δr	0.0014	–7.7	0.0042	–86.6	$<10^{-4}$	0.0002	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\alpha$	**0.7006**	**180.0**	**0.1340**	**–94.1**	0.0022	0.0022	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\beta$	**0.1598**	**82.1**	**0.4848**	**3.5**	0.0012	0.0009	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\sigma$	**0.6327**	**–7.0**	**0.2320**	**145.7**	**0.8830**	**0.6561**	$<10^{-4}$	0.0	$<10^{-4}$



(a) Imaginary versus real parts

(b) Detail centred around the origin

(c) Imaginary parts

(d) Real parts

Fig. 6.17 Variation of the eigenvalues along the nominal trajectory

In principle, we can trace the five motions, which we discussed before for conventional aircraft. However, due to the larger speed regime and due to the distinct nature of the nominal trajectory (large bank angles as compared with the steady cruise flight of subsonic aircraft), we find some differences. Starting with time point 1, we find six eigenmodes. The first mode we have designated the short-period oscillation, because of the large-amplitude oscillations found in the angle of attack and pitch rate (note that the most important components of the eigenmotion have been printed bold).

Whereas in the aircraft case the height and velocity remained constant, it seems that here this is not the case, basically because of the much slower character of the re-entry short-period mode. However, it should be noted that despite the fact that the height component is the largest, only the relative difference is important. An amplitude of 0.0226 rad for the angle of attack means a 1-m amplitude for the height (or: $\Delta\alpha = 10°$ gives $\Delta R \approx 6.7$ m). Moreover, we do not have such a rapid oscillation here: the period is about 850 s, whereas for conventional aircraft this is in the order of seconds. Also the motion is hardly damped, because of the absence of aerodynamic forces in the upper layers of the atmosphere. If it had not been for the angle of attack,

we might also call this a phugoid-kind of motion. However, we have reserved that name for the second mode.[10]

The dominating components of the phugoid mode are the height and velocity, with the angle of attack more or less constant. Furthermore, looking at the period, we find that this eigenmotion is indeed much slower than the short-period oscillation (P = 5,183 s). A difference with the aircraft phugoid is that this mode is unstable, although the doubling time is more than 17,000 s. The next mode, the lateral oscillation, compares well with its aircraft counterpart. Main components are the angle of sideslip and the bank angle, together with the corresponding angular rates. This mode is unstable throughout the flight, albeit only marginally. The three remaining modes are all aperiodic. The first two have been given the common name *aperiodic roll mode* for obvious reasons: the bank angle is by far the largest component. Surprising is to see that in one mode the influence of the angle of sideslip is many orders larger.[11] This means that the first aperiodic roll mode can be compared with the roll convergence of conventional aircraft. Both modes are stable. The last aperiodic mode has been given the name *height mode*. This mode represents a lightly damped aperiodic motion, which has also been found by Sachs (1993), while studying stability and control problems in hypersonic flight.

Now that we have established the basic modes of the eigenmotion, we can focus on the time history of these modes. The short period oscillation changes its nature and comes closer to the aircraft mode. Because at lower altitudes the aerodynamic forces are higher, the damping of this mode increases (although the damping remains low). The period decreases from $P \approx 850$ s at $t = 0$ s, down to $P \approx 40$ s at $t = 96$ s down to $P \approx 9$ s at $t = 996$ s. Then, at time point 300 the periodic mode breaks into two aperiodic modes, of which one is strongly damped ($T_{\frac{1}{2}} = 1.4$ s) and the other one is strongly diverging ($T_2 = 1.9$ s). The half and doubling time decrease further for the last time point. Last but not least, the influence of the bank angle becomes stronger towards the end of the flight, a phenomenon which we will also see with the next mode, the phugoid (for the re-entry flight; not for conventional aircraft).

The phugoid damping ratio increases in time (apart from a small decrement at time point 150). It becomes a very well-damped motion, with, for instance, a half time of 254 s at time point 200. The period of the phugoid decreases from over 5000 s ($t = 0$ s) to 430 s ($t = 1252$ s). One remark remains to be made and that is that at time point 100 the influence of the bank angle has changed a couple of orders of magnitude. This can be explained by the fact that initially the nominal trajectory is flown with zero bank angle, whereas later large bank angles are applied for lateral direction control. Because of this the influence of the bank angle in the phugoid becomes apparent, and a coupling (although small) exists between the symmetrical and asymmetrical motion).

---

[10] As we already find out, at this altitude it is not altogether useful to compare the characteristic modes of re-entry vehicles with those of aircraft, because of the completely different flight regime. However, we will stick to it, because it is currently the only comparison method.

[11] In principle, the eigenvalue of the second aperiodic roll mode is so small that for practical reasons it can be considered to be zero.

As we mentioned before, the lateral oscillation is an unstable, periodic motion and this continues to be during the whole trajectory. But, some characteristics of the motion change. Towards time point 150, the negative damping coefficient gets smaller, which means that the motion becomes less unstable. From that point, the damping coefficient becomes increasingly negative again. The period changes drastically during the first 100 s of flight (P changes from 373 s down to 19.3 s). It continues to decrease till time point 250 ($P = 4.2$ s), after which it varies only slightly. At $t =$ 396 s, we see that the contribution of $\Delta\alpha$ has changed many orders of magnitude, indicating that due to a large bank angle a small coupling exists between the symmetrical and asymmetrical motion. Besides, the ratio between $\Delta\sigma$ and $\Delta\beta$ changes from 1.39 to 1.36 indicating a slightly weaker coupling between the bank angle and the angle of sideslip at higher bank angles. Δh becomes the major component, although when we see this in proportion to the attitude angles it does not seem that major (for time point 100, an amplitude of 17.3° in the bank angle corresponds with an amplitude of 1 m in the height). The relative difference between the height and the attitude angles, however, becomes larger towards the end of the flight.

The aperiodic roll modes changes into a single unstable periodic roll mode with a small coupling to the angle of attack (an amplitude of $\Delta\sigma = 10°$ gives $\Delta\alpha = 0.33°$) at time point 100. Moreover, the height becomes the major component. Again, this begins at the moment the vehicle starts banking. It should be noted that a small oscillation in the bank angle will result in a large amplitude in the height ($\Delta\sigma = 0.1°$ gives $\Delta h \approx 3{,}300$ m). However, at the next time point this mode changes again into two aperiodic modes, which are similar to the aperiodic height mode. For time points 200 and 250, a (very) stable periodic roll mode is back, while at time point 300 we have again two aperiodic modes. In this case, the stable mode is similar to the height mode, whereas the unstable one shows a coupling between the bank angle and the angle of attack, with the angle of attack the larger one.

The height mode already entered the discussion, when the focus was on the several roll modes. We will briefly finalise the discussion on characteristic modes. Initially, the height mode is stable. For time point 50, it has become unstable with an amplification ratio which is as large as the prior damping ratio. When the banking begins (time point 100), the eigenvalue related to the height mode is very small, and stays very small during the rest of the flight, being alternately positive and negative. For this reason, the height mode is not interesting to discuss any further, because for the short time of the flight this mode is practically indifferent. However, as we discussed before, for some time points the height mode consists of more than one aperiodic component. These additional modes have much larger eigenvalues and become the dominating modes for the height mode, although the damping and/or amplification remains very small.

In Figs. 6.18 and 6.19 the eigenvalues have been separated for translational and rotational motion, and some of the discussed eigenmotions have been indicated.

The free response according to Eq. (6.80) by itself is not that interesting. It will confirm what we have already seen in the tables with numerical results. Basically, the graphs will show us both periodic and aperiodic motions, either diverging or converging. Due to the sometimes long periods, the curves will not give us real

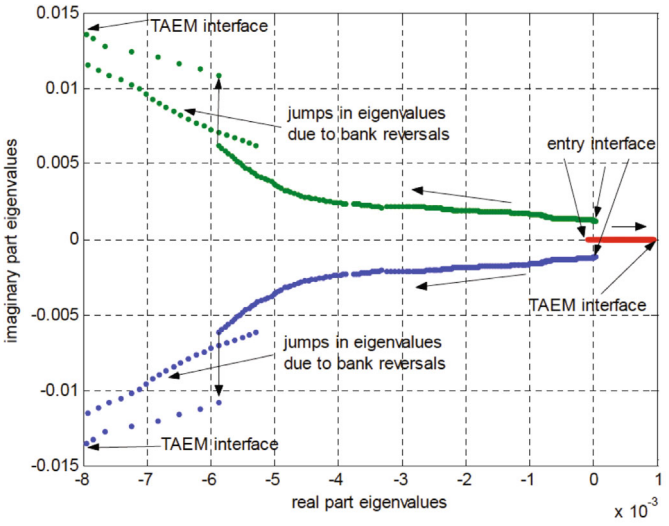


Fig. 6.18 Eigenvalues for translational motion

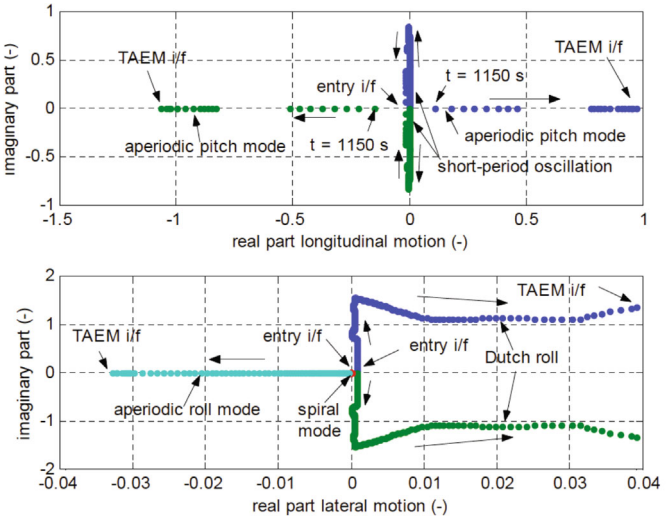


Fig. 6.19 Eigenvalues for rotational motion

insight in the eigenmotion at one particular moment in time. Only when the motion is, for instance, reasonably damped (or unstable, for that matter) and has a sufficiently short period, the resulting eigencurve can be quite informative. As an example, the curves for the angle of sideslip and the bank angle (the major variables for the lateral

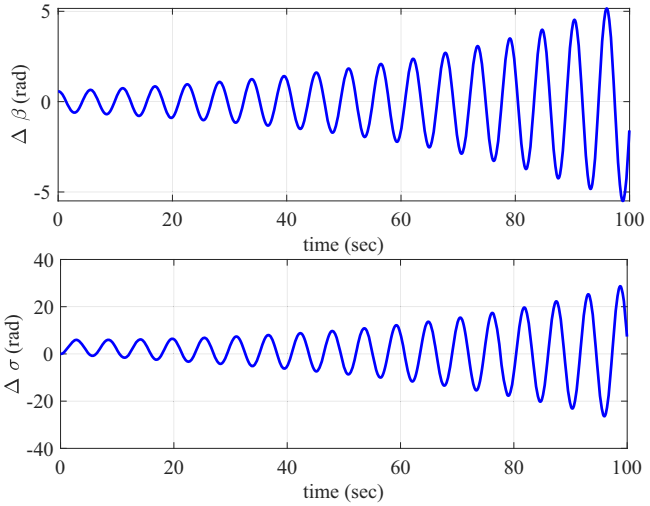


Fig. 6.20 The variation of the angle of sideslip and the bank angle with time (unstable lateral oscillation) for t = 1196 (time point 300). The period of this mode is 5.6 s, whereas the doubling time is 30.3 s

oscillation) have been plotted as a function of time for time point 300, see also Fig. 6.20.

There is another more important reason for not looking into more detail at the curves representing the eigenmotion. The eigenvalues are computed for a certain point in time. When we compute the eigenmotion for a certain time interval (100 s in Fig. 6.20), we assume that the dynamics of the vehicle are not changing (given by a constant system matrix and therefore constant eigenvalues). For a subsonic aircraft in steady (cruise) flight this can be a reasonable approximation, but for a re-entry vehicle this is most of the time not the case.

However, it is illustrative to observe the eigenmotion by simulating the descent of the HORUS-2B in a so-called open-loop 6-d.o.f. mode, i.e., a free-fall re-entry including attitude dynamics, but without attitude control. For the same initial conditions as given in Sect. 6.4.2,[12] the results are shown in Fig. 6.21.

Two conditions are compared, i.e., one starting with a (nominal) angle of attack of 40°, and one starting from $\alpha = 0°$. In addition, it is important to note that the body-flap has been fixed at its initial trim value of $\delta_{b,trim} = 15°$, which is the nominal setting for an entry at $\alpha = 40°$, until the trim law is activated when the dynamic pressure is large enough. From the plots its is clear that the vehicle enters a skipping flight, mainly because the bank angle remains close to zero, and thus the vertical lift force is too

[12] Besides the initial conditions for position and velocity, also the initial values for attitude and angular rotation of the vehicle are required. The nominal attitude angles at t = 0 are α = 40°, β = 0°and σ = 0°, whereas the nominal pitch rate is 0.072663°/s (the nominal roll and yaw rate are considered to be zero).

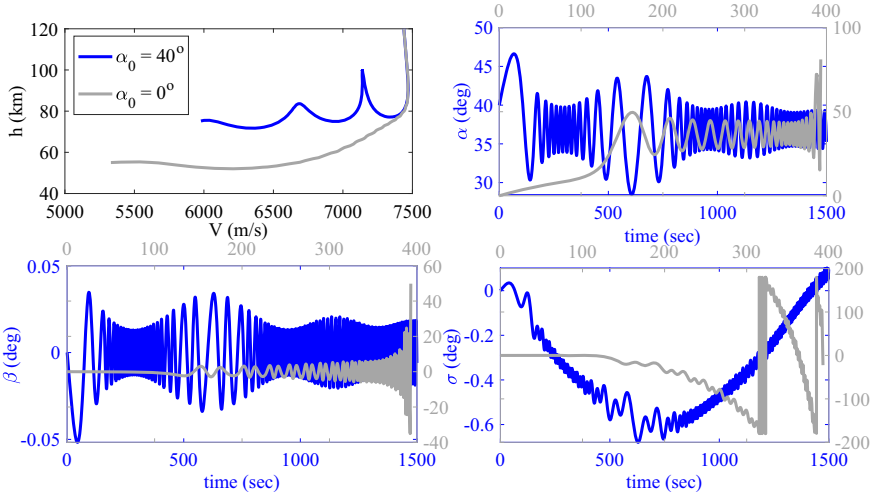


Fig. 6.21 Open-loop simulation ($\alpha_0 = 40°$ and $\alpha_0 = 0°$), fixed body-flap deflection of $\delta_{b,trim} = 15°$

large for entering the atmosphere in a gliding flight. The simulation is cut off after 1,500 s, when the angular rates become so large that the integrator cannot handle this anymore, without going into very small step sizes. The angle of attack starts oscillating right from the beginning, slowly at first, but when the dynamic pressure increases, also the oscillation frequency picks up. Due to the deflected body flap, α is oscillating around some average value of about 37°. The oscillation in angle of sideslip is much faster, although the amplitude remains small at all times. Also the bank angle remains small in absolute sense, but the roll-yaw coupling shows as a high-frequency oscillation. Coupling between the longitudinal and lateral rotational motion is small to non-existent, mainly because of the non-zero body-flap deflection.

In case the vehicle starts with a zero angle of attack—but still with a fixed $\delta_{b,trim} = 15°$—the oscillations are much more vehement. The vehicle has the natural tendency to move towards the trim angle of attack of 40°, but once it reaches that attitude it has a(n undesirably) large angular rate. The vehicle tries to stabilise, but keeps overshooting. This is mainly due to the larger dynamic pressure; from the altitude-velocity diagram it is clear that the absence of lift in the initial phase forces the vehicle deeper into the atmosphere and preventing skipping flight. After 400 s the simulation "explodes" due to excessive angular rates. The angle of sideslip and bank angle both diverge, and acquire large values. The bank angle "wiggles" around $t = 310$ s, is a minor oscillation around $\sigma = \pm 180°$, but due to the fact it is constrained to the interval [–180°,180°> it shows as jumps of 360°.

Figure 6.22 shows the three angular-rate components p, q, and r. For the simulation starting from $\alpha_0 = 40°$, the amplitudes remain small, although the frequency of p and r is relatively large. Again, after $t = 1{,}500$ s the angular rates increase rapidly, leading to problems for the integrator. For the situation starting from $\alpha_0 = 0°$, it is

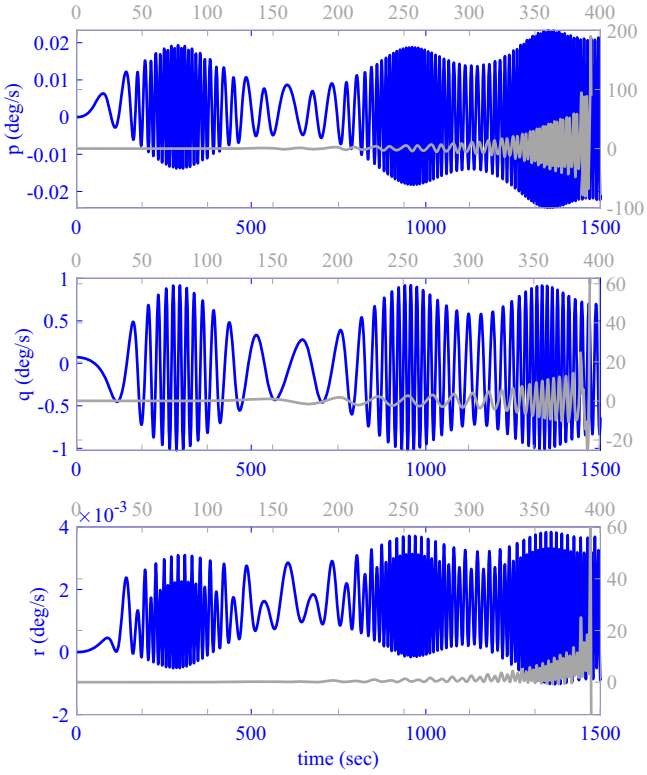


Fig. 6.22 Angular rates for the open-loop simulation ($\alpha_0 = 40°$ and $\alpha_0 = 0°$), fixed body-flap deflection of $\delta_{b,trim} = 15°$

evident that the vehicle is indeed rapidly rotating, with, for instance, a roll rate getting close to 200°/s. Finally, in Fig. 6.23 the *g-load* is shown as a function of time. For the skipping flight, the three peaks with increasing amplitude are clear, but since these skips take place in the higher layers of the atmosphere the loads are not so large. This is not true for the second case: due to the deeper dive into the atmosphere the *g-load* peaks at 8 g; the smaller oscillations are linked to the angle-of-attack oscillations that directly affect the aerodynamic forces. It is clear that there is a strong coupling between translational and rotational motion in either case.

To show the effect of absence of trim, a third simulation is done with a zero deflection of the body flap. The initial angle of attack is taken as $\alpha_0 = 40°$. The corresponding results are shown in Fig. 6.24, where it should be noted that both angle of attack and bank angle have been plotted as a continuation, rather than constrained to their respective ranges. Due to the absence of stabilising moment coming from the body flap, the pitches to a backwards orientation in about 150 s, experiencing

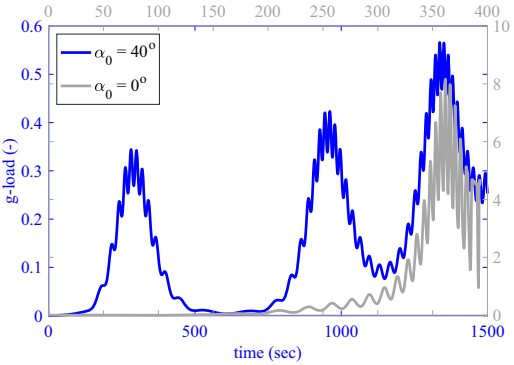


Fig. 6.23 Mechanical load for the open-loop simulation ($\alpha_0 = 40°$ and $\alpha_0 = 0°$), fixed body-flap deflection of $\delta_{b,trim} = 15°$

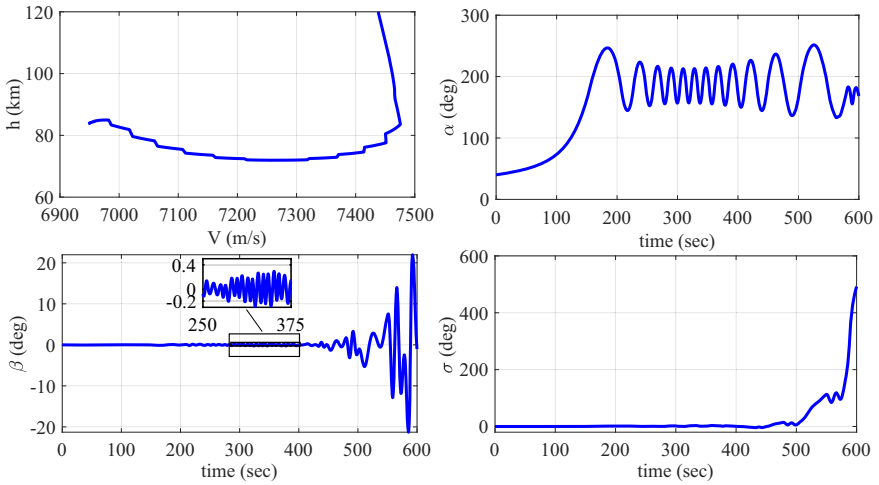


Fig. 6.24 Open-loop simulation ($\alpha_0 = 40°$), fixed body-flap deflection of $\delta_{b,trim} = 0°$

large oscillations after that.[13] The trajectory looks like that of a skipping flight, but with almost discrete jumps in it, induced by the large angle-of-attack oscillations. Initially, the bank angle remains close to zero, but towards the end of the simulation interval (around $t = 500$ s), it rapidly increases and reaches a value of about 500° (a full rotation and then some). The simulation crashes at 600 s, when the angular rates have become excessively large.

The obvious conclusion from the above discussion is that the individual eigenmotions cannot be readily discerned. Due to the very non-linear nature of the open-loop

[13] It should be understood that the aerodynamic database does not include data for these very large angle-of-attack values. The real-life oscillations will clearly be different. However, one may expect violent and out-of-control oscillations around all three axes.

motion, the vehicle is never in equilibrium and the deviations from any nominal state (if that could already be determined) are far too large to be compared to the linearised eigenmotion discussed before. And, as a very trivial conclusion, the vehicle cannot do without an attitude-control system, if it is expected to have a stable flight and reach the landing area safely. However, that discussion will not be made here, but in the next chapter, where the guidance, navigation, and control system for entry vehicles is treated in detail.

The discussion on eigenmotion of winged re-entry vehicles is concluded with two examples. These will show how analytical expressions for the eigenvalues of decoupled rotational motion can lead to design criteria for the aerodynamic shape of the vehicle. Even though these examples are limited in detail, they do provide a good first impression.

Example 6.6 (*Eigenvalues for Longitudinal Motion*) In this example we will look at the longitudinal rotational motion, defined by Δq and $\Delta\alpha$, and try and derive analytical expressions for the eigenvalues. From the related coefficients stated above, notably Eqs. (6.187), (6.186), (6.193), and (6.194), we can set up the corresponding equations of motion, assuming that there is no coupling with the translational motion.

$$\Delta\dot{q} = \frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref} \Delta\alpha + \frac{\Delta M_{T,y}}{I_{yy}} \tag{6.231}$$

$$\Delta\dot{\alpha} = \Delta q - \frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref} \Delta\alpha \tag{6.232}$$

The eigenvalue problem is stated as $\mathbf{A}\boldsymbol{\mu} = \lambda\boldsymbol{\mu}$ and is solved by

$$\mathbf{A}\boldsymbol{\mu} - \lambda\boldsymbol{\mu} = (\mathbf{A} - \lambda\mathbf{I})\,\boldsymbol{\mu} = 0 \tag{6.233}$$

The condition for a non-trivial solution is

$$|\mathbf{A} - \lambda\mathbf{I}\,| = 0 \tag{6.234}$$

$$\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} -\lambda & \frac{1}{I_{yy}} \frac{\partial C_m}{\partial\alpha} \bar{q}_0 S_{ref} c_{ref} \\ 1.0 & -\frac{1}{mV_0} \frac{\partial C_L}{\partial\alpha} \bar{q}_0 S_{ref} - \lambda \end{bmatrix} \tag{6.235}$$

Calculating the determinant of this matrix gives the following second-order polynomial:

$$\lambda^2 + \frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref} \lambda - \frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref} = 0 \tag{6.236}$$

Solution to this equation is obtained from the quadratic formula (the "*abc*" formula, i.e., $\lambda_{1,2} = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$):

$$\lambda_{1,2} = \frac{-\frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref} \pm \sqrt{\left(\frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref}\right)^2 + 4 \frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref}}}{2} \tag{6.237}$$

Inspecting the above equation leads to the following conclusions. As the first term under the square root is positive, and for a statically stable vehicle (with $C_{m_\alpha} < 0$, the second term is negative, it means that when

$$\left(\frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref}\right)^2 > \left| 4 \frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref} \right| \tag{6.238}$$

the argument of the square root is positive and the eigenvalues are real. In the limit case, i.e., the second term is zero, the eigenvalues will be:

$$\lambda_{1,2} = -\frac{1}{2\,mV_0} C_{L_\alpha} \bar{q}_0 S_{ref} \tag{6.239}$$

which represents a stable aperiodic motion for positive values of C_{L_α}. However, since C_{m_α} is not zero there are two possible cases. For negative C_{m_α} the square root will be smaller than $\frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref}$, such that both eigenvalues are negative, and thus both components represent a stable aperiodic motion. The situation changes when the second term is, in absolute sense, larger than the first one. In that case the argument to the square root is negative, and the eigenvalues become complex, though with a negative real part. So, in that case the eigenmotion becomes a damped periodic motion. The second case is that of a positive C_{m_α}, where the square root will be larger than $\frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref}$. In that case, the eigenmotions are aperiodic, with one stable and one unstable motion.

For HORUS, the first term under the square root is much smaller than the second one; the combination of large mass and velocity, mV_0, drives this term to zero, much more than I_{yy} can do, despite the much larger value of C_{L_α}. Consequently, for the damped periodic motion, the eigenvalues can also be written as

$$\lambda_{1,2} = -\frac{1}{2\,mV_0} C_{L_\alpha} \bar{q}_0 S_{ref} \pm i \sqrt{\frac{1}{I_{yy}} \left| C_{m_\alpha} \right| \bar{q}_0 S_{ref} c_{ref}} \tag{6.240}$$

where the resulting minus sign of the argument to the square root should be ignored, since it has been taken out as $i = \sqrt{-1}$. With Eqs. (6.51) and (6.52) the damping and natural frequency of the short-period oscillation can be explicitly written as:

$$\zeta = \frac{\frac{1}{2mV_0} C_{L_\alpha} \bar{q}_0 S_{ref}}{\sqrt{\left(-C_{L_\alpha} \frac{\bar{q}_0 S_{ref}}{2mV_0}\right)^2 - \frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref}}} \quad \Rightarrow \tag{6.241}$$

$$\zeta \approx \frac{\frac{1}{2mV_0} C_{L_\alpha} \bar{q}_0 S_{ref}}{\sqrt{-\frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref}}} \approx \frac{1}{2mV_0} \frac{C_{L_\alpha}}{\sqrt{-C_{m_\alpha}}} \sqrt{\frac{I_{yy} \bar{q}_0 S_{ref}}{c_{ref}}} \tag{6.242}$$

and

$$\omega_n \approx \sqrt{-\frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref}} \tag{6.243}$$

assuming a negative C_{m_α} (so the minus sign replaces the absolute value). So, from vehicle properties and velocity profile the longitudinal eigenmotion can be derived, or, *vice versa* and given design freedom, by specifying some of these properties the eigenmotion can be enforced. ■

Example 6.7 (*Eigenvalues for Lateral Motion*) To isolate the eigenvalues for the lateral motion, we will assume that the longitudinal and lateral motion are decoupled. This can easily be achieved by putting all "symmetrical" values to zero (ΔV, Δh, Δq, and $\Delta\alpha$, and reformulate the state-space model with just the states Δp, Δr, $\Delta\beta$, and $\Delta\sigma$. In that case the state matrix is reduced to

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & \frac{1}{I_{xx}} C_{l_\beta} \bar{q}_0 S_{ref} b_{ref} & 0 \\ 0 & 0 & \frac{1}{I_{zz}} C_{n_\beta} \bar{q}_0 S_{ref} b_{ref} & 0 \\ \sin\alpha_0 & -\cos\alpha_0 & -\frac{1}{mV_0} C_{S_\beta} \bar{q}_0 S_{ref} & -\frac{g_0}{V_0} \cos\gamma_0 \cos\sigma_0 \\ -\cos\alpha_0 & -\sin\alpha_0 & \frac{\tan\gamma_0 \cos\sigma_0}{mV_0} C_{S_\beta} \bar{q}_0 S_{ref} + \frac{g_0}{V_0} \cos\gamma_0 \cos\sigma_0 - \frac{L_0}{mV_0} & \frac{\tan\gamma_0}{mV_0} \cos\sigma_0 L_0 \end{bmatrix} \tag{6.244}$$

where we have also introduced the short-hand notation $C_{S_\beta} = \frac{\partial C_S}{\partial \beta}$, $C_{l_\beta} = \frac{\partial C_l}{\partial \beta}$ and $C_{n_\beta} = \frac{\partial C_n}{\partial \beta}$. To facilitate the analysis, the above matrix will be further simplified. Inspecting the individual elements shows that some coefficients are quite a bit smaller than others, especially those involving terms $\frac{1}{mV_0}$. After removing the affected terms from a_{33}, a_{43} and a_{44}, $\mathbf{A}$ is reduced to:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & \frac{1}{I_{xx}} C_{l_\beta} \bar{q}_0 S_{ref} b_{ref} & 0 \\ 0 & 0 & \frac{1}{I_{zz}} C_{n_\beta} \bar{q}_0 S_{ref} b_{ref} & 0 \\ \sin\alpha_0 & -\cos\alpha_0 & 0 & -\frac{g_0}{V_0} \cos\gamma_0 \cos\sigma_0 \\ -\cos\alpha_0 & -\sin\alpha_0 & \frac{g_0}{V_0} \cos\gamma_0 \cos\sigma_0 & 0 \end{bmatrix} \tag{6.245}$$

Evaluating the determinant gives a fourth-order characteristic polynomial, with obviously four eigenvalues. However, as some coefficients are zero only three non-zero eigenvalues are found. Thus, the polynomial can be written as

$$P(\lambda) = A\lambda^3 + B\lambda^2 + C\lambda + D \tag{6.246}$$

with

$$A = 1 \qquad B = 0$$

$$C = \left(\frac{g_0}{V_0}\right)^2 + \left(\frac{C_{n_\beta}\cos\alpha_0}{I_{zz}} - \frac{C_{l_\beta}\sin\alpha_0}{I_{xx}}\right)\bar{q}_0 S_{ref} b_{ref}$$

$$D = -\left[\frac{C_{l_\beta}\cos\alpha_0}{I_{xx}} + \frac{C_{n_\beta}\sin\alpha_0}{I_{zz}}\right]\frac{g_0}{V_0}\bar{q}_0 S_{ref} b_{ref}$$

Solving for the three roots is easily done numerically, but to obtain analytical expressions is laborious, even though the *cubic formula* to provide the roots has been known since the 16th century. Abramowitz and Stegun (1972) give the solution for the three roots, but even then, substituting the coefficients and simplifying the final expressions is a tedious task. Inspecting the values of the coefficients A, B, C, and D, indicates that the three eigenvalues consist of a real one and a complex pair, the latter of which describes the Dutch roll. Without further derivation, this complex pair is given by[14]:

$$\lambda_{Dutch\ roll} = \pm i\sqrt{\left(C_{n_\beta} - C_{l_\beta}\tan\alpha_0\frac{I_{zz}}{I_{xx}}\right)\frac{\bar{q}_0 S_{ref} b_{ref}}{I_{zz}}\cos\alpha_0} \tag{6.247}$$

In the above expression, the term

$$C_{n_{\beta,dyn}} = C_{n_\beta} - C_{l_\beta}\tan\alpha_0\frac{I_{zz}}{I_{xx}} \tag{6.248}$$

represents the lateral aerodynamic stiffness or the restoring force of the vehicle's motion (Yonemoto and Inatani 1988). For the open-loop analysis, it has an approximate meaning of a sufficient stability parameter.

$C_{n_{\beta,dyn}}$ is an important design parameter, because its sign predicts whether the vehicle restores or enlarges a sideslip disturbance, $\Delta\beta$. The desired sign, and thus a proper design criterion, is

$$C_{n_{\beta,dyn}} > 0 \tag{6.249}$$

Of course, this criterion cannot just be based on the individual moment derivatives, C_{n_β} and C_{l_β}, as for different flight phases the angle of attack may be substantially

[14] A complex eigenvalue with zero real part means that the Dutch roll is a purely oscillatory motion without damping or amplification. Earlier, we have seen that the numerical counterparts have a non-zero real part. Even though only small, this positive component shows that the Dutch roll is slightly unstable, especially towards the end of the flight.

different, and thus have an impact not only on the derivatives themselves, but also through $\tan \alpha_0$. Last but not least, also the vehicle's mass properties, i.e., the ratio of the two moments of inertia, will contribute to the restoring moment as a scaling factor.

Commonly one strives for a positive C_{n_β} and a negative C_{l_β}, although the former cannot always be guaranteed. In that case, the following design restriction comes into play:

$$C_{n_\beta} > C_{l_\beta} \tan \alpha_0 \frac{I_{zz}}{I_{xx}} \tag{6.250}$$

To allow for a reasonable design margin, the following restriction for the ratio of C_{n_β} over C_{l_β} should be net:

$$\forall C_{l_\beta} < 0 \quad \Rightarrow \quad \frac{C_{n_\beta}}{C_{l_\beta}} < \tan \alpha_0 \frac{I_{zz}}{I_{xx}} \tag{6.251}$$

In Fig. 6.25 the related aspects for this design criterion has been plotted for HORUS. In the left figure, one can see that C_{n_β} is negative, which is less favourable. For most of the flight, $C_{n_\beta} > C_{l_\beta}$, but towards the end the situation reverses. In combination with the decreasing angle of attack this gives a drop in $C_{n_{\beta,dyn}}$, although it does not become negative. However, it is shown in the right figure that the design margin has decreased significantly. It may be expected that the flying characteristics and controllability of HORUS become more of a challenge in the low hypersonic, high supersonic flight range. ■

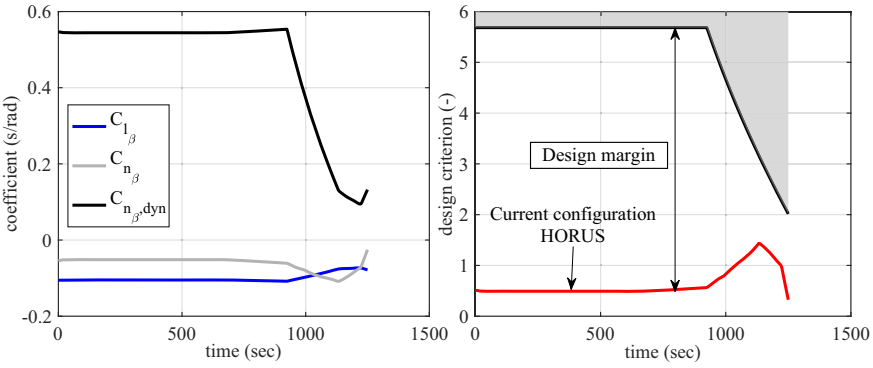


Fig. 6.25 Stability derivatives and lateral design criterion for HORUS

6.5 Eigenmotion of Entry Capsules

In the previous section we extensively described the eigenmotion of a winged entry vehicle and tried to relate it to the eigenmotion of conventional aircraft, flying at conventional speeds. It was concluded that:

1. There are similar characteristic modes compared with conventional aircraft, yet with several differences.
2. The absence of the atmosphere in the early re-entry phase makes everything "slower".
3. There is a coupling between symmetric and asymmetric motion.
4. The open-loop, non-linear simulations show the necessity of control due to the severe oscillatory nature.
5. Due to the time-variant and non-linear nature of re-entry missions, the eigenmodes are not distinguishable in an open-loop simulation.

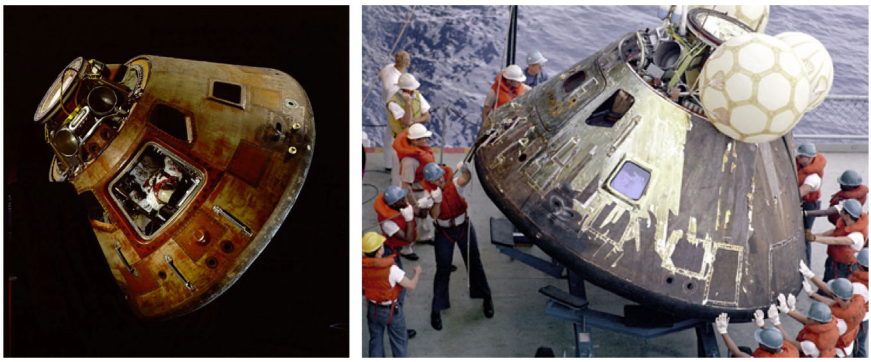
The next step in the analysis of eigenmotion of space vehicles is to move away even further from conventional aircraft by looking at different vehicle shapes, albeit in similar velocity and altitude regimes as before. To identify the influence of the vehicle shape on the eigenmotion we have selected the Apollo capsule as the object of study, and consider the entry in the atmosphere after its return flight from the Moon.

The Apollo program was the spaceflight effort carried out by the United States' National Aeronautics and Space Administration (NASA) to land the first humans on Earth's Moon. Initiated in the early sixties of the previous century, Apollo followed the Mercury and Gemini programs and ran from 1961 to 1972. Its goal was accomplished with the Apollo-11 mission when astronauts Neil Armstrong and Buzz Aldrin landed their Lunar Module on the Moon on July 20, 1969, and walked on its surface. They safely landed on Earth on July 24 in the so-called Command Module (see Fig. 6.26), a blunt entry capsule with an ablative heat shield. After this pioneering mission, another five Moon landings were carried out in a similar fashion.

To study the characteristic motion of the Apollo Command Module, the mission and vehicle will be introduced in the next Sect. 6.5.1. Since the state-space model for Apollo will be different from the model used for HORUS, in Sect. 6.5.2 the differences will be highlighted. Then, in Sect. 6.5.3 the time history of the eigenvalues is shown and the corresponding eigenmotion discussed.

6.5.1 Apollo Nominal Mission

The model that is used in this study is that of an Apollo-like entry capsule (mass m = 4976 kg), which is on a return leg from the Moon. It enters the atmosphere at 220 km altitude with a relative velocity of V = 11 km/s, and a corresponding flight-path



(a) Apollo Entry capsule. (b) Apollo-13 recovery.

Fig. 6.26 Typical shape of a manned entry capsule. *Images* NASA

angle of $\gamma = -9.536°$. To avoid excessive fuel consumption for attitude control, the z-location of the c.o.m. has been moved up from the symmetry axis to provide a natural hypersonic trim condition with an angle of attack of about $\alpha = -24.5°$. Therefore, angle-of-attack control is only used to stabilise the angle of attack around this trim condition; bank-angle-control is used to orient the lift in a lift-down condition to avoid skipping out of the atmosphere,[15] and is effectively the only guidance means. For attitude control the capsule is equipped with reaction-control thrusters only.

Apollo is a semi-ballistic capsule that flies with a low lift-to-drag ratio ($L/D \approx 0.3$), in a trajectory that is characterised by a high heat and mechanical load. The high heat load drove the design of the thermal protection system (TPS) towards the use of ablative materials, i.e., materials that absorb the heat and consequentially slowly burn away, thus rejecting the heat from the vehicle.

Apollo enters the atmosphere with the heat shield facing the flow. Let us define a body frame with the origin in the theoretical apex of the vehicle, i.e., if we would extend the smallest portion of the vehicle it would end in a point, which is called the *apex*. The x_B-axis is positive in the direction of the heat shield and intersects it in the centre of the circular cross section. The z_B-axis is positive pointing downwards for a nominal entry and the y_B-axis completes the right-handed system. In this frame of reference, the location of the centre of mass is given by $\mathbf{r}_{\mathbf{cm}} = (2.57, 0.0, -0.137)^T$. Due to the (significant) shift of the centre of mass in z-direction, the capsule is not rotationally symmetric in mass, which means that one of the products of inertia is

[15] When the entry velocity of a vehicle entering the atmosphere is significantly larger than the local circle velocity (about 7.8 km/s), the kinetic-energy of the vehicle is so large that the vehicle may bounce off the atmosphere, or *skip*. On one hand, on could use this skipping concept to reduce the energy level and enter the atmosphere after the second or third skip, but with the risk that the vehicle bounces off for good, when this is not carefully planned. On the other hand, when the vehicle should remain in the atmosphere during the initial entry, the lift should be oriented in a downward direction to "pull" the vehicle into the atmosphere. The latter guidance concept was chosen for Apollo, to avoid the chance of losing the vehicle and its crew.

non-zero. The corresponding numerical values of moments and products of inertia are $I_{xx} = 5617.61\,\text{kg/m}^2$, $I_{yy} = 4454.62\,\text{kg/m}^2$, $I_{zz} = 4454.80\,\text{kg/m}^2$, $I_{xy} = I_{yz} = 0.0\,\text{kg/m}^2$ and $I_{xz} = 1752.00\,\text{kg/m}^2$.

The aerodynamic database used here (see also see Appendix B) is based on extensive wind-tunnel measurements done in the mid-sixties as part of the Apollo-capsule design (NAA 1965; Mosely et al. 1967). What we need for our analysis are expressions for the aerodynamic forces and moments, as well as the first derivatives to be substituted in the linearised system discussed in Sect. 6.4.1. The drag, side and lift force appear to be functions of the states h and V, and the attitude angles α and β:

$$D(h, V, \alpha, \beta) = C_D(M(h, V), \alpha, \beta)\frac{1}{2}\rho(h)V^2 S_{ref} \tag{6.252}$$

$$S(h, V, \beta) = C_S(M(h, V), \beta)\frac{1}{2}\rho(h)V^2 S_{ref} \tag{6.253}$$

$$L(h, V, \alpha) = C_L(M(h, V), \alpha)\frac{1}{2}\rho(h)V^2 S_{ref} \tag{6.254}$$

where $\rho(h)$ is the altitude-dependent atmospheric density and $M(h, V)$ is the Mach number, through the speed-of-sound dependency with altitude a function of both h and V. S_{ref} is the aerodynamic reference area, with $S_{ref} = 12\,\text{m}^2$.

The aerodynamic roll, pitch and yaw moments, $\mathcal{L}$, $\mathcal{M}$ and $\mathcal{N}$ are functions of the states h, V, α, β, q and r (NAA 1965; Mosely et al. 1967) and include the dynamic damping coefficients C_{m_q} and C_{n_r}. However, the quoted references do not give any data on the roll-moment coefficient. At most it will be a small value, so we can safely assume it to be zero:

$$\mathcal{L} = 0 \tag{6.255}$$

$$\mathcal{M}(h, V, \alpha, \beta) = \left[C_m(M(h, V), \alpha, \beta) + C_{m_q}(M(h, V), \alpha, \beta)\frac{q d_{ref}}{V}\right]\frac{1}{2}\rho(h)V^2 S_{ref} d_{ref} \tag{6.256}$$

$$\mathcal{N}(h, V, \alpha, \beta) = \left[C_n(M(h, V), \alpha, \beta) + C_{n_r}(M(h, V), \alpha, \beta)\frac{r d_{ref}}{V}\right]\frac{1}{2}\rho(h)V^2 S_{ref} d_{ref} \tag{6.257}$$

where d_{ref} is the aerodynamic reference length (i.e., the vehicle diameter). The actual value is $d_{ref} = 3.9\,\text{m}$. Note that the force and moment coefficients are defined with respect to the apex—during the wind-tunnel campaign the mounting point of the wind-tunnel model—so the aerodynamic moments will also include a contribution of the forces to compensate for the distance between apex and centre of mass.

Knowing the state dependency of forces and moments will serve as input for obtaining the state-space formulation required to do the eigenmotion analysis. Before we move on to this we will briefly discuss the reference trajectory. Since we do not have the actual Apollo reference trajectory at our disposal, nor the models for the guidance, navigation, and control system, we will generate the reference trajectory by simulating the mission in open loop. This means that given the initial conditions as stated before, we "drop" Apollo in the atmosphere and let it "fall" to the Earth's

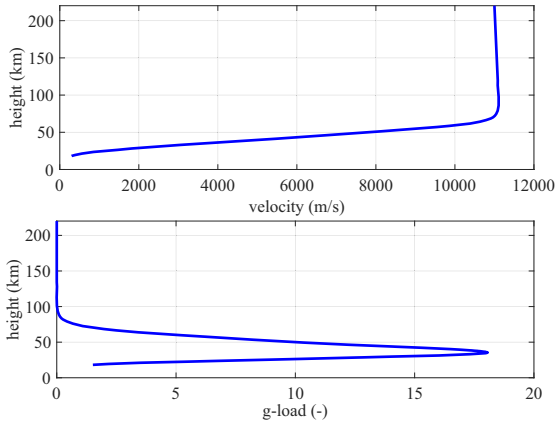


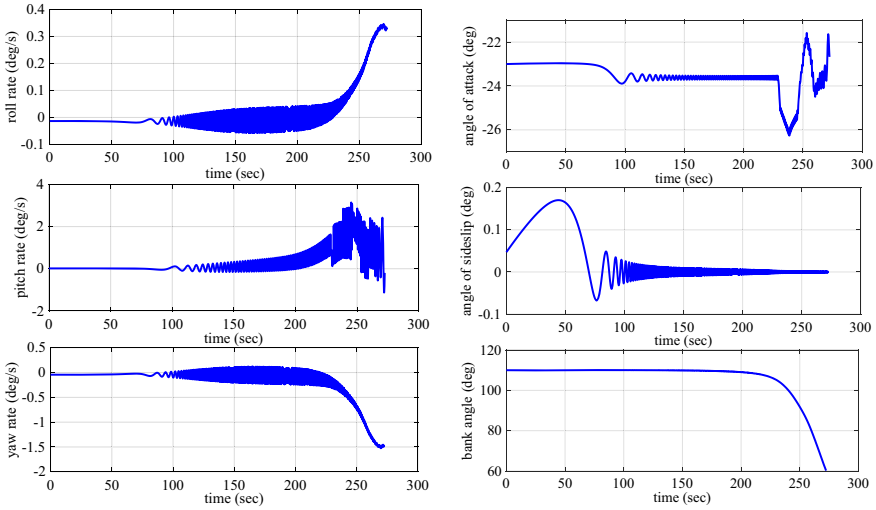
Fig. 6.27 Altitude-velocity profile for a free-fall Apollo entry (top), and occurring *g-load* (bottom)

surface. Thus, the resulting motion includes the very eigenmotion that we want to study in the next sub-section. However, we saw in the previous section, while discussing the open-loop motion of HORUS, that the individual components of the eigenmotion cannot be discerned, so we will postpone that discussion until later.

In Fig. 6.27 the altitude-velocity profile is shown. The entry begins at 220 km altitude and the first phase is characterised by an increase in velocity. This is always the case for entry vehicles, since the atmosphere is too thin to generate significant aerodynamic forces. Thus, the gravitational force dominates and pulls the vehicle down, thereby increasing the vehicle's velocity. At a certain combination of altitude and velocity the dynamic pressure starts building up and the retarding force is now larger than the gravitational pull, thus decreasing the vehicle's velocity.

The velocity rapidly decreases over just a small altitude range, resulting in a large mechanical load, or *g-load*. The *g-load* peaks at 18 g, which clearly shows the necessity to carefully control the flight path. Without a guidance system to properly plan the path and an attitude controller to maintain the commanded attitude the astronauts will not easily survive this particular entry. The final velocity shown in the plot ($V \approx 300$ m/s or $M = 1$) is attained after only 275 s of flight. One can easily understand that the thermal environment is challenging to say the least, if one knows that the only mechanism to reduce the kinetic energy is friction.

In Fig. 6.28 the rotational states are shown. Its is clear that the angle of attack stabilises around a trim value, dictated by the location of the centre of mass (notably, the location in z-direction). Around $t \approx 230$ s the vehicle behaviour becomes quite erratic, due to a combination of aerodynamic moments and non-linear couplings. Since the location of the centre of mass was determined to give a trim angle in the hypersonic regime, where the Mach effect is very small, a different trim equilibrium is found in the low hypersonic and supersonic regime. Here, the Mach effect is relatively strong.

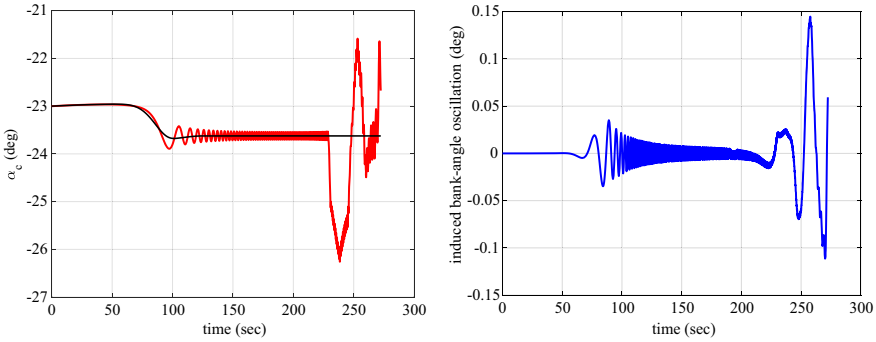


(a) Attitude angles (b) Angular rates

Fig. 6.28 State variables for the rotational motion of a free-fall Apollo entry

However, the motion is not completely unstable as the angle of attack does not diverge more than a few degrees, with only a minor local (high-frequency) oscillation. The angle of sideslip is only marginally affected, although its behaviour is also characterised by a high-frequency oscillation. The bank angle seems to be unstable as it keeps on decreasing. Main reason is that due to the roll-yaw coupling a bank-angle rotation is induced, and since there is no (roll) damping modelled, the rotation is undamped. The angular rates are oscillating rapidly, indicating that the aerodynamic moments keep on changing sign. This is consistent with the angle-of-attack oscillation around its trim value, representing a zero moment.

To do the eigenmotion analysis it would be better to have a smooth representation of the attitude angles, as then the change of eigenvalues will not suffer from the high-frequency oscillations. This seems justified as the induced eigenmotion should not have an influence on itself. Therefore, we will do a curve fit of the angle-of-attack and bank-angle profile and set the angle of sideslip to zero. This does not have an effect on the actual trajectory, because the trajectory dynamics are at least an order of magnitude slower, so the high-frequency attitude oscillations will average out. The resulting, smoothed angle of attack profile is shown in Fig. 6.29a. One might expect a larger influence towards the end of the flight, as the difference between actual and smoothed angle of attack is large. However, one should realise that the dynamic pressure has dropped significantly thus reducing the magnitude of the aerodynamic forces. At that moment, the forces have only a small influence on the shape of the trajectory. The induced oscillation in the bank angle is only small (Fig. 6.29b), but is removed nonetheless to avoid any effect on the eigenmotion analysis.



(a) Actual and smoothed angle-of-attack profile

(b) Difference between actual and smoothed bank-angle profile

Fig. 6.29 Smoothing applied to actual attitude profiles to obtain reference profile

6.5.2 State-Space Model

To calculate the eigenvalues and study the eigenmotion of the entry capsule we will, of course, follow the same approach as for the winged entry vehicle HORUS in the previous section. Starting from the non-linear equations of motion the equations are linearised with respect to state and control variables and, as before, put into state-space form:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{6.258}$$

In terms of state variables, there is no reason to change the ones we used before, so (see also Sect. 6.4.1):

$$\mathbf{x} = (\Delta V, \Delta\gamma, \Delta R, \Delta p, \Delta q, \Delta r, \Delta\alpha, \Delta\beta, \Delta\sigma)^T$$

However, there are two fundamental differences with respect to HORUS that affect the definition of **A** and **B**. In the first place, the Apollo(-like) entry capsule does not have aerodynamic control surfaces. The control system can only command the thrusters of the reaction control system to modify (or maintain) its attitude. This means that the control vector is now defined as

$$\mathbf{u} = (\Delta M_{T,x}, \Delta M_{T,y}, \Delta M_{T,z})^T$$

with

$M_{T,x}$ = roll-thruster moment (Nm)
$M_{T,y}$ = pitch-thruster moment (Nm)
$M_{T,z}$ = yaw-thruster moment (Nm)

As a consequence, **B** reduces to a 9×3 matrix (rather than 9×6). In the second place, we cannot assume that the vehicle is rotationally symmetric in mass, as we did for HORUS. It may seem strange that we could assume this for an aircraft-like vehicle, but we did this after inspecting the values of moments and products of inertia; for HORUS I_{xz} is much smaller than the moments of inertia. In case of Apollo, I_{xz} is on average about 30% of the value of I_{xx}, I_{yy} and I_{zz} due to the shift in centre of mass. This value is far too large to be ignored, so it will introduce additional (coupling) terms in both **A** and **B**. Further, since the aerodynamic database of HORUS is more extensive, for Apollo some of the components disappear.

The linearisation is done in exactly the same way as in the previous section and will not be repeated here. It is noted that the inclusion of the (aerodynamic) damping terms due to C_{m_q} and C_{n_r} do not give additional coefficients in **A**, since these terms are of higher order. This has already been proven in Sect. 6.4.1. The non-zero coefficients of **A** and **B** are shown in Eq. (6.166). Below, we will only list those coefficients that are different from the earlier definition.

The related elements of matrix **A**, i.e., a_{ij}, are given by

$$a_{pp} = I_{p_1} q_0 \tag{6.259}$$

$$a_{pq} = I_{p_1} p_0 + I_{p_2} r_0 \tag{6.260}$$

$$a_{pr} = I_{p_2} q_0 \tag{6.261}$$

$$a_{p\beta} = \frac{I_{zz}}{I^*} \frac{\partial C_l}{\partial \beta} \bar{q}_0 S_{ref} b_{ref} + \frac{I_{xz}}{I^*} \frac{\partial C_n}{\partial \beta} \bar{q}_0 S_{ref} b_{ref} \tag{6.262}$$

$$a_{qp} = -2 I_{q1} p_0 + I_{q2} r_0 \tag{6.263}$$

$$a_{qr} = I_{q2} p_0 + 2 I_{q1} r_0 \tag{6.264}$$

$$a_{rp} = I_{r_1} q_0 \tag{6.265}$$

$$a_{rq} = I_{r_1} p_0 + I_{r_2} r_0 \tag{6.266}$$

$$a_{rr} = I_{r_2} q_0 \tag{6.267}$$

$$a_{r\beta} = \frac{I_{xx}}{I^*} \frac{\partial C_n}{\partial \beta} \bar{q}_0 S_{ref} b_{ref} + \frac{I_{xz}}{I^*} \frac{\partial C_l}{\partial \beta} \bar{q}_0 S_{ref} b_{ref} \tag{6.268}$$

In the above coefficients, the following auxiliary variables have been used:

$$I_{p_1} = \frac{(I_{xx} - I_{yy} + I_{zz}) I_{xz}}{I^*} \quad \text{and} \quad I_{p_2} = \frac{\left(I_{yy} - I_{zz}\right) I_{zz} - I_{xz}^2}{I^*}$$

$$I_{q_1} = \frac{I_{xz}}{I_{yy}} \quad \text{and} \quad I_{q_2} = \frac{I_{zz} - I_{xx}}{I_{yy}}$$

$$I_{r_1} = \frac{\left(I_{xx} - I_{yy}\right) I_{xx} - I_{xz}^2}{I^*} \quad \text{and} \quad I_{r_2} = \frac{\left(-I_{xx} + I_{yy} - I_{zz}\right) I_{xz}}{I^*}$$

$$I^* = I_{xx} I_{zz} - I_{xz}^2$$

As discussed earlier, the equilibrium angular rate components p_0, q_0 and r_0 follow from the condition $\dot{\alpha} = \dot{\beta} = \dot{\sigma} = 0$, and are given by Eqs. (6.103)–(6.107)—or, equivalently, by Eqs. (6.108)–(6.111) for the manoeuvring condition.

The non-zero elements of matrix **B**, i.e., b_{ik}, are:

$$b_{px} = \frac{I_{zz}}{I^*} \tag{6.269}$$

$$b_{pz} = \frac{I_{xz}}{I^*} \tag{6.270}$$

$$b_{qy} = \frac{1}{I_{yy}} \tag{6.271}$$

$$b_{rx} = \frac{I_{xz}}{I^*} \tag{6.272}$$

$$b_{rz} = \frac{I_{xx}}{I^*} \tag{6.273}$$

By inspecting the differences between the two system matrices **A** we can expect for Apollo an even stronger coupling between the three axes of motion than was the case for HORUS. This will show not only between longitudinal and lateral motion, but also as a stronger roll-yaw coupling in the lateral motion. Looking at **B** it is clear that also here the roll and yaw axes have a stronger coupling, which makes it harder to control the individual axes of the entry capsule. This means that a control moment exerted on the X-axis will induce a perturbing motion around the Z-axis and vice versa. Control, however, is not part of the current discussion, so this remains an observation only.

6.5.3 Apollo Characteristic Motion

After linearising the non-linear equations of motion and formulating the result in state-space form, we can evaluate **A** at a number of discrete time points, as we already did in Sect. 6.4. For each of the resulting LTI systems we can calculate the eigenvalues and eigenvectors, and derive a number of performance indicators: the

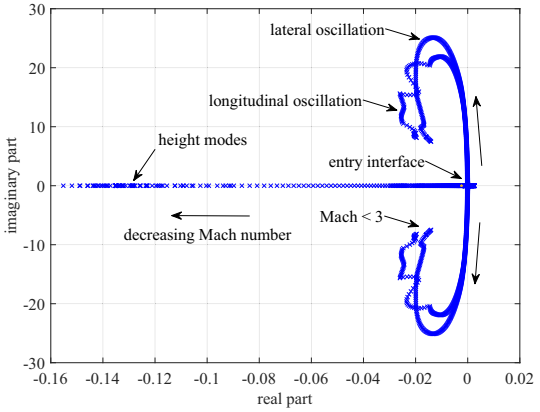


Fig. 6.30 Eigenvalues for coupled rotational and translational motion. Evaluation done for discrete time points

period P, Eq. (6.49), the *halving time* $T_{\frac{1}{2}}$, Eq. (6.50), the *damping ratio* ζ, Eq. (6.51), and the *natural frequency* ω_n, Eq. (6.52). From the eigenvectors we calculate the modulus z and argument θ, Eq. (6.78). These parameters will help us to identify the type of eigenmode and the state variables involved in the eigenmotion.

Before we get to the detailed discussion of the eigenmotion, we will first have a look at the global picture, i.e., the variation of all eigenvalues throughout the mission. In Fig. 6.30 the eigenvalues for the coupled rotational and translational motion is shown (calculated from the 9×9 system matrix $\mathbf{A}$). The main observations are:

- there are two complex pairs, representing a longitudinal and lateral oscillation. The complex part is close to zero at entry but reaches large values (around 25) when the dynamic pressure (and thus the aerodynamic forces and moments) reach large values. This indicates a high-frequency oscillation when the vehicle is left to itself, and this is exactly what we saw earlier in Fig. 6.28.
- The remaining five eigenvalues seem to be all real (i.e., not complex), and what we will see later is that they are mainly related to altitude (and occasionally velocity). Therefore, we have indicated them by "height modes".
- Apart from one aperiodic eigenmode that is marginally unstable, all eigenmodes have negative real parts and therefore, are stable.

An aid in analysing the time history of the eigenvalues is to assume that the translational and rotational motion are decoupled, an assumption that usually holds because the translational dynamics are at least an order of magnitude slower than the rotational dynamics. Also, similar to what we do when analysing aircraft motion, at first we can can assume that the longitudinal and lateral motion are decoupled and verify this later. So, in Figs. 6.31 and 6.32 the eigenvalues have been separated for translational and rotational motion. The shown eigenmodes confirm the ones indicated in Fig. 6.30.

Additional information shown is that for the translational motion, there is first an oscillatory height motion, which later separates into two aperiodic modes. Further, one of the aperiodic modes for translational motion is marginally unstable, whereas the height oscillation has a low frequency and is marginally stable. The other height modes are stable, and the damping increases with decreasing Mach number. For the lateral motion, there are also two aperiodic modes that are a combination of bank and height modes. These modes are either poorly damped or slightly unstable, an effect that becomes more apparent with decreasing Mach number. More details will follow from the inspection of the characteristic values.

These characteristic values have been calculated for 7 time points ($t = 0$, 50, 100, 150, 200, 250 and 275 s) and are listed, together with eigenvalues and eigenvectors, in Table 6.10 through Table 6.16 (to safeguard the reading flow, again placed at the end of the chapter). As we did for HORUS, we have labelled each of the eigenmotions with a name that closely resembles the nature of the eigenmotion. However, as we will see, there will be less relation with the corresponding modes for winged re-entry vehicles (and thus also for conventional aircraft).

We will begin the discussion by looking at the so-called *longitudinal* oscillation. In the tables we have put "longitudinal" between quotes, as it is clear that even though the angle of attack is the driving state in this motion, there is a very strong coupling with the bank angle, and to a lesser extent with the angle of sideslip. Note that this coupling mainly arises from the non-zero bank angle and the inertia coupling. In principle, this motion is a combined symmetric and asymmetric motion. However, at the entry interface when the dynamic pressure (and thus the aerodynamic forces and moments) is practically zero, the motion is driven by inertia only, so the coupling will not be so apparent. The eigenvalue shows small real and imaginary parts, yielding a period of $P = 885$ s and a very small damping ratio leading to an amplitude half time of about 75,000 s (!). Effectively this means that the oscillation is undamped.

However, gradually with increasing dynamic pressure the aerodynamic moments increase, giving rise to faster oscillations; P decreases to 155 s at $t = 50$ s, 4.3 s at $t = 100$ s to the minimum value of $P = 0.4$ s between $t = 200$ and $t = 250$ s. So the motion becomes very high frequency, and if it were a true longitudinal mode, it could possibly be compared with the short-period oscillation of conventional aircraft. However, in the short-period oscillation the velocity is not a "player", contrary to the current eigenmode. From a marginal contribution for $t < 50$ s, ΔV becomes a significant component at $t = 100$ s, and is the second largest component between $t = 200$ s and 250 s, indicating a significant coupling between translational and rotational motion. This coupling also manifests through the coupling with altitude for $t < 100$ s. The coupling with the lateral motion remains active up to $t = 100$ s, after which this coupling reduces to a minor one only. However, close to the end of flight ($t = 275$ s) the coupling is back, whereas the link with the translation motion has disappeared. Throughout the flight the oscillation is damped, although this effect is not so strong; towards the end of the flight the amplitude half time has become $T_{\frac{1}{2}} \approx 35$ s with a maximum damping ratio of $\zeta = 2.4 \cdot 10^{-3}$.

The second oscillatory eigenmode that we will discuss has been labelled *lateral oscillation*. Also here, in the table the "lateral" has been put in quotes, because it is not

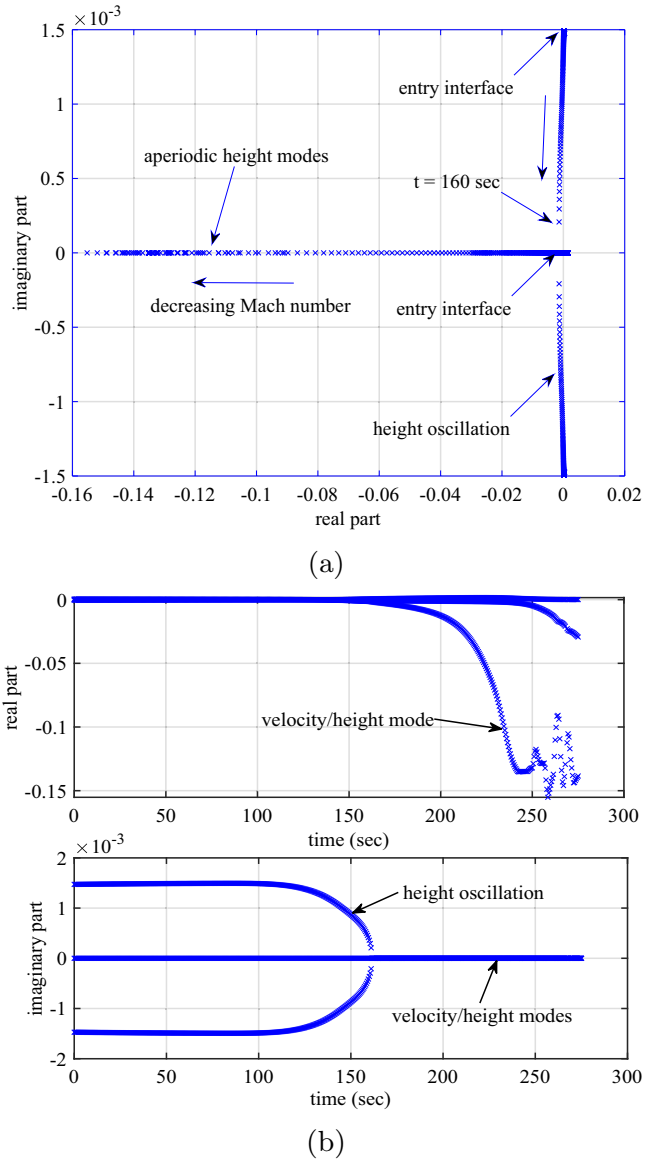


Fig. 6.31 Eigenvalues for translational motion. Evaluation done for discrete time points

a "pure" lateral oscillation. In fact, up to $t = 100$ s, the altitude is the dominating state component, representing a strong coupling with a symmetric, translational motion. Also from the beginning, the angle of attack represents a significant component besides the angle of sideslip and bank angle that made us decide to call this mode lateral oscillation. So, also in this case the longitudinal and lateral motion are coupled.

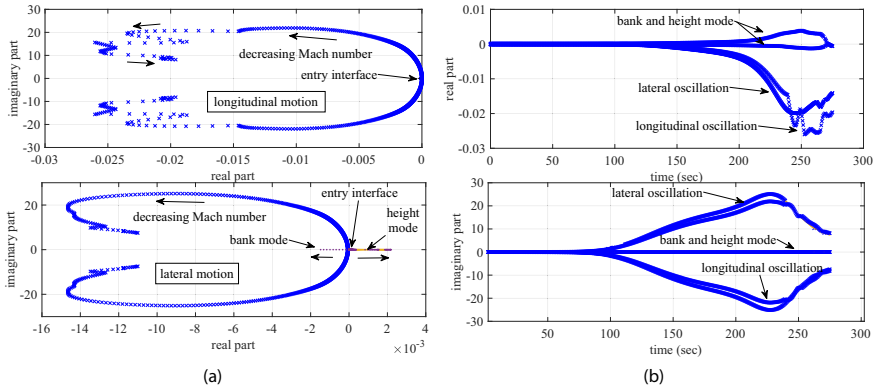


Fig. 6.32 Eigenvalues for rotational motion. Evaluation done for discrete time points

At $t = 100$ s, this coupling has disappeared and the yaw rate has popped up to be the main component (together with ΔR). The yaw rate becomes the only main component until the very end of the flight, and only then the angle of sideslip has become a significant contributor. So, basically, the lateral oscillation could be called yaw oscillation. With a little bit of imagination, one might compare this mode with the Dutch roll of conventional aircraft, as the main components are the roll and yaw rate, as well as the angle of sideslip and bank angle. In terms of period and damping, this mode is comparable to the longitudinal oscillation, slightly better damped (but still marginal) and with a shorter period. The fastest oscillation at $t = 200$ s has a period of $P = 0.3$ s. At $t > 200$ s the damping and period of the longitudinal and lateral oscillation are more-or-less the same, the former being only slightly better damped.

The third and last oscillation that we need to discuss is the so-called height oscillation, an eigenmode of translational motion that begins right at the entry interface. Initially unstable (again, with a very small negative damping ratio, so effectively indifferent) the real part of the eigenvalue slowly becomes negative (somewhere between $t = 100$ and 150 s) until it reaches its most negative value at $t \approx 160$ s (see also Fig. 6.31). This eigenmode is a "pure" height mode, as the only state component involved is ΔR. It is a slow motion with a period P well in excess of 4,000 s. The smallest amplitude half time is at $t = 160$ s and is $T_{\frac{1}{2}} = 994$ s. At $t = 160$ s, the periodic eigenmode splits into two aperiodic height modes. Since there are more aperiodic height modes it becomes difficult to exactly pinpoint those two modes. It seems that the three modes, which have a period of $P_1 = 55.4$ s (stable mode), $P_2 = 509.3$ s (unstable mode) and $P_3 = 550.4$ s (stable mode) at $t = 200$ s are the modes related to translational motion, in particular if we study Fig. 6.31. This figure shows the eigenvalues of a decoupled translational motion.

One interesting feature shows itself in these two aperiodic height modes. One of the two modes has both the velocity and the altitude as driving components, and could have been called a phugoid if it were not for the fact that this is an aperiodic mode

rather than a periodic one. To distinguish from the height modes we have designated this one the *height/velocity* mode. It is the most stable of the discussed height modes with an amplitude half time reaching a value of as low as 5 s.

To conclude the discussion on height modes related to translational motion, we already mentioned that the third mode is an aperiodic height mode (the last column in each of the tables). This initially (marginally) stable mode becomes slightly unstable at $t = 150$ s, and after the separation of the periodic height mode shows an almost discrete jump in damping (the amplitude half time changes from 3,209 s to 509.3 s at $t = 200$ s). For the remaining part of the flight this mode is marginally unstable. All-in-all, the height modes are very slow eigenmotions not worth discussing any further.

The last eigenmodes that we discuss consist of a pair of (aperiodic) *bank modes*, one stable and one unstable. These modes begin at $t = 0$ s with only the bank angle as dominating state component. At $t = 50$ s, the height has become the dominating component (note the coupling between rotational and translational motion). Of course, at $t = 0$ the aerodynamic lift is practically zero, so banking would not affect the altitude. When the lift becomes larger, banking results in a smaller lift component in the vertical (symmetry) plane, which does have an effect on the altitude. After $t = 100$ s the height has, in fact, become the only state component driving these eigenmotions (hence the name change to *height mode*). This is actually strange, since these modes originate from the lateral rotational motion and they have turned into modes linked with translational motion. The height modes also consist of one stable and one unstable mode, and represent (again) a very slow motion, and ultimately one with an amplitude half time almost equal to the flight duration ($T_{\frac{1}{2}} = 267.8$ s).

A final remark on the height modes is that in principle one cannot have a change in height unless there is also a change in velocity, i.e., exchange of potential and kinetic energy. However, most of the time the height modes are so slow that also the velocity change is marginal, even though this effect is still there. Only in case of the height/velocity mode with an amplitude half time below 100 s, the motion is fast enough to show. The coupling is a bit weaker towards the end of the flight, even though $T_{\frac{1}{2}} \approx 5$ s. The reason may be that, because the nominal velocity has reduced significantly, the aerodynamic drag (through the dynamic pressure) is also much smaller, and will have less influence on the velocity when the altitude (and thus the density) changes.

In summary, the eigenmotion of the entry capsule presents some fundamental differences with the winged re-entry vehicle discussed earlier. The longitudinal and lateral motion show a stronger coupling, and also include a coupling with the translational motion. These periodic modes are at least an order of magnitude faster compared to the corresponding modes of the winged re-entry vehicle discussed in the previous section, in fact, as fast as the short-period oscillation of conventional aircraft. However, other than the frequency of the oscillation there are no further similarities.

Table 6.11 Apollo eigenvalues and corresponding characteristic values for time point $t = 50$ s

	"Longitudinal" oscillation		"Lateral" oscillation		Bank mode	Bank mode	Height oscillation		Height mode
λ_i Re	$-3.9798 \cdot 10^{-07}$		$-1.2851 \cdot 10^{-4}$		$-4.5155 \cdot 10^{-6}$	$1.3373 \cdot 10^{-4}$	$2.4081 \cdot 10^{-4}$		$-3.8067 \cdot 10^{-4}$
Im	$\pm 4.0473 \cdot 10^{-2}$		$\pm 4.5483 \cdot 10^{-2}$		–	–	$\pm 1.4881 \cdot 10^{-3}$		–
P (s)	155.3		138.2		∞	∞	4,222.3		∞
$T_{\frac{1}{2}}$ (s)	1,741,700		5,393.9		153,500	5,183.1	2,878.4		1,820.9
ζ (–)	$9.8333 \cdot 10^{-6}$		$2.8254 \cdot 10^{-3}$		–	–	$-1.5975 \cdot 10^{-1}$		–
ω_n (rad/s)	$4.0473 \cdot 10^{-2}$		$4.5483 \cdot 10^{-2}$		–	–	$1.5074 \cdot 10^{-3}$		–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	z	z	θ (°)	z
ΔV	0.0088	–89.6	0.0010	–0.6	0.0035	0.0019	0.0009	169.5	0.0001
$\Delta\gamma$	$<10^{-4}$	–52.1	$<10^{-4}$	–89.7	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	84.9	$<10^{-4}$
ΔR	**0.0731**	**–129.7**	**0.8395**	**180.0**	**0.9986**	**0.9996**	**1.0000**	**0.0**	**1.0000**
Δp	$< 10^{-4}$	29.3	0.0010	91.5	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
Δq	**0.0297**	**90.0**	0.0012	179.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
Δr	0.0017	178.4	**0.0243**	**90.5**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\alpha$	**0.7332**	**0.0**	**0.0275**	**88.9**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\beta$	**0.0332**	**–91.6**	**0.4999**	**–179.5**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\sigma$	**0.6745**	**178.6**	**0.2097**	**–7.5**	**0.0522**	**0.0284**	$<10^{-4}$	0.0	$<10^{-4}$

Table 6.12 Apollo eigenvalues and corresponding characteristic values for time point $t = 100\,\mathrm{s}$

	"Longitudinal" oscillation		Lateral oscillation		Height mode	Height mode	Height oscillation		Height mode
λ_i Re	$-1.4960\cdot10^{-5}$		$-1.5326\cdot10^{-4}$		$-4.8072\cdot10^{-6}$	$1.4592\cdot10^{-4}$	$1.4339\cdot10^{-4}$		$-2.3837\cdot10^{-4}$
Im	±1.4492		1.6539		–	–	$\pm1.4875\cdot10^{-3}$		–
P (s)	4.3		3.8		∞	∞	4,224.0		∞
$T_{\frac{1}{2}}$ (s)	46,333.0		4,522.8		144,190	4,750.2	4,834.1		2,907.9
ζ (–)	$1.0323\cdot10^{-5}$		$9.2665\cdot10^{-5}$		–	–	$-9.5951\cdot10^{-2}$		–
ω_n (rad/s)	1.4492		1.6539		–	–	$1.4944\cdot10^{-3}$		–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	z	z	θ (°)	z
ΔV	**0.2252**	**–180.0**	0.0008	–90.0	0.0030	0.0015	0.0009	173.7	$<10^{-4}$
$\Delta\gamma$	$<10^{-4}$	–178.9	0.0001	180.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	86.9	$<10^{-4}$
ΔR	0.0404	91.3	**0.6478**	**90.0**	**1.0000**	**1.0000**	**1.0000**	**0.0**	**1.0000**
Δp	$<10^{-4}$	–90.0	0.0264	0.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
Δq	**0.7116**	**0.0**	0.0008	90.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
Δr	0.0010	90.0	**0.6488**	**0.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\alpha$	**0.4910**	**–90.0**	0.0005	–0.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\beta$	0.0005	180.0	**0.3640**	**90.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\sigma$	**0.4473**	**90.0**	**0.1618**	**–90.2**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$

Table 6.13 Apollo eigenvalues and corresponding characteristic values for time point $t = 150\,\mathrm{s}$

	"Longitudinal" oscillation		Lateral oscillation		Height mode	Height mode	Height oscillation		Height mode
λ_i Re	$-7.8191 \cdot 10^{-4}$		$-1.1030 \cdot 10^{-3}$		$-1.3343 \cdot 10^{-5}$	$2.2203 \cdot 10^{-4}$	$-6.9734 \cdot 10^{-4}$		$2.1600 \cdot 10^{-4}$
Im	$\pm 1.0113 \cdot 10^{1}$		$1.1589 \cdot 10^{1}$		–	–	$\pm 8.6942 \cdot 10^{-4}$		–
P (s)	0.6213		0.5422		∞	∞	7,226.8		∞
$T_{\frac{1}{2}}$ (s)	886.5		628.4		51,947.0	3,121.8	994.0		3,209.0
ζ (–)	$0.7314 \cdot 10^{-5}$		$9.5176 \cdot 10^{-5}$		–	–	$6.2568 \cdot 10^{-1}$		–
ω_n (rad/s)	$1.0113 \cdot 10^{1}$		$1.1589 \cdot 10^{1}$		–	–	$1.1145 \cdot 10^{-3}$		–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	z	z	θ (°)	z
ΔV	**0.3035**	**–180.0**	0.0002	–90.0	0.0001	0.0002	0.0009	–114.1	0.0002
$\Delta\gamma$	$<10^{-4}$	–179.8	0.0002	180.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	130.1	$<10^{-4}$
ΔR	0.0091	90.2	**0.1410**	**90.0**	**1.0000**	**1.0000**	**1.0000**	**0.0**	**1.0000**
Δp	$<10^{-4}$	–90.1	**0.0400**	**0.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
Δq	**0.9444**	**0.0**	0.0002	90.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
Δr	0.0002	90.0	**0.9854**	**0.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\alpha$	**0.0934**	**–90.0**	$<10^{-4}$	–0.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	180.0	**0.0787**	**90.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$
$\Delta\sigma$	**0.0848**	**90.0**	**0.0355**	**–90.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0	$<10^{-4}$

Table 6.14 Apollo eigenvalues and corresponding characteristic values for time point $t = 200\,\mathrm{s}$

	"Longitudinal" oscillation		Lateral oscillation		Height mode	Height mode	Height/Velocity mode	Height mode	Height mode
λ_i Re	$-3.1788 \cdot 10^{-3}$		$-4.1027 \cdot 10^{-3}$		$-2.4495 \cdot 10^{-6}$	$4.8550 \cdot 10^{-4}$	$-1.2517 \cdot 10^{-2}$	$-1.2593 \cdot 10^{-3}$	$1.3609 \cdot 10^{-3}$
Im	$\pm 1.7019 \cdot 10^{1}$		$\pm 1.9502 \cdot 10^{1}$		–	–	–	–	–
P (s)	0.4		0.3		∞	∞	∞	∞	∞
$T_{\frac{1}{2}}$ (s)	218.1		169.0		282,970	1,427.7	55.4	550.4	509.3
ζ (–)	$1.8678 \cdot 10^{-4}$		$2.1037 \cdot 10^{-4}$		–	–	–	–	–
ω_n (rad/s)	$1.7019 \cdot 10^{1}$		$1.9502 \cdot 10^{1}$		–	–	–	–	–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	z	z	z	z
ΔV	**0.3051**	**–180.0**	0.0001	90.0	$<10^{-4}$	0.0001	**0.3605**	0.0001	0.0001
$\Delta\gamma$	$<10^{-4}$	–179.9	0.0002	0.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
ΔR	0.0051	90.1	**0.0826**	**–90.0**	**1.0000**	**1.0000**	**0.9328**	**1.0000**	**1.0000**
Δp	$<10^{-4}$	–90.1	0.0404	–180.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δq	**0.9493**	**0.0**	0.0001	–90.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δr	0.0001	90.0	**0.9944**	**180.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\alpha$	**0.0558**	**–90.0**	$<10^{-4}$	180.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	180.0	**0.0472**	**–90.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\sigma$	**0.0507**	**90.0**	0.0213	89.9	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$

Table 6.15 Apollo eigenvalues and corresponding characteristic values for time point $t = 250\,\mathrm{s}$

	"Longitudinal" oscillation		Lateral oscillation		Height mode	Height mode	Height/Velocity mode	Height mode	Height mode
λ_i Re	$-1.9744 \cdot 10^{-2}$		$-1.9696 \cdot 10^{-2}$		$2.6770 \cdot 10^{-4}$	$2.5881 \cdot 10^{-3}$	$-1.3210 \cdot 10^{-1}$	$-4.6706 \cdot 10^{-3}$	$5.6919 \cdot 10^{-4}$
Im	$\pm 1.5534 \cdot 10^{1}$		$\pm 1.5936 \cdot 10^{1}$		–	–	–	–	–
P (s)	0.4		0.4		∞	∞	∞	∞	∞
$T_{\frac{1}{2}}$ (s)	35.1		35.2		2,589.2	267.8	5.2	148.4	1,217.8
ζ (–)	$1.2710 \cdot 10^{-3}$		$1.2359 \cdot 10^{-3}$		–	–	–	–	–
ω_n (rad/s)	$1.5534 \cdot 10^{1}$		$1.5936 \cdot 10^{1}$		–	–	–	–	–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	z	z	z	z
ΔV	**0.1810**	**0.5**	0.0008	90.4	$<10^{-4}$	0.0002	**0.3071**	0.0004	0.0001
$\Delta\gamma$	0.0001	–0.2	0.0008	0.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
ΔR	0.0002	–111.1	**0.0460**	**–90.1**	**1.0000**	**1.0000**	**0.9517**	**1.0000**	**1.0000**
Δp	0.0001	90.1	**0.0411**	**–179.8**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δq	**0.9797**	**180.0**	0.0021	–90.1	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δr	0.0002	–91.8	**0.9961**	**180.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\alpha$	**0.0631**	**90.1**	0.0001	180.0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\beta$	$<10^{-4}$	–8.0	**0.0595**	**–90.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\sigma$	**0.0591**	**–90.1**	0.0217	89.4	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$

Table 6.16 Apollo eigenvalues and corresponding characteristic values for time point $t = 275\,\mathrm{s}$

	"Longitudinal" oscillation		Lateral oscillation		Height mode	Height mode	Height/Velocity mode	Height mode	Height mode
λ_i Re	$-1.9600\cdot10^{-2}$		$-1.4059\cdot10^{-2}$		$1.1446\cdot10^{-4}$	$-1.3974\cdot10^{-3}$	$-1.3847\cdot10^{-1}$	$-2.9344\cdot10^{-2}$	$5.8057\cdot10^{-5}$
Im	±8.1447		±7.4652		–	–	–	–	–
P (s)	0.8		0.8		∞	∞	∞	∞	∞
$T_{\frac{1}{2}}$ (s)	35.4		49.3		6,055.6	496.0	5.0	23.6	11,939.0
ζ (–)	$2.4064\cdot10^{-3}$		$1.8833\cdot10^{-3}$		–	–	–	–	–
ω_n (rad/s)	8.1448		7.4652		–	–	–	–	–
$\boldsymbol{\mu}_i$	z	θ (°)	z	θ (°)	z	z	z	z	z
ΔV	0.0544	–178.8	0.0012	–89.0	$<10^{-4}$	0.0003	**0.1559**	0.0071	$<10^{-4}$
$\Delta\gamma$	0.0002	178.8	0.0010	–179.8	$<10^{-4}$	$<10^{-4}$	0.0001	0.0001	$<10^{-4}$
ΔR	0.0002	–10.3	0.0267	89.8	**1.0000**	**1.0000**	**0.9878**	**1.0000**	**1.0000**
Δp	0.0002	–90.0	**0.0439**	**–179.7**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δq	**0.9845**	**0.0**	0.0043	89.3	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Δr	0.0024	89.4	**0.9896**	**180.0**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\alpha$	**0.1209**	**–89.9**	0.0006	–0.5	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\beta$	0.0003	179.4	**0.1278**	**–89.9**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
$\Delta\sigma$	**0.1149**	**89.5**	**0.0412**	**90.3**	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$

6.5.4 Huygens Characteristic Motion

To conclude the discussion on eigenmotion of entry capsules, we will briefly look at the natural motion of Huygens, the planetary probe that entered the Titan atmosphere back in 2005. Huygens was an uncontrolled, ballistic probe with stable aerodynamic characteristics down to a Mach number of about $M = 1.5$ at which point its parachute was deployed. Prior to entry it was spun up to an angular roll rate of 10 RPM ($p = 60°/s$). The current analysis will show the effect of this spin rate on the eigenvalues of the vehicle. The eigenmotion will not be discussed in great detail, as it will not bring new insight compared to the discussion in the previous section about Apollo. However, as the Huygens probe was spinning, its effect on the eigenvalues will be illustrative.

The reader is reminded that the Huygens mission has been discussed in detail in Chap. 5, Sect. 5.3.4. Moreover, in Chap. 9, the complete parachute descent phase will be introduced. For either analysis, the probe's characteristics as summarised in Appendix B.3, have been used, and will also be the basis for the current discussion. The reference trajectory has been obtained from a non-linear simulation, which included a rotating central body with a central-field gravity, and the nominal Yelle atmosphere model. Wind was not considered, and the atmosphere was considered to be co-rotating with Titan. Initial conditions were selected to be $h = 1{,}250\,\text{km}$, $\tau = 0°$, and $\delta = 45°$, with a relative velocity of $V = 6{,}040\,\text{m/s}$, $\gamma = -64°$, and $\chi = 0°$ (due north). In any case, the probe's initial attitude is $\alpha = \beta = \sigma = 0°$, whereas the spin rate (= angular roll rate) $p = 60°/\text{s}$. Because of the non-zero p, the bank angle will continuously increase (but is scaled in between $\pm 180°$). In the linear analysis, because $\alpha = 0°$, it holds that $\dot{\sigma} = p$. Stop condition for all simulations is $M = 1.5$, which is the moment the parachute descent phase is initiated and also the moment the aerodynamic configuration of Huygens became more and more unstable.

In Fig. 6.33 the eigenvalues have been plotted. Translational motion starts out as a marginally unstable motion, but becomes more and more stable during entry. Note that the label "full" means that the eigenvalues have been computed from the full 9×9 **A**-matrix; as the eigenvalues from the 3×3 sub-matrix for translational motion are overlapping, this shows that the translational and rotational eigenmotion are decoupled. This only holds, because the lift force is zero at all times ($\alpha = 0°$), and cannot rotate due to to the spin. For non-zero angle of attack there will be a stronger coupling; however, this is not further analysed.

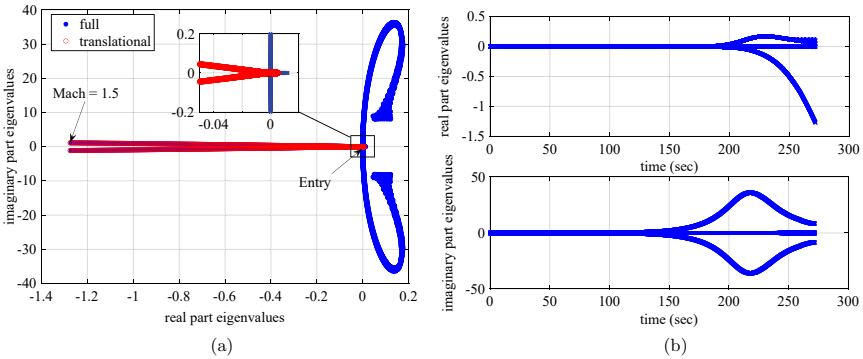


Fig. 6.33 Huygens' eigenvalues for coupled translational and rotational motion (probe spin $p = 60°/s$). Evaluation done for discrete time points

The eigenmodes of longitudinal and lateral rotational motion are more or less overlapping, due to the (almost) rotational symmetry.[16] Throughout entry, the eigenmodes are unstable albeit with relative low amplification, especially up to $t \approx 200\,s$. At that moment, the complex pair becomes more unstable and also its period rapidly increases. However, after some 60 s the reverse occurs. Towards the end of entry, the eigenvalues seem to jump around a little, which is caused by a combination of the particular aerodynamic characteristics and the spin rate. Figure 6.34 gives a more detailed perspective on this. This figure shows the overlaying decoupled longitudinal and lateral modes, and the eigenvalues from the 6×6 matrix representing the coupled rotational motion. The (inertial) coupling terms are obviously the responsible terms for the kind of zig-zag pattern, as they do not appear in the decoupled patterns. Also, while being decoupled the longitudinal and lateral eigenmodes are clearly overlapping, but the coupled results shows, in fact, two slightly shifted patterns. Finally, the two aperiodic (lateral) modes close to the origin, i.e., one stable and one unstable, have shrunken to an indifferent point (almost) at the origin.

From the linearised model, most notably Eqs. (6.260), (6.263), and (6.266) the inertial coupling between the different rotational modes can be discerned. To see the actual effect of having a spin rate, we redo the analysis, but now with $p = 0°/s$, the result of which is shown in Fig. 6.35. This plot confirms the earlier observations, as the longitudinal and lateral eigenmodes appear basically decoupled and show no irregularities towards the end of entry. It also shows that the aperiodic lateral modes have disappeared from the full-matrix results, in the sense that they have converged to (more or less) a single point.

To conclude the discussion on Huygens' eigenmotion, we have executed a few non-linear open-loop six-degrees-of-freedom simulations to address the free-fall

[16] The moments of inertia I_{yy} and I_{zz} are slightly different, and there is a small offset in centre-of-mass location in Y- and Z-direction, which introduces (small) non-zero products of inertia.

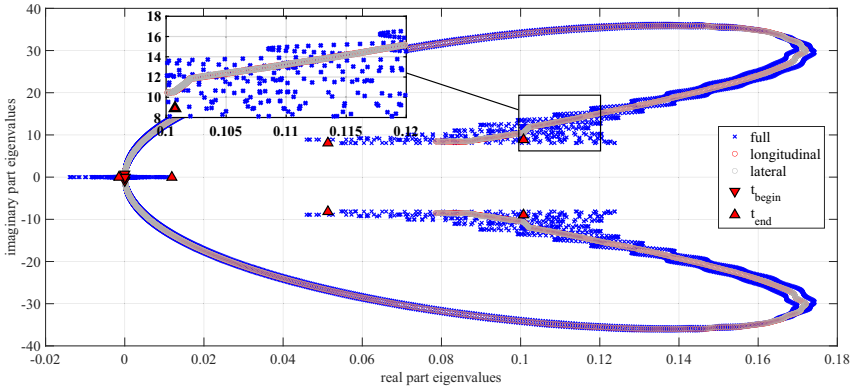


Fig. 6.34 Huygens' eigenvalues for rotational motion (probe spin $p = 60°/s$). Evaluation done for discrete time points

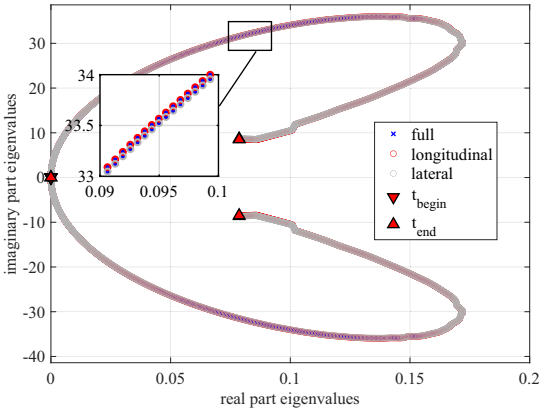


Fig. 6.35 Huygens' eigenvalues for coupled rotational motion (no probe spin). Evaluation done for discrete time points

rotational motion. Figure 6.36 shows the angles of attack and sideslip for the two case of either having spin or not, with the corresponding angular-rate components plotted in Fig. 6.37. Due to the (small) inertial coupling, there is an exchange of rotational energy between all three axes, for both spin cases. When there is no spin the unstable behaviour shows as a slowly diverging angle of attack, followed by the angle of sideslip. This slow increase takes place in the upper, thin layer of the atmosphere, where the dynamic pressure is small, which is also the reason that the oscillatory nature of the modes does not show. After reaching a certain minimum level of dynamic pressure, the motion becomes a little more dynamic and the the periodic nature of the eigenmodes shows. In any case, α and β stabilise at a non-zero value due to the offset in centre-of-mass location, similarly to what we saw for Apollo: the Z-offset gives rise to an α_{trim}, whereas the Y-offset results in a β_{trim}. The

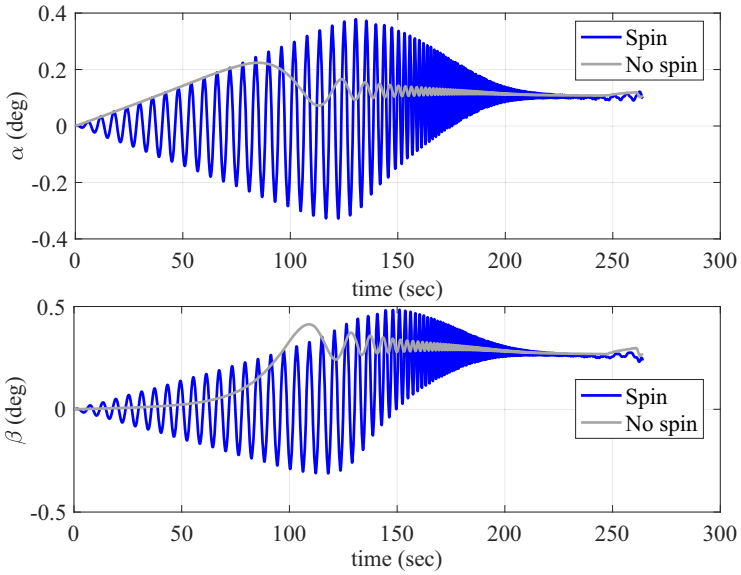


Fig. 6.36 Huygens' open-loop attitude motion during entry until parachute deployment: angles of attack and sideslip

picture changes a bit when the vehicle has a non-zero spin. Even though the amplitude for α is the same, due to the spin α oscillates with a period that resembles that of the eigenmotion. When the dynamic pressure increases the period of oscillations gets smaller and smaller. Due to the dynamic damping in the system the amplitude reduces.

It is noted that despite the increasing real part of the eigenvalues after $t \approx 200\,\mathrm{s}$, still the amplitude decreases. This counter-intuitive effect is caused by a decreasing dynamic pressure, peaking just before $t = 200\,\mathrm{s}$, which has a stabilising (dynamic) effect on the oscillatory motion.

The last simulation executed is one where the initial angle of attack is perturbed with an offset of $\Delta\alpha = 10°$. Initially, the angle of attack oscillates in between $\pm 10°$, with the same open-loop period as before (Fig. 6.38). When the probe is spinning, there is a continuous exchange between α and β, contrary to the non-spinning case, where β effectively remains zero (but is oscillating with the very small amplitude, also seen in Fig. 6.36). In either case, the probe stabilises at its trim values close to zero, and the oscillations have effectively damped out, despite the unstable motion predicted by the eigenvalues. Without trying to generalise, it can be concluded for this particular analysis that the study of eigenvalues is not complete without open-loop non-linear simulations. Dynamic stability and other transient (= time dependent) effects may play an important role, especially when the eigenvalues only have small positive real parts.

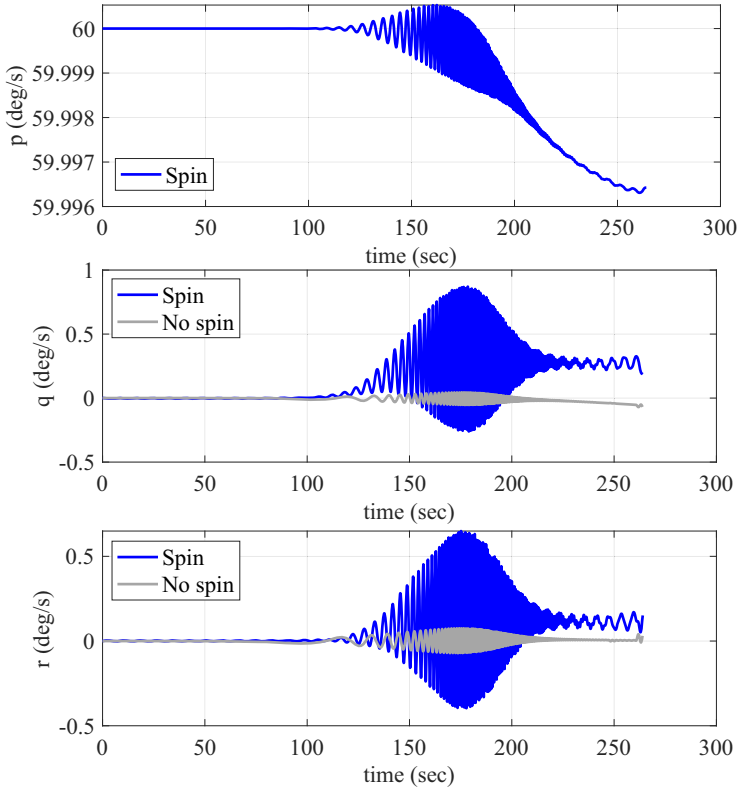


Fig. 6.37 Huygens' open-loop attitude motion during entry until parachute deployment: angular-rate components

Problems

Judge whether the following statements are TRUE or FALSE (Problems 6.1 to 6.9).

6.1 The eigenmotion of a system can only be determined for Linear Time Invariant systems.

6.2 Supersonic and hypersonic winged vehicles typically exhibit similar eigenmotions to subsonic aircraft due to their similar aerodynamic characteristics.

6.3 Eigenmotions of supersonic and hypersonic winged vehicles are primarily dominated by longitudinal modes, such as pitch and heave, rather than lateral modes.

6.4 Eigenmotions refer to the natural modes of oscillation or motion exhibited by a vehicle without external perturbations.

6.5 Eigenmotions of supersonic and hypersonic winged vehicles may involve coupling between different degrees of freedom, such as pitch-roll or pitch-yaw coupling.

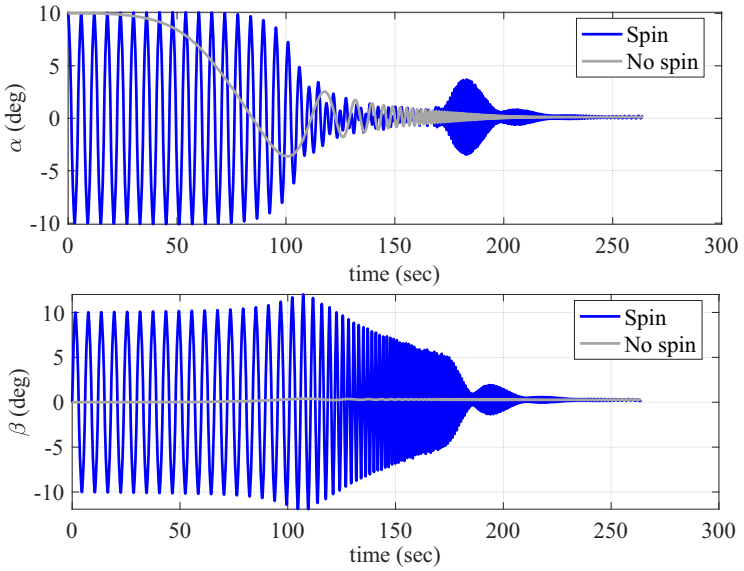


Fig. 6.38 Huygens' response to an angle-of-attack error of $\Delta\alpha = 10°$

6.6 Eigenmotions of supersonic and hypersonic winged vehicles are less influenced by aerodynamic forces compared to subsonic aircraft.
6.7 Analysing eigenmotions of hypersonic winged vehicles is primarily important for understanding structural dynamics rather than aerodynamic performance.
6.8 Eigenmotions of supersonic and hypersonic winged vehicles are typically unstable and more difficult to control due to their high-speed nature.
6.9 The eigenvalues of a non-linear system describe the characteristic motion of this system.

Linearisation and eigenmotion of entry capsules (Problems 6.10 to 6.14).

The open-loop flight behaviour of an entry capsule is characterised by very strong oscillations around all three axes. Individual components of motion are hard to distinguish, because of a strong dynamic coupling, partially the result of an offset in the location of the centre of mass in Z direction. This offset gives a product of inertia I_{xz} that is too large to be ignored.

The pitch motion is given by:

$$\dot{q} = \frac{M_y}{I_{yy}} + \frac{I_{xz}}{I_{yy}}\left(r^2 - p^2\right) + \frac{I_{zz} - I_{xx}}{I_{yy}} pr \tag{6.274}$$

$$\dot{\alpha} \approx q - \frac{L}{mV} \tag{6.275}$$

The external moment M_y contains an aerodynamic moment $\mathcal{M}$ and a reaction-control moment $M_{T,y}$.

6.10 List *all* state variables $\mathbf{x}$ and control variables $\mathbf{u}$ in Eqs. (6.274)–(6.275), even though remaining state equations are not shown here.

6.11 Linearise Eqs. (6.274)–(6.275). Write your answer in the form $\Delta\dot{\mathbf{x}} = \mathbf{A}\Delta\mathbf{x} + \mathbf{B}\Delta\mathbf{u}$ as two scalar equations. Consider the following assumptions:

- Both pitch moment $\mathcal{M}$ and lift L are a function of the angle of attack, α, and Mach number, M.
- For the rotational motion considered, the atmospheric properties are constant.
- Due to the oscillatory nature of the rotational motion, the nominal angular rates p_0, q_0 and r_0 cannot be considered small.
- The Mach number is *not* a state and has to be properly linearised.
- Translational and rotational motion are *not* decoupled.
- Consider *all* state and control variables (even though remaining state equations are not shown here!)

6.12 To analytically study the open-loop pitch motion, we make some simplifications. In that case, the state-space model is given by:

$$\Delta\dot{q} = \frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref} \Delta\alpha + \frac{\Delta M_{T,y}}{I_{yy}}$$

$$\Delta\dot{\alpha} = \Delta q - \frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref} \Delta\alpha$$

Derive an analytical expression for the eigenvalues related to the longitudinal oscillation.

6.13 A numerical representation of the eigenvalues found under (b) is the complex pair $\lambda_{1,2} = -1.9744 \cdot 10^{-2} \pm 1.5534 \cdot 10^{1} i$. You are asked to calculate the *dimensional* values of the damping factor, ζ, period, P, and amplitude half (or double) time, $T_{\frac{1}{2}}$ or T_2, depending on the nature of the eigenmotion. First state the used equation, then your numerical value.

6.14 Is the mode under Problem 6.13 stable or unstable? Explain why.

Kinematic attitude equations formulated with Euler angles φ, θ, and ψ (Problems 6.15 to 6.16).

Alternatively, the attitude motion of the entry capsule can be evaluated with

$$\dot{\varphi} = p + \sin\varphi \tan\theta q + \cos\varphi \tan\theta r \tag{6.276}$$

$$\dot{\theta} = \cos\varphi q - \sin\varphi r \tag{6.277}$$

$$\dot{\psi} = \frac{\sin\varphi}{\cos\theta} q + \frac{\cos\varphi}{\cos\theta} r \tag{6.278}$$

6.15 Linearise the above equations.
6.16 Formulate the above results in state-space form.

References

Abramowitz M, Stegun IA (eds) (1972) Handbook of mathematical functions with formulas, graphs, and mathematical tables, 10th printing. US Government Printing Office, Washington
Donaldson BK (2006) Introduction to structural dynamics. Cambridge University Press
D'Souza AF (1988) Design of control systems. Prentice Hall Inc., Englewood Cliffs, New Jersey
Duke EL, Antoniewicz RF, Krambeer KD (1988) Derivation and definition of a linear aircraft model. NASA Reference Publication 1207
Etkin B (1972) Dynamics of atmospheric flight. Wiley, New York
Kreyszig E (2011) Advanced engineering mathematics, 10th edn. Wiley, New Jersey
Kuo BC (1987) Automatic control systems. Prentice Hall Inc., Englewood Cliffs, New Jersey
MBB Space Communication and Propulsion Systems Division (1988) Study on re-entry guidance and control, ESA report reference: ESA CR (P) 2652
Meirovitch L (2001) Fundamentals of vibrations. McGraw-Hill, New York
Mooij E (1998) Space plane flight dynamics. Analysis of Guidance and Control Concepts, PhD Thesis, Delft, The Netherlands
Mooij E (1995) The Horus-2B reference vehicle, Memorandum M-692, Delft University of Technology
Moseley Jr, WC, Moore Jr, RH, Hughes JE (1967) Stability characteristics of the Apollo command module, NASA-TN-D-3890
North American Aviation, Inc., "Aerodynamic Data Manual for Project Apollo", NASA-CR-82907, 1 January 1965
Ogata K (2004) System dynamics, 4th edn. Pearson Prentice Hall, New Jersey
Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes: the art of scientific computing, 3rd edn. Cambridge University Press
Sachs G (1993) Stability and control problems in hypersonic flight - Paper 35. Space Course Munich
Trujillo BM (1986) Determination of Lift and drag characteristics of Space Shuttle Orbiter using maximum likelihood estimation technique, AIAA-1986-2225. In: AIAA atmospheric flight mechanics conference. Williamsburg, Virginia
Yonemoto K, Inatani Y (1988) Analytical interpretation on lateral-directional stability and controllability of high angle-of-attack re-entry flight, IAS-report no. 630
Vinh NX (1981) Optimal trajectories in atmospheric flight. Elsevier Scientific Publishing Company, Amsterdam

Chapter 7
Guidance, Navigation, and Control



7.1 Introduction

Two important aspects of the entry and descent of space vehicles are the tactical aspect of having control over the time and location of landing, and the severe mechanical and thermal loading on the vehicle. For manned missions, for instance, the maximum deceleration is usually limited by an upper bound of 3 g to save the occupants from discomfort or worse. Also in case of unmanned (scientific) missions a too strong deceleration might be harmful to the on-board instruments. The thermal load, e.g., the maximum heat flux, the wall temperature distribution over the surface, and the integrated heat load, define to a large extent the design and also the mass of the thermal protection system. Obviously, we want to have the most favourable load on the vehicle. However, complete freedom over the trajectory is usually also not possible: a winged re-entry vehicle like the Space Shuttle is supposed to land on an air strip, so its trajectory should be targeted to the landing location right on from entry. In case of the parachute descent of the Apollo capsules, it was important that they would splash down in the ocean not too distant from the recovery ships.

During the process of mission analysis and mission design, starting from specified initial conditions an optimal trajectory is computed, which usually satisfies trajectory constraints (e.g., a maximum allowable thermal load) and end conditions (e.g., the landing place in case of a re-entry mission). Once this so-called nominal trajectory has been defined it must be verified that the vehicle can actually fly this trajectory, or, in other words, whether the vehicle can execute the required manoeuvres without violating any constraints. Furthermore, it must be guaranteed that the vehicle will still be able to fulfil its mission when it encounters (unforeseen) disturbances, which force the vehicle away from its nominal path. To ensure mission success, the space vehicle is equipped with a so-called guidance, navigation, and control system (GNC system), which has been conceptualised in Fig. 7.1.

The task of the guidance system is to generate steering commands, e.g., a commanded attitude or thrust level, taking a reference state, trajectory constraints and/or

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_7

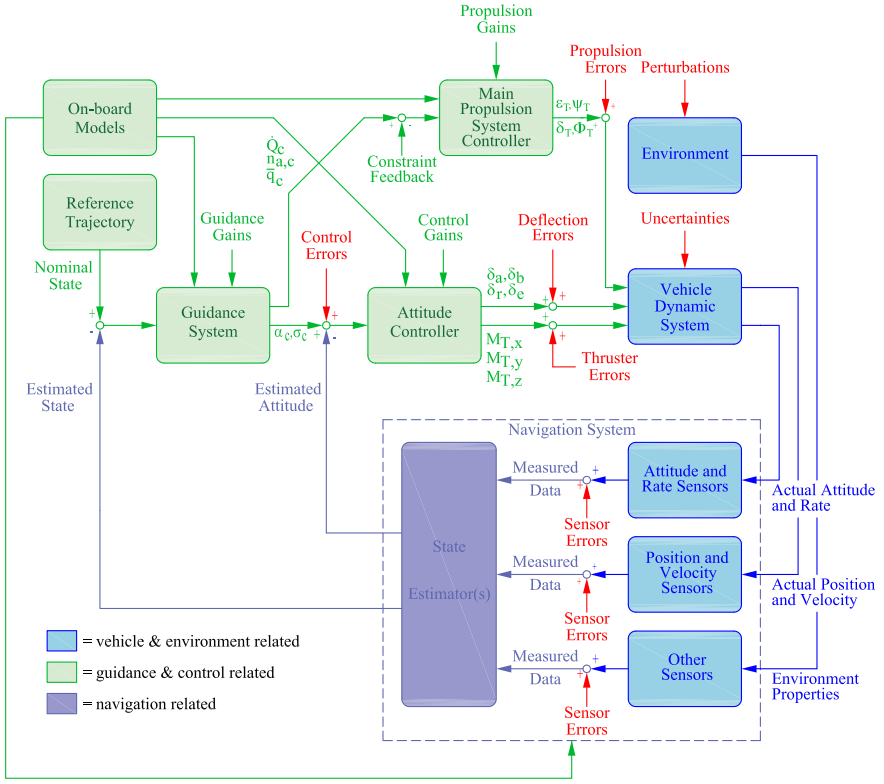


Fig. 7.1 Schematic overview of a typical GNC system

a final state into account. For this task, the system needs input from the outside world, for instance, the current actual state. These data have to be provided by the navigation system, using sensor information and predefined theoretical models. The control system has to guarantee that the steering commands are carried out, such that, for example, the actual attitude equals the commanded attitude in a reasonably short time and that this attitude is dynamically stable (trim stability). To achieve this, the control system may drive aerodynamic control surfaces, reaction-control wheels, and thrusters, etc.

Since in this book we do not only focus on unpowered re-entry systems, but also on (powered) test vehicles or vehicles that execute aero-assisted manoeuvres possibly supported by propulsion, the GNC system may also include a propulsion controller. This controller regulates the magnitude and orientation of the propulsion force, which becomes a major player in the force equilibrium in that case. A typical usage of the propulsion force could be to keep the vehicle safe from the trajectory constraints, such as heat flux, normal load and dynamic pressure. Another function could be to guarantee flight at a constant Mach number when aerodynamic means are no longer sufficient.

The design of a GNC system is centred around a nominal mission that is free from disturbances. The environment is modelled at a certain level of complexity only and also the description of the vehicle is of course not infinitely accurate. An important question is whether the GNC system will be able to steer and control the vehicle in the presence of all kinds of uncertainties that it is bound to encounter during the actual mission. One way to assure this is to design a very robust system.[1] But the next question that arises then is: how robust should our system be? A too robust system might negatively influence the flying qualities and manoeuvrability, and if the system is not robust enough, we can still end up with a severe control problem. Therefore, it is very important that already during the design process we study all possible uncertainties to see how our guidance and control system will deal with them. Usually what one does, once the GNC system has been developed for the nominal mission, is to simulate a number of test cases with different error sources included, and with all dynamic, vehicle and environment models as accurately as possible.

In Fig. 7.1 the lines between the blocks indicate data flows. At several places, errors may enter the system. For instance, a typical sensor measures (part of) the vehicle state. However, these sensors are not perfect, so sensor or measurement errors will be added to the true vehicle state. These errors are, at best, estimated by the state estimators, so an estimated state will be fed back to the guidance and control system. The control system commands the actuators, which are also not perfect, so at that point actuator errors in, for instance, the control-surface deflection or the thruster-performance parameters will enter the system.

The underlying notion of the guided and controlled flight of a re-entry vehicle is that the control forces can be manipulated such that their magnitude and direction result in a required acceleration or deceleration. To use the aerodynamic and propulsion force for guidance and control purposes, it is necessary to study their origin in more detail. The aerodynamic force, with as main components lift and drag, is basically a function of the aerodynamic shape of the vehicle, its velocity and attitude with respect to the incoming flow, and atmospheric properties such as air pressure and temperature. So to change lift and drag we can, for instance, vary the aerodynamic shape by deflecting control surfaces, although the effect is not so large unless the control surfaces are large. An efficient and fast way to control the magnitude (and at a slower time scale also the direction) of the aerodynamic force is by changing the vehicle's attitude and thus to make use of the aerodynamic dependency on the angle of attack. In addition, the vehicle can be rolled around its velocity vector to change the orientation of the lift component, a concept that is known as banking.

This means that in principle we can establish two guidance variables: the angle of attack (α_c) and bank angle (σ_c) to control the aerodynamic force. Note that the subscript c stands for *commanded* (by the guidance system). Usually a separate

[1] Robustness is defined as the particular property that a control system must possess for it to operate properly in realistic situations. Mathematically, it means that the controller must not only work for the system that it has been designed for, but for a whole family of (similar) systems. If the controller is stable for a set of systems whose parameters deviate substantially from the nominal system, then the controller is said to have robust stability.

(but interfaced) controller is used to execute them: an attitude-control system. The operating principle is simple: generate a moment such that an attitude change is initiated and guarantee that this attitude is stable once the commanded attitude has been reached. The moments can be acquired with several means, e.g.:

- Aerodynamic-control surfaces; depending on the vehicle, these surfaces may consist of canards, rudder(s), elevators, ailerons, elevons (combined elevator and aileron function) and body flap(s), to name but a few. A deflection of these surfaces results in a change of the aerodynamic shape. The induced forces are usually small, but since the moment arms can be large, also the moments can be significant. Of course, the control surfaces are only efficient when the dynamic pressure is sufficiently high.
- Reaction-control jets; in case the control surfaces are not efficient to guarantee a fast response, e.g., because the atmosphere is too thin (or even absent), cold- or hot-gas thrusters (also called reaction-control system) can be used to generate moments around the three body axes.
- Moving mass points; by displacing a mass inside the vehicle, relative to the c.o.m., a gravitational moment is induced. It should be noted that when the vehicle has a large inertia and roll damping, either large displacements or a large mass is required to generate any moments of significance.

Of course there may be other means of control—a well-known example is the parachute system, which is in principle a large drag actuator—so the above listed actuators only serve as an example.

To discuss the key elements of the schematic of Fig. 7.1 we simplify it to the one shown in Fig. 7.2. As can be seen, the only "external" input provided to the system is originating from the *Mission Manager*. This input may include guidance and control commands in any possible form, i.e., including switching logic, manoeuvre sequencing, target information, trajectory and steering constraints, etc. The control of the entry vehicle is divided into three parts, analogous with aircraft guidance, navigation, and control. The output of the Mission Manager serves as input to the *Guidance Logic*, which purpose is to guide the motion of the centre of mass of the vehicle by, for instance, determining the magnitude of the lift vector or the instance when to initiate a bank manoeuvre. The second input channel is formed by the output of the *State Estimator*, which gives information about the real-life performance. This performance is, on one hand, measured by *Sensors*, and on the other hand derived by a state estimator. The difference between the commanded and measured state is also known as the state error for guidance, or guidance error. Note that we can refer to the state estimators (in combination with some of the functionality of the Mission Manager) as the *Navigation System*. Commonly used sensors for entry vehicles are accelerometers and gyroscopes, pressure sensors and GPS receivers. Other sensors that are used, also for planetary landers, are star trackers, horizon sensors and lidar (light detection and ranging[2]), to name but a few.

[2] An alternative acronym is *laser imaging, detection, and ranging*.

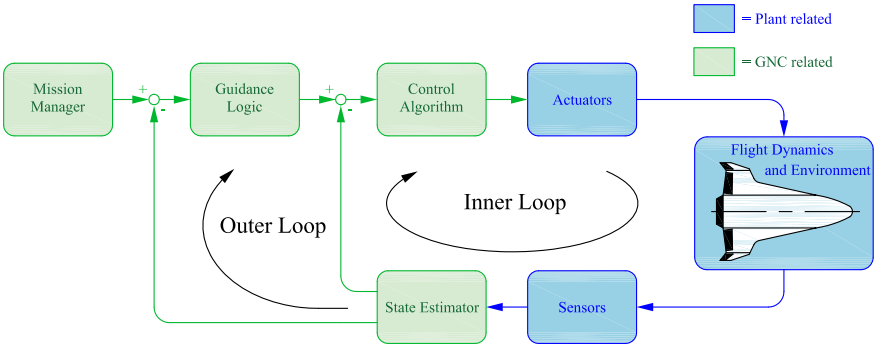


Fig. 7.2 Simplified schematic of a GNC system

The output of the guidance logic serves as input to the actual *Control Algorithm*, that determines the motion *around* the centre of mass. Also in this case, the second input channel comes from the navigation system, in this case the measured state is of course related to the control variables. Again, the difference between the commanded (by the guidance logic) and measured (by the sensors) state is known as the state error for control, or control error. The control algorithm determines the actuator commands based on some control concept, e.g., a combination of P(roportional), I(ntegral) and D(erivative) manipulation of the control error, adaptive, fuzzy-logic, or H_∞ control, or yet another concept. The *Actuators* are the actual hardware elements that convert the control commands to forces and moments. Typical examples for entry vehicles are control surfaces, reaction control thrusters, parachutes and retro rockets. The latter typically assist during the landing, when the atmosphere is too thin or even absent, or when the landing loads have to be significantly reduced for the occupants' comfort. Note that any actuator-specific control logic should not be put in the control-algorithm sub-system (high-level control), but in the actuator sub-system (low-level control). As mentioned before, the output of the actuator sub-system is a control force and moment, which is input to *Flight Dynamics and Environment*. This block is actually the entry vehicle in its environment, and it is the hardware element to be guided and controlled. It will behave according to the laws of physics under the influence of forces and moments. In a simulation environment this part would have to be modelled, of course. Output of this block is a vehicle state, consisting of position, velocity, attitude and angular rate, and any state-derived parameters, such as a heat flux or a normal load.

So, summarised, the sensors and actuators are the interfaces with the vehicle, the former measuring the state and the latter changing it. In combination with state estimators and the mission manager, the navigation system provides the related input to the guidance logic, which in turn provides the control system with the right commands. As both the navigation system (input side) and control system (output side) are interfacing with hardware (sensors and actuators) the guidance logic is "nothing else" but a mathematical entity in between, albeit as crucial as all the other elements.

Usually the control loop is executed at a higher frequency than the guidance loop, because the rotational dynamics are typically an order of magnitude faster than the translational dynamics. Because the control loop is also associated with stability of the complete system, the control loop is usually known as the inner loop, whereas the guidance loop is called the outer loop.

Based on this simplified schematic, it is relatively easy now to structure this chapter. In Sect. 7.2 the sensors, commonly used for entry systems, are discussed. These are the Inertial Measurement Unit (IMU) and Global Positioning System (GPS), though a brief description of other sensors will be given as well. Section 7.3 introduces the navigation system, or to be more precise, it touches on the topic of filters that could be used as state estimators. A typical example of flight-path reconstruction is included as a case study. The next Sect. 7.4 gives an overview of the most commonly used actuators, focussing on aerodynamic control surfaces and reaction-control thrusters; also included is a brief discussion about other actuators that are—or could be—used in entry vehicles. The control algorithms that drive the actuators are discussed next (Sect. 7.5). Classical control theory and the so-called Proportional-Integral-Derivative (PID) controllers are covered, as well as the more general concept of (optimal) state and output feedback. An example design of a controller for a winged re-entry vehicle concludes that section. The final section of this chapter, i.e., Sect. 7.6 covers the system that links navigation with control: the guidance system. This discussion is primarily based on the guidance system of the Space Shuttle, as it is a representative system. Finally, an example of the tracking of a pre-defined heat-flux profile is included.

7.2 Sensors

In this section some of the most commonly used sensors on entry, descent, and landing vehicles will be discussed. In line with the nature of this text, the discussion will not so much focus on the physical aspects of the sensor, but more on the application. Then in particular on how one could model the sensors, combine them with state estimators, and include them in navigation systems.

Intuitively, it may be obvious that flying in an atmosphere will open up more possibilities to measure (part of) the state of the vehicle. One can think of, for instance, pressure sensors that the navigation system can use to relate atmospheric pressure with altitude. On the other hand, this same atmosphere may also hinder the proper sensing of, for example, signals of the GPS, because the propagation of these signals is perturbed. GPS measurements are typically used in Earth-bound missions, so a landing on Mars would require different sensors to obtain the same state information.

A good alternative when atmospheric measurements are not possible are the different radar systems that are available, although these systems could also be used in the atmosphere. Star trackers and Sun sensors can also be used in the atmosphere, but the success of doing so is (partially) depending on the opacity of this atmosphere. Other aspects of the planetary environment that can be used are, for instance, the

magnetic field. Measuring the magnetic field lines with a magnetometer allows one to estimate the attitude of the vehicle, provided, of course, that an accurate on-board model of the magnetic field is available.

The Earth provides many more navigation means, of course, because of the extensive infrastructure that is available to track satellites and to guide aircraft to a safe landing. The Space Shuttle, for instance, used three tactical air navigation units (TACAN) and three microwave scan beam landing systems for the approach and landing phase. In 1993, though, the Space Shuttle Program selected an off-the-shelf GPS receiver to eventually replace the three TACAN units on each space shuttle orbiter. The flight certification took until 2002 before the Shuttles started flying with one GPS receiver, in combination with the TACAN units (Goodman 2003).

In the remainder of this section we will pay particular attention to two commonly used navigation sensors. The first one is the IMU, which can measure linear and angular motion, and can thus provide us with the means to estimate position, velocity, attitude and angular rates. The IMU is discussed in Sect. 7.2.1. This section also includes an overview of the errors that are associated with these measurements and how these errors can be modelled. The next "sensor" is the GPS (Sect. 7.2.2), although by itself GPS is not a sensor, but rather a signal generator that allows sensors to use these signals to estimate the translational and rotational state of the vehicle. In the last Sect. 7.2.3 a brief discussion about other sensors is included.

7.2.1 Inertial Measurement Unit

An IMU is a very useful instrument for navigation and it is also referred to as an Inertial Navigation System (INS). Inertial navigation systems were originally developed for rockets. American rocket pioneer Robert Goddard experimented with rudimentary gyroscopic systems. This idea inspired Wernher von Braun to include them in the V-2. Most other measurement techniques require some external reference to be measured. Navigation using IMUs, however, is in theory fully self-sufficient. Thus, no knowledge about the surroundings is necessary. In reality this is not the case, and not only because of the presence of instrument errors that are to be discussed shortly. An IMU also needs an estimate of its initial position and velocity (and attitude and angular rate) from another source (e.g., a human operator or a GPS receiver) before it can start integrating its measurements. Note that an IMU allows in principle for the open-loop propagation of position, velocity, attitude and angular rates. The larger the uncertainty in the initial conditions, the worse this propagation will be. Also, the instrument errors contribute to a diverging solution of the propagation.

IMUs consist of a set of accelerometers and gyroscopes. The accelerometers measure the (non-conservative) acceleration with respect to the ambient acceleration due to gravity. Thus, a set of three non-collinear accelerometers can measure the acceleration experienced in all three dimensions. Gyroscopes are capable of measuring angular velocity. The physical effects used by these two measurement devices can be seen in Table 7.1.

Table 7.1 Basic inertial sensor technologies, as stated by Grewal et al. (2001)

Sensor	Gyroscope			Accelerometer		
Physical effect used	Conservation of angular momentum	Coriolis effect	Sagnac effect	Gyroscopic precession	Electro-magnetic force	Strain under load
Sensor implementation methods	Angular displacement	Vibration	Ring laser	Angular displacement	Drag cup	Piezo-electric
	Torque rebalance	Rotation	Fiber optic	Torque rebalance	Electro-magnetic	Piezo-resistive

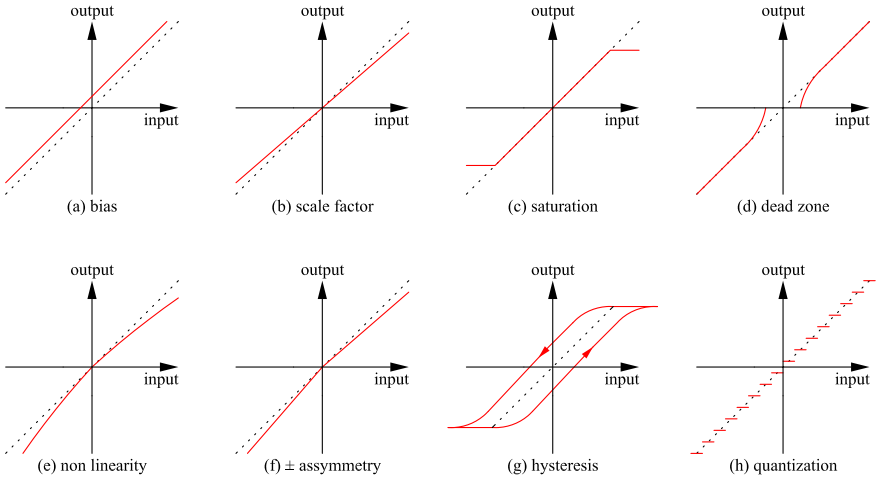


Fig. 7.3 Common errors found in IMUs. (The dotted line is reality, while the solid line is the behaviour due to the error)

No sensor can measure with infinite precision, so also measurements from the accelerometers and the gyroscopes are not perfect. There is a number of errors that cause the measurement to differ from the true value at that point. These errors have been schematised in Fig. 7.3, and are:

(a) *Bias* is a constant error independent from the underlying specific force or angular rate that adds an offset to the true output.

(b) *Scale Factor* is caused by tolerances in manufacturing or due to aging, and adds an error to the output proportional to the specific force or angular rate.

(c) *Saturation* occurs when the dynamic range of the measured quantity exceeds the capabilities of the sensor, the output will be either the minimum or maximum value.

(d) *Dead zone* is a limited region where no measurements can be done usually occurring due to stiction (which is like static friction), or lock-in (e.g., gimbal lock).

(e) *Non-linearity* is always present in electric sensors, and is expressed as the variation of the scale factor over the operating range of the sensor.

(f) *Sign asymmetry* occurs in conjunction with scale factor, and is caused due to push-pull amplifiers that are mismatched (i.e., the difference in response due to either positive or negative current in the electronic circuit).

(g) *Hysteresis* represents the dependency of the sensor not only on the current excitation, but also on its past. This dependency arises, because the system can be in more than one internal state. If a given input alternately increases and decreases, the output tends to form a loop.

(h) *Quantisation* is the result of using a finite number of digits to represent the (analogue) measurements. As a consequence, the accuracy of all digital measurements is limited and the output changes step-wise.

(i) Not shown, but nonetheless a common instrument error: *Misalignments* or cross-coupling errors arise from the misalignments (i.e., non-orthogonality) of the sensitive axes; usually they are combined with the scale factors in a single error matrix.

A realistic simulator of just the hardware would try to incorporate as many of the errors shown in Fig. 7.3 as possible. This may be useful for the manufacturers of the IMU, but not practical for the end users, as not all parameters associated with these errors will be available. These users should preferably have a simpler model for ease of use and to facilitate the analysis of the performance, which is of practical use in re-entry studies. An additional simplification could be to assume that the accelerations and angular velocities of the various individual accelerometers and gyroscopes are all given in the same body-fixed reference frame. When the output of all the devices is combined, a potential error could then be misalignments of the three axes.

Thus, in navigation studies one usually begins with a simplified model for the accelerometers and the gyroscopes, where only the scale factors s, bias b, and the misalignments m, are modelled. For the accelerometers, this model is given by

$$\mathbf{a}_m = \mathbf{a} + \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} + \begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix} \mathbf{a} + \begin{bmatrix} 0 & m_{xy} & m_{xz} \\ m_{yx} & 0 & m_{yz} \\ m_{zx} & m_{zy} & 0 \end{bmatrix} \mathbf{a} = \mathbf{b_a} + (\mathbf{I} + \mathbf{S_a})\,\mathbf{a} \tag{7.1}$$

with

$$\mathbf{S_a} = \begin{bmatrix} s_x & m_{xy} & m_{xz} \\ m_{yx} & s_y & m_{yz} \\ m_{zx} & m_{zy} & s_z \end{bmatrix} \tag{7.2}$$

In Eq. (7.1), the acceleration, $\mathbf{a}$, is actually the difference between the real acceleration and the acceleration due to gravity expressed in the body frame (index B). The bias has a certain value when the accelerometers are switched on, but it also tends to have a random walk. A random walk is, as the name suggests, a mathematical formalisation of a trajectory that consists of taking successive random steps. To express a random walk as a differential equation, one can write:

$$\begin{pmatrix} \dot{b}_x \\ \dot{b}_y \\ \dot{b}_z \end{pmatrix} = \begin{pmatrix} w_{bx} \\ w_{by} \\ w_{bz} \end{pmatrix} \tag{7.3}$$

with w_{bx}, w_{by}, and w_{bz} being normally distributed white-noise parameters. A similar formulation holds for the misalignments and scale factors, since they are also commonly modelled as random walks.

The gyroscopes can be modelled in a manner comparable to that of the accelerometers. The measurement equation is:

$$\boldsymbol{\omega}_m = \boldsymbol{\omega} + \begin{pmatrix} b_p \\ b_q \\ b_r \end{pmatrix} + \begin{bmatrix} s_p & m_{pq} & m_{pr} \\ m_{qp} & s_q & m_{qr} \\ m_{rp} & m_{rq} & s_r \end{bmatrix} \boldsymbol{\omega} = \mathbf{b}_\omega + (\mathbf{I} + \mathbf{S}_\omega)\,\boldsymbol{\omega} \tag{7.4}$$

The bias, also known as gyroscope drift, behaves similarly to the one of the accelerometer

$$\begin{pmatrix} \dot{b}_p \\ \dot{b}_q \\ \dot{b}_r \end{pmatrix} = \begin{pmatrix} w_{bp} \\ w_{bq} \\ w_{br} \end{pmatrix} \tag{7.5}$$

The instrument errors are the cause that the measurements from the IMU are not fully trustworthy. This means that some other kind of measurement device will have to used for high-precision applications, so that the biases and other errors can be accounted for. This other device could be a star tracker or GPS. On-board calibration of the error parameters is one of the tasks of an integrated navigation system. This means that before the actual mission starts, a best estimate of the instrument errors should be obtained. With aiding GPS measurements, the integrated navigation system should be able to accurately estimate not only position, velocity, and attitude state variables, but also the above unknown instrument errors. However, since the observability of those parameters also depends on system inputs, some specific manoeuvres for the re-entry vehicle may have to be designed. Such manoeuvres are not always possible or feasible, though, unlike those for aircraft. All of the manoeuvres have to be performed before the actual re-entry starts, since the so-called blackout phase of the GPS receiver during the re-entry will occur due to atmosphere ionisation and it may otherwise jeopardise the mission. The effect of the blackout on external sensors—all RF transmissions to and from the vehicle are suppressed—is still difficult to assess, and conservatively, the vehicle has to be considered autonomous during this phase. Unfortunately, the vehicle manoeuvrability is maximum, thus one has to rely on the IMU as the core of the navigation solution to offer the necessary autonomy.

7.2.2 Global Positioning System

The GPS is a constellation of satellites maintained by the United States military that has between 24 and 32 satellites occupying five orbital planes. These orbital planes can be seen in Fig. 7.4. The operational altitude of the GPS satellites is approximately 20,200 km.

The GPS satellites are basically beacons floating in space. If distances from a point to three separate satellites is known, it is in theory possible to find the position in three-dimensional space. In reality, the situation is more complex due to errors that will be elaborated further. The distance to the beacon is based on the time-of-flight principle of a signal from the GPS satellite to the user. The signal sent by a GPS satellite contains the following information:

1. Time of transmission
2. Ephemeris of the transmitting satellite, i.e., position and velocity
3. Almanac with some information about the other satellites.

The time of transmission is important, because the time of flight can only be found from the difference between the time of transmission and the time of reception. The ephemeris of the transmitting satellite allows the location of the "beacon" to be fixed. The general almanac allows the user to track other GPS satellites quicker.

As mentioned previously, GPS navigation is not as simple as just using three satellites. The main reason is that because the signal travels at the speed of light, the time of flight needs to be found to a very high level of accuracy. Most GPS receivers are not capable of measuring the time neither to the required precision, nor to the required accuracy. Thus, time has to be one of the variables that has to be solved for.

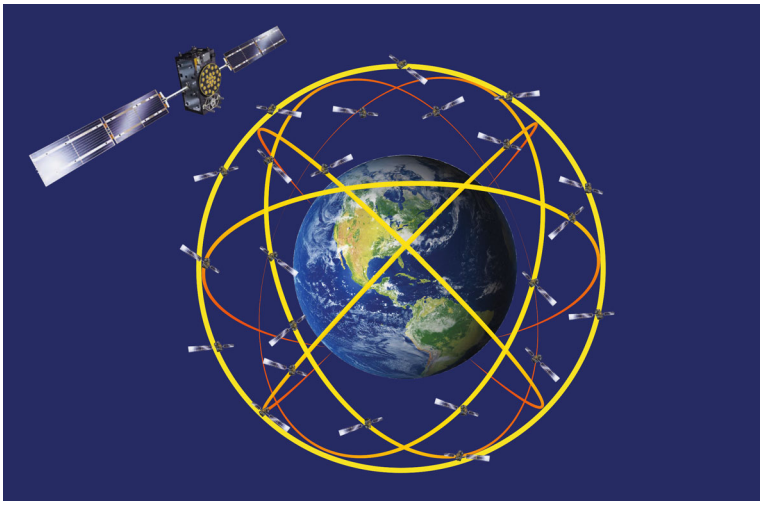


Fig. 7.4 The five orbital planes occupied by GPS satellites

This adds one more variable to the three position coordinates and consequently four satellites are required to determine the position of the user. Apart from determining the position, the frequency of the carrier signal can be used to determine the velocity due to the Doppler effect. Also, the phase of the signal can be used for very precise applications, like attitude determination.

Errors in Signals

Unfortunately, the GPS signals are not perfect, as there is some attenuation from different sources. For instance, there is always some noise present in the receivers. In case the signals travel through some medium like the atmosphere, the speed of light changes. It is also possible that there is some uncertainty in the ephemeris of the GPS satellite. This translates into the position of the beacon not being fixed, and thus, any estimated position would also be wrong.

The atmospheric delays are a problem not only for users on Earth, but also for spacecraft. The ionosphere of the Earth begins at an altitude of 1000 km and thus, the GPS signals to satellites below this altitude will all be affected by it. For re-entry applications, the troposphere, which begins at an altitude of 40 km, also affects the signals. There are models available that can be used to estimate the signal path through the different atmosphere strata at different conditions. A few examples are the *Saastamoinen* model and the *Hopfield* model that can be used for the troposphere (Farrell and Barth 1999). For the ionosphere, the delay differs for different frequencies of the GPS signal. Thus, if dual-frequency measurements are used, this delay can be accounted for. The standard deviation of the noise due to the atmosphere is usually around 5 m. This and all subsequent standard-deviation values are taken from Misra and Enge (2006). These errors in the pseudo-range, which is the range from the satellite to the receiver, are for receivers at ground level and will be different for spacecraft in orbit.

At the receiver, there is the noise at the receiving antenna. There is also a problem known as *multipath*. Multipath occurs when the GPS signal reaches the receiver via more than one path. This can occur for a spacecraft if the signal also bounces off a solar panel before reaching the antenna. The standard deviation of the error due to these two noise sources is usually in the order of 1 m.

The final category of errors are due to the GPS satellites themselves. The position of the GPS satellite and the time are known to a very high accuracy, but there is still some uncertainty here. The standard deviation of the error due to this is approximately 3 m. Another contribution to the positioning uncertainty is that the quality of the position determination is dependent on distribution of the GPS satellites in view of the user. A good distribution of GPS satellites gives a better fix than if the satellites are all in one plane or lined up behind each other. The uncertainty corresponding to this distribution is known as *Dilution of Precision* (DOP). GPS computations are normally carried out in the north-east-zenith coordinate system and this easily translates into a horizontal and vertical position. This is a very useful classification for users who are on Earth. DOP can then be split into a *Vertical Dilution of Precision* (VDOP), *Horizontal Dilution of Precision* (HDOP), and *Geometrical Dilution*

of Precision (GDOP). HDOP is a measure of the uncertainty in the north and east directions, and VDOP is a measure for the zenith direction. GDOP is a combination of HDOP and VDOP, and is a measure of the uncertainty in the position. The uncertainty in the time of the clock can be expressed with the *Time Dilution of Precision* (TDOP). These parameters were very useful before, when GPS receivers were only capable of tracking four GPS satellites, but now GPS receivers are able to track all satellites in view. In case only a limited number of satellites needs to be tracked, then a configuration can be chosen that minimises the chosen DOPs.

GPS Observable Models and Characteristics

GPS receivers reconstruct either the code, the carrier, or both of an incoming satellite signal. The code-tracking loop correlates the incoming code with a code replica generated by the receiver till maximum correlation is achieved. Then, the phase of the code replica is a measurement of the received code phase at the receiver clock time. In the carrier-tracking loop the relative phase between the reconstructed carrier signal and a reference signal generated by the receiver is measured. Both the code and the carrier-phase measurements can be treated as observations of the satellite time at the corresponding time of phase transmission. Differences between the two phases are:

1. The code phase can be measured without ambiguity, while the carrier phase is ambiguous;
2. The carrier phase will be advanced in the ionosphere: the ionospheric refraction correction term has a negative sign due to phase velocity;
3. The code phase will be delayed in the ionosphere: the ionospheric delay has a positive sign due to group velocity;
4. The noise level of the measurements is different. For example the noise of P-code measurements is about 2 ns (0.6 m) while the noise of the carrier phase is about 0.01 ns (3 mm).

For re-entry vehicle navigation GPS carrier-phase measurements will be applicable only for attitude determination, since the accuracy of code phase measurements is sufficient for position determination. Several basic GPS observables for re-entry vehicle position, velocity and attitude determination can now be identified (Mulder et al. 1999):

1. pseudoranges from code phase measurements;
2. Doppler shift from the carrier frequency; and
3. carrier phase or carrier-phase difference from different GPS receiver antennas.

The pseudo-range ρ_i, which is the distance from the transmitting ith GPS satellite to the user, can be computed in the following way

$$\rho_i = c\left(t_r - t_{s_i}\right) = \sqrt{\left(\mathbf{r}_{sv,i} - \mathbf{r}\right)^{\mathrm{T}}\left(\mathbf{r}_{sv,i} - \mathbf{r}\right)} + c \cdot b_r + v_{\rho i} \tag{7.6}$$

where

c = speed of light, which is 299,792,458 m/s
t_r = receiver clock time
t_{s_i} = broadcast clock time of the ith GPS satellite
$\mathbf{r_{sv,i}}$ = $(x_{sv,i}, y_{sv,i}, z_{sv,i})^{\mathrm{T}}$ is the position of the ith GPS satellite
$\mathbf{r}$ = $(x_R, y_R, z_R)^{\mathrm{T}}$ is the user or receiver position in Earth-centred, Earth-fixed (ECEF) co-ordinates
b_r = the bias of the receiver clock
$v_{\rho i}$ = white measurement noise.

From pseudorange measurements to four satellites the user position x_R, y_R, z_R and receiver clock bias b_r can be determined.

It is well known that, if an observer has a non-zero velocity component in the line-of-sight direction with respect to a radio source (which is, in this case, also moving) transmitting a signal of a certain frequency, the observer will receive a signal with a frequency not equal to the transmitted frequency. The difference between received and transmitted frequencies is called the Doppler shift of the signal. Many navigation systems are based on this effect. The relationship between the Doppler shift (Δf) and the range rate $\dot{\rho}$ is:

$$\Delta f = f_s - f_r = \frac{\dot{\rho} f_s}{c} \tag{7.7}$$

where

Δf = the frequency change due to the Doppler shift
f_s = the frequency of the original transmitted signal and is 1575.42 MHz for the L_1 signal
f_r = the frequency of the received signal
$\dot{\rho}$ = is the range rate
c = the speed of light.

The range rate equation with respect to the ith GPS satellite can be written as

$$\dot{\rho}_i = \frac{\left(\mathbf{r}_{sv,i} - \mathbf{r}\right)^{\mathrm{T}} \left(\dot{\mathbf{r}}_{sv,i} - \dot{\mathbf{r}}\right)}{\sqrt{\left(\mathbf{r}_{sv,i} - \mathbf{r}\right)^{\mathrm{T}} \left(\mathbf{r}_{sv,i} - \mathbf{r}\right)}} + c \cdot d_r + v_{\dot{\rho}i} \tag{7.8}$$

where

$\dot{\mathbf{r}}_{sv,i}$ = the velocity of the GPS satellite
$\dot{\mathbf{r}}$ = the velocity of the user (in ECEF components)
d_r = the receiver clock drift
$v_{\dot{\rho}i}$ = the white measurement noise of the ith satellite.

In the case of carrier-phase measurements (difference in carrier phase between a master antenna and one or more slave antennas), the observation equation (in cycles) has the following form:

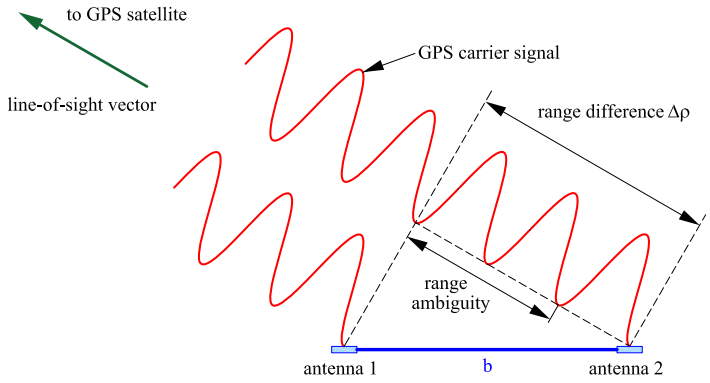


Fig. 7.5 Carrier phase difference measurement using two GPS receiver antennae

$$\Phi_m(t_r) = \Phi_{CR}(t_r) - \Phi_0(t_r) - \frac{f_s}{c}(\rho - \dot{\rho} b_r) + N + v \tag{7.9}$$

where $\Phi_{CR}(t_r)$ is the received carrier phase, $\Phi_0(t_r)$ is the reference phase, both at receiver time t_r, N is the integer ambiguity, which needs to be resolved first and, finally, v is the measurement noise.

In real applications, the derivation of the above observable is extended to single and double difference schemes to eliminate observation bias and integer ambiguity (Chu and Van Woerkom 1997).[3] As the accuracy of carrier-phase measurements is much higher than the accuracy of code-phase measurements, this observable is widely applied to precise positioning, navigation, satellite geodesy, and attitude determination, the latter of which will be used in the current discussion. It is assumed that the integer ambiguity can be solved correctly and is therefore not included in the current model.

The measurement concept of carrier-phase difference can be seen in Fig. 7.5. The range difference is just the projection of the baseline vector **b** (directed from the master to the slave), onto the user-to-GPS satellite line-of-sight unit vector $\mathbf{e_i}$, i.e.:

$$\Delta\rho_{s_i} = \mathbf{b}^{\mathrm{T}}\mathbf{e_i} \tag{7.10}$$

The line-of-sight unit vector $\mathbf{e_i}$ in Eq. (7.10) can be calculated from user and ith GPS satellite position, and may be expressed in the ECEF frame. Therefore, a transformation from this reference frame to the spacecraft body-fixed reference frame is

[3] The range difference is caused not only by the phase difference, but also due to a range ambiguity. The range ambiguity is caused by an integer number of wavelengths occurring in range difference, ΔN. There is also a bias known as the electrical line bias, β.

necessary. The line-of-sight vector from the user position to a GPS satellite is given by:

$$\mathbf{r_{LOS,i}} = \mathbf{r_{sv,i}} - \mathbf{r} \tag{7.11}$$

To normalise this vector, it has to be divided by its norm, which is in this case equal to the pseudorange. So

$$\mathbf{e_{i,R}} = \frac{\mathbf{r_{LOS,i}}}{\sqrt{(\mathbf{r_{sv,i}} - \mathbf{r})^{\mathrm{T}}(\mathbf{r_{sv,i}} - \mathbf{r})}} \tag{7.12}$$

The GPS spacecraft position can be obtained from orbital parameters broadcast by the GPS spacecraft itself, while the position of the user (re-entry vehicle) is estimated from the measurements of pseudoranges and Doppler shifts (i.e., from the solution of the GPS position determination).

In practice, the double differencing technique is applied to carrier-phase observations for integer-ambiguity identification purposes; therefore, the line biases from multiple antennas are eliminated. Assuming that the integer ambiguity difference and the electrical line bias have been correctly estimated, the phase difference measurement equation (for satellite i and slave antenna j) can be written as:

$$\Delta\Phi_{ij} = -\frac{f_s}{c}\mathbf{b}^{\mathrm{T}}_{\mathbf{ij_B}}\mathbf{C_{B,I}}\mathbf{C_{I,R}}\mathbf{e_{i,R}} + \Delta N + \beta + v \tag{7.13}$$

where v is the measurement noise and the transformation matrix $\mathbf{C_{B,I}}$ contains the vehicle attitude information. So, rearranging Eq. (7.13) and isolating $\mathbf{C_{B,I}}$ will allow us to extract the attitude angles.

Unlike the position determination, attitude determination using carrier phase measurements needs in principle only two GPS spacecraft, as the rotation about the line of sight to the GPS spacecraft cannot be determined. Moreover, the line-of-sight vector is known from some other source. If the line-of-sight vector needs to be determined from the difference between GPS spacecraft positions and receiver position as well, in that case at least four GPS spacecraft are necessary to determine the attitude angles of the user spacecraft.

7.2.3 Other Sensors

Even though there are many sensor techniques and related instruments, in this section we will only mention a few due to space limitations. The discussion will focus on lidar, star trackers and sun sensors, magnetometers, and pressure sensors.

Lidar (light detection and ranging, sometimes also called ladar) is an optical remote sensing technology that can measure the distance to (or other properties of) a target by illuminating the target with light, often using pulses from a laser. The working principle of the so-called pulse lidar is based on measuring the time of flight

of a light pulse from an emitter to an object and back to determine position. It is sensitive to atmospheric effects such as dust and aerosols, which scatter the laser, thereby reducing the signal strength of the laser before it reaches the target. And the laser that is reflected off dust and detected by the receiver introduces noise. Lidar is minimally sensitive to temperature variation, which changes the path length inside the receiver and the clock-oscillator rate. Other error sources are the measurement error of the rotating mirror angle, specular (i.e., non-smooth) surfaces, and the clock error (similar to the GPS clock error). Other types of lidar include Doppler lidar, which uses a continuous beam to measure the speed differential of the target and the emitter/receiver. The measured frequency change of the reflected light indicates when the target and lidar device are moving closer together (blue shift) or moving apart (red shift).

The last variation we mention is the 3D-flash lidar, which readout semiconductors enable each pixel in the focal plane array to act independently and measure range and intensity of every pixel (point) in the camera's field of view. This sensor provides all required data types, i.e., altimetry, velocimetry, terrain relative navigation, hazard detection, and precision navigation when close to the surface. Even though there is no operational-mission heritage yet, the technology readiness level (TRL) of the system is advanced through other space-based lidar applications, e.g., the 3D flash lidar video camera has been tested on three Space-Shuttle missions, STS-127, STS-133, and and STS-134 (SpaceX 2011). Moreover, Amzajerdian et al. (2016) describe how such a system has been flight-tested on the rocket-propelled free-flyer vehicle, *Morpheus*, specifically with the objective of autonomous safe landing. Therefore, it will make an excellent candidate for future entry, descent-and-landing missions including autonomous landing and hazard avoidance, collision avoidance, and situational awareness. Further use could include autonomous rendezvous-proximity and docking/capture operations, with applications in asteroid sample return, satellite servicing, and active space-debris removal.

The recently launched Mars Science Laboratory (November 26, 2011) has besides an IMU also a so-called terminal descent sensor (TDS), which is a pulse Doppler radar design (see also Fig. 7.6). It produces line-of-sight range and velocity measurements on 6 independent beams. It has a single antenna per beam, with 3 beams canted 20° off nadir, 2 beams canted 50° off nadir, and 1 nadir beam, and employs Ka-band (35.75 GHz) centre frequency (San Martin 2010). Due to its dedicated design it is capable of measuring the terrain relative velocity with high precision, even in the presence of uncertain terrain slopes, thus fulfilling the low touchdown velocity requirements.

A variation of the lidar is the radar altimeter, amongst others used on the Space Shuttle. The two radar altimeters on board the orbiter measure absolute altitude with respect to the nearest terrain within the beam width of the orbiter's antennas. The radar altimeters constitute a low-altitude terrain-tracking and altitude-sensing system based on the precise time it takes for an electromagnetic energy pulse to travel from the orbiter to the nearest object on the ground below and back. It can operate during altitude rate changes of as much as 2,000 feet per second. This enables tracking of mountain or cliff sides ahead or alongside the orbiter if these obstacles are nearer than the ground below, and warns of rapid changes in absolute altitude.

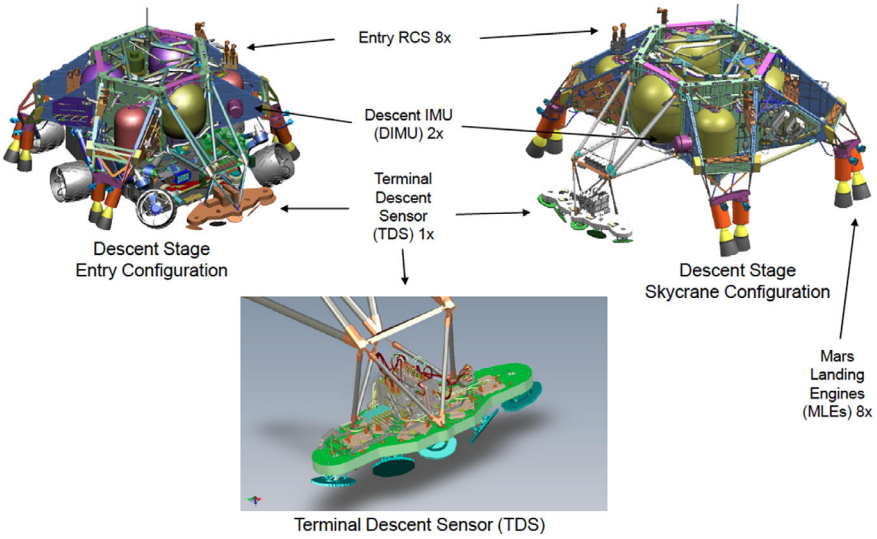


Fig. 7.6 Sensor and actuator configuration of the Mars Science Laboratory. Highlighted is the terminal descent sensor. Image: NASA (San Martin 2010)

A star tracker is an optical device that measures the positions of stars using photocells or a camera. The operation of the star trackers, which require high sensitivity, can become influenced by sunlight reflection, e.g., from the spacecraft, or by exhaust gas plumes from the RCS thrusters that contaminate the star-tracker window. Star trackers are also susceptible to a variety of errors (e.g., low spatial frequency, high spatial frequency, and temporal) in addition to a variety of optical sources of error (for instance, spherical and chromatic aberration). A typical star catalogue for high-fidelity attitude determination originates from a standard base catalogue and then filtered to remove problematic stars, for example, due to apparent magnitude variability or colour index uncertainty. These types of star catalogues can have thousands of stars stored in memory on-board the spacecraft. Use of star trackers in the atmosphere depends on the visibility of the stars, of course. A larger opacity of the atmosphere works favourable; for instance, a dense atmosphere like the one on Titan may prohibit the proper use of star trackers. In that case a Sun sensor, a device that senses the direction to the Sun, may be better due to the larger brightness of the Sun. A Sun sensor can be as simple as some solar cells and shades, or as complex as a steerable telescope, depending on mission requirements. Final remark: the Space Shuttle had two star trackers for celestial navigation (in particular during rendezvous and docking missions). Its navigation system could not use them during entry and descent, because the star trackers were shielded from the hot airflow by protective covers. This aspect should be taken into account when designing a navigation system for atmospheric flight.

A magnetometer is a device that senses magnetic field strength and, when used in a three-axis triad, magnetic-field direction. As a spacecraft navigational aid, the sensed field strength and direction is compared to a map of the Earth magnetic field stored in the memory of an on-board or ground-based guidance computer. If the spacecraft position is known then with the known mounting of the magnetometer attitude can be derived.

In atmospheric flight, pressure sensors can be used to derive altitude information if an accurate atmosphere model is stored on board. This information can, for instance, be used to initiate parachute deployment. The Space Shuttle has two air data probes, located on the left and right sides of the orbiter's forward lower fuselage. During the ascent, on-orbit, deorbit, and initial entry heat-load environment phases, for protection the probes are stowed inside the forward lower fuselage. The air data probe is covered by thermal protection system tiles while in the stowed position. At approximately Mach 3, the air data probes are deployed, such that they can provide information on the movement of the orbiter. It senses air pressures related to spacecraft movement through the atmosphere to update the navigation state vector in altitude, to provide guidance in calculating steering and speed-brake commands, and to update flight-control law computations and provide display data for the commander's and pilot's alpha Mach indicators, and altitude/vertical velocity indicators (Jenkins 2008).

A non-intrusive concept (versus the intrusive pitot tube) is the so-called flush airdata sensing (FADS) system concept, in which air data are inferred from nonintrusive surface pressure measurements. This systems was developed to circumvent many of the difficulties with intrusive air-data systems. This method does not require probing of the local flow field to compute air-data parameters, and allows for air-data measurements in flight applications where probe-based systems are no longer viable (Cobleigh et al. 1999). Originally tested up to Mach numbers of 5, the on-going development opened up hypersonic applications, amongst others for the X-33, X-34, and X-38 vehicles, although these vehicles are no longer under development. The FADS was wind-tunnel tested for calibration purposes, and the three calibration parameters are well-behaved from Mach 0.25 to Mach 5.0, an angle-of-attack range extending to greater than 30°, and an angle-of-sideslip range extending to greater than 15° (Cobleigh et al. 1999).

7.3 Navigation

In simple wording, navigation is concerned with the age-old question: where am I? To answer this question in general we need information from the surroundings, and also from the state in which we are and possibly were: where was I just moments before, how fast am I moving and in what direction? Theoretical equations of motion always give exact answers. In real life, however, not everything is so straightforward. When a system is modelled, there are always some discrepancies compared to real life. It is not possible to predict everything exactly. The goal of navigation is to

be able to accurately estimate the state of a spacecraft. The word *estimate* is very important here, because it is also not possible to exactly *measure* the state. Despite all their efforts, engineers are only able to estimate the state of the spacecraft, albeit with a very high accuracy. If the estimation does not have to be carried out in real time, a method such as least squares can be used. There are plenty of cases where it is necessary to get estimated parameters in real time. For this purpose, it is much more efficient to use recursive filtering techniques.

These filters can fuse measurement data from different sensors and give a best estimate of state variables of interest, thereby reducing the errors that are present in each of the individual sensor-data components. The field of filtering techniques is huge and we can only cover the basics of it. Therefore, we restrict to the original filter, i.e., the Kalman filter. Named after Rudolf Kálmán,[4] he is commonly acknowledged of being the first to develop this filter for linear systems (Kalman 1960). Therefore, we will discuss the linear Kalman filter (LKF) in Sect. 7.3.1. An often used extension to the linear Kalman filter is the extended Kalman filter (EKF), which provides a first-order correction for non-linear systems. This filter is considered to be the de facto standard in the theory of non-linear state estimation. The EKF will be discussed in Sect. 7.3.2. Further, in Sect. 7.3.3 the integration of different sensors is introduced. An actual application of this can be found in Sect. 7.3.4, where we discuss a possible implementation of the HORUS navigation system used for flight-path reconstruction.

7.3.1 Linear Kalman Filter

A Kalman filter is an extremely effective and versatile procedure for combining noisy sensor outputs to estimate the state of a system with uncertain dynamics. As the LKF is the original Kalman filter, it will be derived in this section. This derivation is based on the one found in Welch and Bishop (2001), but with some added detail. In fact, the derivation can be found in many text books on state estimation, e.g., Farrell and Barth (1999) and Groves (2008), to name but a couple.

The goal of a Kalman filter is to estimate the state of a linear system

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_{k+1} + \mathbf{w}_k \tag{7.14}$$

with measurements in the form of

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{7.15}$$

In Eqs. (7.14) and (7.15), $\mathbf{w}_k$ and $\mathbf{v}_k$ represent the system and measurement noise, respectively. They are Gaussian noises, so they have a normal distribution with zero mean. The variables $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{z}$ are the state, control input, and measurement,

[4] Rudolf Emil Kálmán (May 19, 1930–July 2, 2016) was an Hungarian-American electrical engineer, mathematician, and inventor.

respectively. **A** is the system matrix, in the field of navigation commonly called the state-transition matrix, and it defines how the state at the next step is related to the current state. **B** describes the relation between the input and the state at the next step. **H** is the measurement matrix and shows how the measurements can be computed (or *predicted*) from the states.

Both noises are assumed to be uncorrelated to each other. Their probability distributions are

$$\mathrm{p}(\mathbf{v}_k) \sim N(0, \mathbf{R}_k) \tag{7.16}$$

$$\mathrm{p}(\mathbf{w}_k) \sim N(0, \mathbf{Q}_k) \tag{7.17}$$

where **R** is the noise present in the measurement device and **Q** is the uncertainty present in the modelling of the system. **R** is usually available from the manufacturer of the measurement device. **Q**, on the other hand, can only be derived by testing, or from the design experience of the engineer.

The estimation process can be split into two steps. The first step is to get an a-priori estimate, and the second step is to get an a-posteriori estimate. The a-priori estimate only takes into account the system noise, this means that the next state is estimated using only the model. For navigation purposes, the time derivative of the state elements can sometimes be measured and then it also occurs in the state transition. This occurs, for example, when an IMU, or velocity measurements are used. In this case the measurement noise is incorporated into the system noise. The a-posteriori estimate then updates and refines the a-priori estimate using the measurement. The a-priori and a-posteriori estimates are expressed as $\hat{\mathbf{x}}_k^-$ and $\hat{\mathbf{x}}_k$ respectively. The a-priori and a-posteriori errors can be defined as

$$\mathbf{e}_k^- \equiv \mathbf{x}_k - \hat{\mathbf{x}}_k^- \tag{7.18}$$

$$\mathbf{e}_k \equiv \mathbf{x}_k - \hat{\mathbf{x}}_k \tag{7.19}$$

The error between the true state and the estimated state can be expressed as an uncertainty in terms of a covariance matrix. These error covariance matrices are

$$\mathbf{P}_k^- = \mathrm{E}\left[\mathbf{e}_k^- \mathbf{e}_k^{-T}\right] \tag{7.20}$$

$$\mathbf{P}_k = \mathrm{E}\left[\mathbf{e}_k \mathbf{e}_k^{\mathrm{T}}\right] \tag{7.21}$$

Since the a-priori estimate of the state is not perfect and there is noise present in the measurements, it is not possible to predict the measurements with full accuracy. To predict the measurement, noise is not taken into account as the expected value of noise is 0. The predicted measurement is

$$\hat{\mathbf{z}}_k = \mathbf{H}_k \hat{\mathbf{x}}_k^- \tag{7.22}$$

The difference between the actual measurement and the predicted measurement is known as the measurement innovation, $\delta\mathbf{z}_k^-$.

$$\delta\mathbf{z}_k^- = \mathbf{z}_k - \hat{\mathbf{z}}_k = \mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^- \tag{7.23}$$

Given an initial, or equivalently an a-posteriori estimate, the a-priori state estimate of the next state is found via

$$\hat{\mathbf{x}}_{k+1}^- = \mathbf{A}_k\hat{\mathbf{x}}_k + \mathbf{B}_k\mathbf{u}_k \tag{7.24}$$

The a-priori covariance can be found using the proper linear covariance propagation law.

$$\mathbf{P}_{k+1}^- = \mathbf{A}_k\mathbf{P}_k\mathbf{A}_k^{\mathrm{T}} + \mathbf{Q}_k \tag{7.25}$$

The aim of the Kalman filter is to improve the state estimation by optimally incorporating the information found in the measurement. The update can be expressed in terms of the sum of the a-priori state and the product of a gain with the measurement innovation.

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}\delta\mathbf{z}_k^- = \hat{\mathbf{x}}_k^- + \mathbf{K}\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-\right) \tag{7.26}$$

As can be seen, it is important to find the value of the gain, $\mathbf{K}$, which shall now be derived.

The a-posteriori error can be expanded as

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k^- - \mathbf{K}_k\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-\right) \tag{7.27}$$

where Eq. (7.26) has been substituted. Grouping terms yields

$$\mathbf{e}_k = -\left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\hat{\mathbf{x}}_k^- - \mathbf{K}_k\mathbf{z}_k + \mathbf{x}_k \tag{7.28}$$

and with the (noisy) measurement $\mathbf{z}_k$, Eq. (7.15), substituted this gives

$$\mathbf{e}_k = -\left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\hat{\mathbf{x}}_k^- - \mathbf{K}_k\left(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k\right) + \mathbf{x}_k \tag{7.29}$$

Further expanding the above equation results in

$$\begin{aligned}\mathbf{e}_k &= -\hat{\mathbf{x}}_k^- - \mathbf{K}_k\mathbf{H}_k\hat{\mathbf{x}}_k^- - \mathbf{K}_k\mathbf{H}_k\mathbf{x}_k - \mathbf{K}_k\mathbf{v}_k + \mathbf{x}_k \\ &= \left(\mathbf{x}_k - \hat{\mathbf{x}}_k^-\right) + \mathbf{K}_k\mathbf{H}_k\left(\hat{\mathbf{x}}_k^- - \mathbf{x}_k\right) - \mathbf{K}_k\mathbf{v}_k \\ &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^-\right) - \mathbf{K}_k\mathbf{v}_k\end{aligned} \tag{7.30}$$

Using the definition of the a-priori error from Eq. (7.19), the a-posteriori error can thus be written as

$$\mathbf{e}_k = \left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]\mathbf{e}_k^- - \mathbf{K}_k\mathbf{v}_k \tag{7.31}$$

The a-posteriori covariance matrix can be written as

$$\mathbf{P}_k = \mathrm{E}\left[\mathbf{e}_k \mathbf{e}_k^{\mathrm{T}}\right] = \left(\mathbf{I}_k - \mathbf{K}_k \mathbf{H}_k\right) \mathbf{P}_k^- \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right)^{\mathrm{T}} + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^{\mathrm{T}} \tag{7.32}$$

In the covariance matrix, the diagonal terms express the variance of the state elements. The off-diagonal terms express the covariance between the state elements. To make sure that the estimation is optimal, the variance has to be brought to a minimum value. A measure of the variance of all the state elements is the trace of the covariance matrix. Thus, the trace has to be brought to a minimum value. To get the trace, the relation for the covariance found in Eq. (7.32) has to be expanded.

$$\begin{aligned}
\mathbf{P}_k &= \mathrm{E}\left[\mathbf{e}_k \mathbf{e}_k^{\mathrm{T}}\right] = \left(\mathbf{I}_k - \mathbf{K}_k \mathbf{H}_k\right) \mathbf{P}_k^- \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right)^{\mathrm{T}} + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^{\mathrm{T}} \\
&= \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right) \mathbf{P}_k^- \left(\mathbf{I} - \mathbf{H}_k^{\mathrm{T}} \mathbf{K}_k^{\mathrm{T}}\right) + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^{\mathrm{T}} \\
&= \left(\mathbf{P}_k^- - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^-\right) \left(\mathbf{I} - \mathbf{H}_k^{\mathrm{T}} \mathbf{K}_k^{\mathrm{T}}\right) + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^{\mathrm{T}} \\
&= \mathbf{P}_k^- - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^- - \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^{\mathrm{T}}
\end{aligned}$$

The trace is

$$\begin{aligned}
\mathrm{Tr}\left(\mathbf{P}_k\right) = \; & \mathrm{Tr}\left(\mathbf{P}_k^-\right) - 2\mathrm{Tr}\left(\mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^-\right) + \mathrm{Tr}\left(\mathbf{K}_k \left(\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}}\right) \mathbf{K}_k^{\mathrm{T}}\right) \\
& + \mathrm{Tr}\left(\mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^{\mathrm{T}}\right)
\end{aligned} \tag{7.33}$$

To find the gain that minimises the trace of the covariance matrix, the derivative has to be 0:

$$\frac{\partial\left(\mathrm{Tr}\left(\mathbf{P}_k\right)\right)}{\partial \mathbf{K}_k} = \mathbf{0} \tag{7.34}$$

To compute the derivative, the following two trace calculation rules are used

$$\frac{\partial \mathrm{Tr}\left(\mathbf{M}_1 \mathbf{M}_2 \mathbf{M}_1^{\mathrm{T}}\right)}{\partial \mathbf{M}_1} = 2\mathbf{M}_1 \mathbf{M}_2 \tag{7.35}$$

which is only applicable for symmetric $\mathbf{M}_2$, and

$$\frac{\partial \mathrm{Tr}\left(\mathbf{M}_1 \mathbf{M}_2\right)}{\partial \mathbf{M}_1} = \mathbf{M}_2^{\mathrm{T}} \tag{7.36}$$

This results in the following condition for the gain

$$\frac{\partial \mathrm{Tr}\left(\mathbf{P}_k\right)}{\partial \mathbf{K}_k} = -2\left(\mathbf{H}_k \mathbf{P}_k^-\right)^{\mathrm{T}} + 2\mathbf{K}_k \left(\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}}\right) + 2\mathbf{K}_k \mathbf{R}_k = 0 \tag{7.37}$$

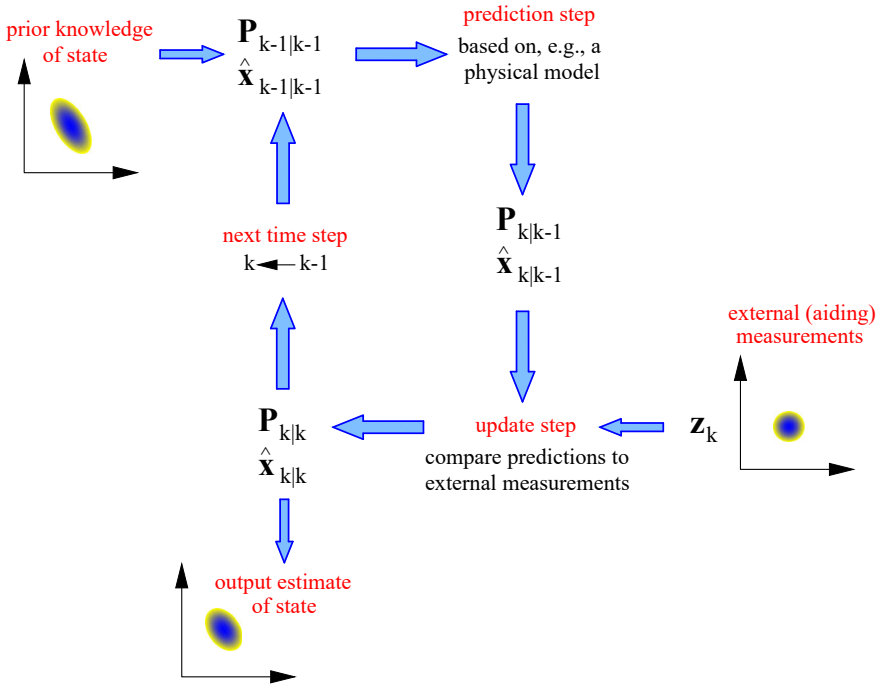


Fig. 7.7 Implementation of the linear Kalman filter

Equation (7.37) can then be rearranged to give the optimal gain.

$$-2\left(\mathbf{H}_k\mathbf{P}_k^-\right)^{\mathrm{T}} + 2\mathbf{K}_k\left(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\right) + 2\mathbf{K}_k\mathbf{R}_k = 0 \quad \Rightarrow$$

$$\mathbf{K}_k^{-1}\left(\mathbf{H}_k\mathbf{P}_k^-\right)^{\mathrm{T}} = \left(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\right) + \mathbf{R}_k \quad \Rightarrow$$

$$\mathbf{K}_k^{-1} = \left(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)\left(\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\right)^{-1}$$

Taking the inverse, finally, gives the value of the optimal gain

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right) \tag{7.38}$$

The optimal gain found in Eq. (7.38) provides a more compact version of the a-posteriori covariance matrix, Eq. (7.32).

$$\mathbf{P}_k = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_k^- \tag{7.39}$$

This concludes the derivation for the optimal recursive filter for linear systems, the Kalman filter. The way the filter works has been visualised in Fig. 7.7.

7.3.2 Extended Kalman Filter

The linear Kalman filter can, as its name implies, in principle only be applied to linear systems. Most systems that are modelled are non-linear, thus rendering the LKF less efficient. The EKF is one way to adapt the Kalman-filtering techniques to non-linear systems. The EKF works by using a first-order Taylor series approximation of the non-linear system around the current estimate.

The derivation of the EKF can be found in many textbooks, e.g., Farrell and Barth (1999), Welch and Bishop (2001) and Groves (2008). The current description provides a summary of this derivation. The non-linear system is expressed by the following non-linear equation

$$\mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k\right) \tag{7.40}$$

with the measurement

$$\mathbf{z}_k = \mathbf{h}\left(\mathbf{x}_k, \mathbf{v}_k\right) \tag{7.41}$$

The noise parameters are the same as in Eqs. (7.16) and (7.17).

$$\begin{aligned} \mathrm{p}(\mathbf{v}_k) &\sim N(0, \mathbf{R}_k) \\ \mathrm{p}(\mathbf{w}_k) &\sim N(0, \mathbf{Q}_k) \end{aligned}$$

The a-priori state estimate can be derived from the previous a-posteriori estimate using

$$\hat{\mathbf{x}}_{k+1}^{-} = \mathbf{f}\left(\hat{\mathbf{x}}_k, \mathbf{u}_{k+1}, \mathbf{0}\right) \tag{7.42}$$

The measurement can be estimated using

$$\hat{\mathbf{z}}_k = \mathbf{h}\left(\hat{\mathbf{x}}_k^{-}, \mathbf{0}\right) \tag{7.43}$$

Finally, the state and the measurement can be approximated by the linear equations

$$\mathbf{x}_{k+1} \approx \hat{\mathbf{x}}_{k+1}^{-} + \mathbf{A}_k\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right) + \mathbf{W}_k\mathbf{w}_k \tag{7.44}$$

$$\mathbf{z}_k \approx \hat{\mathbf{z}}_k + \mathbf{H}_k\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{-}\right) + \mathbf{V}_k\mathbf{v}_k \tag{7.45}$$

In Eqs. (7.44) and (7.45), the terms $\mathbf{A}$, $\mathbf{H}$, $\mathbf{V}$, and $\mathbf{W}$ are Jacobians that are used to linearise the system. The related coefficients are obtained with

$$\mathbf{A}_k = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{(\hat{\mathbf{x}}_k, \mathbf{u}_k, \mathbf{0})} \tag{7.46}$$

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{(\hat{\mathbf{x}}_k^{-}, \mathbf{0})} \tag{7.47}$$

$$\mathbf{V}_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{v}}\right|_{(\hat{\mathbf{x}}_k^-, \mathbf{0})} \tag{7.48}$$

$$\mathbf{W}_k = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{w}}\right|_{(\hat{\mathbf{x}}_k, \mathbf{u}_{k+1}, \mathbf{0})} \tag{7.49}$$

The Jacobians should preferably be derived analytically beforehand, such that the values can be computed at each step. It is also possible to establish them numerically by using numerical differentiation techniques, such as complex-step differentiation (Martins et al. 2000). Numerical differentiation obviates the need to derive the Jacobians, but is less accurate and less stable for application.

A prediction error, which is equivalent to the a-priori estimation error of the linear Kalman filter, is defined as

$$\hat{\mathbf{e}}_{xk}^- \equiv \mathbf{x}_k - \hat{\mathbf{x}}_k^- \approx \mathbf{A}_{k-1}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + \varepsilon_k \tag{7.50}$$

The measurement residual, which is equivalent to the measurement innovation of the linear Kalman filter, is defined as

$$\hat{\mathbf{e}}_{zk}^- \equiv \mathbf{z}_k - \hat{\mathbf{z}}_k \approx \mathbf{H}_k\left(\hat{\mathbf{e}}_{xk}^-\right) + \eta_k \tag{7.51}$$

The variables ε and $\boldsymbol{\eta}$ are the linear approximations of the $\mathbf{w}$ and $\mathbf{v}$ respectively. Using the linear variance propagation laws, the distribution of these two noise parameters is

$$\mathrm{p}(\varepsilon_k) \sim N\left(\mathbf{0}, \mathbf{W}_{k-1}\mathbf{Q}_{k-1}\mathbf{W}_{k-1}^{\mathrm{T}}\right) \tag{7.52}$$

$$\mathrm{p}(\boldsymbol{\eta}_k) \sim N\left(\mathbf{0}, \mathbf{V}_k\mathbf{R}_k\mathbf{V}_k^{\mathrm{T}}\right) \tag{7.53}$$

The prediction error estimate is defined as the difference between the a-posteriori and the a-priori errors.

$$\hat{\mathbf{e}}_k \equiv \hat{\mathbf{x}}_k - \hat{\mathbf{x}}_k^- \tag{7.54}$$

This prediction error estimate should be equivalent to the state update. Using a Kalman gain, this becomes

$$\hat{\mathbf{e}}_k = \mathbf{K}\left(\hat{\mathbf{e}}_{zk}^-\right) = \mathbf{K}\left(\mathbf{z}_k - \hat{\mathbf{z}}_k^-\right) \tag{7.55}$$

The a-posteriori state estimate can now be written as

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}\left(\mathbf{z}_k - \hat{\mathbf{z}}_k^-\right) \tag{7.56}$$

This results in the same kind of equations as the linear Kalman filter with an important difference. The difference being that the noise parameters are replaced by their linearised counterparts. The equation for the covariance propagation is

$$\mathbf{P}_{k+1}^{-} = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^{\mathrm{T}} + \mathbf{W}_k \mathbf{Q}_k \mathbf{W}_k^{\mathrm{T}}. \tag{7.57}$$

The gain can then be found by

$$\mathbf{K}_k = \mathbf{P}_k^{-} \mathbf{H}_k^{\mathrm{T}} \left(\mathbf{H}_k \mathbf{P}_k^{-} \mathbf{H}_k^{\mathrm{T}} + \mathbf{V}_k \mathbf{R}_k \mathbf{V}_k^{\mathrm{T}} \right)^{-1} \tag{7.58}$$

Example 7.1 (*Extended Kalman Filter for One-Dimensional Motion*) As an example application of the extended Kalman filter, the estimate of position and velocity of a (one degree-of-freedom) falling object will be discussed. This example is based on a similar example provided by Zarchan (2009), with the difference that here we will provide the data in SI units rather than imperial units. The problem is formulated as follows.

A high-velocity object is falling towards a tracking radar, as shown in Fig. 7.8. It has an initial position and an initial velocity of 65 km and 1,900 m/s. Gravity and drag are the only forces acting on the body, and are aligned but opposite in direction. Its position is measured with a sample rate of 10 Hz. Given the measurement data, the object's position and velocity have to be estimated. Thus, the state vector is defined as $\mathbf{x} = \begin{pmatrix} h & V \end{pmatrix}^{\mathrm{T}}$. The time derivative of the position is simply the velocity, whereas the derivative of the velocity is:

$$\dot{V} = \frac{\rho V^2}{2K} - g_0 \tag{7.59}$$

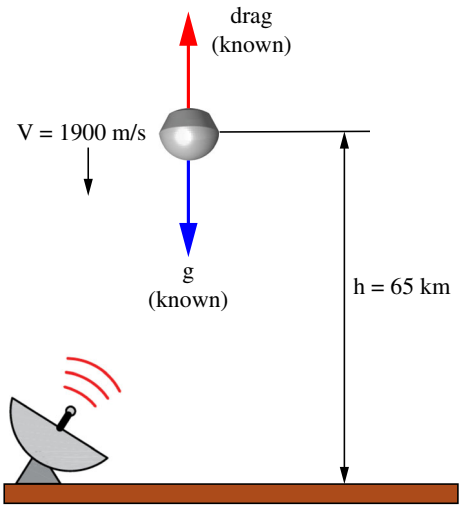


Fig. 7.8 Tracking of a falling body

where K is the ballistic coefficient and g_0 is the gravitational acceleration (assumed to be constant and equal to 9.81 m/s^2). The atmospheric density, ρ, is assumed to follow an exponential profile

$$\rho = \rho_0 \exp^{-h/H_s} \tag{7.60}$$

with $\rho_0 = 1.225$ kg/m^3 and the scale height $H_s = 6{,}950$ m.

The dynamics function, $\mathbf{f}$ is thus written as:

$$\mathbf{f} = \begin{pmatrix} V \\ \frac{\rho V^2}{2K} - g_0 \end{pmatrix} \tag{7.61}$$

We can linearise Eq. (7.59) by saying that:

$$\Delta\mathbf{x} = \begin{pmatrix} \Delta\dot{h} \\ \Delta\dot{V} \end{pmatrix} = \begin{bmatrix} \frac{\partial\dot{h}}{\partial h} & \frac{\partial\dot{h}}{\partial V} \\ \frac{\partial\dot{V}}{\partial h} & \frac{\partial\dot{V}}{\partial V} \end{bmatrix} \begin{pmatrix} \Delta h \\ \Delta V \end{pmatrix} + \begin{pmatrix} 0 \\ \eta \end{pmatrix} \tag{7.62}$$

Here, a process noise η for the acceleration is introduced, which is useful for tuning the filter. The partial derivatives are expressed below:

$$f_{11} = \frac{\partial\dot{h}}{\partial h} = 0 \tag{7.63}$$

$$f_{12} = \frac{\partial\dot{h}}{\partial\dot{h}} = 1 \tag{7.64}$$

$$f_{21} = \frac{\partial\dot{V}}{\partial h} = -\frac{\rho_0 g_0 \exp\left(-h/H_s\right)\dot{h}^2}{2H_s K} \tag{7.65}$$

$$f_{22} = \frac{\partial\dot{V}}{\partial\dot{h}} = \frac{\rho_0 g_0 \exp\left(-h/H_s\right)\dot{h}}{K} \tag{7.66}$$

Having the partial derivatives, the state-transition matrix can be formulated, which is evaluated at each current estimate:

$$\mathbf{F} = \begin{bmatrix} f_{12} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}_{\mathbf{x}=\hat{\mathbf{x}}} \tag{7.67}$$

Since it is expected that the filter will be evaluated in an on-board setting with a discrete GNC system, its discrete time counterpart can be approximated by the first two terms of a Taylor series:

$$\mathbf{\Phi} \approx \mathbf{I} + \mathbf{F}\Delta t \tag{7.68}$$

where Δt is the time step for the filter. The discrete-time process noise covariance matrix, given by Zarchan (2009), is defined as follows:

$$\mathbf{Q} = \sigma_\eta^2 \begin{bmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} + f_{22}\frac{\Delta t^3}{3} \\ \frac{\Delta t^2}{2} + f_{22}\frac{\Delta t^3}{3} & \Delta t + f_{22}\Delta t^2 + f_{22}^2\frac{\Delta t^3}{3} \end{bmatrix} \tag{7.69}$$

The radar measures the position each 0.1 s, so its measurement equation is given as:

$$z = \tilde{h} = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x} + \nu \tag{7.70}$$

Thus, the measurement matrix is simply:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{7.71}$$

whereas the measurement covariance, R, is a scalar:

$$R = \sigma_\nu^2 \tag{7.72}$$

Now that everything that is needed to formulate the filter has been defined, the steps that are required to execute the filter can be formulated. These steps are, after initialisation, the standard Kalman filter steps discussed earlier. First, the state is propagated with an Euler step method[5]:

$$\mathbf{x}_{k+1/k} = \mathbf{x}_{k/k} + \mathbf{f}\left(\mathbf{x}_{k/k}\right) \Delta t \tag{7.73}$$

Then the discrete time transition matrix is evaluated and the error covariance matrix is propagated:

$$\mathbf{P}_{k+1/k} = \mathbf{\Phi} \mathbf{P}_{k/k} \mathbf{\Phi}^{\mathrm{T}} + \mathbf{Q} \tag{7.74}$$

Each time an external measurement is available (here, with a 0.1 s sample rate), the update stage starts. The steps are as follows, beginning with the evaluation of the Kalman gain, $\mathbf{K}$:

$$\mathbf{K} = \mathbf{P}_{k+1/k} \mathbf{H}^{\mathrm{T}} \left(\mathbf{H} \mathbf{P}_{k+1/k} \mathbf{H}^{\mathrm{T}} + R \right)^{-1} \tag{7.75}$$

$$\hat{\mathbf{x}}_{k+1/k+1} = \hat{\mathbf{x}}_{k+1/k} + \mathbf{K} \left(z_{k+1} - x_{k+1/k} \right) \tag{7.76}$$

with $x_{k+1/k} = h_{k+1/k}$. Finally,

$$\mathbf{P}_{k+1/k+1} = \left(\mathbf{I} - \mathbf{K}\mathbf{H} \right) \mathbf{P}_{k+1/k} \left(\mathbf{I} - \mathbf{K}\mathbf{H} \right)^{\mathrm{T}} + \mathbf{K} R \mathbf{K}^{\mathrm{T}} \tag{7.77}$$

[5] The Euler integrator is in principle very inaccurate, and requires a high sample rate (here, 1000 Hz, i.e., $\Delta t = 0.001$ s, is used. For numerical stability it would be better to use, for instance, a fourth-order, fixed-step Runge–Kutta method. However, in doing so one should realise that in case IMU (= acceleration) measurements are used to propagate the state, a zero-order hold on these measurements should be applied. In other words, the measurements are constant for a single integration step, despite the fact that four function evaluations are done. This is easily understood that at the midpoints of the integration interval (the minor time steps), no new measurements are available.

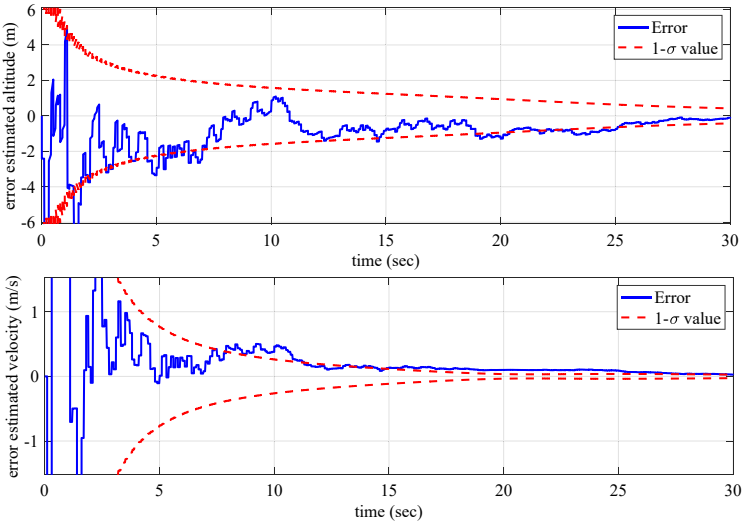


Fig. 7.9 Altitude (top) and velocity (bottom) estimates, $\sigma_\nu = 8$ m, $\sigma_\eta^2 = 0$, $\Delta t = 0.001$

Figure 7.9 shows the EKF results for an Euler integration with $\Delta t = 0.001$ s and a measurement update of 0.1 s. The filter is able to estimate position and velocity, with errors converging to within the boundaries of theoretical standard deviations. ■

7.3.3 Integrated Navigation System

As stated earlier, the IMUs have a tendency to drift that can create considerable deviations in the estimate from the truth in the long run. Thus, it is imperative to use another navigation method that can give absolute measurements. A method that has been treated in this chapter is GPS, thus it can be used to compensate for the biases found in IMU measurements. This method is known as GPS/IMU coupling and will be explained in this section. The reader is again referred to Farrell and Barth (1999), Grewal et al. (2001) and Groves (2008) for a more detailed discussion on sensor coupling and navigation-system integration.

Basically, there are two ways to couple the IMU and GPS, i.e., loose and tight coupling. Loose coupling is the simplest method of coupling. It treats the GPS and IMU measurements separately. The minimum of four GPS pseudo-range measurements are used to find the position and the velocity of the user. The (estimated) position and velocity are then treated as a measurement and fed to the main navigation filter along with IMU measurements. The estimation of the position using the pseudo-ranges involves some filtering, which makes the noise on the estimated position to

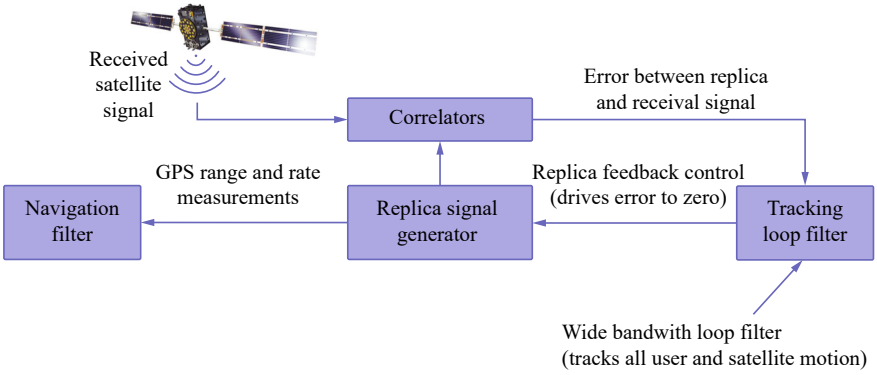


Fig. 7.10 Loosely coupled GPS/IMU

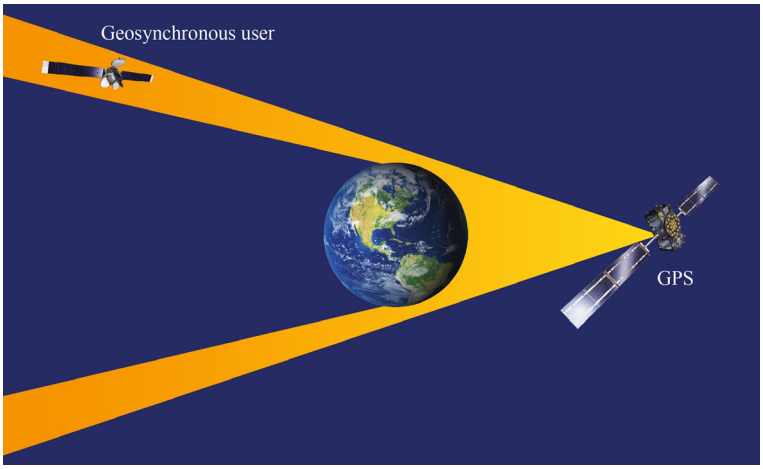


Fig. 7.11 GPS spillover for GEO satellites

be coloured versus the white noise on the pseudoranges. An example of a loosely coupled system can be seen in Fig. 7.10.

Tight coupling feeds the raw measurements from the GPS receiver directly into the navigation filter. This means that new state parameters, which estimate the bias of the receiver clock, have to be added. Also, each measurement from a GPS satellite has to be treated as an individual measurement for the filter. All this increases the computational load, but at the same time, the estimation accuracy also increases. Another advantage is that the GPS measurements can be processed even if fewer than four satellites are in view. This can be very beneficial for the case of geostationary satellites that can receive measurements from GPS satellites on the other side of the Earth, as can be seen in Fig. 7.11. An example of a tightly coupled system is shown in Fig. 7.12.

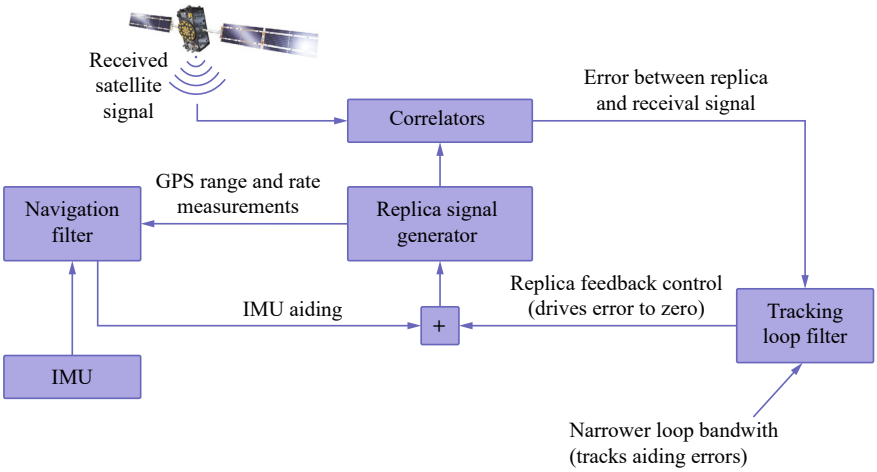


Fig. 7.12 Tightly coupled GPS/IMU

7.3.4 Case Study: HORUS Navigation System

To illustrate the implementation of a tightly coupled navigation system in combination with an extended Kalman filter we will use the reference mission of HORUS (see Appendix C, Sect. B.2) to see how well we can reconstruct the trajectory and attitude. This case study is largely based on results published by Mooij and Chu (2002). It is stressed that the navigation system has not been actually used to fly the reference mission, so the effect of navigation errors on the flight performance is not included here. For this case study, HORUS will be equipped with four GPS antennas, such that the GPS carrier-phase measurements can be used to determine the vehicle's attitude. The location of the antennae is the following:

- The master antenna, index m, is located in the geometric plane of symmetry, at 16 m from the nose and on top of the vehicle. From Fig. B.4, it can be seen that the body-frame location will be $\boldsymbol{r_m} = (-3, 0, -2.7)^{\mathrm{T}}$ m.
- The first slave antenna, index s_1, is located 7 m before the master antenna, i.e., $\boldsymbol{r_{s1}} = (4, 0, -2.7)^{\mathrm{T}}$ m.
- The second and third slave antenna, indices s_2 and s_3, are located on both sides of the master antenna, at $y = \pm 1$ m, i.e., $\boldsymbol{r_{s2}} = (-3, 1, -2.5)^{\mathrm{T}}$ m and $\boldsymbol{r_{s3}} = (-3, -1, -2.5)^{\mathrm{T}}$ m.

The nominal re-entry mission of HORUS starts at the atmospheric entry interface at an altitude of $h = 122$ km and a relative velocity of $V = 7435.5$ m/s. The vehicle is assumed to be heading towards a runway in Kourou. Some 80 km from the landing site guidance switches to the Terminal Area Energy Management logic. In Fig. 7.13, the altitude-velocity profile and nominal control history are plotted.

Mathematical Model

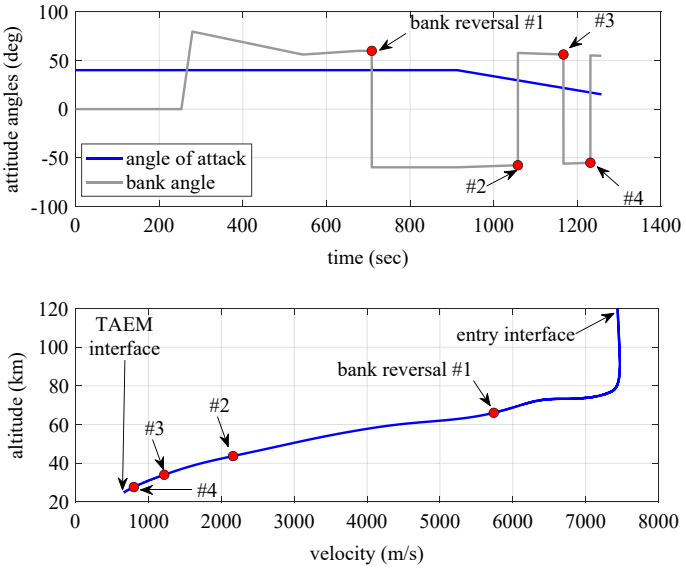


Fig. 7.13 Nominal control history (top) and altitude-velocity profile (bottom)

During the course of this study it appeared that the implementation of the so-called tangent-plane equations[6] that were used in an earlier study (Mooij and Chu 2001) gave numerical problems. The output matrix appeared to be badly scaled, which resulted in a poor filter performance. Also it was concluded that the large errors in the roll angle (up to 0.5°) for an otherwise converged estimate showed whenever a bank reversal was executed, could be traced back to the use of these equations. Therefore, the Cartesian position $\mathbf{r} = (x_R, y_R, z_R)^{\mathrm{T}}$ and velocity $\mathbf{V_R} = (u, v, w)^{\mathrm{T}}$, defined with respect to the ECEF reference frame will be used.

The dynamic equations for translational motion are given by:

$$\dot{u} = a_{R,x_e} + g_x + \omega_{cb}^2 x_R + 2\omega_{cb} v \tag{7.78}$$

[6] The tangent-plane equations are the differential equations of motion based on the north, east and down velocity, as well as latitude, longitude, and altitude.

$$\dot{v} = a_{R,y_e} + g_y + \omega_{cb}^2 y_R - 2\omega_{cb} u \tag{7.79}$$

$$\dot{w} = a_{R,z_e} + g_z \tag{7.80}$$

In the above definition, a_{R,x_e}, a_{R,y_e}, and a_{R,z_e} are the estimated specific-force components (non-gravitational acceleration, as measured by the IMU) expressed in the ECEF frame (index R, for rotating planetocentric frame), g_x, g_y, and g_z are the Cartesian components of the gravitational acceleration and ω_{cb} is the rotational rate of the Earth. The kinematic (position) equations are:

$$\dot{x}_R = u \tag{7.81}$$

$$\dot{y}_R = v \tag{7.82}$$

$$\dot{z}_R = w \tag{7.83}$$

To improve the computational efficiency, the kinematic attitude equations will be based on quaternions rather than local Euler angles (Mooij and Chu 2001). Since the GPS carrier-phase difference measurement is a line-of-sight detector, it offers the possibility to measure quaternions so no implementation problems will be introduced. Moreover, for the current tightly-coupled navigation system the use of quaternions will make the rotational kinematic equations linear, since the rotational-rate components are system inputs. This results in a simpler model, which is expected to be numerically more stable when unprocessed GPS measurements are going to be used. In fact, anticipating the results later on in this section, two observations were made. In the first place, a gain in computational speed was obtained and in the second place, after converting the quaternions to local Euler angles, the corresponding values appeared to be identical. The former observation will be important at a later stage, once a real-time implementation of the navigation system is considered, whereas the second observation confirms a proper implementation of the kinematic model.

Let the quaternion vector be defined as $\mathbf{q}_{\mathbf{I},\mathbf{B}} = (q_1, q_2, q_3, q_4)^{\mathrm{T}}$ where the index I, B means that the quaternions define the attitude of the body frame with respect to the inertial geocentric frame. The time derivative of $\mathbf{q}_{\mathbf{I},\mathbf{B}}$, $\dot{\mathbf{q}}_{\mathbf{I},\mathbf{B}}$, is expressed as

$$\dot{\mathbf{q}}_{\mathbf{I},\mathbf{B}} = \frac{1}{2} \begin{bmatrix} -q_4 & -q_3 & q_2 \\ q_3 & -q_4 & -q_1 \\ -q_2 & q_1 & -q_4 \\ q_1 & q_2 & q_3 \end{bmatrix} \begin{pmatrix} p_e \\ q_e \\ r_e \end{pmatrix} \tag{7.84}$$

with p_e, q_e, and r_e being the estimated roll, pitch and yaw rate. The corresponding transformation matrix from body to inertial frame is given by:

$$\mathbf{C}_{\mathbf{I},\mathbf{B}} = \begin{bmatrix} q_1^2 + q_4^2 - q_2^2 - q_3^2 & 2(q_1 q_2 + q_3 q_4) & 2(q_1 q_3 - q_2 q_4) \\ 2(q_1 q_2 - q_3 q_4) & q_2^2 + q_4^2 - q_1^2 - q_3^2 & 2(q_2 q_3 + q_1 q_4) \\ 2(q_1 q_3 + q_2 q_4) & 2(q_2 q_3 - q_1 q_4) & q_3^2 + q_4^2 - q_1^2 - q_2^2 \end{bmatrix} \tag{7.85}$$

The estimated quaternions can easily be transformed to local Euler angles ϕ (roll), θ (pitch) and ψ (yaw) by the following relations. Let the transformation matrix from the body to the navigation frame (or local tangent plane) be given by:

$$\mathbf{C_{N,B}} = \begin{bmatrix} \mathrm{c}\theta\mathrm{c}\psi & \mathrm{s}\phi\mathrm{s}\theta\mathrm{c}\psi - \mathrm{s}\psi\mathrm{c}\phi & \mathrm{c}\psi\mathrm{c}\phi\mathrm{s}\theta + \mathrm{s}\psi\mathrm{s}\phi \\ \mathrm{c}\theta\mathrm{s}\psi & \mathrm{s}\psi\mathrm{s}\theta\mathrm{s}\phi + \mathrm{c}\psi\mathrm{c}\phi & \mathrm{s}\psi\mathrm{c}\phi\mathrm{s}\theta - \mathrm{c}\psi\mathrm{s}\phi \\ -\mathrm{s}\theta & \mathrm{s}\phi\mathrm{c}\theta & \mathrm{c}\phi\mathrm{c}\theta \end{bmatrix} \tag{7.86}$$

This matrix can also be written as

$$\mathbf{C_{N,B}} = \mathbf{C_{N,I}}\mathbf{C_{I,B}} \tag{7.87}$$

with $\mathbf{C_{I,B}} = \mathbf{C_{B,I}^{T}}$ from Eq. (7.85) and

$$\mathbf{C_{N,I}} = \begin{bmatrix} -\mathrm{c}\tilde{\tau}\mathrm{s}\delta & -\mathrm{s}\tilde{\tau}\mathrm{s}\delta & \mathrm{c}\delta \\ -\mathrm{s}\tilde{\tau} & \mathrm{c}\tilde{\tau} & 0 \\ -\mathrm{c}\tilde{\tau}\mathrm{c}\delta & -\mathrm{s}\tilde{\tau}\mathrm{c}\delta & -\mathrm{s}\delta \end{bmatrix} \tag{7.88}$$

In $\mathbf{C_{N,I}}$, $\tilde{\tau} = \tau + \omega_{cb}t$ is the celestial longitude, and τ and δ are the geocentric longitude and latitude, respectively.

Looking at each of the matrix elements, the following expressions for ϕ, θ, and ψ are found:

$$\phi = \arctan\left(\frac{\mathbf{C_{N,B}}(2,3)}{\mathbf{C_{N,B}}(3,3)}\right) \tag{7.89}$$

$$\theta = -\arcsin\left(\mathbf{C_{N,B}}(3,1)\right) \tag{7.90}$$

$$\psi = \arctan\left(\frac{\mathbf{C_{N,B}}(1,2)}{\mathbf{C_{N,B}}(1,1)}\right) \tag{7.91}$$

using the elements from Eq. (7.87).

In Sect. 7.2.1 the Inertial Measurement Unit has been discussed, including (simplified) models for measured acceleration and angular rate, Eqs. (7.1) and (7.4). Substitution of these equations into Eqs. (7.78) to (7.84), and taking into account the body components of specific forces measured by accelerometers strapped down to the re-entry vehicle, the navigation equations are formed. The estimated specific force components input to Eqs. (7.78)–(7.80) are given by:

$$\begin{pmatrix} a_{R,x_e} \\ a_{R,y_e} \\ a_{R,z_e} \end{pmatrix} = \mathbf{C_{R,B}} \begin{bmatrix} 1+s_x & m_{xy} & m_{xz} \\ m_{yx} & 1+s_y & m_{yz} \\ m_{zx} & m_{zy} & 1+s_z \end{bmatrix}^{-1} \left\{ \begin{pmatrix} a_{x_m} \\ a_{y_m} \\ a_{z_m} \end{pmatrix} - \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} \right\} \tag{7.92}$$

where

$$\mathbf{C}_{\mathbf{R},\mathbf{B}} = \mathbf{C}_{\mathbf{R},\mathbf{I}}\mathbf{C}_{\mathbf{I},\mathbf{B}} \tag{7.93}$$

$$\mathbf{C}_{\mathbf{R},\mathbf{I}} = \begin{bmatrix} \cos(\omega_{cb}t) & \sin(\omega_{cb}t) & 0 \\ -\sin(\omega_{cb}t) & \cos(\omega_{cb}t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.94}$$

and $\mathbf{C}_{\mathbf{I},\mathbf{B}}$ being the transpose of Eq. (7.85). Of course, the instrument errors in Eq. (7.92) are the best available estimates.

Similarly, the estimated angular rates input to Eq. (7.84) are given by:

$$\begin{pmatrix} p_e \\ q_e \\ r_e \end{pmatrix} = \begin{bmatrix} 1+s_p & m_{pq} & m_{pr} \\ m_{qp} & 1+s_q & m_{qr} \\ m_{rp} & m_{rq} & 1+s_r \end{bmatrix}^{-1} \left\{ \begin{pmatrix} p_m \\ q_m \\ r_m \end{pmatrix} - \begin{pmatrix} b_p \\ b_q \\ b_r \end{pmatrix} \right\} \tag{7.95}$$

In vector form, the above equations can be written as:

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] + \mathbf{G}[\mathbf{x}(t), t]\mathbf{w}(t) \tag{7.96}$$

As aiding measurements we will use the GPS pseudorange, range rate, and carrier-phase differences. The related measurement Eqs. (7.6), (7.8), and (7.13) can also be written in vector form.

$$\mathbf{z}(t) = \mathbf{h}[\mathbf{x}(t), \mathbf{u}(t), t] + \mathbf{v} \tag{7.97}$$

In state Eq. (7.96), the state vector $\mathbf{x}$ consists of the following state variables:

1. vehicle position components (x_R, y_R, and z_R);
2. vehicle velocity components (u, v, and w);
3. vehicle attitude quaternions (q_1, q_2, q_3, and q_4);
4. unknown IMU calibration parameters (b_x, b_y, b_z, s_x, s_y, s_z, m_{xy}, m_{xz}, m_{yx}, m_{yz}, m_{zx} and m_{zy}, and b_p, b_q, b_r, s_p, s_q, s_r, m_{pq}, m_{pr}, m_{qp}, m_{qr}, m_{rp}, and m_{rq});
5. receiver clock bias and clock drift (c_{b_r} and c_{d_r}).

Furthermore, the input vector $\mathbf{u}$ consists of the following measured variables:

1. Specific-force measurements from accelerometers (a_{x_m}, a_{y_m}, and a_{z_m})
2. Rotational-rate measurements from rate gyros (p_m,q_m, and r_m).

In the measurement Eq. (7.97) the observation vector $\mathbf{z}$ consists of the following measurement variables:

1. GPS pseudorange measurements (ρ_1, ρ_2, ρ_3, and ρ_4);
2. GPS Doppler shift (or, equivalently, pseudorange rate) measurements ($\dot{\rho}_1$, $\dot{\rho}_2$, $\dot{\rho}_3$, and $\dot{\rho}_4$);
3. GPS carrier difference measurements from two satellites between the master and three slave antennas ($\Delta\Phi_{1,1}$, $\Delta\Phi_{1,2}$, $\Delta\Phi_{1,3}$, $\Delta\Phi_{2,1}$, $\Delta\Phi_{2,2}$, and $\Delta\Phi_{2,3}$).

v and **w** are uncorrelated white-noise vectors. As compared to the measurement equations of a loosely coupled system (Mooij and Chu 2001), the tightly coupled integrated GPS/IMU has a highly non-linear observation model.

The state estimator applied to the integrated navigation system will be the EKF, which performs a first-order approximation of non-linear filters. When the non-linearity of the integrated navigation system is high—as is the case with re-entry flights—the EKF may diverge. In that case, improvements in performance of the EKF can be realised by introducing local iterations. Since these algorithms are recursive, they are suitable for on-board navigation. However, these extensions are beyond the scope of the current text.

Measurements

To obtain raw GPS measurements, a simulator can be developed that provides position and velocity information on the available GPS satellites at a certain desired location, thereby using the ICD-GPS-200 algorithm, see, for instance, Farrell and Barth (1999). The simulator will also provide the combination of four satellites that yields the best Geometric Dilution of Precision (GDOP), since for user-position (x_R, y_R, and z_R) and -velocity (u, v, and w) determination at least four satellites are required, as the receiver clock bias and drift need to be determined as well. Of course, we could also decide to use more satellites, but with this case study we want to determine the minimum performance.

The algorithm can briefly be described in a number of steps:

1. Determine receiver time, based on date and time information (so-called computer clock tagging);
2. Using the GPS satellites algorithm (ICD-GPS-200), and taking the (weekly updated) ephemeris data from US NAWCWD as input (NOAA 2011), the position and velocity of all GPS satellites is computed and expressed in the ECEF reference frame;
3. Given the user location and a mask angle, only those satellites, which line of sight results in an angle larger than the mask angle, are tagged "visible";
4. Compute the Geometric Dilution of Precision GDOP;
5. Determine the four satellites that gives the best GDOP;
6. For the actual position, velocity, and attitude of the re-entry vehicle, compute the pseudorange, pseudorange rate, and carrier-phase difference using Eqs. (7.6), (7.8), and (7.13).

The validation of the simulator has been done by comparing the results with US NAWCWPNS GPS/INS Section using their *Interactive GPS Satellite Prediction Utility*, at the time of this study available on the internet (but now offline). The comparison has shown a close-match result, including the GDOP prediction.

The (ideal) IMU measurements are obtained from a non-linear flight simulation with the accelerations in three orthogonal directions as output. Originally, the IMU measurement frequency was set at 25 Hz, because a lower data rate may be better from a computational point of view. However, it can be disputed whether this will

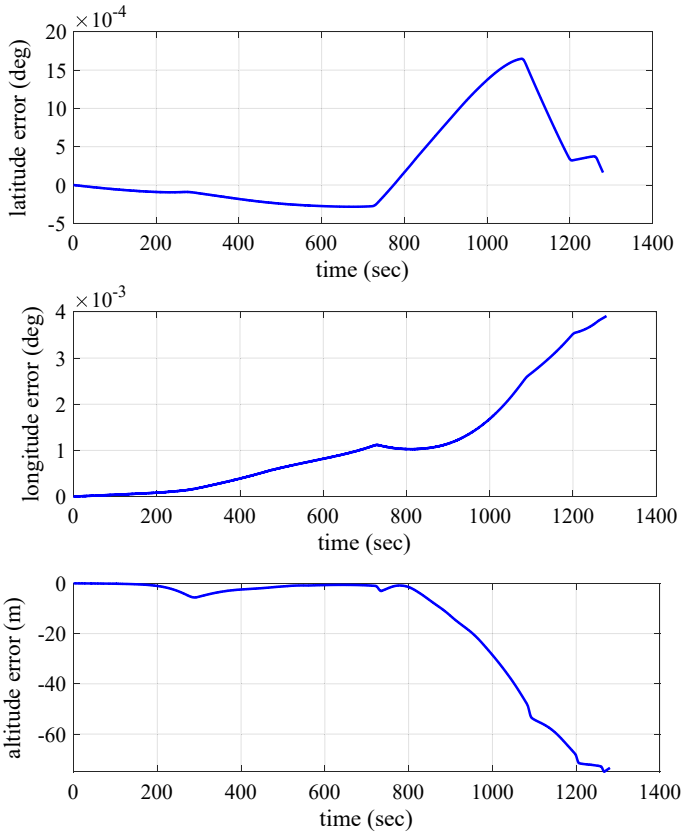


Fig. 7.14 Relative position errors due to 25 Hz nominal-state propagation (Euler integration)

not introduce unnecessary errors in the nominal state propagation, due to truncation errors of the applied integration method. Therefore, this frequency has also been increased to 200 Hz (yielding an integration step size of $dt = 0.005$ s), and the results of both nominal state propagations will be compared with the results from the actual re-entry simulation, which is considered to be the truth model.

When the frequency is set at 200 Hz, the resulting nominal state is very close to the truth model, and can be considered to be identical for practical purposes. However, the differences become larger when the frequency is decreased. To show the impact of lowering the frequency, in Figs. 7.14 and 7.15 the difference between propagating at 200 Hz (truth model) and 25 Hz has been plotted for the complete re-entry trajectory. It is obvious, that by setting the IMU measurement frequency at 200 Hz, the error source due to the nominal state propagation will be removed. It is noted that the accuracy improvement for the roll, pitch and yaw angle is only in the order of 0.01° for each of the three angles. It is interesting to note, that the

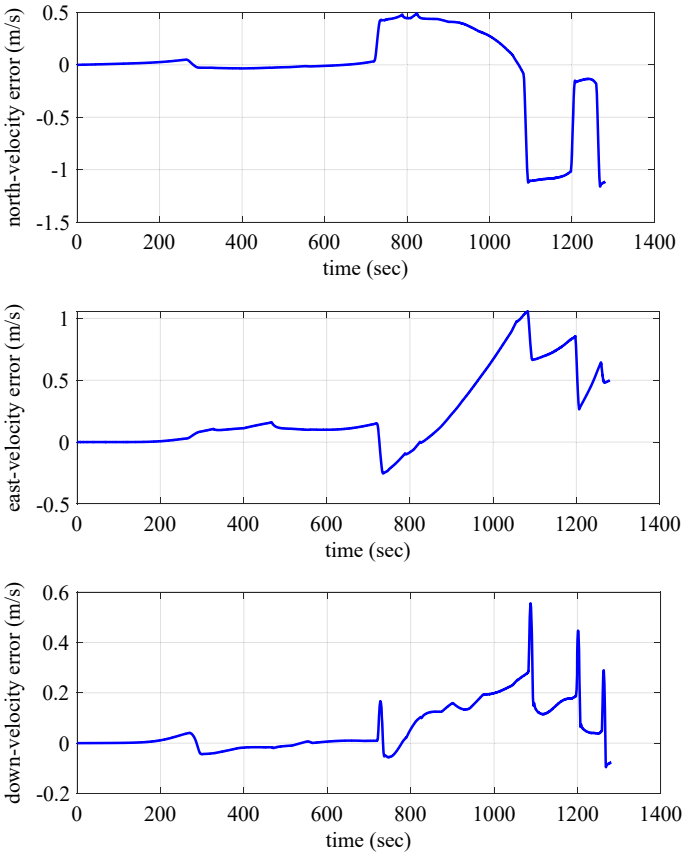


Fig. 7.15 Relative velocity errors due to 25 Hz nominal-state propagation (Euler integration)

velocity errors show discrete jumps at the moment that the bank reversals take place, indicating the need for more accurate measurements during these manoeuvres.

In-Flight Filter Calibration

To test the proper implementation of the model, as well as to check the observability of the system, the same manoeuvres as discussed by Mooij and Chu (2001) will be simulated. The considered manoeuvres are the angular doublets,[7] the yaw reversal to align for the deorbit burn and the de-orbit burn itself. Doublets are executed in roll, pitch and yaw, although the yaw manoeuvre is also covered in the reversal. For the sake of simulation, the doublets are executed simultaneously, whereas all three

[7] An angular doublet consists of (i) a commanded angle rate, followed by (ii) a constant attitude, (iii) a negative angle rate, (iv) a constant attitude, and finally (v) a commanded angle rate, which brings the attitude back to zero. In principle the doublet resembles a sine wave, with linear and constant segments.

Table 7.2 Input instrument errors, after ground calibration

Accelerometer error	Value	Gyroscope error	Value
b_x	$3 \cdot 10^{-4}$	b_p	$5 \cdot 10^{-6}$
b_y	$-3.5 \cdot 10^{-4}$	b_q	$-5 \cdot 10^{-6}$
b_z	$3 \cdot 10^{-4}$	b_r	$5 \cdot 10^{-6}$
s_x	$2 \cdot 10^{-4}$	s_p	$2 \cdot 10^{-4}$
s_y	$-1.7 \cdot 10^{-4}$	s_q	$-1.7 \cdot 10^{-4}$
s_z	$2.3 \cdot 10^{-4}$	s_r	$2.3 \cdot 10^{-4}$
m_{xy}	10^{-6}	m_{pq}	10^{-6}
m_{xz}	-10^{-6}	m_{pr}	-10^{-6}
m_{yx}	-10^{-6}	m_{qp}	-10^{-6}
m_{yz}	$2 \cdot 10^{-6}$	m_{qr}	$2 \cdot 10^{-6}$
m_{zx}	10^{-6}	m_{rp}	10^{-6}
m_{zy}	$2 \cdot 10^{-6}$	m_{rq}	$2 \cdot 10^{-6}$

manoeuvre types are executed sequentially with only little spacing. Of course, in an actual mission the manoeuvre timing may be different.

The timing and execution logic is as follows. After 10 s from t = 0 s, the doublets are initiated with $\dot{\phi} = \dot{\theta} = \dot{\psi} = 3°/\text{s}$. When the amplitude is 30° ($t = 20$ s), the attitude is stabilised for 10 s, after which $\dot{\phi} = \dot{\theta} = \dot{\psi} = -3°/\text{s}$ is commanded ($t = 30$ s) until the amplitude is −30° ($t = 50$ s). Again, after a stabilising interval of 10 s, the attitude is commanded back to zero with $\dot{\phi} = \dot{\theta} = \dot{\psi} = 3°/\text{s}$. Hence, the total duration is 60 s. At $t = 70$ s, the beginning of the yaw turn is commanded with $\dot{\psi} = -3°/\text{s}$, so after 60 s the turn is completed. Finally, after another 10 s ($t = 150$ s), the deorbit burn starts. With a thrust level of 8000 N, a specific impulse of 300 s and a total ΔV requirement of 120 m/s, the manoeuvre lasts 398.5 s. Note that the acceleration level is 0.31 m/s^2, and only in x-direction.

As input to the filter, the mentioned manoeuvres are simulated using the instrument errors listed in Table 7.2. In addition, the accelerometer measurements have white-noise errors with a 1σ value of 10^{-5} m/s^2, whereas the gyroscope measurements have white-noise errors with $1\sigma = 10^{-6}$ rad/s. The GPS data are considered to have white-noise errors with 1σ values of $\sigma_\rho = 25$ m, $\sigma_{\dot{\rho}} = 0.1$ m/s, and $\sigma_{\Delta\Phi} = 0.1$ cycle ($\approx$19 cm). The applied errors in initial conditions are 30 m, −30 m, and 35 m for the position, −0.1 m/s, −0.1 m/s, and 0.1 m/s for the velocity, and 0.05, −0.05, −0.05, and 0.05 for the quaternions. Initial instrument-error estimates were negated with respect to their respective input values of Table 7.2.

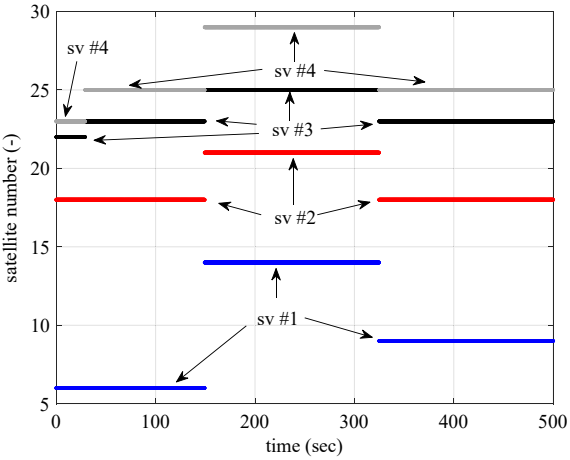


Fig. 7.16 Best combination of 4 visible satellites during exo-atmospheric phase

For the duration of the simulation (500 s), the four GPS satellites that were used to generate the measurements have been depicted in Fig. 7.16.[8] It is clear from this figure, that during the course of the simulation, the combination of satellites changes. In Fig. 7.17, an example of the (ideal) pseudorange measurements has been plotted, whereas in Fig. 7.18 an example of the carrier-phase differences can be found. It is noted that the integer ambiguity is assumed to be solved, which explains that the plotted values are not scaled between 0 and 1 (cycles). In both figures, the change in satellite constellation is represented by discrete jumps in the measurements.

After tuning the EKF, it was found that the position, velocity and quaternions can be estimated fairly well. In Fig. 7.19, the estimated ECEF position is plotted as a function of time. For x_R and z_R it is clear that during the first 30 s, the estimation error is rather large, for x_R even between 25 and 35 m. However, the error covariance almost discretely changes at $t \approx 30$ s, as well as the estimation error. From that moment onwards, the error remains very small.

This sudden jump in error and error covariance can easily be explained, when Fig. 7.16 is studied. At $t \approx 30$ s, the constellation of visible satellites changes. This change introduces new information in the system, which aids the filter convergence. Also at $t \approx 150$ s and $t \approx 330$ s, the constellation changes. The effect can be particularly well observed in Fig. 7.20, where the estimated ECEF velocity is plotted versus time. It is evident that the filter estimate of the x-velocity component slightly diverges (the filter is saturated) in between $t \approx 150$ s and $t \approx 330$ s. After the last constellation change, the filter rapidly converges again.

A similar effect is observed in the clock bias and clock drift, see also Fig. 7.21. The constellation changes are clear, but also the fact that the filter can very well estimate these parameters. The estimated quaternions are all very stable, with almost constant estimation errors of $\Delta q_1 = 4 \cdot 10^{-4}$, $\Delta q_2 = -3 \cdot 10^{-5}$, $\Delta q_3 = 7 \cdot 10^{-4}$,

[8] For the sake of reproducibility of the results, during the course of this study the date and time were frozen to 13-02-2002 and 10:55:27.277, respectively.

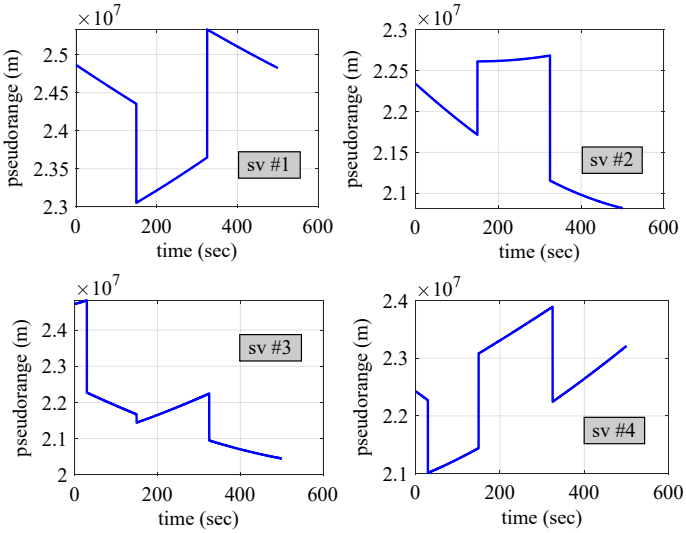


Fig. 7.17 Ideal GPS measurements for exo-atmospheric phase: pseudorange from four satellites

and $\Delta q_4 = 2 \cdot 10^{-4}$. Although it seems that for this reason the quaternions are not interesting enough to discuss in more detail, the opposite is true. The very fact that they are stable implies that, once converted to Euler angles, also these parameters will be stable. This means that the source of the attitude-angle error that was discussed at the beginning of this case study, has been removed by basing the model on the ECEF equations of motion rather than the tangent-plane equations. Apparently, not only the output matrix **H** was badly scaled, but also the system matrix **F** suffered (at least a bit) from the combination of position angles, altitude, and Cartesian velocity components.

In Fig. 7.22, some of the estimated instrument errors have been plotted, i.e., the difference between estimated and input value. It is obvious from the plots that the manoeuvres have a significant impact on the observability of the states, e.g., the scale factor for the x-accelerometer, s_x. Although the results seem to be acceptable, in fact only the accelerometer bias and gyroscope drift can be estimated accurately. The scale factors s_y and s_z cannot be estimated at all, since the corresponding accelerations are zero, which also holds for most of the accelerometer misalignments. The same was observed by Mooij and Chu (2001), where it was decided to keep only b_x, b_y, b_z, s_x, b_p, b_q, b_r, s_p, s_q, and s_r in the filter model. Although no deterioration of the results has been observed, also an improvement was not found. Therefore, the same reduced model will be used in the subsequent analysis of the actual re-entry. From this it is assumed that for the unobservable parameters the calibrated values from ground tests can be used.

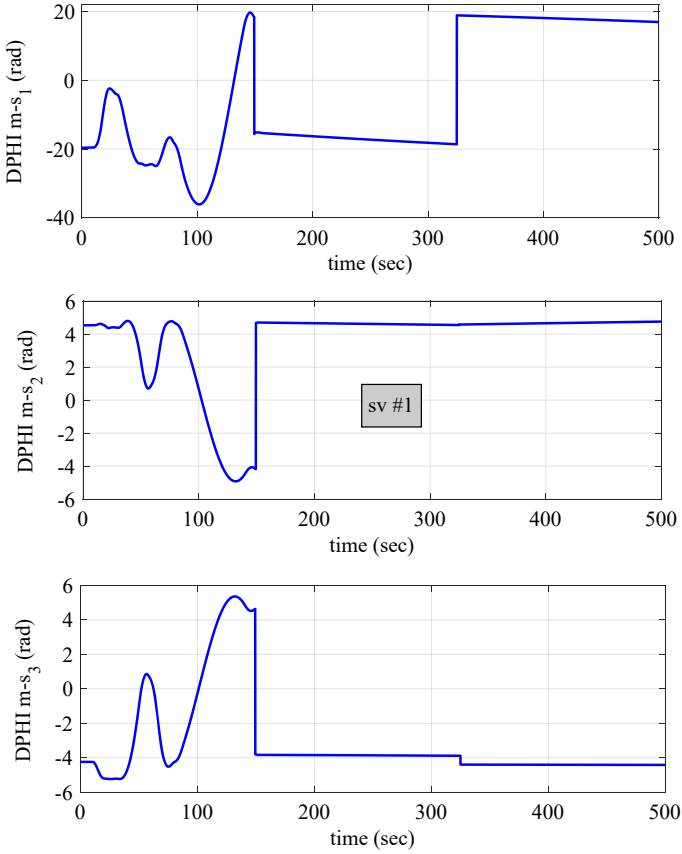


Fig. 7.18 Ideal GPS measurements for exo-atmospheric phase: carrier-phase differences from satellite #1 (integer ambiguity has been solved)

Re-entry Trajectory Simulation Results

The actual re-entry trajectory (Fig. 7.13) is divided into three parts. The first part deals with the flight phase starting at 122 km altitude until the blackout phase starts. This point has been set at an altitude of 88 km ($t = 190.5$ s). The second part concerns the blackout phase between 88 and 42 km ($t = 1066.7$ s). The third and final part begins right after leaving the blackout phase and continues down to an altitude of about 25 km ($t = 1280.5$ s), which marks the end of the hypersonic descent. The constellation of 4 visible GPS satellites is shown in Fig. 7.23.

In the first flight phase the gravitational acceleration is largely dominating. The aerodynamic forces and moments start increasing only close to the beginning of the blackout phase, which means that this phase will not give much extra information as compared with the exo-atmospheric phase. Adding s_z to the filter model—the high entry angle of attack makes the lift force the larger component, which would

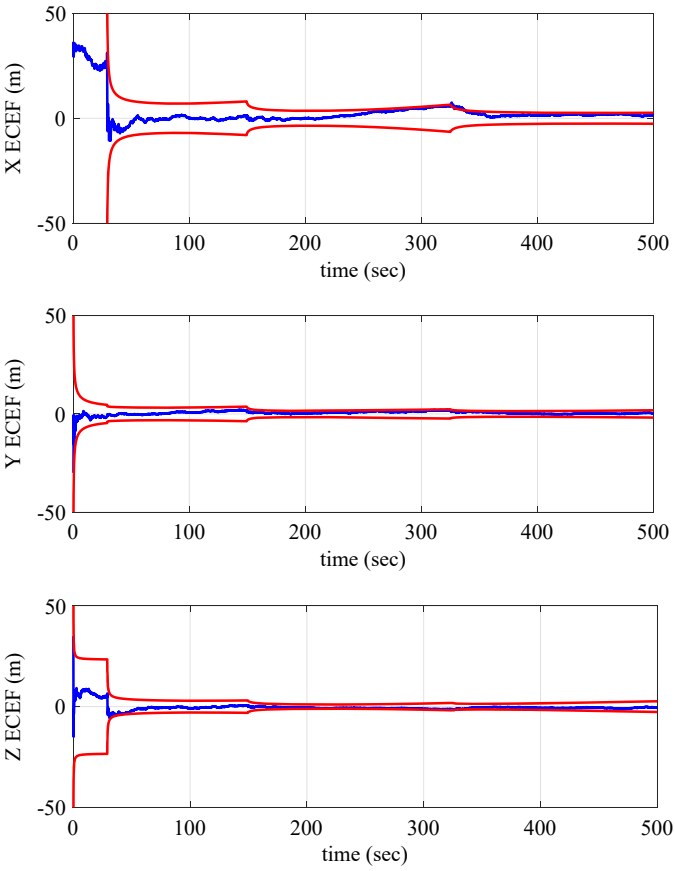


Fig. 7.19 Estimated ECEF position errors, including error standard-deviation bounds, for the exo-atmospheric phase

excite this parameter—shows convergence, albeit at the moment the blackout phase starts, no accurate estimate is obtained. Therefore, we restrict to the reduced model, as discussed before. The remaining "calibrated" scale factors and misalignments are zero initialised, which means a 100% uncertainty since they are excluded from the estimation process. The IMU measurement frequency is 200 Hz, whereas the GPS "update" frequency is 10 Hz. The carrier-phase noise is reduced to 0.026 cycle ($\approx$5 mm), to have a more realistic value for this parameter.

Since the current analysis concerns a new estimation process, the initial errors have been taken quite large, thereby discarding the information for the related states from the exo-atmospheric phase. The filter converges rapidly resulting in accurate navigation information. Converged errors are in the order of a few meters for the position components, 0.02 m/s or less for the velocity and less than 10^{-4} for the quaternions. The clock-drift and clock-bias errors are 0.04 m/s and 2 m, respectively.

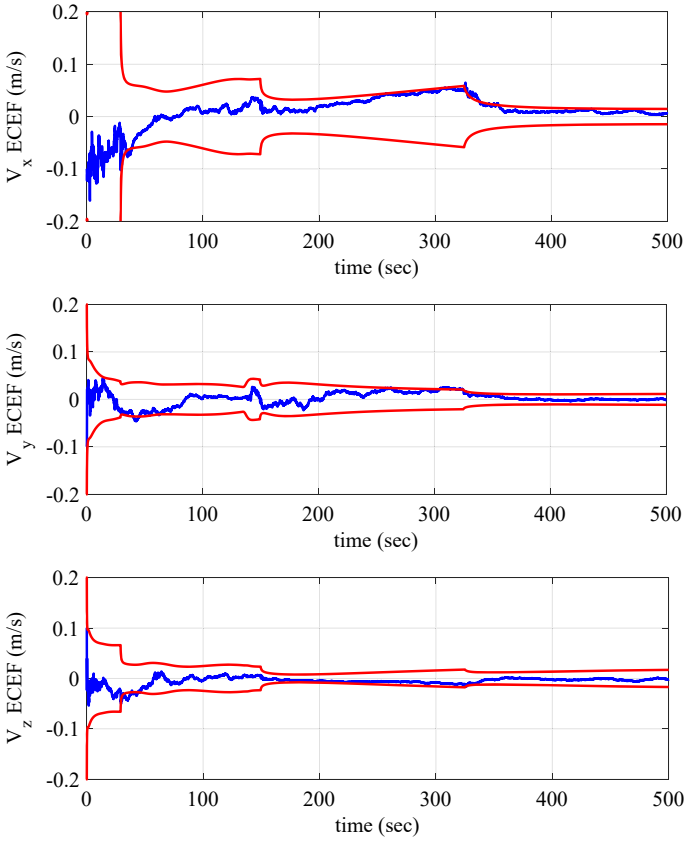


Fig. 7.20 Estimated ECEF velocity errors, including error standard-deviation bounds, for the exo-atmospheric phase

The estimated Cartesian position and velocity, as well as the quaternions, are transformed to spherical position h, τ, δ, local tangent-plane velocity v_N, v_E, and v_D, and local Euler angles ϕ, θ, and ψ, and compared with the actual values, resulting from the non-linear simulation. In Fig. 7.24, a subset of the results is plotted, i.e., the altitude error, the error in relative velocity (which is the modulus of the ECEF velocity components) as well as the pitch-angle error. It can be concluded that the filter has an excellent performance. An accurate pitch-angle estimate provides a good basis for accurate angle-of-attack control that is required for this type of re-entry (heat-flux minimisation). Final errors for the other parameters are $\tau_{err} = -2 \cdot 10^{-5\circ}$, $\delta_{err} = -10^{-5\circ}$, $v_{N,err} = -0.01$ m/s, $v_{E,err} = -0.025$ m/s, $v_{D,err} = 0.02$ m/s, $\phi_{err} = -1.5 \cdot 10^{-3\circ}$, and $\psi_{err} = -10^{-3\circ}$.

The black-out phase is characterised by the lack of aiding GPS measurements, which means that the navigation system should rely on the estimated instrument

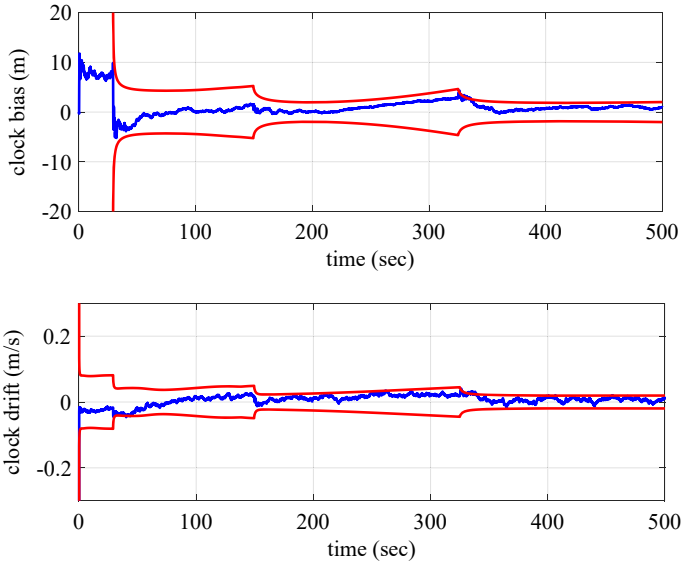


Fig. 7.21 Estimated clock-bias and drift, including error standard-deviation bounds, for the exo-atmospheric phase

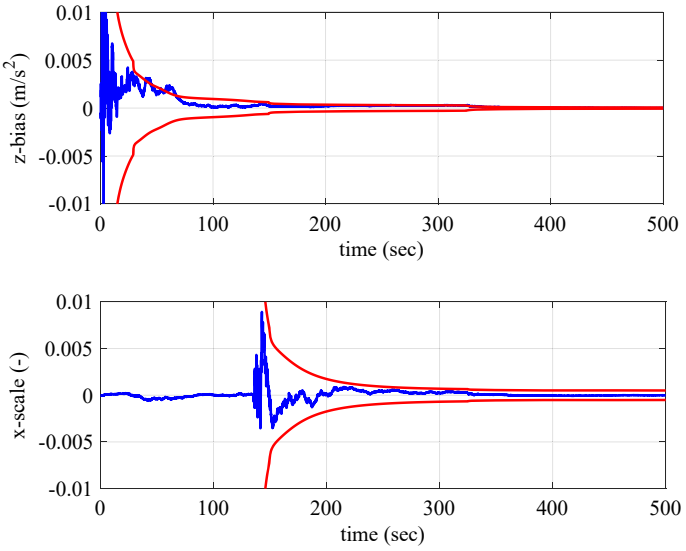


Fig. 7.22 Estimated z-bias and x-scale factor, including error standard-deviation bounds, for the exo-atmospheric phase

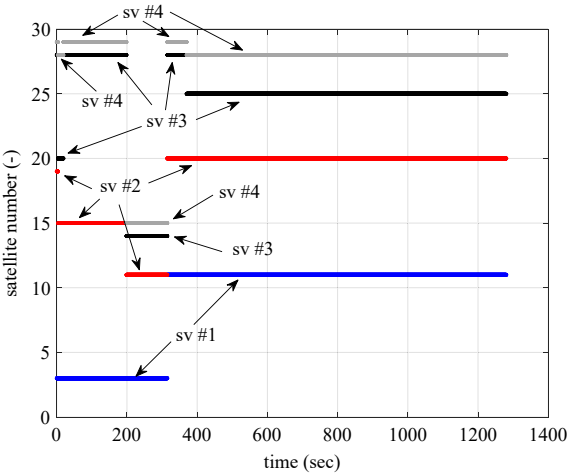


Fig. 7.23 Best combination of 4 visible satellites during re-entry (including blackout phase and final descent)

errors from the calibration phase and the previous first part of the entry. In the following, a dead-reckoning simulation will be done for the blackout phase with the best estimate of the instrument errors obtained from the previous phases as well as ground calibration. This means that the scale factors and misalignments will all have a 30% uncertainty as compared with their input values (Table 7.2), whereas the errors in b_x, b_p, and b_r are put at 25%. The remaining bias and drift values are assumed to be known. Initial conditions are set equal to the final values of the (first) re-entry phase.

In Fig. 7.25, the altitude, relative-velocity and roll-angle error have been plotted. Despite the relatively large IMU sample frequency, the error in, for instance, the relative velocity is rather large. It remains to be studied what the impact on the performance of the guidance system will be. At the moment that a bank reversal is initiated ($t \approx 730$ s), the error in attitude angles changes almost discretely, indicating that the IMU sample frequency is still too low for these fast dynamics, although the error is small. In turn, the attitude error results in a transformation and hence a small acceleration error, such that there is also an almost discrete change in the pattern of velocity error. Obviously, the decreasing rate in altitude error is a direct consequence of this. The remaining (maximum) errors are $\tau_{err} = -7.5 \cdot 10^{-3\circ}$ and $\delta_{err} = 4 \cdot 10^{-3\circ}$, which are gradually increasing with time, as is $v_{E,err}$ (final error of -4.4 m/s), $v_{N,err} = 1.6$ m/s, but with a rather constant error after the reversal, and $v_{D,err} = 1.25$ m/s at the moment of the reversal, but which decreases again to zero afterwards. Finally, $\theta_{err} = 0.05^\circ$ and $\psi_{err} = 0.025^\circ$.

Initialising the Kalman filter with the above state-variable errors ($x_{R,err} = 940$ m, $y_{R,err} = 115$ m, $z_{R,err} = -423$ m, $u_{err} = 3.6$ m/s, $v_{err} = 2.65$ m/s and $w_{err} = -1.65$ m/s; the quaternion errors are all very small) and taking large initial values for the error covariance matrix, the estimation process for the final descent is started. As for the first phase, the filter converges rapidly, again showing a good per-

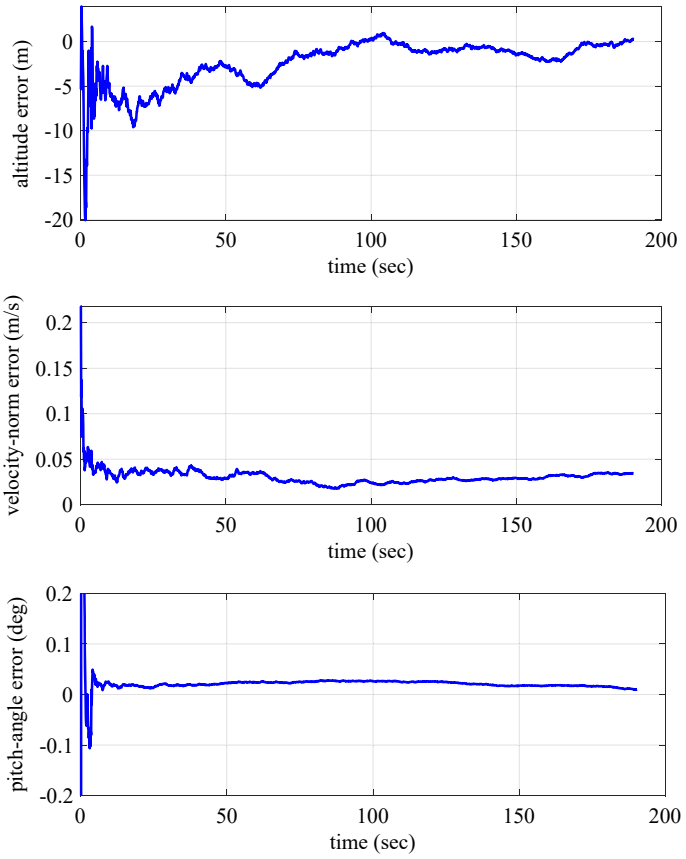


Fig. 7.24 Error in estimated altitude (top), relative velocity (middle) and (local) pitch angle (bottom) as a function of time since re-entry

formance. Figure 7.26 shows a subset of errors. The altitude and velocity error rapidly converge to small values, which also holds for the errors that are not shown here. After inspecting the roll-angle error, it is clear that there are no large errors during the bank reversals anymore. The filter behaviour is very stable, which is promising for an actual on-board implementation of the presented navigation system.

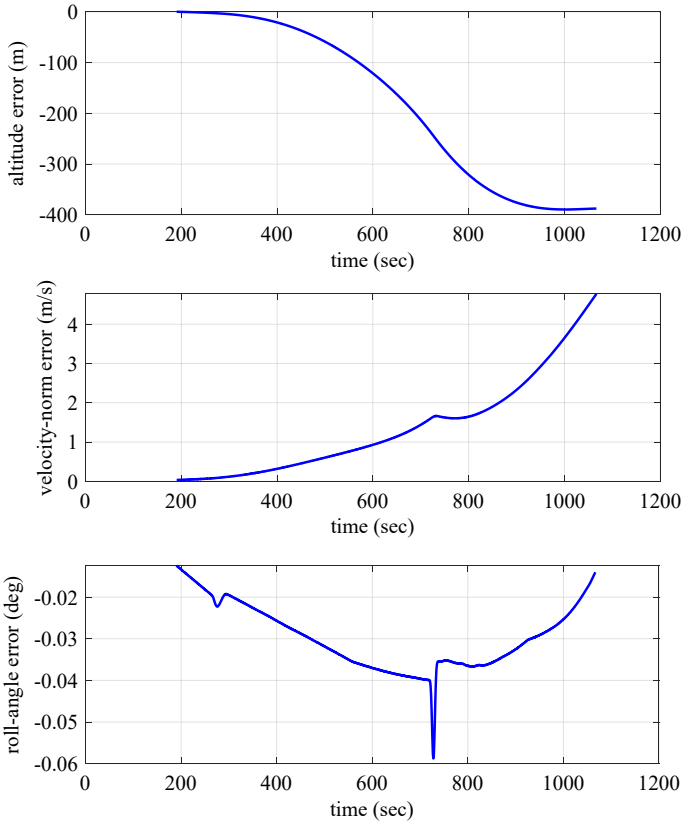


Fig. 7.25 Error during blackout phase in estimated altitude (top), relative velocity (middle) and (local) roll angle (bottom) as a function of time since re-entry

Conclusions

In this case study a tightly-coupled IMU/GPS integrated navigation system for re-entry vehicles has been presented. Despite the limited manoeuvre possibilities, the on-board calibration of inertial sensors has been addressed. Not all modelled uncertain parameters of the IMU can be estimated, although the accelerometer bias and gyroscope drift can be estimated at a satisfactory level. Position, velocity and attitude can be accurately estimated, with a filter performance that is much better than the loosely coupled approach presented in a previous paper. An implementation based on ECEF position and velocity, as well as quaternions rather than Euler angles, gives a numerically stable filter which is relatively fast.

Simulating the hypersonic entry and descent yields accurate position, velocity, and attitude estimates in all phases of the re-entry, including the blackout phase.

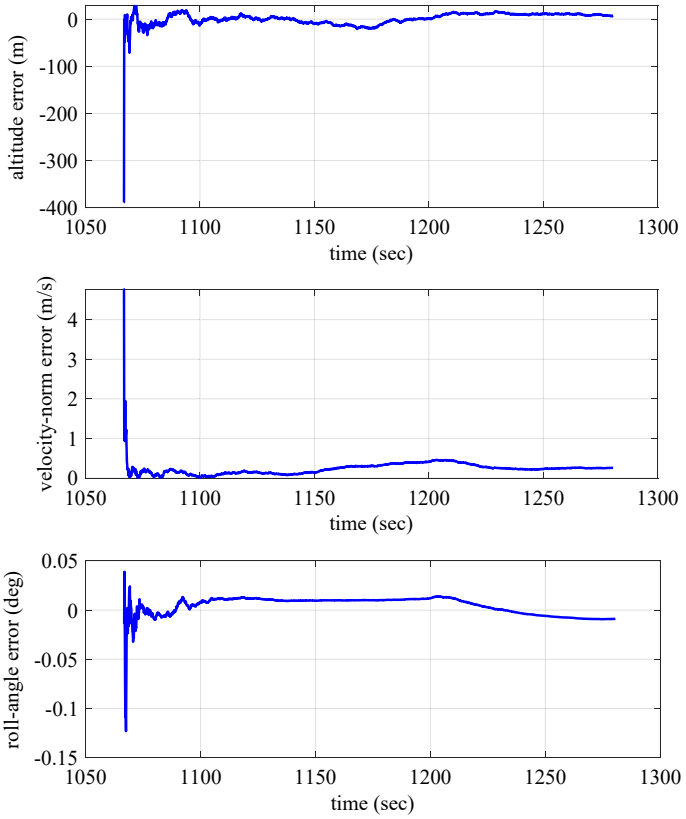


Fig. 7.26 Error during final descent in estimated altitude (top), relative velocity (middle) and (local) roll angle (bottom) as a function of time since re-entry

Also during the (excessive) bank reversals, the attitude-angle error remains small and does not induce significant acceleration errors.

The tightly coupled GPS/IMU integrated system offers flexibility to improve the performance of the navigation system by remodelling errors of the GPS receiver, while in the loosely coupled system the position, velocity and attitude are already estimated by the receiver. However, the latter approach is less realistic, because only white-noise errors in the GPS measurements are assumed. In reality, the processed GPS data will be corrupted with coloured measurement noise, because pre-Kalman filters have already been applied to process the GPS data. Compared to loosely coupled or uncoupled systems, a tightly coupled system offers the highest accuracy.

Currently, the real-time implementation of the integrated navigation system has not yet been addressed. In addition, more attention should be paid to the manoeuvre

design, to allow for a more accurate calibration of the inertial sensor. Finally, future work could focus on the combination with the guidance and control system of the re-entry vehicle, to allow for a realistic performance analysis.

7.4 Actuators

Actuators can be defined as actively controlled devices that are capable of exerting forces (and resulting moments) on a vehicle and so altering its state, whether it be orientation, position, or velocity. In some situations, though, a change in vehicle configuration can be counted as actuation, even though subsequently this configuration remains constant. One can think of, for instance, parachute systems or spin vanes, which only appear after, for instance, a heat shield has been deployed. In this text we will call these actuators passive actuators, although we are fully aware that parachutes can also be actively controlled (steerable parachutes called parafoils, or the stepwise opening of conventional parachutes, a process called reefing). Most entry vehicles are equipped with actuators, to follow up on the guidance commands for load control or targeting the landing site. In either case, this will ensure that the vehicle will survive the severe entry, descent, and landing such that its occupants can safely conclude their mission or science instruments can be fully operational to start theirs. The Space Shuttle is by far the most familiar vehicle when it comes to returning astronauts from space, even though the last Shuttle has flown. Its aircraft-like features hints at aerodynamic control surfaces as actuators, but since the vehicle also needs to be controlled outside and in the upper layers of the atmosphere it needs to have alternative actuators that are independent of the interaction with the atmosphere. A propulsion system is a common choice for such an actuator. In Fig. 7.27 the actuators that were used by the Space Shuttle (or, equivalently, the Russian shuttle Buran) are shown.

Planetary entry probes usually need not only to survive the hypersonic entry and descent, but also the landing on the surface of the planet or moon they are targeted for. In that way the scientific instruments can start (or continue) to work for a longer period of time. Depending on the density of the atmosphere, active control of the vehicle state is required to achieve this. The European Huygens probe that descended in the Titan atmosphere was an uncontrolled, ballistic entry vehicle, but required a parachute to prolong its descent to increase the measurement time, despite the relatively thick atmosphere. Because of this dense atmosphere, the parachute could also not be too large, as a too long descent time would result in loss of contact with the Cassini orbiter. Mars probes, on the other hand, should have a maximum parachute drag area, because the atmosphere is very thin and does not decelerate the vehicle nearly enough to survive landing. Therefore, these kind of probes usually has so-called retro rockets to reduce the landing velocity to an acceptable level.

It is clear that actuators form an important subsystem of the entry vehicle. This section will discuss the most commonly used active actuators, i.e., control surfaces, reaction control thrusters and retrorockets, in Sects. 7.4.1, 7.4.2, and 7.4.3.

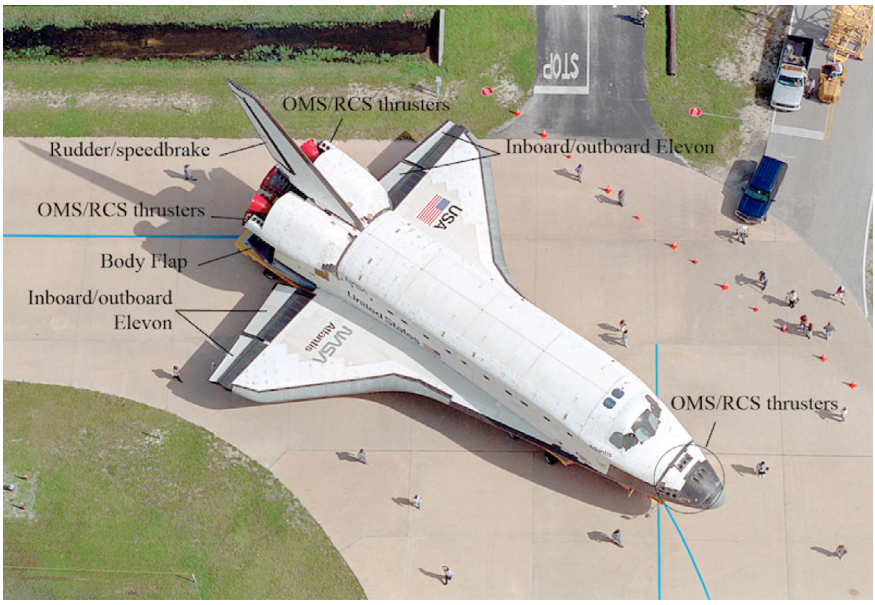


Fig. 7.27 Space Shuttle Atlantis (STS-86) Rollover at Kennedy Space Center with control surfaces and thrusters indicated (OMS = orbital maneuvering system, RCS = reaction control system). Image adapted from: NASA

In Sect. 7.4.4 we have grouped the remaining actuators, i.e., reaction wheels, moving mass systems and passive actuators.

7.4.1 Control Surfaces

Control surfaces are the most commonly used actuators on aircraft, both civil and military. Looking at any winged re-entry vehicle (e.g., Fig. 7.27) shows that such a configuration has been adapted for spacecraft as well. But not only aircraft-like entry vehicles, also for so-called lifting bodies control surfaces are the actuators of choice, albeit sometimes the implementation is less conventional. A typical example is the HL-10, shown in Fig. 7.28. The indicated subsonic control-surface configuration includes rudders and speed brakes, tip-fin flaps, elevons, and elevon flaps. Elevons are control surfaces that combine the so-called *elevator* (=pitch control) and *aileron* (=roll control) function. It is noteworthy to mention that all control surfaces are so-called split surfaces, i.e., they consist of two surfaces that can move in opposite direction. For example, if the (split) rudders both move outward they only generate drag, and the rudder is in speed-brake mode.

Another example of a lifting vehicle is the X-33 (Fig. 7.29), which has canted fins, vertical fins and body flaps to control its motion. Also in this case, the aerodynamic

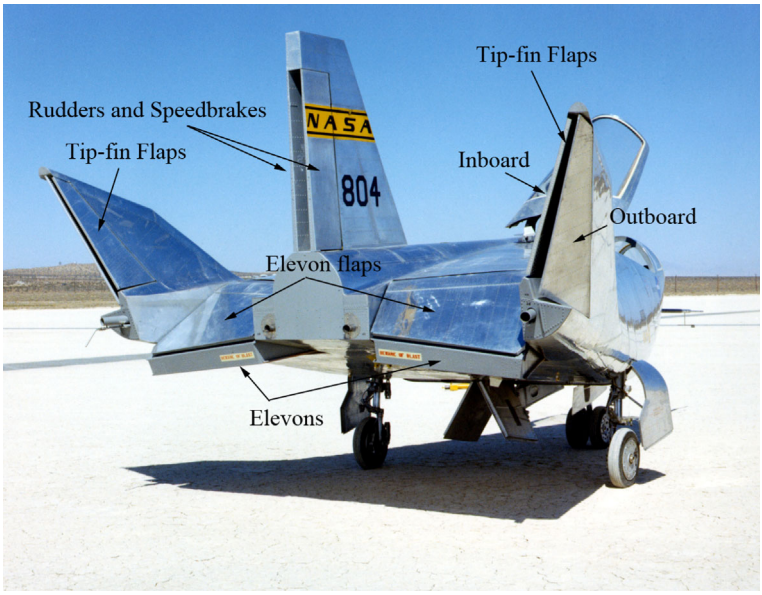


Fig. 7.28 HL-10 lifting body on lakebed showing subsonic control surface configuration. Image adapted from: NASA

configuration is not so conventional that one could have used only the conventional control surfaces elevator, aileron and rudder. Due to the vehicle's aerodynamic characteristics and typical control-surface orientation the control function of the surfaces extends to more than one axis. For instance, the body flaps are not only used for trimming the vehicle's longitudinal motion, but can also control the yaw motion by asymmetrically deflecting them and thus creating a moment due to differential drag.

And last but not least—towards the other side of the spectrum—experimental vehicles can sometimes suffice with only a subset of the available control surfaces. As an example, in Fig. 7.30 the Intermediate eXperimental Vehicle, IXV, is shown. This vehicle was developed to validate technology for future European reusable-launchers. It was a lifting body with only two elevon-like control surfaces that were used to rotate the lift vector out of plane and thus inducing lateral motion.

In the remainder of this section we will discuss the control surfaces of the Space Shuttle in somewhat more detail. The Space Shuttle was using hydraulic actuators for the actuation of not only the aerodynamic control surfaces, but also the engine gimballing, engine control, landing gear control, and nose-wheel steering. The actuator hydraulic pressure was provided by the three auxiliary power unit (APU) driven hydraulic systems. The hydraulic systems were using a nominal pressure of 3000 psi (or 20.8 MPa). The hydraulic/actuation system was designed to provide safe vehicle operation of the vehicle after the loss of one hydraulic system. After the loss of two hydraulic systems the maximum actuator deflection rate would be approximately reduced to 50%, which would still provide safe operation. There would be limitations

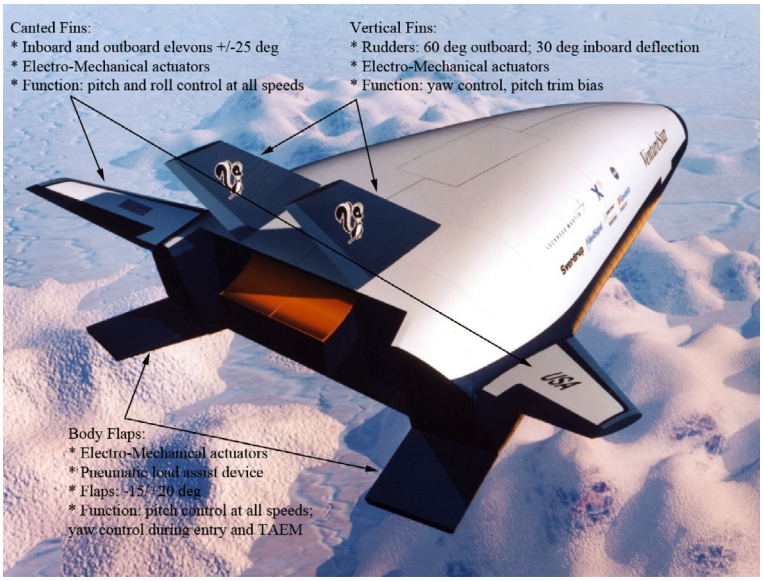


Fig. 7.29 The X-33 with its control surfaces. Image: NASA



Fig. 7.30 The Intermediate eXperimental Vehicle, IXV. Image: ESA

in terms of turbulence and crosswind at approach and landing, though. Although the hydraulic/actuation system is maintenance heavy, the system has proven itself in the many years of aircraft operation and is very reliable.

The aerodynamic surfaces included the body flap, two inboard and two outboard elevons, and a single split rudder (see Fig. 7.31). The body flap was used for "long-term" pitch trimming of the vehicle and to offload the elevons, so they could be kept in low hinge-moment conditions. This improved the elevon-actuators dynamic. The

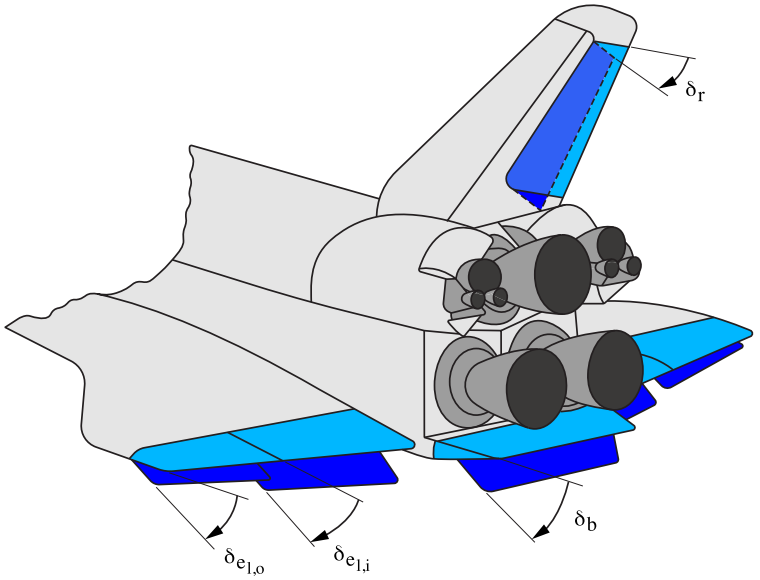


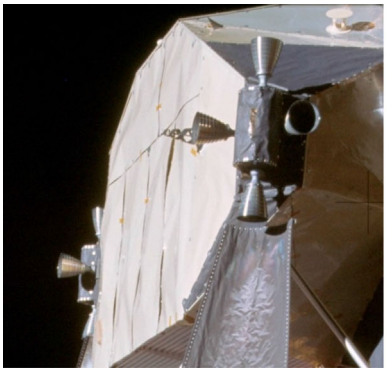
Fig. 7.31 The Space Shuttle control surfaces. Image inspired by Rosello (1995)

body flap was also used to shield the main engines from extreme thermal conditions during re-entry. The body flap was controlled by three hydraulic motors each connected to one hydraulic system. Each hydraulic motor had a brake to hold the motor position when the flap was at the commanded position. It provided shaft power to a different gearbox, which drove a rotary shaft and four rotary actuators. These would position the body flap. The body flap could be deflected between 12° down and 22° up. The deflection rate was unknown, but is expected to be on the order of 1–2°/s. Both sides of the body flap were covered with high-temperature reusable-surface insulation (RSI) tiles.

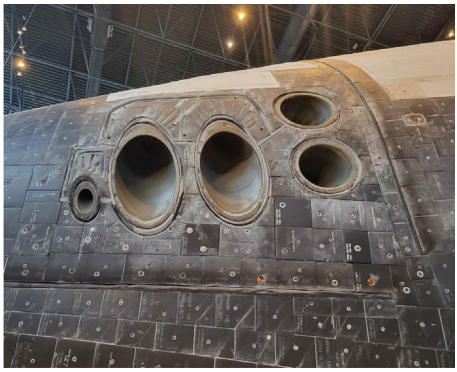
The elevons (also called wing flaps) were split in one outboard and one inboard elevon per side (a total of four surfaces). Each elevon was actuated by a linear hydraulic actuator, which was supplied with hydraulic pressure from the three hydraulic system. A switching valve was used to control the hydraulic system that became the source for the actuator. Either elevon deflected 40° upward and 25° downward, and had a maximum deflection rate of 20°/s.

The rudder was used both for directional control and also as a speed brake during the descent and approach phase. The surface was actuated by a hydraulic power unit installed at the lower trailing edge area of the fin. The maximum deflection when acting as a rudder was ±27° and ±61° when in speed-brake mode.

In terms of modelling the control surfaces, it is usually not necessary to include the actuator dynamics. If there would be a need one could stick to a simple second-order system with appropriate values for the natural frequency, ω_n, and damping, ζ. Doyle et al. (1986) have defined second-order models for the elevons with $\omega_n = 14$



(a) Reaction-control thrusters mounted on the Apollo Lunar Module.



(b) RCS thrusters on the nose of the Space Shuttle.

Fig. 7.32 Examples of RCS thrusters. Image left: NASA, image right: author (Space Shuttle *Atlantis*, on display at the Kennedy Space Center)

rad/s and $\zeta = 0.72$, and for the rudder with $\omega_n = 21$ rad/s and $\zeta = 0.75$. However, in most cases it is sufficient to monitor (and restrict to) the minimum and maximum deflection angles, as well as the maximum deflection rate.

7.4.2 Reaction Control System

A reaction control system (RCS) is a subsystem of a spacecraft whose purpose is attitude control and steering by the use of thrusters. An RCS system is capable of providing small amounts of thrust in any desired direction or combination of directions. An RCS is also capable of providing torque to allow for control of rotation around each of the three axes (roll, pitch, and yaw). This is in contrast to a spacecraft's main engine (e.g., used for the deorbit burn or the landing boost), which is only capable of providing thrust in one direction, but is orders of magnitude more powerful. RCS systems often use a combination of larger and smaller (vernier) thrusters, to allow for different levels of response from the combination. A typical example of RCS thrusters mounted on the Apollo Lunar Module is shown in Fig. 7.32a. Note the orthogonal mounting to allow for perpendicular axis-control; also, each axis direction has (at least) two thrusters in opposite directions to control the rotation direction.

Reaction control systems are used for:

1. for attitude control during re-entry, when control authority of the aerodynamic control surfaces is too low,
2. for stationkeeping in orbit,
3. for close manoeuvring during docking procedures,
4. for control of orientation, or "pointing the nose" of the craft, and
5. as a backup means of deorbiting.

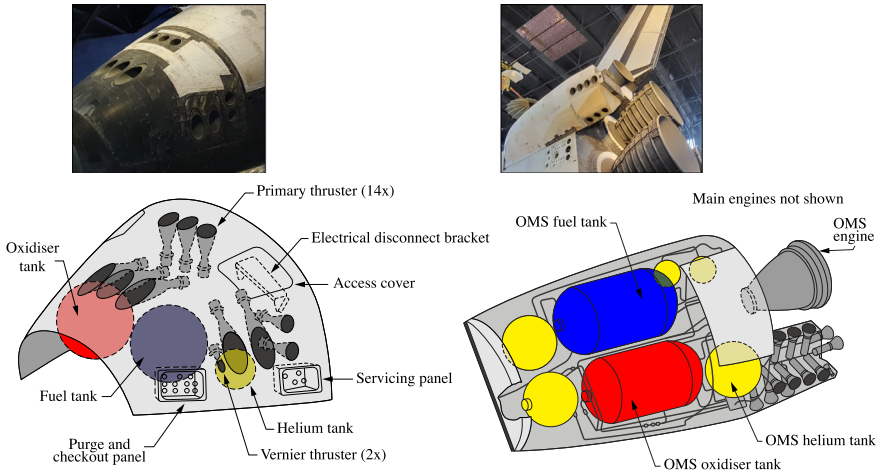


Fig. 7.33 RCS pods on the Space Shuttle. Drawings inspired by Rosello (1995), image inserts: author (Left image: Space Shuttle *Atlantis*, on display at the Kennedy Space Center; right image: Space Shuttle *Discovery*, on display at the Smithsonian Air and Space Museum Steven F. Udvar-Hazy Center)

Compared to the number of axes the Space Shuttle had to control, it seems to have many more thrusters than strictly necessary. However, the Shuttle was also required to carry out docking manoeuvres in orbit. Shuttle thrusters were grouped in the nose of the vehicle and on each of the two aft Orbital Maneuvering System (OMS) pods. No nozzles interrupted the heat shield on the underside of the craft, instead, the nose RCS nozzles that controlled positive pitch were mounted on the side of the vehicle, and were canted downward (Fig. 7.32b). The downward-facing negative pitch thrusters were located in the OMS pods mounted in the tail/afterbody. See for a schematic layout of both front and rear thrusters Fig. 7.33, and for some detail on the rear thruster pods Fig. 7.34.

In terms of modelling, four factors alter the effect of jet firings during reentry: plume interaction with the aerodynamic flow field, plume self-impingement, the increase in atmospheric pressure, and thruster valve on/off transients. Rosello (1995) provides models for each of these effects, of which plume interaction with the flow field is the most important one. For the aft jets, the effect may be so large that the net resulting moment opposes the nominal moment expected from the vacuum thrust alone. So, modelling the RCS thrusters seems to be important enough, yet for re-entry control studies this is not always done, possibly because the phase where only thrust control is used, is relatively short. In that case one can model the thruster as a (scalable) moment that can be achieved, taking only the minimum and maximum values (typically the same, albeit with opposite signs) into account. A higher level of detail would be to only account for transient effects due to thruster on/off transient effects, which can be modelled as a second-order transfer function. Rosello (1995) used the following values for the natural frequency and damping factor: $\omega_n = 200$ rad/s and $\zeta = 0.9$.

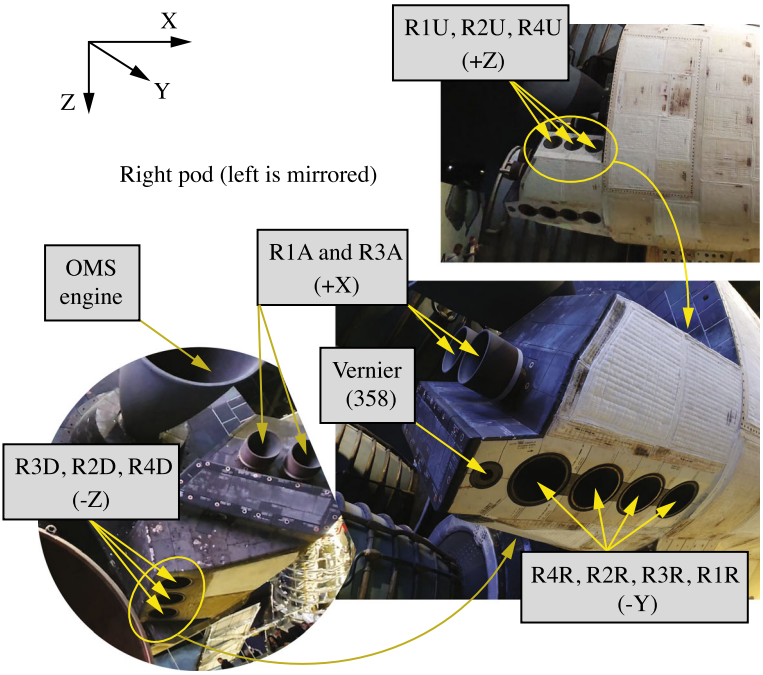


Fig. 7.34 Space Shuttle rear RCS pods. Images: author (Insert images are of Space Shuttle *Atlantis*, on display at the Kennedy Space Center)

7.4.3 Retro Rockets

With the absence of an atmosphere, or an atmosphere that is too thin to provide sufficient deceleration by aerodynamic means, an entry vehicle requires additional means to control the landing velocity. Therefore, such systems are equipped with so-called terminal descent propulsion systems. To support the Apollo program science and technology development, in the 1960s NASA undertook the Surveyor programme, a series of robotic missions to land on the Moon. On June 2, 1966, the first Surveyor performed a soft landing on the Moon, just three years prior to the first manned Moon landing, and was the very first American propulsive lander on another planetary body. To do so, Surveyor had three so-called vernier thrusters (i.e., smaller thrusters commonly used for attitude control). The vernier propulsion system was one of the most difficult developments undertaken by the Surveyor project. The system consisted of three thrust-chamber assemblies, three pairs of fuel and oxidiser tanks, a high-pressure helium tank, and a host of lines and valves needed for loading, test, and operation of the system (see also Fig. 7.35). One of the three engines could be gimbaled $+6°$, about an axis allowing its use for roll control of the spacecraft. Each of the engine's thrust level was individually throttleable between 133 and 471 N (JPL 1966).

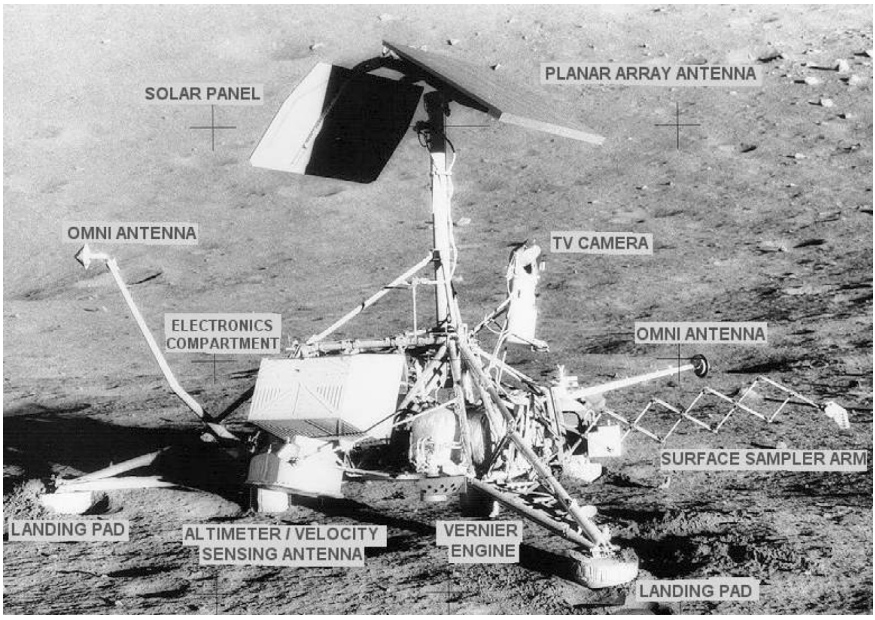
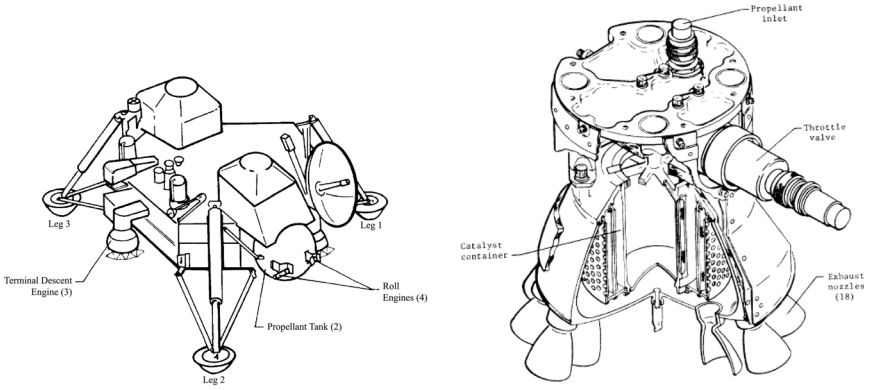


Fig. 7.35 Surveyor spacecraft configuration. Photo taken by taken by Apollo 12 astronauts. Image: NASA



(a) Location of components (b) Engine

Fig. 7.36 Viking terminal descent propulsion system. Images: NASA (Holmberg et al. 1980)

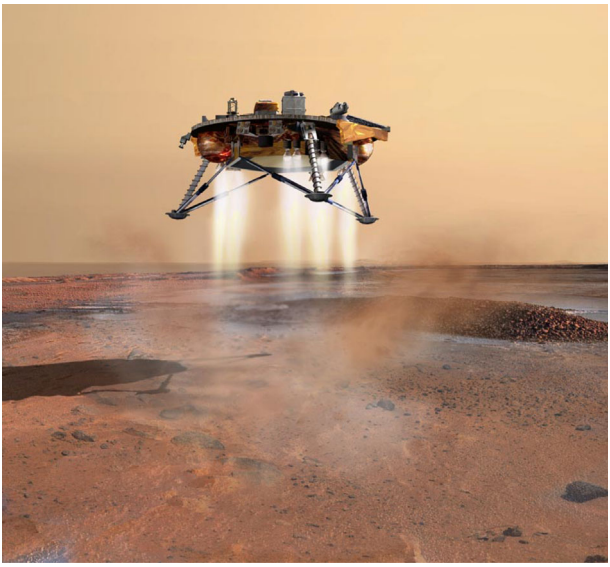
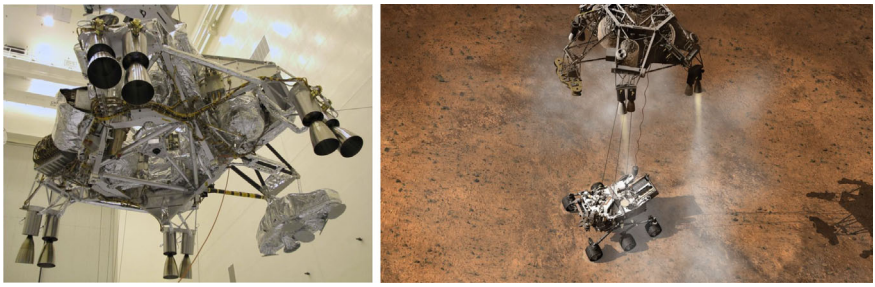


Fig. 7.37 The Phoenix Mars lander. Image: NASA

For missions to Mars, which has an atmosphere, albeit not as dense as compared to Earth, in addition to a parachute a similar terminal descent system (TDS) is commonly used. For instance, the Viking spacecraft used a TDS that provided three-axis attitude control and the impulse required to reduce the lander velocity to approximately 2.4 m/s at touchdown (Holmberg et al. 1980). The TDS was activated at aeroshell separation, and consisted of three large (high-thrust) terminal descent engines, four small roll engines, and two propellant tanks along with the necessary feed lines and pyrotechnic valves. All components of the TDS were mounted on the lander (Fig. 7.36). The landing rockets used an 18-nozzle design to spread the hydrogen and nitrogen exhaust over a large area. NASA calculated that this approach would mean that the surface would not be heated by more than one degree Celsius, and that it would move no more than 1 mm of surface material. Since most of Viking's experiments focussed on the surface material, a more straightforward design would not have served. Many of the missions that followed Viking employed a similar configuration. In Fig. 7.37 an artist impression of the Phoenix Mars lander is shown. After the lander had decelerated to Mach 1.7 it opened its parachute to continue the descent. This phase lasted until it came within 1 km of the Martian surface. At that point, the parachute was jettisoned and the landing engines throttled up. At 12 m altitude the velocity had reduced to 2.4 m/s, which was kept constant until landing. The landing engines were turned off when sensors located on the footpads of the lander detected touchdown (NASA 2011).

A more specialised application of retro rockets is found in the US mission Mars Science Laboratory (MSL). Launched in November 2011 it reached Mars in August



(a) Descent stage. (b) Sky crane.

Fig. 7.38 Mars Science Laboratory. Images: NASA/JPL

2012. The sheer size of the Mars Science Laboratory rover (775 kg) would preclude it from taking advantage of an airbag-assisted landing (JPL 2011). Instead, MSL used the sky-crane touchdown system (Fig. 7.38), which is capable of delivering a much larger rover onto the surface. It has placed the rover on its wheels, ready to begin its mission. The spacecraft was controlled by small rockets during descent through the Martian atmosphere, toward the surface. Like Viking, Pathfinder, and the Mars Exploration Rovers, the Mars Science Laboratory was slowed by a large parachute. Rockets controlled the spacecraft's descent until the moment the rover separated from its final delivery system, the sky crane. Like a large crane on Earth, the sky crane system has lowered the rover to a "soft landing" with wheels down on the surface of Mars. This sky crane was hovering 20 m above the surface, when the rover was lowered down, and descended to about 7 m when the rover was landing. After detaching the rover the sky crane flew away for a crash landing at a safe distance. This very complex landing scenario turned out to be very successful, and as of September 2020 Curiosity is still active.

7.4.4 Other Actuators

Of course there are many other actuation devices, one may think of reaction wheels, magneto torquers, parachutes, etc. However, not all of them are applicable to entry systems, because the required torques can be quite demanding due to the aerodynamic moments. Therefore, reaction wheels and magneto torquers will not be considered here; maybe they could be used during the exoatmospheric phase, but that phase is excluded from the current discussion. Therefore, we will restrict ourselves to three actuator candidates, notably parachute systems (frequently used during planetary entry and descent), moving-mass systems (an internal system with no interaction with the outside world), and so-called passive actuators. A typical example of the latter are spin vanes that are used to keep the spin profile of entry probes within limits.

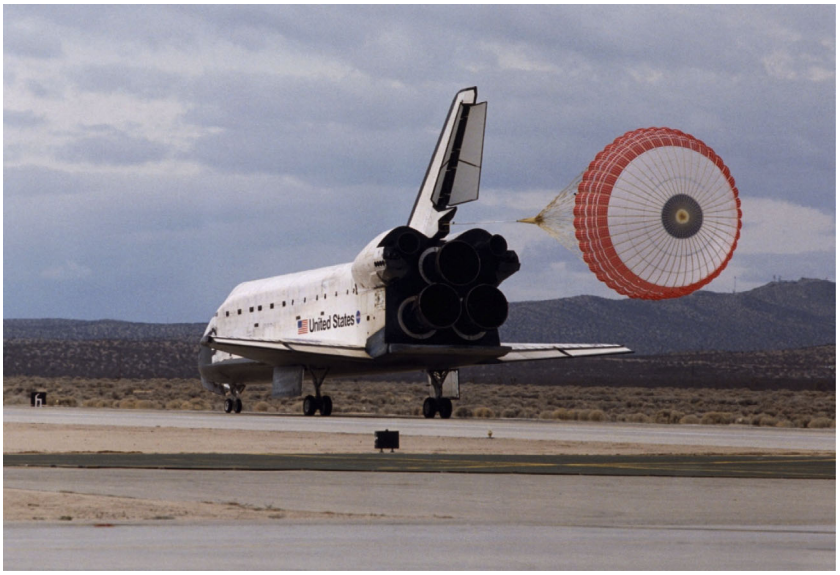


Fig. 7.39 Space Shuttle Atlantis landing on the runway at Edwards Air Force Base, California (February 20, 2001). Use of brake parachute and split rudder as speedbrakes. Image: NASA

7.4.4.1 Parachutes

Parachutes are commonly used decelerators for entry capsules, whether it being for an Earth, Mars, or Titan entry. Depending on the type of system to be decelerated, parachute systems can come in clusters (the Apollo capsule had three main parachutes), or have a number of different parachutes, i.e., a pilot chute that pulls out a stabilising drogue chute, which in turn will pull out the main chute. Deployment can either be at subsonic or at supersonic speeds. The latter is required for a Mars landing, because due to the very thin atmosphere and thus short deceleration time, deployment should take place as early as possible. Chapter 9 is dedicated to entry, descent, and landing systems, where a detailed description of parachute systems for planetary landing can be found. In the current context, we will just mention the use as a brake parachute as used on the Space Shuttle (see Fig. 7.39). Just before landing, a brake parachute is released to assist decelerating the Shuttle once landed on the runway. The Space Shuttle had quite a high landing velocity, which meant a rather large load on the Shuttle's braking system. It can also be seen in this figure, how the rudder is used as speed brake, for the purpose of drag increase to reduce the velocity.

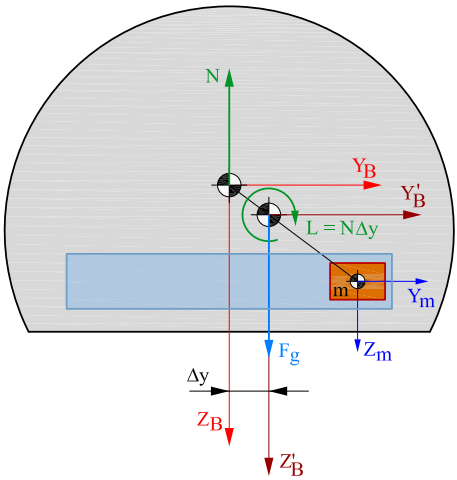


Fig. 7.40 Rolling moment produced by a moving-mass roll-control system (rear view)

7.4.4.2 Moving Mass

The ability of a moving-mass roll control system to control the roll attitude and trajectory of a simple fixed-trim re-entry vehicle has been investigated by Petsopoulos et al. (1996). The simplicity of a moving-mass roll control system, combined with its unique ability to provide roll control from within the re-entry vehicle's protective aeroshell, makes it an attractive alternative to more traditional aerodynamic or thruster-based roll control systems. These systems either require powerful actuators or a complicated system of fuel tanks, piping, and valves. It was indicated in this study that a battery-powered electric motor is sufficient to drive a moving-mass element during re-entry. This mass element slides in a rail, as indicated in Fig. 7.40. Moving the mass to the left or right affects the location of the global centre of mass. The aerodynamic (normal) force, N, which has a fixed point of action, will create a rolling moment, $\mathcal{L}$.

It was shown that a 4.5 kg moving mass with a 2.54 cm full-span travel was all that was needed to provide adequate roll response for the 180-kg re-entry vehicle. The full-span travel was kept small, along with the average actuator slew rates, to minimise the weight, volume, and power requirements of the control system. The conclusion of the study was that despite the limitations of this preliminary investigation, moving-mass roll control appears to be a relatively simple and effective method of modulating the roll angle of a lifting re-entry vehicle, and it deserves serious attention as a viable alternative to traditional aerodynamic flap- and thruster-based roll control systems.

However, as far as re-entry vehicles are concerned, it never seems to have materialised as hardware system or even an in-flight experiment. A few studies can be found in literature, though. Byrne et al. (1996) discuss the design of a moving-mass trim control system for manoeuvring axisymmetric re-entry vehicles. The moving mass trim controller is composed of three equal masses that are independently positioned to deliver a desired centre-of-mass position. For a slowly spinning re-entry

vehicle, the mass offset creates a trim angle-of-attack to generate modest flight-path corrections. Balaram (2005) designed Sherpa, a moving-mass entry, descent, and landing system, which was applied to a ballistic Mars entry vehicle. The Sherpa system takes a small fraction (5–10%) of the total system mass, and is capable of delivering payloads to within the proximity (≈5 km) of a wide variety of landing sites, including those at high-altitude (more than 2.5 km above the reference geoid). Finally, Atkins and Queen (2015) present a moving-mass actuator control system for angle-of-attack and sideslip-based Mars precision entry guidance. With two internal moving-mass actuators to track the guidance commands, they showed that a Mars Phoenix-sized entry vehicle can achieve precision landing, despite errors in the initial entry conditions for flight-path angle, range, cross range, speed, and attitude.

7.4.4.3 Passive Actuators

Many spacecraft sent into planetary atmospheres (including the Earth's) have used spin for stabilisation. This stems from the requirement to maintain a desired orientation of the vehicle for hypervelocity entry. Another reason to spin up an entry probe may be from an instrument point of view: when optical devices need to sample in several directions a spin up of the probe may be the simplest solution rather than designing a complicated motion system for the instrument. The generation of spin on a planetary entry probe is usually performed at, or just before, the moment of separation from the carrier vehicle. Interaction with the atmosphere and asymmetric aerodynamic loading from various protrusions from the descent vehicle has an almost guaranteed effect on the spin rate. The issue gets complicated when (part of) the descent takes place under a parachute. To decouple the motion of the spinning probe from the parachute, commonly a so-called swivel is introduced in between the riser that connects probe with the parachute. However, interaction with the parachute motion can still have an effect on the spin rate.

To influence the spin rate and to make sure that it remains bounded to a range of required values a probe can be equipped with so-called spin vanes. These vanes are small aerofoils, mounted at certain locations of (or even all around) the descent module, with an offset angle with respect to the on-coming airflow. The spin vanes are mounted in a fixed orientation and are only passive actuators. They have been applied in several missions, e.g., the Pioneer Venus, the Galileo entry probe, and most recently, Huygens.

The Pioneer Venus large probe—the mission included one large and four smaller probes—was built as a rigid sphere to resist the hot high-pressure atmosphere (Fig. 7.41a). After separation of the heat shield the spin vanes, which were mounted all around the spheroid body caused a gentle rotation (it must be noted that there was no strict requirement on the spin rate). The Galileo probe was equipped with 3 spin vanes to control its spin rate during parachute descent: two mounted at about 13°, and one at about 3°. To avoid excessive Doppler shift on the radio link the probe's spin design was principally to ensure a spin of less than 50 rpm. The three vanes

were also used to balance asymmetric aerodynamic loads from various protrusions from the probe, which explains the different orientations.

For Huygens, the mission scenario required that the Huygens descent module was spinning throughout the descent phase, to enable an optical camera to continuously scan the Titan horizon (Huygens' flight-path angle $\approx -90°$). In Fig. 7.41b a close up of the spin vanes is shown (the model shown in the picture is the Special Model 2 (SM2) used for drop tests from a stratospheric balloon). The Huygens spin-vane configuration had many vanes with a very shallow setting angle of just 2–3°, to generate a spin nominally in the 1–4 rpm range during most of descent. The large number of vanes is required to provide enough torque to reach the desired rate quickly (Lorenz 2006).

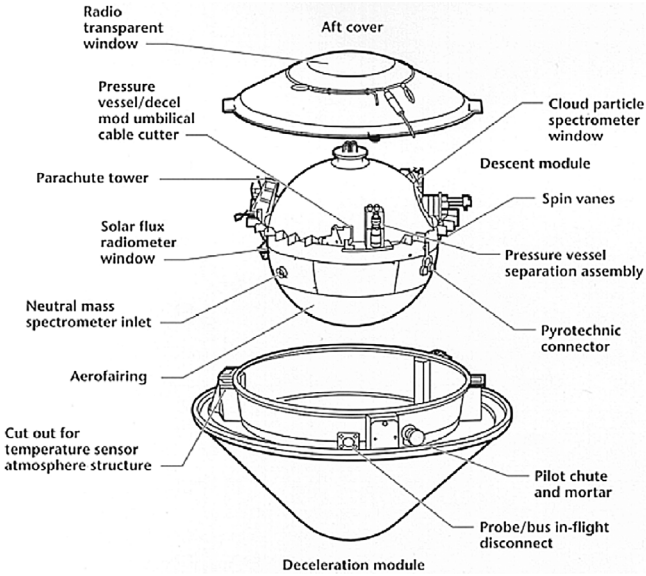
7.5 Control

7.5.1 Introduction

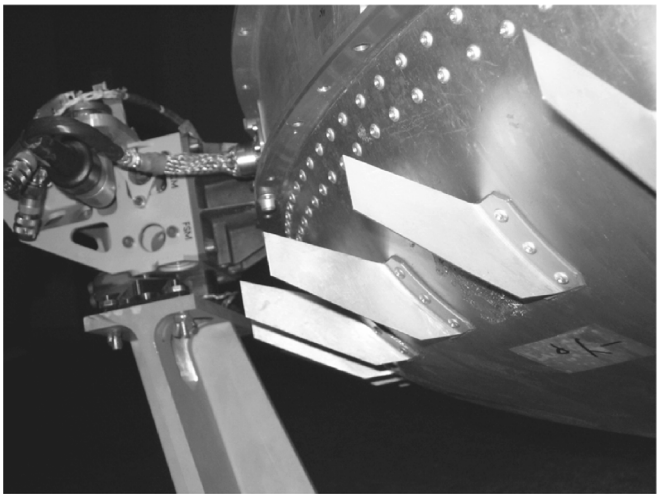
Until today, there has only been one winged re-entry vehicle that has frequently returned from orbit to Earth: the American Space Shuttle. Its Russian counterpart, the Buran, only made one unmanned (atmospheric) test flight before the project was cancelled due to budget problems. Also in Europe budget cuts were the reason that Hermes, a smaller version of the Space Shuttle, did not leave the drawing board and was cancelled. But whether the vehicle has actually been built or not, also the publicly available literature on space vehicles can usually teach us about applied subsystems, such as the attitude-control system. Unfortunately, documentation concerning either the Buran control system, or the proposed Hermes attitude controller, is sparse.

Contrary to what one would expect, looking at the number of reports and papers about the Space Shuttle's sub-systems, detailed publicly-available documentation on the Space-Shuttle attitude controller is only sparsely available, and then, in principle, only for the ascent phase. McHenry et al. (1979) and Schleich (1982) give quite a detailed discussion on the Space-Shuttle ascent guidance, navigation, and control. McDermott et al. (1982) do a linearised stability analysis for this control system, and Schletz (1982) discusses the use of quaternions in the GNC system. The attitude controller for the entry and descent phase is only marginally mentioned in those publications.

The Space Shuttle entered the atmosphere with a large angle of attack of about 40°, to minimise the heat load. Further down the trajectory, the angle of attack was reduced to meet with the crossrange and downrange requirements. Range control throughout the entry was accomplished by control of the bank angle. The flight control system (FCS) had to guarantee a safe and stable flight and thereby take into account wide variations in flight conditions and large model-data tolerances, next to the large attitude changes. To perform its tasks, the FCS could use a reaction control system and aerodynamic control surfaces. Hamilton (1982) states five features of the



(a) Pioneer Venus large probe with spin vanes mounted all around. Image: NASA.



(b) Close-up of the Huygens SM2 model showing the spin vanes. The structure to the left is an attach mechanism for the front shield (Lorenz, 2006)

Fig. 7.41 Typical examples of spin vanes

Space Shuttle that present unique stability problems in combination with the large velocity range: (i) the Shuttle is an unpowered vehicle, (ii) the control of the aileron, rudder and the RCS jets is blended, (iii) the gains of the FCS are scheduled, (iv) the rigid-body stability margins are small, with strong bending modes within an octave of the cross-over frequency, and (v) the control system is multi-rate digital.

Klinar et al. (1975) give a general overview of the Space Shuttle Flight Control System. However, this overview cannot give the final details of the FCS, since the first Space-Shuttle flight was in 1981 and the FCS design was continuously updated when new data became available. Since the general idea did not change, we will use this paper (and others) to give the reader a global impression of the FCS design. The system design goal was to maximise the use of the aerodynamic control surfaces, of course within their power limitations, and to minimise the RCS propellant consumption, while satisfying the handling qualities required for manual operation.

The longitudinal and lateral controllers of the Space Shuttle are described as follows. The longitudinal automatic and manual FCS designs are conventional pitch-rate feedback controllers, with outer-loop closure accomplished by surface-position feedback or by the pilot. Actuators for both trim and control are the symmetric elevons, the body flap and for the low dynamic-pressure region the pitch jets. Also in case of the lateral automatic and manual FCS, the design consists of conventional rate-feedback controllers. Operations depend on the angle of attack. For the higher angle-of-attack operation, the rudder is ineffective, leaving only differential elevons (ailerons) and the RCS thrusters for control. The rudder is only activated below Mach numbers of 3.5. The yaw thrusters are used to control banking, whereas the ailerons are used to damp sideslip. At lower angles of attack, there is a conventional aileron/rudder cross-feed for turn coordination.

A major concern during the design of the FCS was the flexible-body interaction. As we can see in Fig. 7.42, which shows a simplified entry FCS configuration in the all-aerodynamic phase, there are several bending filters included in the design. These filters were added, because due to the high loads, the Shuttle could not be treated as a rigid body. The bending mode stabilisation was considered to be a problem that drove the design of the controller. What we can also see from this figure, is that the longitudinal and lateral motion are not completely decoupled, since the longitudinal controller has a feedback compensation of the yaw rate, which is a lateral-motion component. Note that the commanded attitude that is computed by the guidance logic, is translated into commanded angular rates, which are fed to the attitude controller (attitude-rate control instead of attitude-angle control).

Hamilton (1982) discusses the flexible body stabilisation for the aerosurface control loops. In Fig. 7.43, we have included the simplified block diagram that Hamilton used for the discussion on the lateral FCS. We see that the basic idea is based on rate feedback, with several filters added to account for elasticity and to improve the response of the system. Furthermore, we find scheduled gains. The forward-loop gains are inversely proportional to the dynamic pressure, which raises the gains where aerosurface effectiveness is low. The aileron loop contains a yaw-rate feedback that is scheduled with the angle of attack. This gain is proportional to cot α. Analysis of the entry FCS has been described by a number of authors. Stone and Powell (1976) do

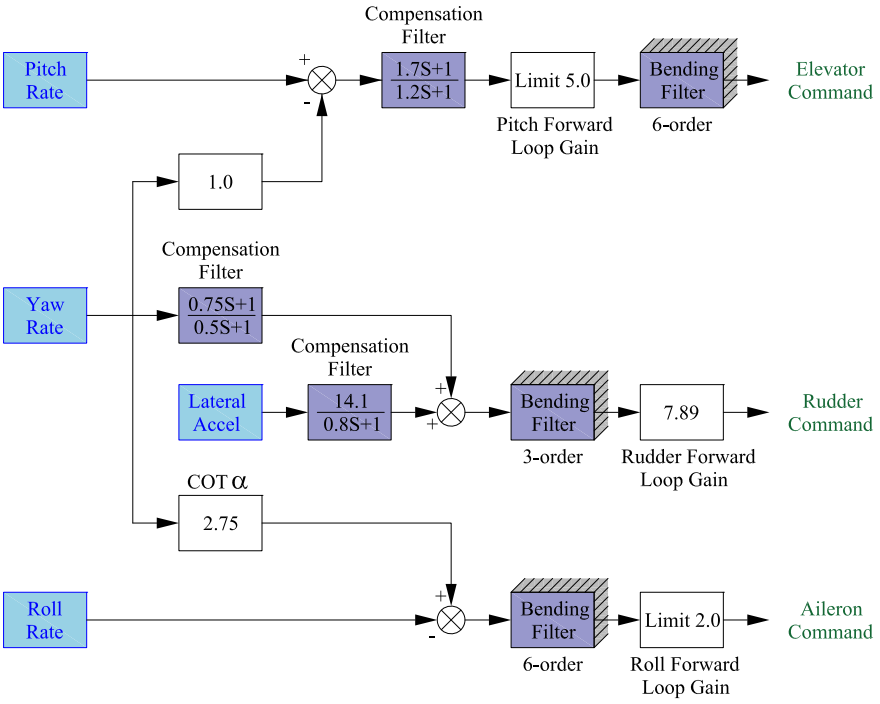


Fig. 7.42 Entry FCS configuration for Mach = 3.5, adapted from Epple and Altenbach (1983)

an analysis of the entry guidance and control system to determine the sensitivity of the Shuttle to off-nominal stability and control aerodynamic parameters. In addition, they identified the boundary values for each of these parameters. The Space Shuttle entry flight control off-nominal design considerations are also the topic of a paper by Bayle (1984). He presents the sensitivity of the flight control stability margins to aerodynamics, discusses the flight control verification process, and compares the predicted performance with the flight-test results of the first Space Shuttle flight, STS-1.

Nguyen et al. (1990) describe the testing methodology that was used for verification of the Shuttle FCS, using simulation (software: linear stability analysis and off-line non-linear simulation programs; hardware: MIL engineering and verification simulators) and flight tests (the first four orbiter missions). Epple and Altenbach (1983) describe the dynamic stability testing of the Space Shuttle Columbia FCS and flexible-body interaction. Input stimuli were applied to the Shuttle vehicle controllers to excite bending while the FCS is powered-up, and the measured responses from the operating FCS were compared with the predicted responses from the flex FCS flight model suitably modified to represent the ground-test configuration. Myers et al. (1987), finally, assess the FCS and the flying qualities of the Space Shuttle during approach and landing.

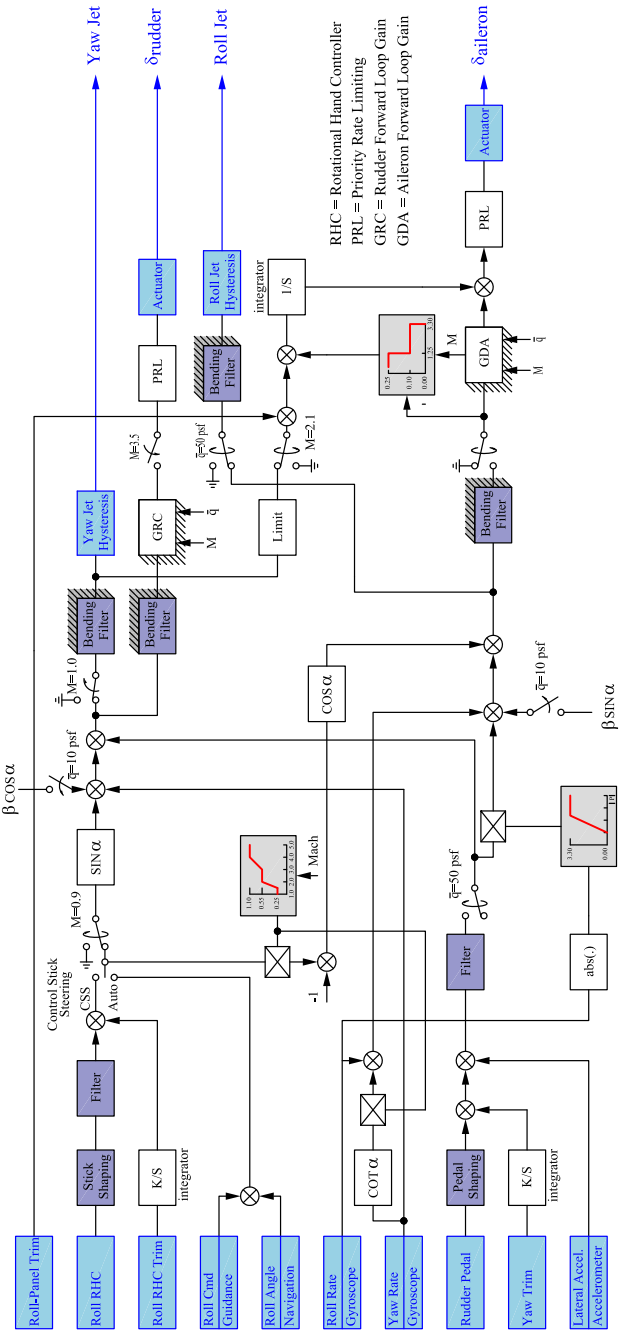


Fig. 7.43 The Space Shuttle lateral entry flight control system, adapted from Hamilton (1982)

A control-system design can be based on a number of underlying theories. The oldest and most widely used concept is the one of feedback, of either the output or state of a process (see Sect. 7.5.2). A technique, which is being used more and more in industrial control systems is adaptive control. Modern, robust control techniques that are subject of many studies are based on H_∞ or μ-synthesis, see, for instance, Doyle et al. (1986). Of course, there are many variations on the different concepts. Since we do not want to do an extensive survey of different control techniques, we will only briefly introduce two of robust control techniques at the end of the section. Other than those, we limit ourselves to the classical control concepts, and leave the more modern techniques for what they are.

Feedback control systems have found widespread use in, amongst others, aeronautical engineering (Bryson 1985). Simple forms of feedback are (a combination of) proportional, integral, and derivative feedback of the output of the system or plant that has to be regulated, see Sect. 7.5.3. Classical control theory of linear systems was based on frequency response and root-locus techniques, see, for instance the books by Kuo (1987), DSouza (1988), Paraskevopoulos (2002) and Ogata (2010). A set of general performance requirements that were not optimal in a mathematical sense, but rather aimed at a reasonable performance, was commonly used.

Initially, the older concepts were not easy to apply to multi-variable plants. State feedback systems, however, are particularly suitable for systems with multiple inputs and multiple outputs, so-called MIMO systems. The parameters that define the control-system performance, the gains, can be obtained by pole placement or, alternatively, based on mathematically defined optimisation criteria. The linear quadratic cost criterion is well known in this respect, resulting in the so-called linear quadratic regulator (LQR), see, for example, the books by Bryson and Ho (1975), Lewis et al. (2012) and Gopal (1989). More recent trends in feedback design are given by Kokotovic (1985), who gives an overview dealing with non-linear feedback, i.e., adaptive and composite control, and with external linearisation. A relatively recent application of the LQR using the method of extended linearisation is given by Wang and Sundararajan (1995), who describe a non-linear longitudinal flight controller for the F-8 aircraft.

An advantage of LQR is that it is a systematic method for designing MIMO systems. Furthermore, the implementation of the control laws in flight-simulation software is fairly simple, and the computational load for on-line simulation is low. The problems dealing with pole assignment linked with MIMO systems have been replaced by an optimisation problem, and pole selection is now changed to the selection of the optimisation parameters (weighting matrices). However, when not all the states of the controlled system are available, then most of the attractive properties of the LQR methodology are lost. In that case, an estimator is introduced to estimate the unavailable states (Shahian and Hassul 1993), which then results in a linear quadratic Gaussian (LQG) controller. However, the LQR seems to be a very appealing concept to illustrate the working principle of attitude controllers, so we will cover this method in Sect. 7.5.4. For more details on modern control techniques, the reader is referred to, e.g., the book by Paraskevopoulos (2002) who gives a good introduction

to robust, adaptive, and fuzzy control techniques. The quoted references therein will help in studying these techniques further.

When a process is dynamic, i.e., the system parameters vary strongly with time, or in case there are many disturbances, a controller with constant parameters is not likely to perform well over the entire operational range. In that case the controller parameters should in some way be adapted to the changing circumstances. Adaptive control, a special type of non-linear feedback control, found its way into use in the early fifties, as an autopilot for high performance aircraft, later on followed by applications in the F-94, F-101, and X-15 research aircraft (Boskovich and Kaufmann 1966). A simple way of adapting to the changing flight regime is to regularly change the controller gains. One implementation to do so could be a simple interpolation on the basis of dynamic pressure. Since the dynamic pressure drives the aerodynamic forces and moments, this seems like the logical thing to do. Varying the gains as a function of some independent parameter is called gain scheduling.

Alternatively, one could consider having some form of model identification on-board that provides the controller with real-time information about the vehicle dynamics. However, that may come with a disadvantage, when the on-board model is not an accurate (enough) description of the actual vehicle. If somehow the model could be removed and the system behaviour can be known in another way, that would make the controller a lot more robust. One such controller is based on incremental non-linear dynamic inversion (INDI), and as an example of modern-day control system designs the INDI controller will be briefly discussed.

Last but not least, there are also controllers where the use of a model without any identification of the model parameters can improve the system's robustness. One example is simple adaptive control (SAC), where the non-linear controller tracks the output of a (simple) reference model. With accurate tracking the vehicle will get the response characteristics of the model, which can be designed according to the performance requirements.

So, in the following five sections we will subsequently discuss state- and output-feedback control (Sect. 7.5.2), PID-control (Sect. 7.5.3), linear quadratic regulators by summarising the related optimal-control theory (Sect. 7.5.4), and finally two non-linear controllers based on Incremental Non-linear Dynamic Inversion (Sect. 7.5.5), and Simple Adaptive Control (Sect. 7.5.6). An application of the LQR is discussed at length in Sect. 7.5.7, where a possible attitude-controller design of HORUS is presented as a case study.

7.5.2 State and Output Feedback Control

Before we start the discussion on state and output feedback we must make clear that *control* is not only related to *attitude* control, i.e., regulating either the attitude, angular rates or a combination of both, but also to position and velocity. In general we will call this type of control *guidance*, which will be discussed in much more detail in Sect. 7.6. It is stressed, though, that similar control theories and control-system

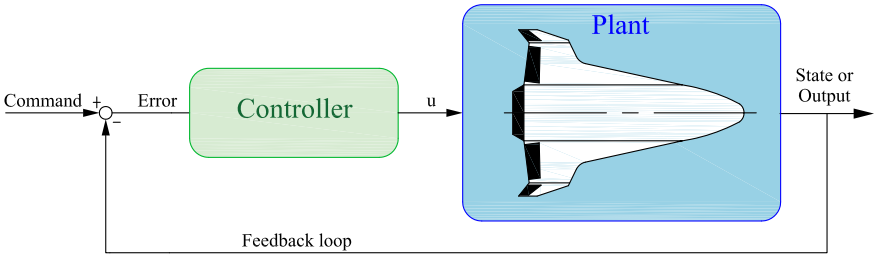


Fig. 7.44 Closed-loop control of an arbitrary plant by means of state or output feedback

designs (control algorithms) are used for either attitude control or guidance. In that sense, the term *control* is somewhat poorly chosen.

Let us begin with the situation as schematically shown in Fig. 7.44, where a to-be-determined controller takes either the state or the output of a system (also called *plant*) as input, and compares it with a command signal. The error drives the controller and provides the input to the system. To discuss either state or output feedback it serves our purpose to write the system to be controlled in so-called state-space form. When equations of motion are written in state-space form, they have the following form in matrix notation, as discussed in the previous Chap. 6:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{7.98}$$

with $\mathbf{x}$ an $n \times 1$ state vector, $\mathbf{u}$ an $m \times 1$ control vector, and $\mathbf{A}$ and $\mathbf{B}$ the $n \times n$ state (or system) and $n \times m$ input (or control) matrices, respectively. Equation (7.98) is called the dynamics or state equation; to complete the description of the state of the vehicle we also provide the so-called output equation:

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} \tag{7.99}$$

where $\mathbf{y}$ is the $k \times 1$ output vector, and $\mathbf{C}$ and $\mathbf{D}$ are the $k \times n$ output and $k \times m$ direct transmission matrices, respectively.

In general, the coefficient matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are obtained from the non-linear equations of motion by linearisation. It is noted, though, that also the non-linear equations of motion can be written in state-space form, but that form is not used in classical control theory, since that theory is centred around linear systems. So, for the remainder of this section we assume that our system is linear; moreover, the coefficient matrices are assumed to be constant, which makes our system a so-called *linear time invariant* (LTI) system. In Chap. 6 we analysed the open-loop behaviour of entry vehicles by calculating the eigenvalues and eigenvectors of state matrix $\mathbf{A}$. The main principle of feedback control is now to change the eigenvalues (and thus eigenmotion) of the open-loop system.

Let us first consider the case of state feedback. Then, the control law is given by

$$\mathbf{u} = -\mathbf{K}\mathbf{x} \tag{7.100}$$

where $\mathbf{K}$ is a time independent feedback or gain matrix. Equation (7.100) substituted into Eq. (7.98) yields

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{B}\mathbf{K})\,\mathbf{x} = \mathbf{A}^{*}\mathbf{x} \tag{7.101}$$

The characteristic Eq. (7.101), which gives us the eigenvalues and corresponding eigenmotion of the closed-loop system, is given by

$$\det|\mathbf{A} - \mathbf{B}\mathbf{K} - \lambda\mathbf{I}| = 0 \tag{7.102}$$

As becomes obvious while studying the above equation, we can change the eigenvalues of the closed-loop system by varying the gain matrix $\mathbf{K}$. Whether the system will be controllable, however, is not only depending on the values of these gains. The system in itself should at least be controllable, and it is not necessarily true that this is always the case. *Controllability* is, in this case, defined as follows (Kuo 1987):

> Given a linear time-invariant system as described by Eqs. (7.98)–(7.99), the state $\mathbf{x}(t)$ is said to be controllable at $t = t_0$ if there exists a piece-wise continuous (and finite) input $\mathbf{u}(t)$ that will drive the state to any final state $\mathbf{x}(t_f)$ for a finite time $t_f \geq t_0$. If every state $\mathbf{x}(t_0)$ of the system is controllable in a finite time interval, the system is said to be completely state controllable or simply state controllable.

The condition of controllability depends on the coefficient matrices $\mathbf{A}$ and $\mathbf{B}$.

To determine whether the system is unstable or not, it is also necessary that (at least the unstable) motions of the system are observable. *Observability* is defined by Kuo (1987) as:

> Given a linear time-invariant system as described by Eqs. (7.98)–(7.99), the state $\mathbf{x}(t)$ is said to be observable if given any input $\mathbf{u}(t)$, there exist a finite time $t_f \geq t_0$ such that the knowledge of (i) $\mathbf{u}(t)$ for $t_0 \leq t < t_f$ (ii) the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ and (iii) the output $\mathbf{y}(t)$ for $t_0 \leq t < t_f$, are sufficient to determine $\mathbf{x}(t_0)$. If every state of the system is observable for a finite time t_f, we say that the system is completely observable, or simply observable.

To check whether a system is controllable and observable, it is necessary and sufficient that the controllability matrix

$$\left[\mathbf{B}\ \mathbf{A}\mathbf{B}\ \mathbf{A}^2\mathbf{B}\ \ldots\ \mathbf{A}^{n-1}\mathbf{B}\right] \tag{7.103}$$

and observability matrix

$$\left[\mathbf{C}^{\mathrm{T}}\ \mathbf{A}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\ (\mathbf{A}^{\mathrm{T}})^2\mathbf{C}^{\mathrm{T}}\ \ldots\ (\mathbf{A}^{\mathrm{T}})^{n-1}\mathbf{C}^{\mathrm{T}}\right] \tag{7.104}$$

have a rank n, i.e., the dimension of the state vector $\mathbf{x}$. When not all of the eigenmotions are part of the observable and controllable state space, then only those eigenmotions that are part of it can be influenced by state feedback.

For output feedback the control law is given by

$$\mathbf{u} = -\mathbf{K}\mathbf{y} = -\mathbf{K}\left(\mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}\right) = -\mathbf{K}\mathbf{C}\mathbf{x} - \mathbf{K}\mathbf{D}\mathbf{u} \tag{7.105}$$

or

$$\left(\mathbf{I} + \mathbf{K}\mathbf{D}\right)\mathbf{u} = -\mathbf{K}\mathbf{C}\mathbf{x} \Rightarrow \mathbf{u} = -\left(\mathbf{I} + \mathbf{K}\mathbf{D}\right)^{-1}\mathbf{K}\mathbf{C}\mathbf{x} \tag{7.106}$$

In most applications, $\mathbf{D} = \mathbf{0}$, which makes the (digital) implementation of the control law easier. In case $\mathbf{D} \neq \mathbf{0}$, sample delays have to be included, which means that the output $\mathbf{y}$ is calculated with the value of $\mathbf{u}$ of the previous sample. Here, we assume that $\mathbf{D} = \mathbf{0}$.

Equation (7.106) substituted in Eq. (7.98) yields

$$\dot{\mathbf{x}} = \left(\mathbf{A} - \mathbf{B}\mathbf{K}\mathbf{C}\right)\mathbf{x} \tag{7.107}$$

The characteristic equation of Eq. (7.107) is then given by

$$\det\left|\mathbf{A} - \mathbf{B}\mathbf{K}\mathbf{C} - \lambda\mathbf{I}\right| = 0 \tag{7.108}$$

As we mentioned earlier, by "cleverly" selecting the elements of gain matrix $\mathbf{K}$ in Eqs. (7.102) and (7.108) the eigenvalues (and thus eigenmotion) of the closed-loop system can be changed to get the required (transient) behaviour of the system. This can be done by using techniques such as pole placement or using visual aids such as Bode plots (Ogata 2010).

In many situations, the performance characteristics of a control system are specified in terms of the transient response to a unit-step input, since it is easy to generate and is sufficiently "drastic". And, if the response to a step input is known, it is mathematically possible to compute the response to any input, since an arbitrary input can be decomposed into a series of sequential step inputs. The transient response of a system to a unit-step input depends on the initial conditions. Therefore, for convenience in comparing transient responses of various systems, one commonly uses the standard initial condition that the system is at rest with the output and all time derivatives thereof zero. The transient response of a practical control system often exhibits damped oscillations before reaching steady state. In specifying the transient-response characteristics of a control system to a unit-step input, it is common to use the following characteristic values (Ogata 2010), see also Fig. 7.45:

1. Delay time, t_d: the delay time is the time required for the response to reach half the final value the very first time.
2. Rise time, t_r: the rise time is the time required for the response to rise from 10 to 90%, 5 to 95%, or 0 to 100% of its final value. For underdamped second-order systems, the 0–100% rise time is normally used, and for overdamped systems, the 10–90% rise time.
3. Peak time, t_p: the peak time is the time required for the response to reach the first peak of the overshoot.
4. Maximum (percent) overshoot, M_p: the maximum overshoot is the maximum peak value of the response curve measured from unity. If the final steady-state

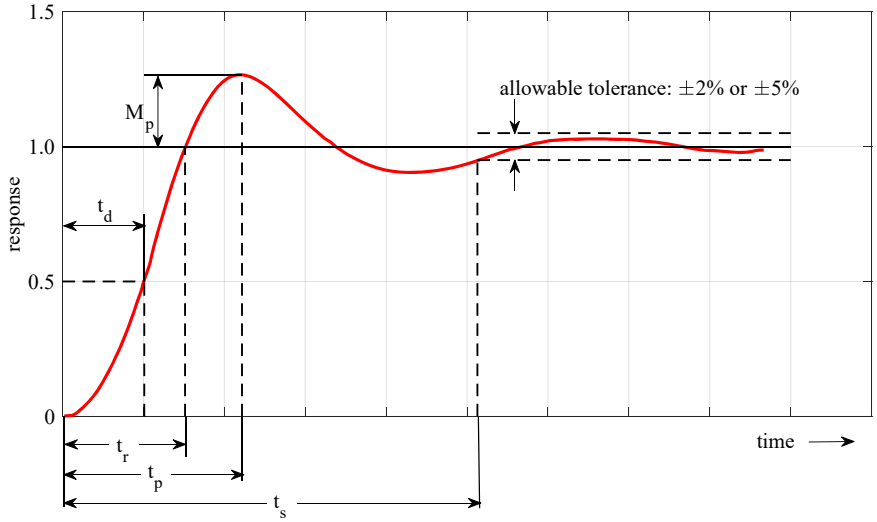


Fig. 7.45 Transient and steady-state response analysis

value of the response differs from unity, then it is common to use the maximum percent overshoot. The amount of the maximum (percent) overshoot directly indicates the relative stability of the system.

5. Settling time, t_s: the settling time is the time required for the response curve to reach and stay within a range about the final value of size, specified by absolute percentage of the final value (usually 2 or 5%). The settling time is related to the largest time constant of the control system. Which percentage error criterion to use may be determined from the objectives of the system design in question.

The above time-domain specifications are quite important, since most control systems are time-domain systems and must have acceptable time responses.

Except for certain applications where oscillations cannot be tolerated, it is desirable that the transient response be sufficiently fast and be sufficiently damped. Thus, for a desirable transient response of a second-order system, the damping ratio, ζ, must be between 0.4 and 0.8. Small values of ζ (i.e., $\zeta < 0.4$) yield excessive overshoot in the transient response, and a system with a large value of ζ (i.e., $\zeta > 0.8$) has a sluggish response. Commonly, the maximum overshoot and the rise time conflict with each other. In other words, both M_p and t_r cannot be made smaller simultaneously.

Due to conflicting requirements and/or when the system to be controlled is of higher order than two, deriving the gain values becomes a tedious job due to the lack of a proper design methodology using the earlier mentioned techniques. Another aspect is that the control effort is not directly taken into account, so that the designer must iterate after inspecting the response (and control-effort) curves. Fortunately, from optimal control theory a relatively simple and good compromise can be found

for the initial specification of feedback controller gains. This method will be discussed in Sect. 7.5.4.2.

7.5.3 PID Control

A proportional-integral-derivative controller is a generic control loop feedback mechanism (controller), widely used in industrial control systems. In fact, the PID controller is the most commonly used feedback controller. A PID controller calculates an "error" value as the difference between a measured state or output variable, and a desired setpoint, or in other words, a *commanded* state or output. The controller attempts to minimise this error by adjusting the control inputs, **u**, to the system.

As the name suggests, the PID algorithm involves three separate parameters, and is accordingly sometimes called three-term control. Heuristically, these values can be interpreted in terms of time: P depends on the present error, I on the accumulation of past errors, and D is a prediction of future errors, based on current rate of change. The weighted sum of these three actions is used to adjust the process. The controller parameters, or *gains*, do not necessarily have to be constant, but can be scheduled as a function of time, or in fact any other—preferably system related—parameter. For flight systems, one can use, for instance, the dynamic pressure, total energy, or Mach number. It is commonly avoided to use the time as scheduling parameter. Due to environmental perturbations or uncertainties in vehicle characteristics a shift in time in one of the trajectory parameters may occur, which will give different time values, but not necessarily different values of, e.g., the dynamic pressure.

In the absence of (accurate) knowledge of the underlying system and/or (flight) environment, a PID controller is the best controller to start with. By tuning the three parameters in the PID controller algorithm, it can provide control actions designed for specific process requirements. The response of the controller can be described in terms of the responsiveness of the controller to an error, using the characteristic values described in the previous section. It should be clear, though, that the use of the PID algorithm does not guarantee optimal control of the system or system stability.

In the following we will address each of the P, I and D control actions. To begin with the *proportional control action*, for such a controller the relationship between the output of the controller, i.e., $u(t)$ and the actuating error signal $e(t)$ is, assuming a single-input single-output (SISO) system for now:

$$u(t) = K_p e(t) \tag{7.109}$$

where K_p is termed the proportional gain. Whatever the actual mechanism may be and whatever the form of the operating power, the proportional controller is essentially an amplifier with an adjustable gain. A high proportional gain results in a large change in the output for a given change in the error. A low gain will obviously result in a slow response and a possibly growth of the error. If the proportional gain is too high, the system can become unstable.

In a controller with *integral control action*, the value of the controller output $u(t)$ is changed at a rate proportional to the actuating error signal $e(t)$ and is also proportional to the duration of the non-zero error. That is,

$$u(t) = K_i \int_0^t e(\tau) d\tau \tag{7.110}$$

where K_i is an adjustable constant. Large gain values or large errors may cause a rapid growth of the integral term, which means that the integrator can get "saturated". This means, that even for small errors the control action will be large. Usually, the use of integral control gives rise to overshoot, but is, on the other hand, very effective in removing residual steady-state errors that occur with a pure proportional controller (a steady-state error is the difference between the desired final output and the actual one). To avoid the rapid growth due to a large error, one could use a so-called limit integrator, i.e., which means that the integral term is only "switched on" within a certain error bandwidth.

A *derivative control action* is proportional to the rate of change of the control error:

$$u(t) = K_d \frac{\mathrm{d}e(t)}{\mathrm{d}t} \tag{7.111}$$

which means that this type is not very effective for slowly varying control errors. The derivative term slows the rate of change of the controller output. Derivative control is used to reduce the magnitude of the overshoot produced by integral control and and to improve the combined controller-process stability. However, the derivative term slows the transient response of the controller. Also, a combined proportional and derivative control action can result in a persistent steady-state error.

The combination of the three individual control actions is termed the proportional-plus-integral-plus-derivative control action. It has the advantages of each of the three individual control actions. The control law of a controller with this combined action is given by

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{\mathrm{d}e(t)}{\mathrm{d}t} \tag{7.112}$$

The block diagram of a proportional-plus-integral-plus-derivative controller is shown in Fig. 7.46. Note to take special care in selecting the proper values of the PID controller parameters, because the wrong choice may lead to system instability. In general, the rules of thumb shown in Table 7.3 can be used to tune the controller gains. Another, more heuristic approach is given by the Ziegler and Nichols tuning rules, see, for instance, the books by Kuo (1987), DSouza (1988) and Ogata (2010). Finally, if one wants to tune the controller using the deviation from a reference (or command) signal, the total control effort, or the combination of both, one can define

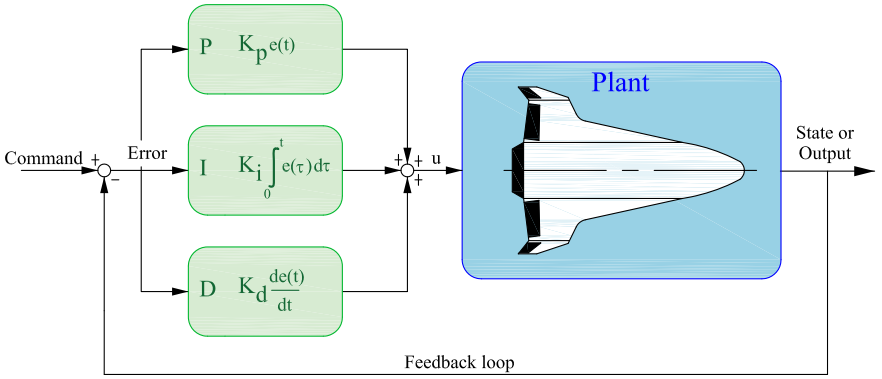


Fig. 7.46 Closed-loop PID control of an arbitrary plant by means of state or output feedback

Table 7.3 Influence on response due to PID-gain tuning

Parameter	Rise time	Overshoot	Settling time	Steady-state error	Stability
K_p	Decrease	Increase	Small change	Decrease	Degrade
K_i	Decrease	Increase	Increase	Decrease significantly	Degrade
K_d	Minor decrease	Minor decrease	Minor decrease	No effect in theory	Improve if K_d small

these as performance metrics and use an optimisation procedure to minimise them. This will be explained in more detail in Example 7.2.

Example 7.2 (*Gain tuning with performance metrics*) As an example of the approach to define performance metrics (integrated state deviation and integrated control effort), let us assume we want to control the attitude of an entry capsule, given by the angle of attack, α, the angle of sideslip, β, and the bank angle, σ, with reaction-control thrusters, represented by the roll, pitch and yaw moment, $M_{T,x}$, $M_{T,y}$ and $M_{T,z}$, respectively. Then, the six performance metrics are defined by:

$$\sum_{\alpha_{err}} = \int_0^t \sqrt{(\alpha_c - \alpha_p)^2}\,dt \quad \sum_{\beta_{err}} = \int_0^t \sqrt{(\beta_c - \beta_p)^2}\,dt \quad \sum_{\sigma_{err}} = \int_0^t \sqrt{(\sigma_c - \sigma_p)^2}\,dt \tag{7.113}$$

$$\sum_{x} = \int_0^t |M_{T,x}|\,dt \quad \sum_{y} = \int_0^t |M_{T,y}|\,dt \quad \sum_{z} = \int_0^t |M_{T,z}|\,dt \tag{7.114}$$

These metrics are the integrated state deviation (attitude angles) and the integrated control effort. It is obvious that these metrics should all be as small as possible. The

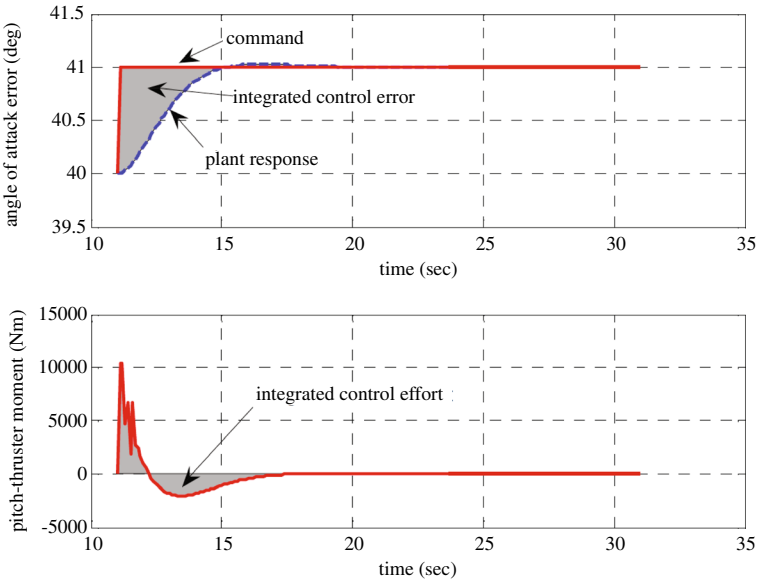


Fig. 7.47 Angle-of-attack response for a reaction-control design point. The shaded areas represent the two performance metrics integrated angle-of-attack error and integrated pitch-thruster activity

subscripts "c" and "p" stand for *command* and *plant*, respectively. A visualisation of two of these metrics is given in Fig. 7.47. The corresponding numerical values are: $\sum_{\alpha_{err}} = 1.903\,^{\circ}\text{s}$ and $\sum_{y} = 10{,}507\,\text{Nm s}$. ■

Example 7.3 (*Proportional Control*) Anticipating the case study in Sect. 7.5.7 we consider the following linearised (state-space) model for the longitudinal motion of a winged re-entry vehicle (see also Chap. 6 for the derivation of the state-space model):

$$\begin{pmatrix} \Delta\dot{q} \\ \Delta\dot{\alpha} \end{pmatrix} = \begin{bmatrix} 0 & \frac{1}{I_{yy}} \frac{\partial C_m}{\partial \alpha} \bar{q}_e S_{ref} c_{ref} \\ 1 & -\frac{1}{mV_e} \frac{\partial C_L}{\partial \alpha} \bar{q}_e S_{ref} \end{bmatrix} \begin{pmatrix} \Delta q \\ \Delta\alpha \end{pmatrix} + \begin{bmatrix} \frac{1}{I_{yy}} \frac{\partial C_m}{\partial \delta_e} \bar{q}_e S_{ref} c_{ref} \\ 0 \end{bmatrix} \Delta\delta_e \quad (7.115)$$

where $\Delta\alpha$ and Δq are the angle of attack and pitch rate, the state variables, and $\Delta\delta_e$ is the elevator deflection angle, the control input. So, this vehicle can control its attitude by aerodynamic means only. The two partial derivatives, i.e., $\frac{\partial C_m}{\partial \alpha}$ and $\frac{\partial C_m}{\partial \delta_e}$ are the vehicle's pitch-moment derivative and elevator effectiveness, respectively.

As a control law we want to use proportional control for a unit step response on the angle of attack. From initial simulations we found that it was not possible to stabilise the angle of attack in this way, no matter what value of K_p we chose. One of the reasons is that the angle of attack is a so-called kinematic variable, i.e., it changes only indirectly due to a change in pitch rate. Only the pitch rate changes under the influence of the elevator moment, but in this case enforced by

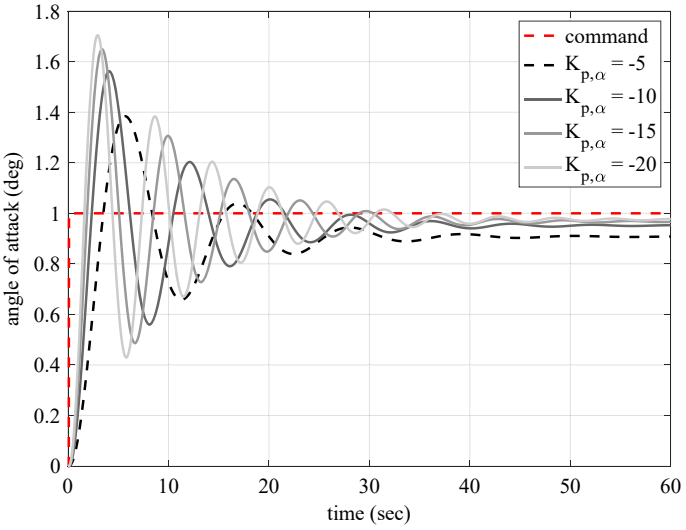


Fig. 7.48 Angle-of-attack response for different values of the proportional gain, K_p

an error in angle of attack. To solve this problem, we have to add a proportional-control component for the pitch rate as well, with a relatively low value for the gain: $K_{p,q} = -1$ s. On one hand this stabilises the angle-of-attack response, but on the other hand it introduces a small steady-state error, on top of any steady-state error that the proportional angle-of-attack control gives. Note that strictly speaking, adding a non-zero K_q resembles to add a derivative term in the angle-of-attack control loop. However, since it is more effective to apply pitch-rate control, we do not want to confuse the issue by introducing derivative control for the angle of attack. That will be done in Example 7.4.

In Fig. 7.48 we have plotted fours response lines for different values of $K_{p\alpha}$, i.e., $K_{p\alpha} = -5, -10, -15$ and -20 (with the unit of rad/rad). It is clear from the results that with increasing gain the response gets faster, albeit at the expense of a larger overshoot. For low gains, the steady-state error is significant, since the gain is not "strong" enough to drive the angle of attack towards its set point. Increasing the gain helps to reduce the steady-state error. ■

Example 7.4 (*Derivative Control*) This example continues the previous one, Example 7.3, in the sense that we use the same model and excitation as before. However, now we focus on the derivative gain, K_d. We set the proportional gain to $K_p = -15$ (and we also keep $K_q = -1$ s), and vary K_d from 0 to -15 s with steps of -5 s. The resulting response curves are shown in Fig. 7.49. It shows that increasing K_d works stabilising, the oscillations are removed for larger gain values. It is also clear that there is a persistent steady-state error, which is slightly larger compared to the error in the previous example. This persistent steady-state error is typical for derivative

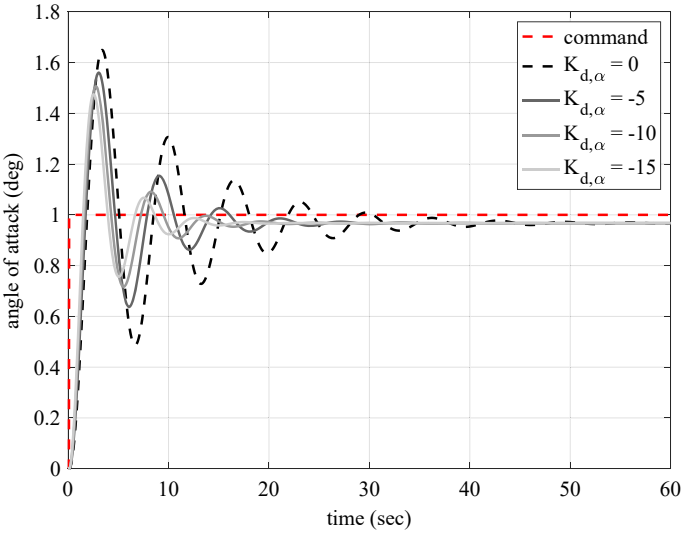


Fig. 7.49 Angle-of-attack response for different values of the derivative gain, K_d. The proportional gain is $K_p = -15$

control, and for that reason an integral component should be added to the controller. This will be discussed in the next example, Example 7.5. ■

Example 7.5 (*Integral Control*) Concluding the three-part example on PID control, we will now add an integral component in the control loop. For the same longitudinal model as in the previous two examples and the same unit-step excitation, we change the integral gain from 0 to -1, -2 and -5 s^{-1}. Both the proportional and derivative gain are fixed at -15 (and $K_{p,q} = -1$ s). The resulting response curves are shown in Fig. 7.50. For non-zero integral-gain values the steady state error reduces to zero. However, having integral control increases the overshoot, not only in the initial response, but also during subsequent oscillations (effectively, the integral gain increases the proportional gain). For too large values of K_i we saw very large oscillations until the system even became unstable. So, care should be taken considering integral control, especially in the presence of large or persistent errors. ■

7.5.4 Optimal Control

In this section, we will give a brief summary of general optimal-control theory that could be used for setting up a (local) trajectory optimisation model, or develop particular (optimal) guidance and/or control systems (Sect. 7.5.4.1). Along the way, Pontragin's maximum principle will be introduced, which constitutes the core of the theory. An example application is provided in Chap. 11, Sect. 11.2.4, when discussing

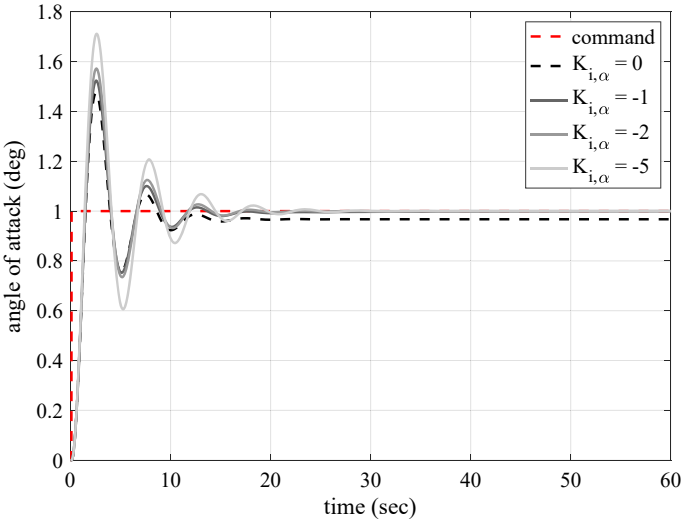


Fig. 7.50 Angle-of-attack response for different values of the integral gain, K_i. The proportional and derivative gains are $K_p = K_d = -15$ s

optimal aerocapture manoeuvres. Note that this summary will only touch upon the surface of optimal control. For those who are interested in all the nitty-gritty details, or those who want to apply the theory to their own problems, the reader is referred to some excellent books that cover the material extensively, i.e., those by Bryson and Ho (1975), Gopal (1989), Betts (2010), and Lewis et al. (2012). A more practical oriented book with many useful spacecraft-trajectory optimisation examples is the one by Conway (2010).

A more hands-on application of optimal-control theory will be provided in Sect. 7.5.4.2, where the so-called Linear Quadratic Regulator, a feedback control system, is introduced. Centred around a quadratic cost criterion, minimising the performance index $\mathcal{J}$ will in theory lead to an optimal-control system design that gives the best compromise between state deviation and control effort.

7.5.4.1 *General Formulation and Pontryagin's Minimum Principle

Let us begin with informally formulating the optimal-control problem. A mathematical statement in that direction consists of:

1. A mathematical description (or model) of the process to be controlled, e.g., a mass-point model doing aerobraking at Mars, or a spacecraft going from low Earth orbit to a geostationary one using low-thrust only,

2. A statement of the physical constraints, e.g., a maximum thermo-mechanical load during atmospheric flight, the maximum time spent in the orbit-raising manoeuvre, or actuator (=control variable!) limitations,
3. Specification of a performance criterion, e.g., the savings in ΔV compared to an all-propulsive capture manoeuvre, or the savings in ΔV compared to high-thrust manoeuvres, and
4. A description of the task to be accomplished, which usually takes the form of initial and/or terminal boundary conditions in the state variables.

In its simplest form, the problem can be formulated as to steer the system, $\mathcal{F}$, by means of its control variables, $\mathbf{u} \in \mathbb{R}^m$, in such a way that it achieves the best value of a specified performance index, $\mathcal{J}$, while the state variables, $\mathbf{x} \in \mathbb{R}^n$, satisfy dynamic constraints, along with any associated boundary conditions. Considering a minimisation problem, although with the introduction of a minus sign this also holds for a maximisation problem, the objective is to minimise $\mathcal{J}$ by integrating the system's "behaviour" over a time interval $[t_0, t_f]$:

$$\min_{\mathbf{u}(t)} \quad \mathcal{J} = \int_{t_0}^{t_f} \mathcal{F}[\mathbf{x}(t), \mathbf{u}(t)]\mathrm{d}t \tag{7.116}$$

$$\text{subject to : } \dot{\mathbf{x}} = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t)]; \quad \mathbf{x}(t_0) = \mathbf{x_0}; \quad \mathbf{x}(t_f) = \mathbf{x_f} \tag{7.117}$$

$$\text{and possibly : } \mathbf{g}[\mathbf{x}(t), \mathbf{u}(t)] = \mathbf{0} \text{ and } \mathbf{h}[\mathbf{x}(t), \mathbf{u}(t)] \geq \mathbf{0} \tag{7.118}$$

In the above formulation, $\mathbf{g}$ are the equality constraints, whereas $\mathbf{h}$ are the inequality constraints. Note that both initial and final state ($\mathbf{x_0}$ and $\mathbf{x_f}$, respectively) can either be specified or part of the optimisation problem and be determined.

The problem is to find functions $\mathbf{u}^*(t)$, i.e., the vector with optimal control(s), that minimise $\mathcal{J}$, while the state variables $\mathbf{x}$ satisfy the constraint(s). These constraints not only include $\mathbf{g}$ and $\mathbf{h}$, but also the so-called dynamic constraints. As the state of the system changes in time according to $\dot{\mathbf{x}} = \mathbf{f}\,[\mathbf{x}(t), \mathbf{u}(t)]$, the system is thus constrained by

$$\mathbf{f}\,[\mathbf{x}(t), \mathbf{u}(t)] - \dot{\mathbf{x}} = \mathbf{0} \tag{7.119}$$

We can now *adjoin* the system of differential equations to $\mathcal{J}$ with multiplier functions $\boldsymbol{\lambda}(t)$ to form the augmented performance index $\bar{\mathcal{J}}$:

$$\bar{\mathcal{J}} = \int_{t_0}^{t_f} \left(\mathcal{F}\,[\mathbf{x}(t), \mathbf{u}(t)] + \boldsymbol{\lambda}^{\mathrm{T}}(t)\,[\mathbf{f}\,[\mathbf{x}(t), \mathbf{u}(t)] - \dot{\mathbf{x}}\}\right) \mathrm{d}t \tag{7.120}$$

The procedure (with multiplier *functions* $\boldsymbol{\lambda}(t)$) to adjoin differential constraints is similar to what is done in constrained static optimisation. There the algebraic con-

straints are adjoined to the function to be optimised by Lagrange *multipliers*, $\boldsymbol{\lambda}$. It is stressed, that even though they looks similar, they are *not* identical.

For convenience, we will define a scalar function, the Hamiltonian $\mathcal{H}$:

$$\mathcal{H} \triangleq \mathcal{F}\left[\mathbf{x}(t), \mathbf{u}(t)\right] + \boldsymbol{\lambda}^{\mathrm{T}}(t)\mathbf{f}\left[\mathbf{x}(t), \mathbf{u}(t)\right] \tag{7.121}$$

Using this definition:

$$\bar{\mathcal{J}} = \int_{t_0}^{t_f} \left\{\mathcal{H}\left[\mathbf{x}(t), \mathbf{u}(t)\right] - \boldsymbol{\lambda}^{\mathrm{T}}(t)\dot{\mathbf{x}}\right\} \mathrm{d}t \tag{7.122}$$

As $\mathcal{F}$ can be a constrained function, the Hamiltonian will also include the (in)equality constraints.

To remove the state derivatives, integrate the last term on the right-hand side by parts[9]

$$\begin{aligned} \int_{t_0}^{t_f} \boldsymbol{\lambda}^{\mathrm{T}}(t)\dot{\mathbf{x}}\mathrm{d}t = & \int_{t_0}^{t_f} \boldsymbol{\lambda}^{\mathrm{T}}(t)\mathrm{d}\mathbf{x} = \boldsymbol{\lambda}^{\mathrm{T}}(t)\mathbf{x}\Big|_{t_0}^{t_f} - \int_{t_0}^{t_f} \mathbf{x}^{\mathrm{T}}\mathrm{d}\boldsymbol{\lambda} \\ = & \ \boldsymbol{\lambda}^{\mathrm{T}}(t)\mathbf{x}\Big|_{t_0}^{t_f} - \int_{t_0}^{t_f} \dot{\boldsymbol{\lambda}}(t)^{\mathrm{T}}\mathbf{x}\mathrm{d}t \end{aligned} \tag{7.125}$$

which, after substitution into Eq. (7.122), yields

$$\bar{\mathcal{J}} = -\boldsymbol{\lambda}^{\mathrm{T}}(t)\mathbf{x}\Big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \left\{\mathcal{H}\left[\mathbf{x}(t), \mathbf{u}(t)\right] + \dot{\boldsymbol{\lambda}}(t)^{\mathrm{T}}\mathbf{x}\right\} \mathrm{d}t \tag{7.126}$$

or, written in a more compact form:

---

[9] Integration by parts or partial integration is a process that finds the integral of a product of functions in terms of the integral of their derivative and anti-derivative

$$\begin{aligned} \int_a^b u(x)v'(x)\mathrm{d}x = & \ [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)\mathrm{d}x \\ = & \ u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)\mathrm{d}x \end{aligned} \tag{7.123}$$

or, in short-hand notation:

$$\int u\mathrm{d}v = uv - \int v\mathrm{d}u. \tag{7.124}$$

$$\bar{\mathcal{J}} = -\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x}\big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \left(\mathcal{H} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{x}\right) \mathrm{d}t \tag{7.127}$$

Let $\mathbf{u}^*$ minimise $\mathcal{J}$, then for any control close to $\mathbf{u}^*$ we may write[10]:

$$\mathbf{u}(t) = \mathbf{u}^*(t) + \delta\mathbf{u}(t) \tag{7.128}$$

with $\delta\mathbf{u}$ being the variation of $\mathbf{u}(t)$. The variation $\delta\mathbf{u}$ results in a (first-order) state variation $\delta\mathbf{x}$, following the "state equation":

$$\delta\dot{\mathbf{x}} = \frac{\partial \mathbf{f}^*}{\partial \mathbf{x}}\delta\mathbf{x} + \frac{\partial \mathbf{f}^*}{\partial \mathbf{u}}\delta\mathbf{u} \tag{7.129}$$

Similarly to what one would do with static optimisation problems, we need to search for the stationary value of the augmented performance index, $\bar{\mathcal{J}}$. Let us therefore consider the variation (to first order) in $\bar{\mathcal{J}}$ due to the variations in the control vector $\mathbf{u}(t)$:

$$\delta\bar{\mathcal{J}} = \bar{\mathcal{J}}\big|_{\mathbf{u}+\delta\mathbf{u}} - \bar{\mathcal{J}}\big|_{\mathbf{u}} \tag{7.130}$$

From Eq. (7.127) it follows that

$$\delta\bar{\mathcal{J}} = -\delta\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x}\right)\big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \delta\left(\mathcal{H} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{x}\right) \mathrm{d}t = -\boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x}\big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \left(\delta\mathcal{H} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\delta\mathbf{x}\right) \mathrm{d}t \tag{7.131}$$

Defining the variation of the Hamiltonian as

$$\delta\mathcal{H} = \frac{\partial \mathcal{H}}{\partial \mathbf{x}}\delta\mathbf{x} + \frac{\partial \mathcal{H}}{\partial \mathbf{u}}\delta\mathbf{u} \tag{7.132}$$

it can easily be seen that

$$\delta\bar{\mathcal{J}} = -\boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x}\big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \left[\left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\right)\delta\mathbf{x} + \frac{\partial \mathcal{H}}{\partial \mathbf{u}}\delta\mathbf{u}\right] \mathrm{d}t \tag{7.133}$$

In principle the above equation determines $\delta\mathbf{x}(t)$ in terms of $\delta\mathbf{u}(t)$, although it would be very tedious to do so. Instead, let us now choose the multiplier functions $\boldsymbol{\lambda}(t)$ to make the coefficients of $\delta\mathbf{x}$ disappear (such that $\delta\mathbf{x}$ is independent from $\delta\mathbf{u}$):

$$\dot{\boldsymbol{\lambda}}^{\mathrm{T}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{x}} = -\frac{\partial \mathcal{F}}{\partial \mathbf{x}} - \boldsymbol{\lambda}^{\mathrm{T}}\frac{\partial \mathbf{f}}{\partial \mathbf{x}} \tag{7.134}$$

[10] For a more fundamental treatment of *Calculus of Variations*, the reader is referred to, for instance, the books by Bryson and Ho (1975) and Schaub and Junkins (2014).

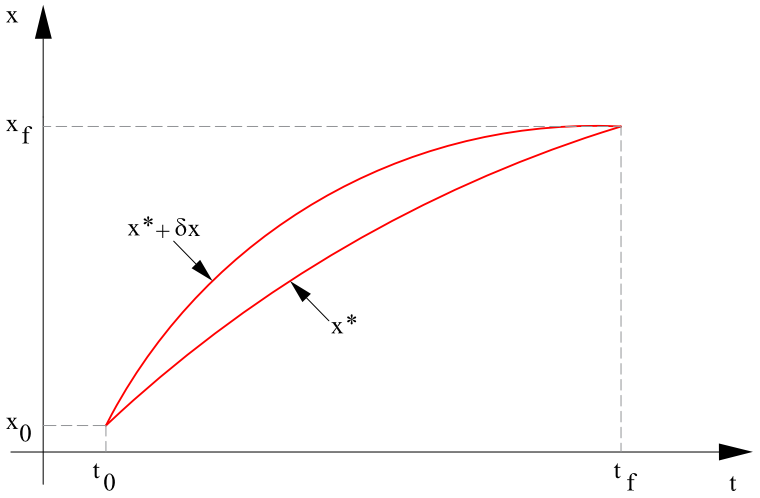


Fig. 7.51 The variation $\delta\mathbf{x}$ for the basic problem with fixed end points

such that Eq. (7.133) reduces to

$$\delta\bar{\mathcal{J}} = -\boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x}\Big|_{t_0}^{t_f} + \int_{t_0}^{t_f}\left(\frac{\partial\mathcal{H}}{\partial\mathbf{u}}\delta\mathbf{u}\right)\mathrm{d}t \tag{7.135}$$

The functions $\boldsymbol{\lambda}(t)$ are usually called the adjoints or co-state functions.

With fixed begin and end points (t_0 and t_f) the boundary conditions are (see Fig. 7.51):

$$\delta\mathbf{x}(t_0) = \delta\mathbf{x}(t_f) = 0 \tag{7.136}$$

In that case, the first term on the right-hand side of Eq. (7.135) disappears, so:

$$\delta\bar{\mathcal{J}} = \int_{t_0}^{t_f}\left(\frac{\partial\mathcal{H}}{\partial\mathbf{u}}\delta\mathbf{u}\right)\mathrm{d}t \tag{7.137}$$

For an optimum, $\delta\bar{\mathcal{J}}$ must be zero for arbitrary $\delta\mathbf{u}$, which is only true if

$$\frac{\partial\mathcal{H}}{\partial\mathbf{u}} = 0, \qquad t_0 \le t \le t_f \tag{7.138}$$

Together with expression for $\dot{\boldsymbol{\lambda}}$, Eq. (7.134), they form the *Euler–Lagrange* equations:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = 0, \qquad t_0 \le t \le t_f \qquad \dot{\boldsymbol{\lambda}}^{\mathrm{T}} = -\frac{\partial \mathcal{F}}{\partial \mathbf{x}} - \boldsymbol{\lambda}^{\mathrm{T}} \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \tag{7.139}$$

Summarised: to find control vector $\mathbf{u}(t)$ that gives a stationary value of performance index $\mathcal{J}$, the following differential equations must be solved:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \qquad \mathbf{x}(t_0) \text{ and } \mathbf{x}(t_f) \text{ are given} \tag{7.140}$$

$$\dot{\boldsymbol{\lambda}}^{\mathrm{T}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{x}} = -\frac{\partial \mathcal{F}}{\partial \mathbf{x}} - \boldsymbol{\lambda}^{\mathrm{T}} \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \tag{7.141}$$

Here, $\mathbf{u}(t)$ is determined by

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = \frac{\partial \mathcal{F}}{\partial \mathbf{u}} + \boldsymbol{\lambda}^{\mathrm{T}} \frac{\partial \mathbf{f}}{\partial \mathbf{u}} = \mathbf{0} \tag{7.142}$$

The boundary conditions are split: half are given for t_0 and half for t_f, which means that the above formulation is a two-point-boundary-value problem (TPBVP). The above optimal-control formulation is known as *Pontryagin's Minimum Principle*[11] (Pontryagin et al. 1962). At the end of this section, we will revisit the minimum principle.

For $\mathcal{J}$ to be a local minimum not only the previously mentioned first-order conditions must apply, but in addition some second-order conditions must hold. For a smooth $\mathcal{H}$ with no constraints on $\mathbf{u}$ it is required that

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = 0, \qquad \frac{\partial^2 \mathcal{H}}{\partial \mathbf{u}^2} = 0, \qquad t_0 \le t \le t_f \tag{7.143}$$

This second condition is also called the Legendre–Clebsch condition.

Example 7.6 (*Optimal Control with Fixed End Points*) This example will address a dynamic-optimisation problem with fixed end points. Minimise

$$\mathcal{J} = \int_{t_0=0}^{t_f=1} \left(\frac{1}{2} u^2 + x \right) \mathrm{d}t$$

subject to

$$\dot{x} = u$$

[11] The original formulation was stated as the so-called maximum principle, in 1956 formulated by the Russian mathematician Lev Pontryagin (September 3, 1908–May 3, 1988). The initial application of Pontryagin and his students was that of maximising the terminal speed of a rocket.

with $x(0) = 0$ and $x(1) = 1$. The Hamiltonian is now formulated as

$$\mathcal{H} = \left(\frac{1}{2}u^2 + x\right) + \lambda_x u$$

which allows us to define the optimal control conditions:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = u + \lambda_x \quad \Rightarrow \quad u^* = -\lambda_x \qquad \frac{\partial^2 \mathcal{H}}{\partial \mathbf{u}^2} = 1 \geq 0$$

The adjoint equation is:

$$\dot{\lambda}_x = -\frac{\partial \mathcal{H}}{\partial x} = -1$$

Substitution of $u^* = -\lambda_x$ in the constraint equation $\dot{x} = u$ and differentiation yields:

$$\ddot{x} = \dot{u}^* = \dot{\lambda}_x = -1$$

Solution of this second-order differential equation is

$$x = c_1 + c_2 t + \frac{1}{2}t^2$$

With the boundary conditions it follows for the constants of integration

$$x(0) = 0 \quad \Rightarrow \quad c_1 = 0 \qquad x(1) = 1 \quad \Rightarrow \quad c_2 = \frac{1}{2}$$

Therefore, the optimal path x^* is found to be:

$$x^* = \frac{1}{2}\left(t^2 + 1\right)$$

with corresponding optimal control:

$$u^* = -\lambda_x = \dot{x} = t + \frac{1}{2}$$

Substitution of the results into $\mathcal{H}$ gives the so-called *first integral (of motion)*:

$$\mathcal{H} = \left(\frac{1}{2}u^2 + x\right) + \lambda_x u = -\frac{1}{2}u^2 + x = -\frac{1}{2}\left(t + \frac{1}{2}\right)^2 + \frac{1}{2}\left(t^2 + t\right) = -\frac{1}{8} = \text{constant}$$

A first integral of (any) boundary-value problem exists if $\mathcal{F}$ and $\mathbf{f}$ are not explicit functions of time t (this is called a stationary system). For such a system, we will

then have $\mathcal{H} = \mathcal{H}(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda})$ (formulated in a general sense). For an optimal trajectory, with $\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = \mathbf{0}$, it is found that

$$\frac{\partial \mathcal{H}}{\partial t} = \frac{\partial \mathcal{H}}{\partial \mathbf{x}}\dot{\mathbf{x}} + \frac{\partial \mathcal{H}}{\partial \mathbf{u}}\dot{\mathbf{u}} + \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}}\dot{\boldsymbol{\lambda}} = \frac{\partial \mathcal{H}}{\partial \mathbf{x}}\dot{\mathbf{x}} - \dot{\mathbf{x}}^{\mathrm{T}}\frac{\partial \mathcal{H}^{\mathrm{T}}}{\partial \mathbf{x}} = 0$$

and consequently along an optimal trajectory, $\mathcal{H} =$ constant. The first integral is not required to solve the TPBVP; the fact that $\mathcal{H}$ is constant along the optimal trajectory is redundant information.

Remark concerning this example: even this (very) simple, single-state single-control problem seems to be rather complicated. Intuitively, one can expect that problems with more states, more controls, and more complicated constraints are difficult (not to say impossible) to solve "by hand", which means that the common approach to solve this kind of problems is to use numerical procedures. ■

Let us have a look again at the augmented performance function:

$$\bar{\mathcal{J}} = \int_{t_0}^{t_f} \left\{\mathcal{H}\left[\mathbf{x}(t), \mathbf{u}(t)\right] + \dot{\boldsymbol{\lambda}}(t)^{\mathrm{T}}\mathbf{x}\right\} \mathrm{d}t = \int_{t_0}^{t_f} \left(\mathcal{H} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{x}\right) \mathrm{d}t \tag{7.144}$$

So far, we considered fixed end points; now they will be kept free. With Leibniz's Rule[12] we get

$$\delta\bar{\mathcal{J}} = \delta\left[\int_{t_0}^{t_f} \left(\mathcal{H} - \boldsymbol{\lambda}^{\mathrm{T}}\dot{\mathbf{x}}\right) \mathrm{d}t\right] = \int_{t_0}^{t_f} \delta\left(\mathcal{H} - \boldsymbol{\lambda}^{\mathrm{T}}\dot{\mathbf{x}}\right) \mathrm{d}t + \left(\mathcal{H} - \boldsymbol{\lambda}^{\mathrm{T}}\dot{\mathbf{x}}\right) \delta t\Big|_{t_0}^{t_f} = 0 \tag{7.145}$$

Similarly as before:

[12] Leibniz's rule for differentiation under the integral sign states that for an integral of the form

$$\int_{a(x)}^{b(x)} f(x,t)\mathrm{d}t$$

where $\infty < a(x), b(x) < \infty$, the derivative of this integral can be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\int_{a(x)}^{b(x)} f(x,t)\mathrm{d}t\right] = f\left[x, b(x)\right]\frac{\mathrm{d}}{\mathrm{d}x}b(x) - f\left[x, a(x)\right]\frac{\mathrm{d}}{\mathrm{d}x}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)\mathrm{d}t$$

where the partial derivative indicates that inside the integral, only the variation of $f(x,t)$ with x is considered in taking the derivative.

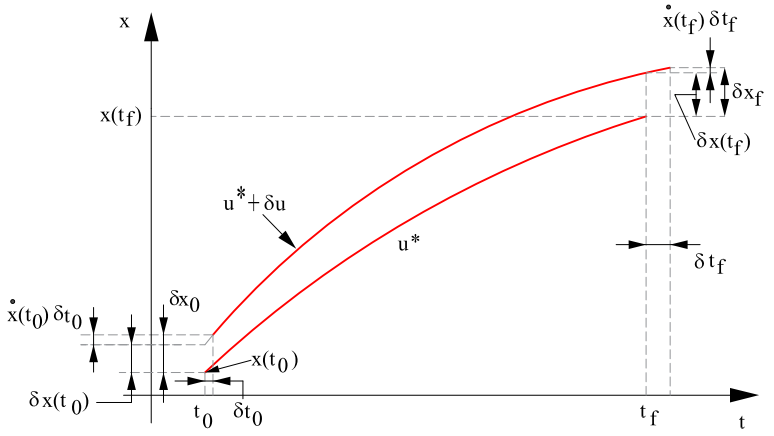


Fig. 7.52 The variation $\delta\mathbf{x}$ for the basic problem with free end points

$$\int_{t_0}^{t_f} \delta\left(\mathcal{H} - \boldsymbol{\lambda}^{\mathrm{T}}\dot{\mathbf{x}}\right) \mathrm{d}t = -\boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x}\Big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \left[\left(\frac{\partial\mathcal{H}}{\partial\mathbf{x}} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\right)\delta\mathbf{x} + \frac{\partial\mathcal{H}}{\partial\mathbf{u}}\delta\mathbf{u}\right] \mathrm{d}t \tag{7.146}$$

so

$$\delta\bar{\mathcal{J}} = -\boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x}\Big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \left[\left(\frac{\partial\mathcal{H}}{\partial\mathbf{x}} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\right)\delta\mathbf{x} + \frac{\partial\mathcal{H}}{\partial\mathbf{u}}\delta\mathbf{u}\right] \mathrm{d}t + \left(\mathcal{H} - \boldsymbol{\lambda}^{\mathrm{T}}\dot{\mathbf{x}}\right)\delta t\Big|_{t_0}^{t_f} \tag{7.147}$$

or

$$\delta\bar{\mathcal{J}} = \int_{t_0}^{t_f} \left[\left(\frac{\partial\mathcal{H}}{\partial\mathbf{x}} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\right)\delta\mathbf{x} + \frac{\partial\mathcal{H}}{\partial\mathbf{u}}\delta\mathbf{u}\right] \mathrm{d}t + \left[\mathcal{H}\delta t - \boldsymbol{\lambda}^{\mathrm{T}}\left(\dot{\mathbf{x}}\delta t + \delta\mathbf{x}\right)\right]_{t_0}^{t_f} \tag{7.148}$$

For the variations at the end points the following transversality conditions holds (Fig. 7.52):

$$\delta\mathbf{x_0} = \delta\mathbf{x}(t_0) + \dot{\mathbf{x}}(t_0)\delta t_0 \qquad \delta\mathbf{x_f} = \delta\mathbf{x}(t_f) + \dot{\mathbf{x}}(t_f)\delta t_f \tag{7.149}$$

Substituting the end-point variations into the equation for $\delta\bar{\mathcal{J}} = 0$ yields:

$$\left[\mathcal{H}\delta t - \boldsymbol{\lambda}^{\mathrm{T}}\left(\dot{\mathbf{x}}\delta t + \delta\mathbf{x}\right)\right]_{t_0}^{t_f} = \left(\mathcal{H}\delta t_f - \boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x_f}\right)_{t=t_f} - \left(\mathcal{H}\delta t_0 - \boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x_0}\right)_{t=t_0} = 0 \tag{7.150}$$

or, with the interior products written out more explicitly:

$$\left[\mathcal{H}\delta t_f - \sum_{i=1}^{n} \lambda_i \left(\delta x_i\right)_f\right]_{t=t_f} - \left[\mathcal{H}\delta t_0 - \sum_{i=1}^{n} \lambda_i \left(\delta x_i\right)_0\right]_{t=t_0} = 0 \tag{7.151}$$

The above expressions constitute the transversality conditions. Since the transversality conditions apply to any admissible variation in the endpoints, each individual term of the total $2n + 2$ terms must be zero. This means that there are $2n + 2$ conditions:

$$\lambda_i(t_0)\left(\delta x_i\right)_0 = 0,\ \lambda_i(t_f)\left(\delta x_i\right)_f = 0\ (i = 1, \ldots, n),\ \mathcal{H}\delta t_0 = 0 \text{ and } \mathcal{H}\delta t_f = 0 \tag{7.152}$$

In generic terms, we have the conditions $\sum_{i=1}^{2n+2} A_i \delta B_i = 0$. This is valid for any admissible variation δB_i, so $A_i \delta B_i = 0$, for $i = 1, 2, \ldots, 2n + 2$. When B_i is given, this means that $\delta B_i = 0$, and thus when B_i is not given this means that $A_i = 0$. More specifically, we have, for $i = 1, 2, \ldots, n$, that the forced boundary conditions are given by prescribed $x_i(t_0)$ and $x_i(t_f)$, versus the (equivalent) natural boundary conditions $\lambda_i(t_0) = 0$ and $\lambda_i(t_f) = 0$. Moreover, for forced boundary conditions both t_0 and t_f are given, whereas in the case of natural boundary conditions $\mathcal{H}|_{t=t_0} = 0$ and $\mathcal{H}|_{t=t_f} = 0$.

Example 7.7 (*Optimal Control with Free End Points*) We will now revisit Example 7.6 and instead of having fixed end conditions, we will leave one of them free. To reintroduce the problem,

$$\mathcal{J} = \int_{t_0=0}^{t_f=1} \left(\frac{1}{2}u^2 + x\right) \mathrm{d}t$$

is minimised, subject to

$$\dot{x} = u$$

with $x(0) = 0$ and $\mathbf{x}(\mathbf{1}) = \mathbf{free}$. The solution is, following similar steps as before

$$x^* = c_1 + c_2 t + \frac{1}{2}t^2 \qquad \dot{x}^* = c_2 + t = u^* = -\lambda_x^*$$

The prescribed boundary condition provides constant of integration, c_1:

$$x(0) = 0 \ \Rightarrow\ c_1 = 0$$

From the natural boundary condition it follows that

$$\dot{x}^* = c_2 + t = u^* = -\lambda_x^*(1) = 0 = u^*(-1) = -(c_2 + 1) \ \Rightarrow\ c_2 = -1$$

Hence, the optimal solution is:

$$x^* = -t + \frac{1}{2}t^2 \qquad u^* = -1 + t. \qquad \blacksquare$$

If an end-cost term is added to the performance index, a more general formulation known as the Bolza problem is obtained:

$$\mathcal{J} = \Phi(\mathbf{x_f}, t_f) + \int_{t_0}^{t_f} \mathcal{F}[\mathbf{x}(t), \mathbf{u}(t)]\mathrm{d}t \tag{7.153}$$

Two special cases follow from the above formulation, i.e., when $\Phi = 0$ Eq. (7.153) reduces to the Lagrange problem, as discussed above, and when $\mathcal{F} = 0$, the problem is known as a *Mayer* problem.

The Bolza (or Mayer) problem can be easily converted into Lagrange problem (and vice versa). Write

$$\Phi(\mathbf{x_f}, t_f) = \int_{t_0}^{t_f} \dot{\Phi}(\mathbf{x}(t), t)\mathrm{d}t, \qquad \text{with } t_0 \text{ given} \tag{7.154}$$

and thus

$$\mathcal{J} = \int_{t_0}^{t_f} \left\{\mathcal{F}[\mathbf{x}(t), \mathbf{u}(t)] + \dot{\Phi}(\mathbf{x}(t), t)\mathrm{d}t\right\} \mathrm{d}t \tag{7.155}$$

which is the formulation of the Lagrange problem.

With no change in the definition of the Hamiltonian, the augmented performance index becomes:

$$\bar{\mathcal{J}} = \int_{t_0}^{t_f} \left\{\mathcal{H}\left[\mathbf{x}(t), \mathbf{u}(t)\right] - \boldsymbol{\lambda}^{\mathrm{T}}(t)\dot{\mathbf{x}} + \dot{\Phi}(\mathbf{x}(t), t)\right\} \mathrm{d}t \tag{7.156}$$

and consequently

$$\delta\bar{\mathcal{J}} = \delta\left(\int_{t_0}^{t_f} \left\{\mathcal{H}\left[\mathbf{x}(t), \mathbf{u}(t)\right] - \boldsymbol{\lambda}^{\mathrm{T}}(t)\dot{\mathbf{x}}\right\} \mathrm{d}t\right) + \delta\left(\int_{t_0}^{t_f} \dot{\Phi}(\mathbf{x}(t), t)\mathrm{d}t\right) \tag{7.157}$$

The first term gives, as before, the Euler–Lagrange equations. The second term expands to

$$\delta\left(\int_{t_0}^{t_f}\dot{\Phi}(\mathbf{x}(t),t)\mathrm{d}t\right)=\delta\left[\Phi(\mathbf{x_f},t_f)\right]=\frac{\partial\Phi}{\partial t_f}\delta t_f+\frac{\partial\Phi}{\partial\mathbf{x_f}}\delta\mathbf{x_f}=0 \tag{7.158}$$

These terms have to be added to the transversality conditions (with fixed initial time t_0):

$$\left[\left(\mathcal{H}+\frac{\partial\Phi}{\partial t_f}\right)\delta t_f+\left(-\boldsymbol{\lambda}^{\mathrm{T}}+\frac{\partial\Phi}{\partial\mathbf{x_f}}\right)\delta\mathbf{x_f}\right]_{t=t_f}-\left(-\boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x_0}\right)_{t=t_0}=0 \tag{7.159}$$

Summarised, the transversality conditions for the Bolza problem are:

$$x_i(t_0)\text{ is given or }\lambda_i(t_0)=0,\qquad(i=1,2,\ldots,n) \tag{7.160}$$

$$x_i(t_f)\text{ is given or }\lambda_i(t_f)=0,\qquad(i=1,2,\ldots,n) \tag{7.161}$$

$$t_f\text{ is given or }\mathcal{H}|_{t=t_f}=\frac{\partial\Phi}{\partial t_f} \tag{7.162}$$

Example 7.8 (*Optimal Control Mayer Problem*) In this example, the objective is to minimise $\mathcal{J}=\Phi(x_f)=x_f$, subject to

$$\begin{aligned}\dot{x}_1&=\frac{1}{2}\left(ax_2^2+bu^2\right)\quad\text{with } a>0, b>0\\ \dot{x}_2&=u\end{aligned}$$

with boundary conditions: $x_1(0)=0, x_2(0)=x_2(t_f)=1$. The final time, t_f is given. Note that the above problem is a Mayer problem, since $\mathcal{F}=0$.

The Hamiltonian is formed as:

$$\mathcal{H}=\lambda_{x_1}\frac{1}{2}\left(ax_2^2+bu^2\right)+\lambda_{x_2}u$$

The optimal-control conditions, Eq. (7.143), are easily found to be

$$\frac{\partial\mathcal{H}}{\partial\mathbf{u}}=\lambda_{x_1}bu+\lambda_{x_2}=0\quad\Rightarrow\quad u^*=-\frac{1}{b}\frac{\lambda_{x_2}}{\lambda_{x_1}}$$

and

$$\frac{\partial^2\mathcal{H}}{\partial\mathbf{u}^2}=\lambda_{x_1}b\geq 0\quad\Rightarrow\quad\lambda_{x_1}\geq 0$$

The adjoint equations are found by differentiating $\mathcal{H}$ with respect to the states, so:

$$\dot{\lambda}_{x_1}=-\frac{\partial\mathcal{H}}{\partial x_1}=0\quad\Rightarrow\quad\lambda_{x_1}=\text{constant}$$

$$\dot{\lambda}_{x_2} = -\frac{\partial \mathcal{H}}{\partial x_2} = \lambda_{x_1} a x_2$$

Finally, the transversality condition results in

$$\lambda_{x_1}(t_f) = \frac{\partial \Phi}{\partial x_{1,f}} = 1 \quad \Rightarrow \quad \lambda_{x_1}(t) = 1 \geq 0$$

Using this result, it thus follows for the control that

$$u^* = -\frac{1}{b}\frac{\lambda_{x_2}}{\lambda_{x_1}} = -\frac{1}{b}\lambda_{x_2}$$

Therefore,

$$\dot{x}_2 = u = u^* = -\frac{1}{b}\lambda_{x_2} \quad \text{and} \quad \dot{\lambda}_{x_2} = -a x_2$$

Differentiation gives then:

$$\ddot{x}_2 = -\frac{1}{b}\dot{\lambda}_{x_2} = \frac{a}{b}x_2 \quad \Rightarrow \quad \ddot{x}_2 - \frac{a}{b} = 0$$

The eigenvalues of this constraint equation are:

$$\lambda_{1,2} = \pm\sqrt{\frac{a}{b}} \triangleq \pm\gamma$$

with the solution for $x_2(t)$ being:

$$x_2 = \alpha e^{-\gamma t} + \beta e^{\gamma t}$$

Here, α and β are integration constants that can be solved from the boundary conditions:

$$x_2(0) = 1 = \alpha + \beta \quad \text{and} \quad x_2(t_f) = 1 = \alpha e^{-\gamma t_f} + \beta e^{\gamma t_f}$$

Solving for α and β yields

$$\alpha = \frac{e^{\gamma t_f} - 1}{e^{\gamma t_f} - e^{-\gamma t_f}} \quad \text{and} \quad \beta = \frac{1 - e^{-\gamma t_f}}{e^{\gamma t_f} - e^{-\gamma t_f}}$$

The optimal "constraint state" is thus formulated as:

$$x_2^* = \alpha e^{-\gamma t} + \beta e^{\gamma t} = \frac{e^{\gamma t_f} - 1}{e^{\gamma t_f} - e^{-\gamma t_f}} e^{-\gamma t} + \frac{1 - e^{-\gamma t_f}}{e^{\gamma t_f} - e^{-\gamma t_f}} e^{\gamma t}$$

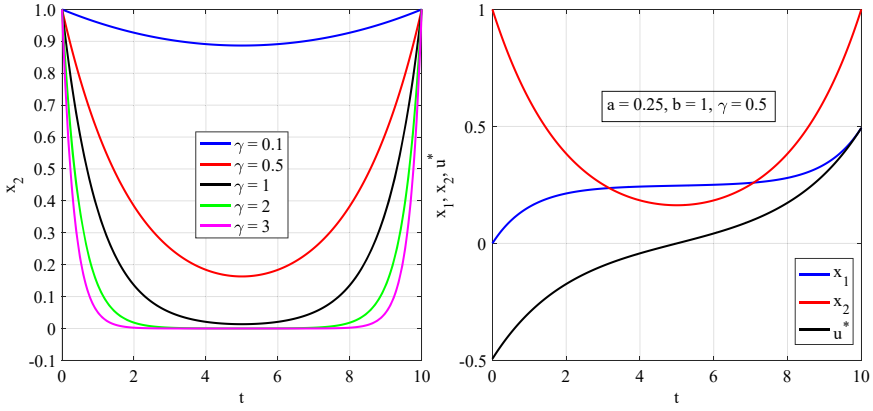


Fig. 7.53 Some solutions for the Mayer problem of Example 7.8, with varying γ (left), and for $\gamma = 0.5$ (right)

The optimal control u^* follows from differentiating x_2,

$$u^* = \dot{x}_2^* = -\alpha\gamma e^{-\gamma t} + \beta\gamma e^{\gamma t}$$

and, finally, the optimal state follows from

$$\dot{x}_1^* = \frac{1}{2}\left[a(x_2^*)^2 + b(u^*)^2\right] \quad \text{with } a > 0, b > 0$$

Some example solutions for different values of $\gamma = \sqrt{\frac{a}{b}}$ are shown in Fig. 7.53. ■

To conclude this section we will briefly elaborate on Pontryagin's Minimum (or, equivalently, Maximum) Principle, which is key to find the best possible control that a dynamical system should take from one state to another, in particular in the presence of constraints for the state or controls. Usually, the control $\mathbf{u}$ belongs to a closed and bounded set, $\mathcal{U}$:

$$\mathbf{u} \in \mathcal{U} \tag{7.163}$$

The boundaries of $\mathcal{U}$ represent the control constraints and are often defined as:

$$\mathbf{u}_{\min} \leq \mathbf{u} \leq \mathbf{u}_{\max} \tag{7.164}$$

The question that needs answering now, is how do control constraints affect the optimal-control problem? From before, Eq. (7.133), we have

$$\delta\bar{\mathcal{J}} = -\boldsymbol{\lambda}^{\mathrm{T}}\delta\mathbf{x}\Big|_{t_0}^{t_f} + \int_{t_0}^{t_f}\left[\left(\frac{\partial\mathcal{H}}{\partial\mathbf{x}} + \dot{\boldsymbol{\lambda}}^{\mathrm{T}}\right)\delta\mathbf{x} + \frac{\partial\mathcal{H}}{\partial\mathbf{u}}\delta\mathbf{u}\right]\mathrm{d}t \tag{7.165}$$

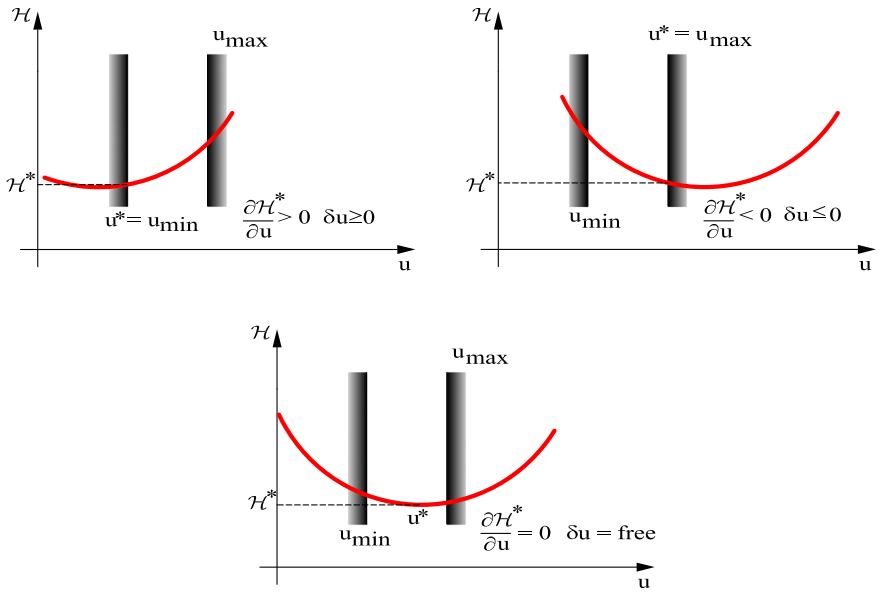


Fig. 7.54 Optimal-control solutions for constrained-control problems

In the current, constrained case, $\delta\mathbf{u}$ can no longer be chosen arbitrary. For $\mathbf{u}(t)$ to be a minimising control, $\delta\bar{\mathcal{J}} \geq 0$ for all admissible $\delta\mathbf{u}$. The optimal $\mathbf{u}^*$ has the property that

$$\delta\mathcal{H}' \triangleq \frac{\partial \mathcal{H}}{\partial \mathbf{u}} \delta\mathbf{u} \geq 0 \tag{7.166}$$

Now, Pontryagin's Minimum Principle states that for a given value of $\mathbf{x}(t)$ and $\boldsymbol{\lambda}(t)$, the optimal control $\mathbf{u}^*$ must be selected as an element of the control space $\mathcal{U}$, such that the Hamiltonian is minimised. Mathematically, this is formulated as

$$\mathbf{u}^* = \arg\left[\min_{\mathbf{u}\in\mathcal{U}} \mathcal{H}(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{u})\right] \tag{7.167}$$

For a scalar control u, this can be readily interpreted as (see also Fig. 7.54):

$$u_{\mathcal{U}} = \arg\left(\frac{\partial \mathcal{H}}{\partial u} = 0\right) \quad \Rightarrow \quad \begin{cases} u^* = u_{max} & : \quad \text{if } u_{\mathcal{U}} > u_{max} \\ u^* = u_{\mathcal{U}} & : \quad \text{if } u_{min} \leq u_{\mathcal{U}} \leq u_{max} \\ u^* = u_{min} & : \quad \text{if } u_{\mathcal{U}} < u_{min} \end{cases} \tag{7.168}$$

If $\mathbf{u_{min}}$ and/or $\mathbf{u_{max}}$ depend on the state $\mathbf{x}$, the problem is more complicated, as one has to use a multiplier technique. This is considered to be beyond the scope of this brief introduction, though.

Example 7.9 (*Optimal Control Problem with Constrained Controls*) In this last example, we will revisit Example 7.6 once more. The problem was formulated as minimising

$$\mathcal{J} = \int_{t_0=0}^{t_f=1} \left(\frac{1}{2}u^2 + x\right) \mathrm{d}t$$

subject to

$$\dot{x} = u$$

with $x(0) = 0$ and $x(1) = 1$.

However, assume now the following control constraint:

$$\frac{3}{4} \leq u \leq \frac{5}{4}$$

The optimal-control condition is now:

$$\begin{array}{lll} u^* = \frac{3}{4} & : & \text{if } u_{\mathcal{U}} < \frac{3}{4} \\ u^* = u_{\mathcal{U}} = -\lambda_x & : & \text{if } \frac{3}{4} \leq u_{\mathcal{U}} \leq \frac{5}{4} \\ u^* = \frac{5}{4} & : & \text{if } u_{\mathcal{U}} > \frac{5}{4} \end{array}$$

Obviously, at any given point in time we do not know a priori which of the three possibilities actually applies. Based on the unconstrained solution, assume the following subsequent phases:

$$\begin{array}{lll} 0 \leq t \leq t_1 & : & u^* = u_{min} = \frac{3}{4} \\ t_1 \leq t \leq t_2 & : & u^* = -\lambda_x \\ t_2 \leq t \leq 1 & : & u^* = u_{max} = \frac{5}{4} \end{array}$$

where t_1 and t_2 are to be determined.

Recall from the unconstrained solution that the adjoint equation is:

$$\dot{\lambda}_x = -1$$

Phase 1 : $0 \leq t \leq t_1, u^* = u_{min} = \frac{3}{4}$

$$\dot{\lambda}_x = -1 \Rightarrow \lambda_x(t) = -t + \lambda_x(0);\ \lambda_x(t_1) = -t_1 + \lambda_x(0) = \frac{3}{4}$$

$$\dot{x} = \frac{3}{4} \Rightarrow x(t) = \frac{3}{4}t;\ x(t_1) = \frac{3}{4}t_1$$

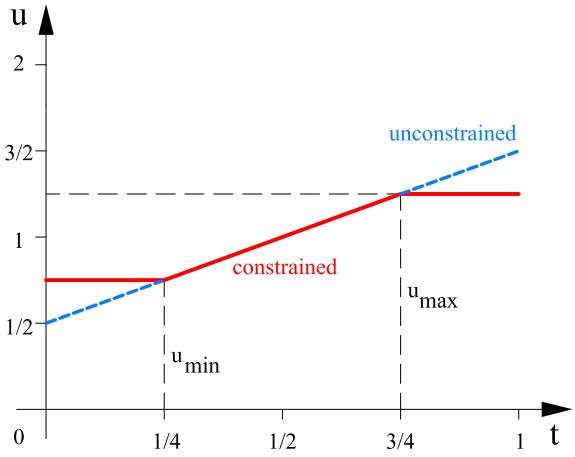


Fig. 7.55 The constrained and unconstrained control solutions for Example 7.9

Phase 2 : $t_1 \leq t < t_2, u^* = -\lambda_x$

$$\dot{\lambda}_x = -1 \Rightarrow \lambda_x(t) = \lambda_x(t_1) - (t - t_1)$$

$$\lambda_x(t_2) = \frac{3}{4} - (t_2 - t_1) = \frac{5}{4} \Rightarrow t_2 - t_1 = \frac{1}{2}$$

$$\dot{x} = -\lambda_x \Rightarrow x(t) = x(t_1) = \lambda_x(t_1)(t - t_1) + \frac{1}{2}(t - t_1)^2$$

$$x(t_2) = x(t_1) - \lambda_x(t_1)(t_2 - t_1) + \frac{1}{2}(t_2 - t_1)^2 = \frac{3}{4}t_1 + \frac{1}{2}$$

Phase 3 : $t_2 \leq t \leq 1, u^* = u_{max} = \frac{5}{4}$

$$\dot{x} = \frac{5}{4} \Rightarrow x(t) = x(t_2) + \frac{5}{4}(t - t_2)$$

$$x(1) = x(t_2) + \frac{5}{4}(1 - t_2) = \left(\frac{3}{4}t_1 + \frac{1}{2}\right) + \frac{5}{4}\left[1 - \left(t_1 + \frac{1}{2}\right)\right] = \frac{1}{2}t_1 + \frac{9}{8} = 1$$

Thus, solving for t_1 and t_2 yields $t_1 = \frac{1}{4}$ and $t_2 = t_1 + \frac{1}{2} = \frac{3}{4}$. The constrained and unconstrained control solutions are shown in Fig. 7.55. ■

7.5.4.2 Linear Quadratic Regulator

In the previous section, we introduced the theory to find a control profile to follow an optimal path, given a cost criterion and possibly some constraints on either the states or the controls. In the current section, a control system will be introduced that will produce the commands to follow such an optimal path, given the so-called *error dynamics* of the system under consideration. Given the error state $\Delta\mathbf{x}$ and the corrective control $\Delta\mathbf{u}$ to return the system to equilibrium—and for simplicity dropping the Δ notation as we did before—from Sect. 7.5.2 we know that the state-space description of the system to be controlled is:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \qquad \mathbf{y} = \mathbf{C}\mathbf{x} \tag{7.169}$$

with $\mathbf{x}(t_0) = \mathbf{0}$. For state feedback, the control law was defined to be

$$\mathbf{u} = -\mathbf{K}\mathbf{x} \tag{7.170}$$

The feedback matrix, $\mathbf{K}$, can be computed in a mathematically closed form. An indirect method is Quadratic Optimal Control, in which a mathematically defined cost criterion is minimised. A direct method is the so-called pole placement, in which $\mathbf{K}$ is solved on basis of the specified poles of the closed-loop system. In this section we will continue with the indirect method, where we will use the following quadratic cost criterion (Gopal 1989):

$$\mathcal{J} = \int_0^\infty (\mathbf{x}^\mathrm{T}\mathbf{Q}\mathbf{x} + \mathbf{u}^\mathrm{T}\mathbf{R}\mathbf{u})\mathrm{d}t \tag{7.171}$$

where the term $\mathbf{x}^\mathrm{T}\mathbf{Q}\mathbf{x}$ represents the state deviation and the term $\mathbf{u}^\mathrm{T}\mathbf{R}\mathbf{u}$ the control effort. $\mathbf{Q}$ is a real positive semi-definite matrix, whereas $\mathbf{R}$ is a real symmetric positive definite matrix, so any $\mathbf{x}$, $\mathbf{u} \neq \mathbf{0}$ cannot give a negative contribution to $\mathcal{J}$. With the proper selection of $\mathbf{Q}$ and $\mathbf{R}$, more weight can be given to the control deviation, resulting in a faster response, or the control effort, giving smaller control signals. By varying each of the elements of $\mathbf{Q}$ and $\mathbf{R}$, each of the corresponding elements of $\mathbf{x}$ and $\mathbf{u}$ can be addressed. Defining $\mathbf{Q}$ and $\mathbf{R}$ is usually done in an iterative manner, and a good first choice is given by "Bryson's Rule" (Bryson and Ho 1975):

$$\mathbf{Q} = \mathrm{diag}\left\{ \frac{1}{\Delta x_{1_{max}}^2} \quad \frac{1}{\Delta x_{2_{max}}^2} \quad \cdots \quad \frac{1}{\Delta x_{n_{max}}^2} \right\} \tag{7.172}$$

with $\Delta x_{i_{max}}$ the maximum allowable amplitude of the ith element of the state vector, and

$$\mathbf{R} = \mathrm{diag}\left\{ \frac{1}{\Delta u_{1_{max}}^2} \quad \frac{1}{\Delta u_{2_{max}}^2} \quad \cdots \quad \frac{1}{\Delta u_{m_{max}}^2} \right\} \tag{7.173}$$

with $\Delta u_{j_{max}}$ the maximum allowable value of the jth control. Note that because $\mathbf{Q}$ is positive semi-definite, this means that any of the diagonal elements can be put to zero, or, equivalently, the corresponding $\Delta x_i = \infty$, thereby removing the particular state variable as a constraint.

Response tests must prove that these weighting matrices have been correctly selected. If not, the values of the diagonal elements should be adjusted. Frangos and Yavin (1992) propose a synthesis procedure that automatically varies the weighting matrices and computes the gains in an iterative manner, based on minimisation of the quadratic cost criterion $\mathcal{J}$. Luo and Lan (1995), finally, describe a systematic method to determine the weighting matrices, so as to produce specified closed-loop eigenvalues. The discussion of either algorithm, however, is considered to be beyond the scope of the current text.

To find an expression for $\mathbf{K}$, we substitute Eq. (7.170) into the cost criterion Eq. (7.171), yielding

$$\mathcal{J} = \int_0^\infty \mathbf{x}^{\mathrm{T}}(\mathbf{Q} + \mathbf{K}^{\mathrm{T}}\mathbf{R}\mathbf{K})\mathbf{x}\mathrm{d}t \tag{7.174}$$

To solve the optimisation problem, we will use the stability analysis according to Lyapunov. The integrand of the above integral equation is considered to be the negative time derivative of the Lyapunov function $V(\mathbf{x})$:

$$V(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x} \tag{7.175}$$

which attaches a mathematically formulated fictitious energy to the system as a function of the state vector (Gopal 1989). This makes the minimisation of $\mathcal{J}$ to be a minimisation of the integrated fictitious power. $\mathbf{P}$ is a positive definite matrix, so every deviation of the state from the equilibrium state is treated as a positive energy. Based on the above, we may write

$$\mathbf{x}^{\mathrm{T}}(\mathbf{Q} + \mathbf{K}^{\mathrm{T}}\mathbf{R}\mathbf{K})\mathbf{x} = -\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x}) \tag{7.176}$$

Differentiating results in

$$\begin{aligned} \mathbf{x}^{\mathrm{T}}(\mathbf{Q} + \mathbf{K}^{\mathrm{T}}\mathbf{R}\mathbf{K})\mathbf{x} &= -\dot{\mathbf{x}}^{\mathrm{T}}\mathbf{P}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{P}\dot{\mathbf{x}} \\ &= -\mathbf{x}^{\mathrm{T}}\left[(\mathbf{A} - \mathbf{B}\mathbf{K})^{\mathrm{T}}\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{B}\mathbf{K})\right]\mathbf{x} \end{aligned} \tag{7.177}$$

Since the above expression should be valid for all $\mathbf{x}$, the following expression holds:

$$(\mathbf{A} - \mathbf{B}\mathbf{K})^{\mathrm{T}}\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{B}\mathbf{K}) = -(\mathbf{Q} + \mathbf{K}^{\mathrm{T}}\mathbf{R}\mathbf{K}) \tag{7.178}$$

or

$$\tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{P} + \mathbf{P}\tilde{\mathbf{A}} = -\tilde{\mathbf{Q}} \tag{7.179}$$

with $\tilde{\mathbf{A}} = \mathbf{A} - \mathbf{BK}$ and $\tilde{\mathbf{Q}} = \mathbf{Q} + \mathbf{K}^{\mathrm{T}}\mathbf{RK}$. Equation (7.179) is called the *Lyapunov Equation*. Gopal (1989) states that the linear system

$$\dot{\mathbf{x}} = \tilde{\mathbf{A}}\mathbf{x} \tag{7.180}$$

is globally asymptotically stable[13] at the origin if and only if for any symmetric positive definite matrix $\tilde{\mathbf{Q}}$, there exists a symmetric positive definite matrix $\mathbf{P}$ that satisfies the Lyapunov Equation. This gives us the possibility to check whether our system can be stabilised or not.

Continuing with Eqs. (7.174) and (7.176), the cost criterion can be written as:

$$\mathcal{J} = -\left.\mathbf{x}^{\mathrm{T}}\mathbf{Px}\right|_0^\infty = \mathbf{x}(0)^{\mathrm{T}}\mathbf{Px}(0) \tag{7.181}$$

assuming that our system is asymptotically stable, or in mathematical terms

$$\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{0} \tag{7.182}$$

Minimising $\mathcal{J}$ is now the same as minimising the right-hand term of Eq. (7.181) by means of $\mathbf{K}$. Because $\mathbf{R}$ is positive definite, we may decompose it as

$$\mathbf{R} = \mathbf{S}^{\mathrm{T}}\mathbf{S} \tag{7.183}$$

Substituting this into the Lyapunov Equation (7.178) yields

$$(\mathbf{A}^{\mathrm{T}} - \mathbf{K}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}})\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{BK}) + \mathbf{Q} + \mathbf{K}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{SK} = \mathbf{0} \tag{7.184}$$

which can be further expanded to

$$\mathbf{A}^{\mathrm{T}}\mathbf{P} + \mathbf{PA} - \mathbf{PBR}^{-1}\mathbf{BP} + \mathbf{Q} + \left[\mathbf{SK} - (\mathbf{S}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}\right]^{\mathrm{T}} \left[\mathbf{SK} - (\mathbf{S}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}\right] = \mathbf{0} \tag{7.185}$$

The last term on the left-hand side can be treated as a function of $\mathbf{K}$, so $\mathcal{J}$ is minimised when

$$\mathbf{x}^{\mathrm{T}}\left[\mathbf{SK} - (\mathbf{S}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}\right]^{\mathrm{T}} \left[\mathbf{SK} - (\mathbf{S}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}\right]\mathbf{x} \tag{7.186}$$

is minimised with respect to $\mathbf{K}$. Since Eq. (7.186) is either zero or greater than zero, the minimum is found when Eq. (7.186) is zero, or

$$\mathbf{SK} = (\mathbf{S}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P} \tag{7.187}$$

so that the optimal feedback matrix $\mathbf{K}$ can be expressed as a function of $\mathbf{P}$:

[13] Kuo (1987) defines asymptotic stability as follows: "If the zero-input response $\mathbf{x}(t)$, subject to the finite initial state $\mathbf{x}(t_0)$, returns to the equilibrium state $\mathbf{x}(t) = \mathbf{0}$ as t approaches infinity, the system is said to be stable; otherwise, the system is unstable. This type of stability is also known as the asymptotic stability."

$$\mathbf{K} = \mathbf{S}^{-1}(\mathbf{S}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P} = \mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P} \tag{7.188}$$

Substituting Eq. (7.188) into Eq. (7.185) gives us $\mathbf{P}$ as a function of the system matrices $\mathbf{A}$ and $\mathbf{B}$, and the weight matrices $\mathbf{Q}$ and $\mathbf{R}$:

$$\mathbf{A}^{\mathrm{T}}\mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}\mathbf{P} + \mathbf{Q} = \mathbf{0} \tag{7.189}$$

Equation (7.189) is also known as the *matrix Riccati equation* or *algebraic Riccati equation*. Solving this equation can commonly be done with standard algorithms, available in control-system design tools such as MATLAB®.

Does the above controller relate in any way to the optimal-control theory we discussed in the previous section? Yes, it actually does, and we will briefly show how. More details can be found, for instance, in the book by Lewis et al. (2012). Starting again with the error-dynamics system Eq. (7.169), we define a slightly altered cost criterion:

$$\mathcal{J} = \frac{1}{2}\int_0^{t_f} (\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x} + \mathbf{u}^{\mathrm{T}}\mathbf{R}\mathbf{u})\mathrm{d}t + \frac{1}{2}\mathbf{x}^{\mathrm{T}}(t_f)\mathbf{P_1}\mathbf{x}(t_f) \tag{7.190}$$

The factor $\frac{1}{2}$ has been added here just for the convenience of the derivation, but does not alter the cost criterion apart from a scaling. The last term on the right-hand side is an end-cost term or *terminal constraint*, conform $\Phi(\mathbf{x_f}, t_f)$ in Eq. (7.153).

The Hamiltonian, Eq. (7.121), is now formulated as

$$\mathcal{H} \triangleq \mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x} + \mathbf{u}^{\mathrm{T}}\mathbf{R}\mathbf{u} + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}) \tag{7.191}$$

where each of the $\mathbf{x}$, $\mathbf{u}$, and $\boldsymbol{\lambda}$ are functions of time, but where we have dropped the "(t)" notation for brevity.

Finding the optimal solution of Eq. (7.190) is done by using Pontryagin's minimum principle, i.e., Eqs. (7.134) and (7.138):

$$-\dot{\boldsymbol{\lambda}}^{\mathrm{T}} = \left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}}\right)^{\mathrm{T}} = \mathbf{Q}\mathbf{x} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} \quad \text{with } \boldsymbol{\lambda}(t_f) = \mathbf{P_1}\mathbf{x}(t_f) \tag{7.192}$$

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = 0 = \mathbf{R}\mathbf{u} + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B} \quad \Rightarrow \quad \mathbf{u} = -\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda} \tag{7.193}$$

The state derivative follows easily from

$$\left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}}\right)^{\mathrm{T}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \quad \text{with } \mathbf{x}(t_0) = \mathbf{0} \tag{7.194}$$

and is, of course, the original constraint equation.

The above set of equations is hard to solve, though. Instead, let us *guess* the form of the solution for the adjoint functions, i.e., $\boldsymbol{\lambda} = \mathbf{P}(t)\mathbf{x}(t)$. It then follows that

$$\dot{\boldsymbol{\lambda}} = \dot{\mathbf{P}}\mathbf{x} + \mathbf{P}\dot{\mathbf{x}} = \dot{\mathbf{P}}\mathbf{x} + \mathbf{P}\left(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}\right) \tag{7.195}$$

or, with Eq. (7.193) and substituting $\boldsymbol{\lambda} = \mathbf{P}\mathbf{x}$:

$$\dot{\boldsymbol{\lambda}} = \dot{\mathbf{P}}\mathbf{x} + \mathbf{P}\left(\mathbf{A}\mathbf{x} - \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}\right)\mathbf{x} \tag{7.196}$$

Reshuffling terms, and substituting Eq. (7.192) for $\dot{\boldsymbol{\lambda}}$, results in

$$-\dot{\mathbf{P}}\mathbf{x} - \mathbf{P}\mathbf{A}\mathbf{x} + \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}\mathbf{x} = \mathbf{Q}\mathbf{x} + \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{x} \tag{7.197}$$

This equation needs to be satisfied for any value of $\mathbf{x}$, so

$$-\dot{\mathbf{P}} = \mathbf{P}\mathbf{A} + \mathbf{A}^{\mathrm{T}}\mathbf{P} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P} + \mathbf{Q} \tag{7.198}$$

with the terminal condition for $\mathbf{P}$ given by $\mathbf{P}(t_f) = \mathbf{P_1}$. Equation (7.198) is called the Riccati ordinary differential equation.

Since the terminal condition for $\mathbf{P}$ is given, we can use that to integrate $\mathbf{P}(t)$ backwards and through Eq. (7.193), i.e., $\mathbf{u}(t) = -\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}(t)$ calculate the time-varying feedback control, $\mathbf{u}(t)$. It will allow the system to move from any state towards the origin. To be in line with the quadratic cost criterion introduced at the beginning of this section, Eq. (7.171), the final time is set to infinity, and the terminal constraint is removed. Going through the steps, the same result as before is obtained:

$$\mathbf{u}(t) = -\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P} \tag{7.199}$$

where it can be shown that $\mathbf{P}$ is constant. Thus, Eq. (7.198) changes to the *algebraic* Riccati equation (7.189):

$$\mathbf{0} = \mathbf{P}\mathbf{A} + \mathbf{A}^{\mathrm{T}}\mathbf{P} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P} + \mathbf{Q}. \tag{7.200}$$

7.5.5 Incremental Non-linear Dynamics Inversion

Fixed-gain controllers will usually only have the best performance for the operating conditions for which they have been designed. In case these conditions are known to vary gain scheduling may be applied to accommodate this. Non-linear dynamic inversion (NDI) controllers were developed to eliminate non-linearities in the model by cancelling them out with state feedback. As a result, a single (classical) linear controller can be used as outer control loop of the system, to remove the need for gain scheduling. However, this can only work when an accurate on-board dynamics model

is available. As such, the desired dynamics of the closed-loop system is achieved by explicit model following. In case of actuator failures or model uncertainties, though, this model dependency may compromise the stability and performance of the system.

To reduce the model dependency several modifications to the NDI framework have been proposed, one of them being the concept of incremental non-linear dynamic inversion (INDI). By making use of actuator output and angular-acceleration measurement feedback the desired closed-loop dynamics does not come from explicit model following, but are the implicit result from closing the feedback loops (Sieberling et al. 2010; Acquatella et al. 2012).

Let us assume a non-linear system, affine in its control and where the output vector $\mathbf{y}$ is equal to the $n \times 1$ state vector $\mathbf{x}$:

$$\dot{\mathbf{y}} = \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u} \tag{7.201}$$

where $\mathbf{f}(\mathbf{x})$ represents the system dynamics, $\mathbf{G}(\mathbf{x})$ is the state-dependent control matrix of dimension $n \times m$, and $\mathbf{u}$ is the $m \times 1$ control vector. To get the approximated dynamics in its incremental form, the first-order Taylor-series expansion of $\dot{\mathbf{x}}$, evaluated around $[\mathbf{x_0}, \mathbf{u_0}]$ is

$$\dot{\mathbf{x}} = \dot{\mathbf{x}}_\mathbf{0} + \underbrace{\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u}] \Big|_{\substack{\mathbf{x}=\mathbf{x_0} \\ \mathbf{u}=\mathbf{u_0}}}}_{\mathbf{F}(\mathbf{x_0},\mathbf{u_0})=\mathbf{F_0}} (\mathbf{x} - \mathbf{x_0}) + \underbrace{\frac{\partial}{\partial \mathbf{u}} [\mathbf{G}(\mathbf{x})\mathbf{u}] \Big|_{\substack{\mathbf{x}=\mathbf{x_0} \\ \mathbf{u}=\mathbf{u_0}}}}_{\mathbf{G}(\mathbf{x_0})=\mathbf{G_0}} (\mathbf{u} - \mathbf{u_0}) \tag{7.202}$$

For a sufficiently high control update (and thus small time increment), $\mathbf{x}$ approaches $\mathbf{x_0}$, so $\mathbf{F_0}(\mathbf{x} - \mathbf{x_0}) << \mathbf{G_0}(\mathbf{u} - \mathbf{u_0})$ and can thus be ignored. The linearised system $\boldsymbol{\nu} = \dot{\mathbf{y}}$ is rewritten as

$$\boldsymbol{\nu} = \dot{\mathbf{x}}_\mathbf{0} + \mathbf{G_0}\delta\mathbf{u} \tag{7.203}$$

where the control *increment* $\delta\mathbf{u} = \mathbf{u} - \mathbf{u_0}$ has been introduced. This increment is obtained by inverting Eq. (7.203), and is to be added to the nominal (or reference) control $\mathbf{u_0}$ to obtain the control $\mathbf{u}$, input to the non-linear system. So

$$\mathbf{u} = \mathbf{u_0} + \delta\mathbf{u} = \mathbf{u_0} + \mathbf{G_0^{-1}} [\boldsymbol{\nu} - \dot{\mathbf{x}}_\mathbf{0}] \tag{7.204}$$

Exact knowledge of the plant dynamics $\mathbf{f}(\mathbf{x})$ is not required, as it is reflected by the measurement of $\dot{\mathbf{x}}_\mathbf{0}$ and would automatically include any change (or uncertainty) in $\mathbf{f}(\mathbf{x})$. In addition, the existing $\mathbf{u_0}$ is taken from the output of the actuators. In case of ideal actuators one can simply take the commanded value of the previous sample, but when actuator dynamics is involved, either an on-board model should be present or the actuator state should be estimated.

The control-law design is based on outer-loop control of the attitude (kinematics) and inner-loop control of the angular rate, $\boldsymbol{\omega}$ (dynamics), as shown in Fig. 7.56. Let us assume an attitude representation $\boldsymbol{\Phi}$, which could be based on Euler angles,

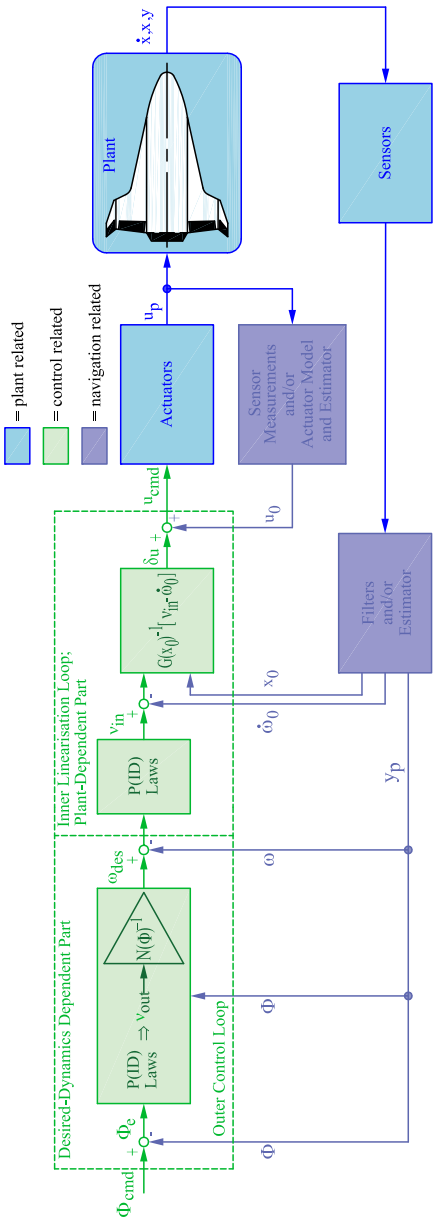


Fig. 7.56 General layout of the Incremental Non-linear Dynamics Inversion control system

quaternions, modified Rodrigues parameters, etc. (in the current scenario only the pitch angle is controlled, but a general formulation will be presented). So, for the outer loop

$$\mathbf{y_{out}} = \mathbf{\Phi} \tag{7.205}$$

with the generic formulation for its derivative given by

$$\dot{\mathbf{y}}_{\mathbf{out}} = \dot{\mathbf{\Phi}} = \mathbf{N}(\mathbf{\Phi})\boldsymbol{\omega} \tag{7.206}$$

The desired angular rate is easily obtained from the above equation through

$$\boldsymbol{\omega}_{\mathbf{des}} = \mathbf{N}(\mathbf{\Phi})^{-1}\left[\boldsymbol{\nu}_{\mathbf{out}}\right] \tag{7.207}$$

with an outer-loop virtual control, $\boldsymbol{\nu}_{\mathbf{out}}$, that can be obtained, for instance, from a simple P(ID) control law, or a more complex one, when the application requires that. Here, a proportional law is used:

$$\boldsymbol{\nu}_{\mathbf{out}} = \mathbf{K}_{\mathbf{p,out}}\mathbf{\Phi}_{\mathbf{e}} \tag{7.208}$$

with $\mathbf{\Phi}_e = \mathbf{\Phi}_{\mathbf{c}} - \mathbf{\Phi}$ being the attitude error, and the index c indicating the *desired* or *commanded* attitude. In this paper it will be the commanded pitch angle, θ_c.

The desired angular rate, $\boldsymbol{\omega}_{\mathbf{des}}$, serves as command for the inner-loop control. As before, we start with the inner-loop output vector:

$$\mathbf{y_{in}} = \boldsymbol{\omega} \tag{7.209}$$

with its derivative given by Eq. (7.203), i.e., $\dot{\mathbf{y}}_{\mathbf{in}} = \boldsymbol{\nu}_{\mathbf{in}}$. The virtual control parameter, $\boldsymbol{\nu}_{\mathbf{in}}$, may yet again be obtained from a simple P(ID) control law, taking $\boldsymbol{\omega}_{\mathbf{des}}$ as commanded value:

$$\boldsymbol{\nu}_{\mathbf{in}} = \mathbf{K}_{\mathbf{p,in}}\boldsymbol{\omega}_{\mathbf{e}} = \mathbf{K}_{\mathbf{p,in}}\left(\boldsymbol{\omega}_{\mathbf{des}} - \boldsymbol{\omega}\right) \tag{7.210}$$

Substituting $\boldsymbol{\nu}_{\mathbf{in}}$ into Eq. (7.204), finally, will give the control input to the plant p:

$$\mathbf{u_p} = \mathbf{u}_{\mathbf{p,0}} + \mathbf{G}_{\mathbf{0}}^{-1}\left[\boldsymbol{\nu}_{\mathbf{in}} - \dot{\boldsymbol{\omega}}_{\mathbf{0}}\right]. \tag{7.211}$$

7.5.6 Simple Adaptive Control

An alternative robust control concept is that of simple adaptive control (SAC) (Kaufman 1998), and is based on the principle of tracking the output of a reference model. Therefore, this system could also be classified as a model reference adaptive control (MRAC) system, although a principal difference from the original MRAC is that full state knowledge of the plant to be controlled is not required. The advantage of

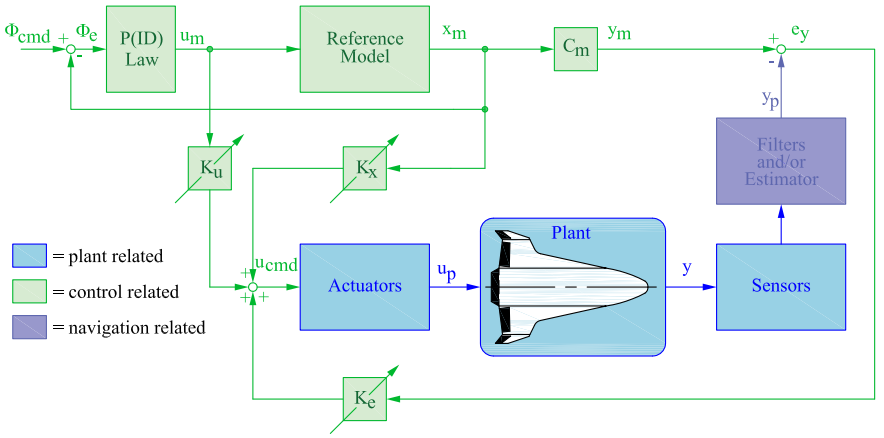


Fig. 7.57 Basic architecture of a Simple Adaptive Control algorithm

this control concept is that when one designs a reference model according to the control-performance requirements, and the non-linear plant accurately tracks the model, then also the plant will meet the requirements. The model represents a simplified (but known!) version of the plant; of course, it cannot be overly simplified, as the transient-response characteristics of the model should be with the performance envelope of the plant. SAC has been studied in attitude-control applications for, amongst others, large, flexible space structures (Barkana and Kaufman 1993), robots (Barkana and Guez 1990), a space telescope (Mehiel and Balas 2004), a single-stage-to-orbit space plane (Mooij 2001), and re-entry vehicles (Mooij 2014, 2018).

A schematic overview of a simple adaptive controller is shown in Fig. 7.57. The control law is given by

$$\mathbf{u_p}(t) = \mathbf{K_r}(t)\mathbf{r}(t) \tag{7.212}$$

where $\mathbf{r}(t) = \left[\mathbf{e_y}(t)\ \mathbf{x_m}(t)\ \mathbf{u_m}\right]^{\mathrm{T}}$ and $\mathbf{K_r}(t) = [\mathbf{K_e}(t)\ \mathbf{K_x}(t)\ \mathbf{K_u}(t)]$. It can be seen that the model input $\mathbf{u_m}$ and model state $\mathbf{x_m}$ are required to form part of the input signal $\mathbf{u_p}$ to the plant. Moreover, the so-called output error $\mathbf{e_y}$ serves as a feedback quantity to form the third element that composes $\mathbf{u_p}$. The three gains, i.e., $\mathbf{K_x}$, $\mathbf{K_u}$ and $\mathbf{K_e}$, are adaptive.

To compute the adaptive gains, $\mathbf{K_r}$ is defined to be the sum of an integral and a proportional component:

$$\mathbf{K_r}(t) = \mathbf{K_i}(t) + \mathbf{K_p}(t) \tag{7.213}$$

with

$$\dot{\mathbf{K}}_\mathbf{i}(t) = \mathbf{e_y}\mathbf{r}^{\mathrm{T}}(t)\mathbf{T_i} \tag{7.214}$$

$$\mathbf{K_p}(t) = \mathbf{e_y}(t)\mathbf{r}^{\mathrm{T}}(t)\mathbf{T_p} \tag{7.215}$$

In Eqs. (7.214) and (7.215), the weighting matrices $\mathbf{T_p}$ and $\mathbf{T_i}$ are positive semi-definite and positive definite, respectively. Note that the proportional-gain component has a direct influence on the transient tracking behaviour, but is strictly speaking not required to enforce asymptotic tracking ($\mathbf{T_p}$ can be zero). This is guaranteed by the integral gain. To improve the transient response by only using an integral gain, a constant gain value can be added to $\mathbf{K_i}$. An advantage over the use of the proportional gain is that this constant value is independent of $\mathbf{e_y}$, and is non-zero even if $\mathbf{e_y}$ is zero. In that case, the integral gain derived from Eq. (7.214) becomes

$$\mathbf{K_i}(t) = \mathbf{K_{i,0}} + \int_0^t \dot{\mathbf{K}}_\mathbf{i}(\tau)\mathrm{d}\tau \tag{7.216}$$

One way to improve the damping of the system is to include the error derivatives in the output error vector. In that case, the error for output $\mathbf{y}$ becomes:

$$\mathbf{e_y}(t) = \mathbf{K_y^T}(\mathbf{y_m}(t) - \mathbf{y_p}(t)) + \mathbf{K_{\dot{y}}^T}(\dot{\mathbf{y}}_\mathbf{m}(t) - \dot{\mathbf{y}}_\mathbf{p}(t)) \tag{7.217}$$

with $\mathbf{K_y^T}$ and $\mathbf{K_{\dot{y}}^T}$ being a proportional and derivative output gain, respectively. However, to avoid calculating the numerical derivative of the outputs and to tune the related gains in multiple-output systems, an alternative expression for the output error can be used. Adjusting the output matrix by pre-multiplying it with $\mathbf{K_c}$, the optimal gain matrix from solving the Algebraic Riccati Equation for a closed-loop linearised version of the plant, sufficient damping is commonly introduced in the system to have a proper response (combining angular rates with attitude angles can be considered as a form of proportional-derivative output, see Mooij 2014). The output error becomes, in that case,

$$\mathbf{e_y} = \mathbf{y_m} - \mathbf{y_p} = \mathbf{K_c}\left(\mathbf{C_m}\mathbf{x_m}(t) - \mathbf{C_p}(\mathbf{x_p}, t)\mathbf{x_p}(t)\right) \tag{7.218}$$

So far, an ideal environment has been considered. To cope with environmental disturbances that lead to a persistent non-zero error and to a continuous change in the integral gain $\mathbf{K_i}$, a robust design can be applied to adjust the integral gain and prevent it from reaching very high values. The integral term of Eq. (7.214) is adjusted as follows:

$$\dot{\mathbf{K}}_\mathbf{i} = \mathbf{e_y}(t)\mathbf{r}^\mathrm{T}\mathbf{T_i} - \sigma_i\mathbf{K_i}(t) \tag{7.219}$$

Without the σ_i-term, $\mathbf{K_i}(t)$ is a perfect integrator and may steadily increase (and even diverge) whenever perfect output-following is not possible. Including the σ_i-term, $\mathbf{K_i}(t)$ is obtained from a first-order filtering of $\mathbf{e_y}(t)\mathbf{r}^\mathrm{T}\mathbf{T_i}$ and cannot diverge, unless the output error diverges.

7.5.7 Case Study: HORUS Attitude Controller Design

To illustrate the use of optimal control theory we will design an attitude controller for the winged re-entry vehicle HORUS (see Appendix C, Sect. B.2 for more details about this reference vehicle). In Chap. 6, the eigenmotion of HORUS has been analyzed. It became clear that even though there is a coupling between the symmetric and asymmetric eigenmotion, this coupling is not strong. Comparing the eigenvalues of the complete system with the simplified and decoupled systems for the symmetric and asymmetric motion, shows that there is not a large difference. Therefore, we can assume that the decoupled systems are a good representation of the complete system.

To simplify the design of the controller, we will use this fact of decoupling by separating the controller into two parts: a longitudinal and a lateral controller. The longitudinal controller consists of an inner and an outer loop. The inner loop takes care of stability augmentation while the outer loop provides the corrective control, i.e., to make the actual attitude approach the commanded one in a finite time. The stability augmentation, or longitudinal trim, is executed by aerodynamic means only, i.e., in the first instance the body flap is used, and in case the maximum deflection has been reached, the elevators (symmetric deflection of the elevons), are used for additional control. For corrective longitudinal control, the elevators and pitch jets will be used. Lateral control consists of only the outer, corrective loop, because only small moments are required to maintain stability. This control is achieved by using the ailerons (asymmetric deflection of the elevons), the rudders, and the roll and yaw jets. The state of the vehicle is fed back into the two controllers. To react to the strongly varying dynamics, varying gains are used in both corrective control loops. In Fig. 7.58, the layout of the controller has been schematically depicted. In this respect, we mention that there are, in principle, three controllers discussed in this case study: a longitudinal and a lateral controller, that are designed independently from each other, and an integrated controller, combining the longitudinal and lateral controllers.

Activation of the different control effectuators is based on the control scheme of the Space Shuttle (Cooke 1982). In the case of HORUS, the rudders are only outward movable, which means that only one rudder can be operated at a time. Since one rudder by itself is not sufficient to control the yaw motion throughout the trajectory, yaw control is supported by the yaw thrusters. In Fig. 7.59 the adapted scheme with control modes for HORUS is shown.

In the following two sections the design of both the longitudinal and lateral controller will be discussed.

Longitudinal Control

Restricting to the attitude subsystem (i.e., $\Delta V = \Delta\gamma = \Delta R = 0$), and neglecting the asymmetric motion by setting $\Delta\beta = \Delta\sigma = 0°$, the following approximation can be derived from the linearised equations of motion:

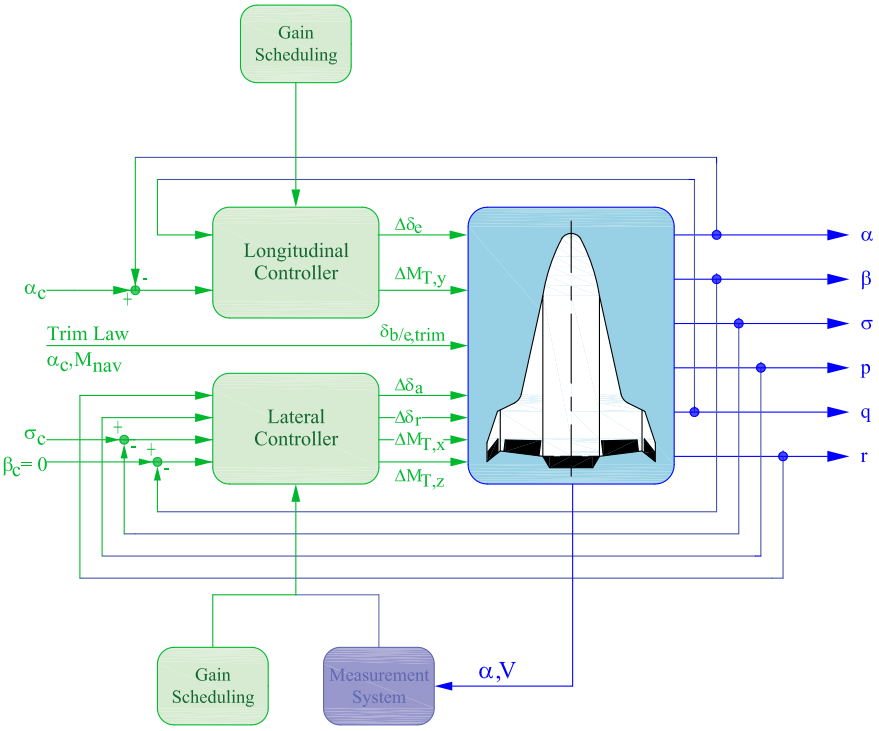


Fig. 7.58 Schematic layout of the HORUS-2B attitude controller

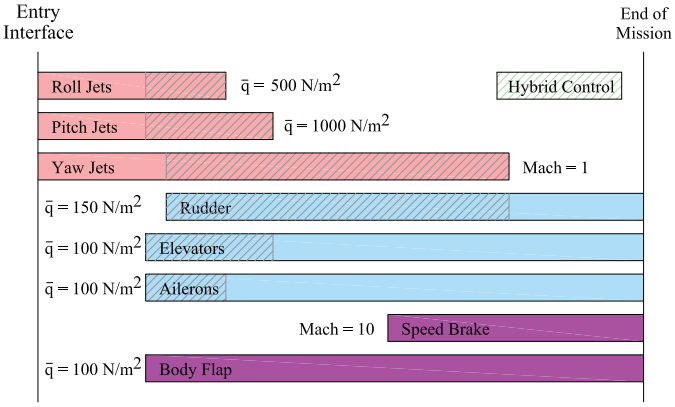


Fig. 7.59 Entry control modes for HORUS

$$\Delta\dot{q} = \frac{1}{I_{yy}}\left(\frac{\partial C_m}{\partial\alpha}\Delta\alpha + \frac{\partial C_m}{\partial\delta_e}\Delta\delta_e\right)\bar{q}_e S_{ref} c_{ref} + \frac{\Delta M_{T,y}}{I_{yy}} \tag{7.220}$$

$$\Delta\dot{\alpha} = \Delta q - \frac{1}{mV_e}\frac{\partial C_L}{\partial\alpha}\bar{q}_e S_{ref}\Delta\alpha \tag{7.221}$$

The static stability mainly depends on $\frac{\partial C_m}{\partial\alpha}$, which changes sign at the end of the re-entry phase. This results in an unstable short-period oscillation (see once more Chap. 6). As control variables, the symmetric elevon deflection angle δ_e (elevator function) and the thruster moment about the pitch axis, $M_{T,y}$, are available. The elevators are activated at $\bar{q} > 100$ N/m^2. The pitch jets will start working at the entry interface (in principle $\bar{q} = 0$ N/m^2) and will continue to do so until $\bar{q} > 1000$ N/m^2. The elevator deflection angle is defined as the symmetric combination of the left and right elevon, i.e.,

$$\delta_e = \frac{\delta_{e,l} + \delta_{e,r}}{2} \tag{7.222}$$

The state-feedback laws for longitudinal control are selected to be simple proportional laws, as has been schematically depicted in Fig. 7.60:

$$\frac{\Delta\delta_e}{\delta_{e_{max}}} = -K_{eq}\Delta q - K_{e\alpha}\Delta\alpha \tag{7.223}$$

$$\frac{\Delta M_{T,y}}{M_{T,y_{max}}} = -K_{yq}\Delta q - K_{y\alpha}\Delta\alpha \tag{7.224}$$

where

$\Delta q =$ $q - q_e = q - q_c$ (rad/s)
$\Delta\alpha =$ $\alpha - \alpha_e \approx \alpha - \alpha_c$ (rad)
q_c = commanded pitch rate (rad/s)
α_c = commanded angle of attack from the guidance system (rad).

These control laws are a linear combination of the state variables, which is the right form for the LQR:

$$\Delta\mathbf{u} = -\mathbf{K}\Delta\mathbf{x} \tag{7.225}$$

Note that due to the relation between q and α, this controller can be considered to be a PD controller. Writing both the equations of motion Eqs. (7.220) and (7.221) and the feedback laws Eqs. (7.223) and (7.224) in matrix form and evaluating it for discrete points along a reference trajectory, the gains can be computed by solving the Riccati Equation (Gopal 1989), as discussed in Sect. 7.5.4.2. The selected weighting matrices $\mathbf{Q}$ and $\mathbf{R}$ are chosen to be

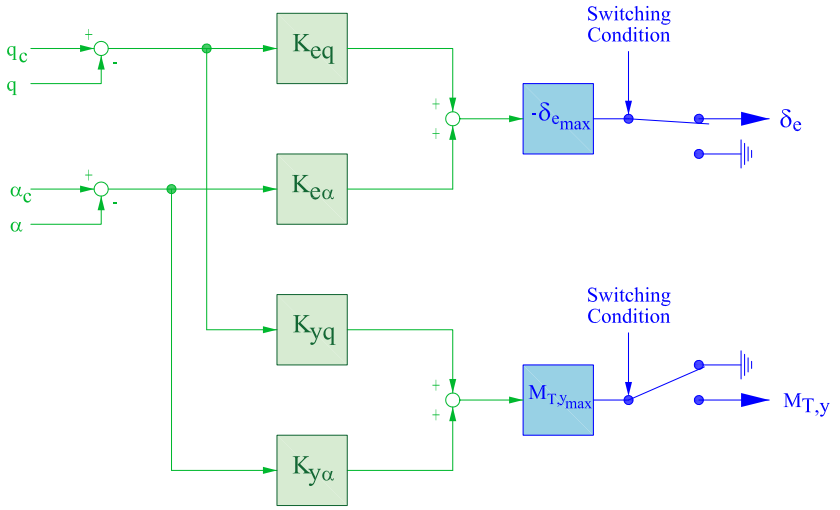


Fig. 7.60 The longitudinal controller

$$\mathbf{Q} = \text{diag}\left\{\frac{1}{\Delta q_{max}^2}, \frac{1}{\Delta \alpha_{max}^2}\right\} \qquad \mathbf{R} = \text{diag}\left\{\frac{1}{\Delta \delta_{e_{max}}^2}, \frac{1}{\Delta M_{T,y_{max}}^2}\right\} \tag{7.226}$$

where Δq_{max} and $\Delta \alpha_{max}$ can be chosen depending on the flight phase (they need not be constant for the complete flight). Typically, the control weights $\Delta \delta_{e_{max}}$ and $\Delta M_{T,y_{max}}$ are chosen in accordance with the actuator limits, which are for HORUS $\Delta \delta_{e_{max}} = 40°$ and $\Delta M_{T,y_{max}} = 10{,}400$ Nm. The maximum allowed state deviation is chosen to be constant for each control mode, i.e., for the reaction control mode they are $\Delta q_{max} = 1.5°/\text{s}$ and $\Delta \alpha_{max} = 1°$, for the hybrid-control mode $\Delta q_{max} = 1.5°/\text{s}$ and $\Delta \alpha_{max} = 2.5°$, and finally for the aerodynamic-control mode $\Delta q_{max} = 4°/\text{s}$ and $\Delta \alpha_{max} = 2°$. The value of $\Delta \alpha_{max}$ has been selected based on a combination of good performance and the small margin that is left, since the nominal α is close to its maximum. Note that the selection of the weighting matrices is in principle an iterative procedure that is not pursued here.

Software packages, such as MATLAB®, have standard routines to solve the Riccati Equation to obtain the gains. Doing so for the longitudinal controller gives the gains of Fig. 7.61. Indicated are the moments when the actuators switch on and off, which yields discrete jumps in the gains. The large variation in gain values indicates that constant-gain control may not yield a good performance in all flight phases. Therefore, it is common practise to interpolate the gains as a function of some parameter, a process called gain scheduling. Because the physical state of the vehicle may remain constant over time, it is advised not to use time as independent variable, but, for instance, dynamic pressure, as this parameter is directly linked with the control-surface effectiveness. To prevent spikes in the control commands, e.g., when actuators are switched on or off, the gain interpolation should allow for smooth transitions.

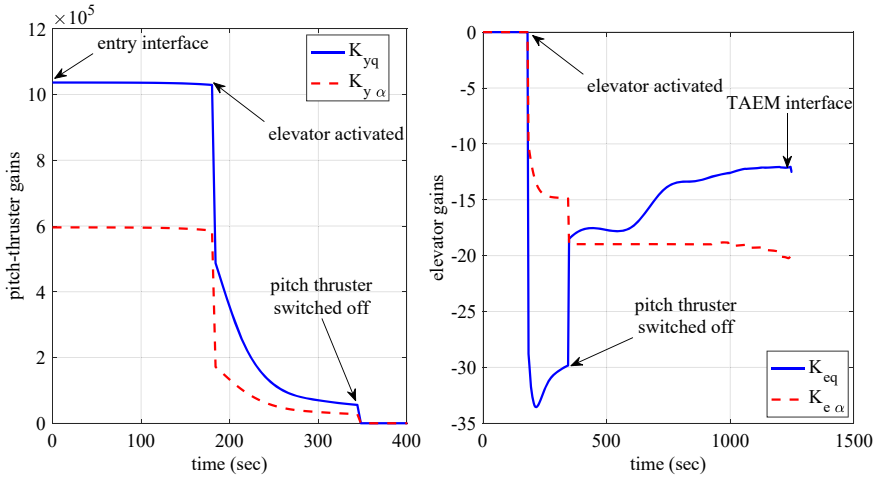


Fig. 7.61 Optimal gains for the HORUS longitudinal attitude controller

Lateral Control

The lateral dynamics consist of a coupled motion of roll and yaw. This motion can be approximated by setting $\Delta\alpha$ to zero and neglecting smaller terms. The following equations can be derived:

$$\Delta\dot{p} = \frac{1}{I_{xx}}\left(\frac{\partial C_l}{\partial\beta}\Delta\beta + \frac{\partial C_l}{\partial\delta_a}\Delta\delta_a\right)\bar{q}_e S_{ref} b_{ref} + \frac{\Delta M_{T,x}}{I_{xx}} \tag{7.227}$$

$$\Delta\dot{r} = \frac{1}{I_{zz}}\left(\frac{\partial C_n}{\partial\beta}\Delta\beta + \frac{\partial C_n}{\partial\delta_a}\Delta\delta_a + \frac{\partial C_n}{\partial\delta_r}\Delta\delta_r\right)\bar{q}_e S_{ref} b_{ref} + \frac{\Delta M_{T,z}}{I_{zz}} \tag{7.228}$$

$$\Delta\dot{\beta} = \sin\alpha_e\Delta p - \cos\alpha_e\Delta r - \frac{g_e}{V_e}\cos\gamma_e\cos\sigma_e\Delta\sigma \tag{7.229}$$

$$\Delta\dot{\sigma} = -\cos\alpha_e\Delta p - \sin\alpha_e\Delta r + \left(\frac{g_e}{V_e}\cos\gamma_e\cos\sigma_e - \frac{L_e}{mV_e}\right)\Delta\beta + \frac{\tan\gamma_e}{mV_e}\cos\sigma_e L_e\Delta\sigma \tag{7.230}$$

Because the damping coefficients such as C_{l_p} and C_{n_r}, are not defined in the current aerodynamic model, the corresponding terms do not appear in the equations and therefore there is no natural damping. The available controls consist of the aileron, rudder, roll jets and yaw jets.

The aileron deflection angle δ_a is defined as the asymmetric combination of the left and right elevon, i.e.,

$$\delta_a = \frac{\delta_{e,l} - \delta_{e,r}}{2} \tag{7.231}$$

The rudder deflection angle δ_r is defined to be equal to $\delta_{r,}$, when $\delta_{r,l}$ is positive, and equal to $-\delta_{r,r}$, when $\delta_{r,r}$ is positive.

In its simplest form, the control laws are defined similarly to Eqs. (7.223) and (7.224), taking the state errors as input and multiplying them with the corresponding gains:

$$\Delta\mathbf{u} = -\mathbf{K}\Delta\mathbf{x} \tag{7.232}$$

with $\Delta\mathbf{x} = (\Delta p, \Delta r, \Delta\beta, \Delta\sigma)^{\mathrm{T}}$, where

$$\begin{array}{lcl}
\Delta p & = & p - p_e = p - p_c \text{ (rad/s)} \\
\Delta r & = & r - r_e = r - r_c \text{ (rad/s)} \\
\Delta\beta & = & \beta - \beta_e \approx \beta - \beta_c = \beta \text{ (rad)} \\
\Delta\sigma & = & \sigma - \sigma_e \approx \sigma - \sigma_c \text{ (rad)} \\
p_c & = & \text{commanded roll rate (rad/s)} \\
r_c & = & \text{commanded yaw rate (rad/s)} \\
\beta_c & = & \text{commanded angle of sideslip from the guidance system, nominally zero} \\
\sigma_c & = & \text{commanded bank angle from the guidance system (rad).}
\end{array}$$

Due to the large angle of attack during a major part of the re-entry flight, the roll-yaw coupling is much stronger than the situation where the angle of attack is zero. That means that also the actuator actions suffer from this coupling, and to anticipate on that at he control laws for aileron and rudder are chosen in the form

$$\frac{\Delta\delta_a}{\delta_{a_{max}}} = -\left[K_{ap}\dot{\sigma} + K_{a\sigma}(\sigma - \sigma_c)\right]\cos\alpha - \left[K_{ar}(\dot{\beta} + \frac{g}{V}\sin\sigma) + K_{a\beta}\beta\right]\sin\alpha \tag{7.233}$$

$$\frac{\Delta\delta_r}{\delta_{r_{max}}} = -\left[K_{rp}\dot{\sigma} + K_{r\sigma}(\sigma - \sigma_c)\right]\sin\alpha - \left[K_{rr}(\dot{\beta} + \frac{g}{V}\sin\sigma) + K_{r\beta}\beta\right]\cos\alpha \tag{7.234}$$

which shows that errors in sideslip and bank angle will be jointly taken care of by both actuators. Of course, for the roll and yaw thruster moment similar laws can be chosen. The lateral controller, which is represented by these equations, is schematically shown in Fig. 7.62. Again, it is noted that the aerodynamic- and reaction-control part are identical, but for the gains.

The (non-linear) control laws are not yet in the right format such that they can be solved for the gains using the approach for the LQR. However, by linearising them, they can be written as

$$\begin{aligned}
\Delta\mathbf{u} = \begin{pmatrix} \Delta\delta_a \\ \Delta\delta_r \\ \Delta M_{T,x} \\ \Delta M_{T,z} \end{pmatrix} &= - \begin{bmatrix} \delta_{a_{max}}K^*_{ap} & \delta_{a_{max}}K^*_{a\sigma} & \delta_{a_{max}}K^*_{ar} & \delta_{a_{max}}K^*_{a\beta} \\ \delta_{r_{max}}K^*_{rp} & \delta_{r_{max}}K^*_{r\sigma} & \delta_{r_{max}}K^*_{rr} & \delta_{r_{max}}K^*_{r\beta} \\ T_{x_{max}}\tilde{k}^*_3 & T_{x_{max}}\tilde{k}^*_4 & T_{x_{max}}\tilde{k}^*_5 & T_{x_{max}}\tilde{k}^*_6 \\ T_{z_{max}}\tilde{k}^*_7 & T_{z_{max}}\tilde{k}^*_8 & T_{z_{max}}\tilde{k}^*_9 & T_{z_{max}}\tilde{k}^*_{10} \end{bmatrix} \begin{pmatrix} \Delta p \\ \Delta r \\ \Delta\beta \\ \Delta\sigma \end{pmatrix} \\
&= -\mathbf{K}^*\Delta\mathbf{x}
\end{aligned} \tag{7.235}$$

with

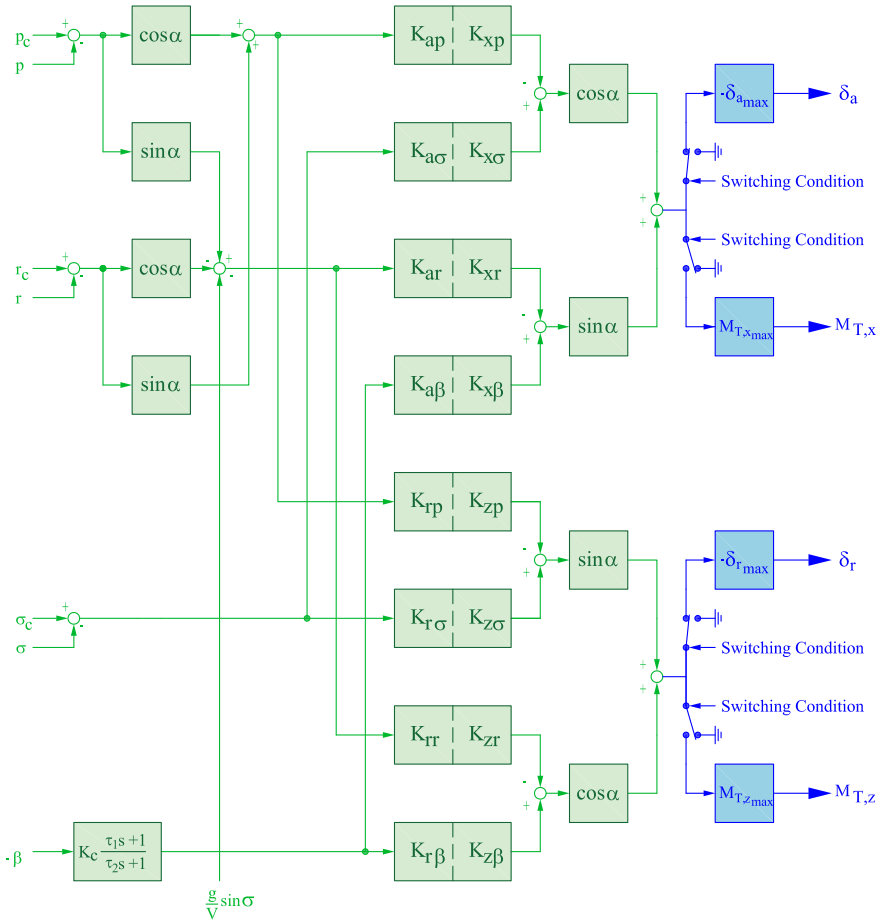


Fig. 7.62 The lateral controller

$$\begin{pmatrix} K^*_{ap} \\ K^*_{a\sigma} \\ K^*_{ar} \\ K^*_{a\beta} \end{pmatrix} = \begin{bmatrix} -\cos^2\alpha_e & 0 & \sin^2\alpha_e & 0 \\ -\sin\alpha_e\cos\alpha_e & 0 & -\sin\alpha_e\cos\alpha_e & 0 \\ 0 & 0 & 0 & \sin\alpha_e \\ 0 & \cos\alpha_e & \frac{g_e}{V_e}\cos\sigma_e\sin\alpha_e & 0 \end{bmatrix} \begin{pmatrix} K_{ap} \\ K_{a\sigma} \\ K_{ar} \\ K_{a\beta} \end{pmatrix}$$

$$\begin{pmatrix} K^*_{rp} \\ K^*_{r\sigma} \\ K^*_{rr} \\ K^*_{r\beta} \end{pmatrix} = \begin{bmatrix} -\sin\alpha_e\cos\alpha_e & 0 & \sin\alpha_e\cos\alpha_e & 0 \\ -\sin^2\alpha_e & 0 & -\cos^2\alpha_e & 0 \\ 0 & 0 & 0 & \cos\alpha_e \\ 0 & \sin\alpha_e & \frac{g_e}{V_e}\cos\sigma_e\cos\alpha_e & 0 \end{bmatrix} \begin{pmatrix} K_{rp} \\ K_{r\sigma} \\ K_{rr} \\ K_{r\beta} \end{pmatrix}$$

Table 7.4 Definition of weighting terms for the HORUS lateral controller

	Roll control		Yaw control	
	Δp_{max} (°/s)	$\Delta\sigma_{max}$ (°/s)	Δr_{max} (°)	$\Delta\beta_{max}$ (°)
RCS	1.5	4	0.5	2
Hybrid	4	4	5	2
Aero	10	10	5	2

The formulation for the roll- and yaw-thruster gains is exactly the same as Eq. (7.235), but with the subscript a replaced by x and (the first) r replaced by z. To solve the Riccati Equation to obtain the gains for the equilibrium points, the following weighting matrices are applied:

$$\mathbf{Q} = \text{diag}\left\{\frac{1}{\Delta p_{max}^2}, \frac{1}{\Delta r_{max}^2}, \frac{1}{\Delta\beta_{max}^2}, \frac{1}{\Delta\sigma_{max}^2}\right\}$$

$$\mathbf{R} = \text{diag}\left\{\frac{1}{\Delta\delta_{a_{max}}^2}, \frac{1}{\Delta\delta_{r_{max}}^2}, \frac{1}{\Delta M_{T,x_{max}}^2}, \frac{1}{\Delta M_{T,z_{max}}^2}\right\}$$

where, as before for the longitudinal controller, the weighting terms can be selected for each control mode separately. The selected values have been listed in Table 7.4. The control-effort weights are again chosen in accordance with the actuator limits, i.e., $\Delta\delta_{a_{max}} = 40°$, $\Delta\delta_{r_{max}} = 40°$, $\Delta M_{T,x_{max}} = 1{,}600$ Nm, and $\Delta M_{T,z_{max}} = 7{,}600$ Nm.

After solving the Riccati Equation and transforming the gains following the linearised control law, yields the gains of Figs. 7.63 and 7.64. Visible are the discrete jumps in gain values whenever an actuator is switched on or off, but is also clear that the changing flight conditions affect not only the magnitude, but also the sign of the gains. Careful interpolation is required, such that no control reversal can occur.

For the sake of comparison, in case only the state errors are fed back, and not pre-processed by some control law, i.e., $\Delta\mathbf{u} = -\mathbf{K}\Delta\mathbf{x}$, the resulting gains are shown in Figs. 7.65 and 7.66. The difference between the two sets is apparent, and the effect of some gains seem to have interchanged. However, without a proper performance analysis it is impossible to judge the effect of these differences. It is left to the reader, though, to assess the effect on the controller performance when either set is used.

A preliminary sensitivity analysis with the controller integrated with the nonlinear system indicated that, although the guidance system had no problems at all for properly guiding the vehicle, the deviations were such that at $t \approx 930$ s, the vehicle could not be controlled any longer and exhibited diverging oscillations around all three axes. These oscillations could be traced back by a combined angle-of-attack and angle-of-sideslip corrective control. The former requires a symmetric deflection of the elevons, whereas the latter needs an asymmetric deflection. As a result, the absolute deflection of the left and right elevon, which is a simple summation of

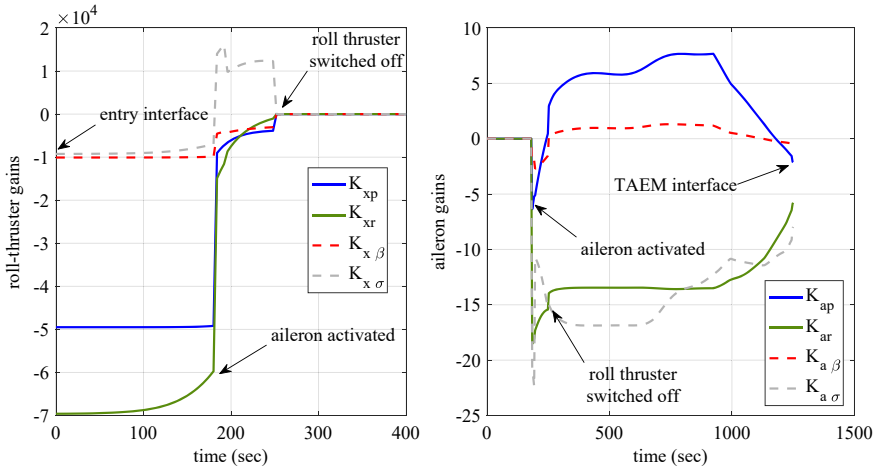


Fig. 7.63 Control-law optimal gains for the HORUS lateral attitude controller (roll thruster and aileron)

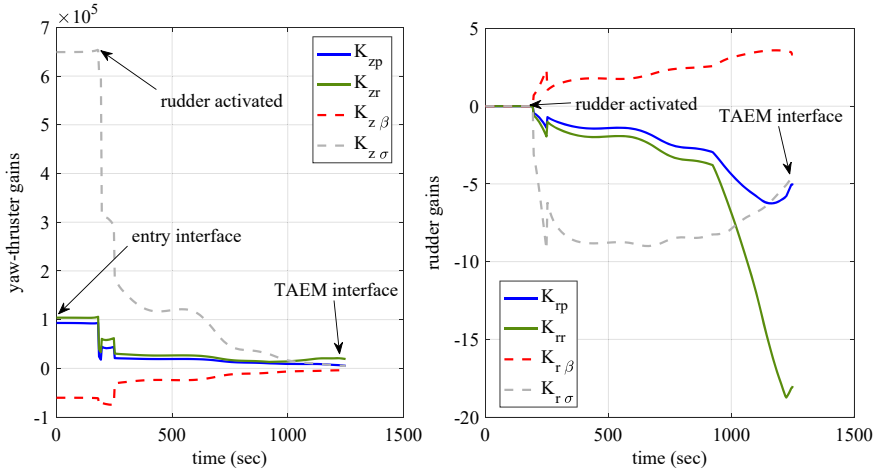


Fig. 7.64 Control-law optimal gains for the HORUS lateral attitude controller (yaw thruster and rudder)

elevator and aileron deflection, differed substantially. Since the aerodynamic moments due to elevon deflection are quite non-linear for larger deflection angles, the attitude controller assumed roll and pitch moments that deviated substantially from the actual moments. The induced differences resulted in diverging, oscillating elevon deflections and an uncontrollable vehicle. This problem clearly indicates the effect of non-linearities on the (decoupled) linearised control-system design.

Without going into great detail of trying to solve this problem, the lateral controller was adjusted such that the above mentioned problem did not occur any more. The

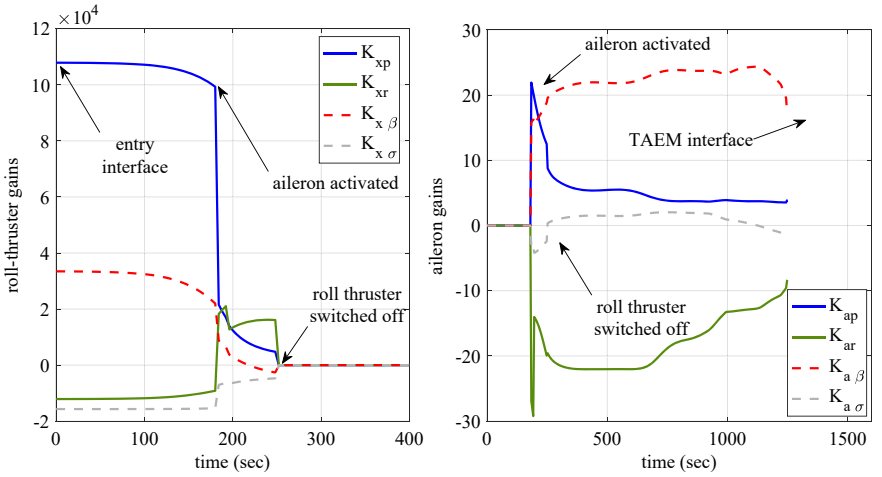


Fig. 7.65 Basic optimal gains for the HORUS lateral attitude controller (roll thruster and aileron)

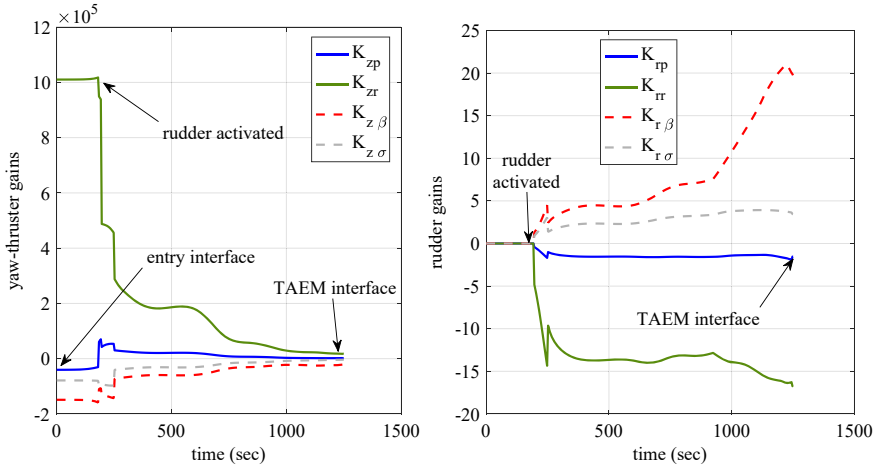


Fig. 7.66 Basic optimal gains for the HORUS lateral attitude controller (yaw thruster and rudder)

angle of sideslip is an induced disturbance that can relatively fast be controlled, since, next to the ailerons, also the rudder and yaw jets are used. Therefore, priority is given to β control by including a phase-lead filter in the loop for β, such that derivative information is taken into account. As a result, a strong variation in β (large derivatives) results in larger deflections of rudder and aileron (side-slip compensation) and therefore faster side-slip control. The transfer function $G_c(s)$ of the phase-lead filter is:

$$G_c(s) = K_c \frac{\tau_1 s + 1}{\tau_2 s + 1} \text{ with } K_c = 10,\ \tau_{c,1} = 1 \text{ and } \tau_{c,2} = 10$$

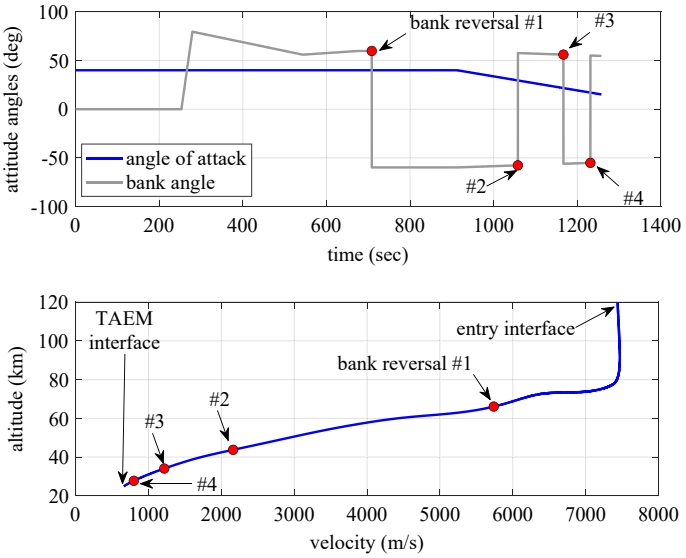


Fig. 7.67 Nominal control history (top) and altitude-velocity profile (bottom)

It is noted that the lead filter was not considered in the computation of the feedback gains with optimal control theory. The filter has only later been added to improve the sideslip response to rudder deflections. For more details on phase-lead (and also phase-lag) filters, one is referred to Ogata (2010).

Results

To test the performance of the controller first a number of step functions in α and σ has been done, as well as a response test due to an error in β and a ramp function for σ to study a bank reversal. All these tests were done for the linearised system and the results were as expected. A more interesting test is the one to use the attitude controller for the non-linear system and to simulate the complete reference trajectory. The nominal mission has been introduced at several places before (e.g., Sects. 6.4.2, 7.3.4, and B.2), but for the sake of convenience the control history and altitude-velocity profile are repeated in Fig. 7.67.

Since we are currently only interested in the performance of the attitude controller, we will restrict the discussion to the attitude-related variables. Figure 7.68 shows two of the three attitude angles, which are controlled by the attitude control system, i.e., α and σ. In this figure we see more of a difference between the controlled value and the nominal one. During the first 200 s, α is slightly diverging, caused by the fact that the trim law is not active. This effect can be better observed in Fig. 7.69 (where the difference between α_c and α is plotted). As discussed earlier, trim is primarily guaranteed by deflecting the body flap, which is only activated when $\bar{q} > 100$ N/m^2.

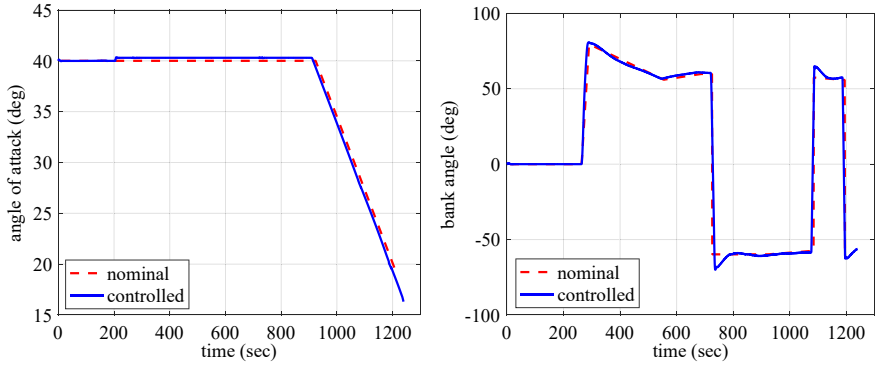


Fig. 7.68 Angle of attack and bank angle for the nominal trajectory (dashed line), flown in closed loop (solid line)

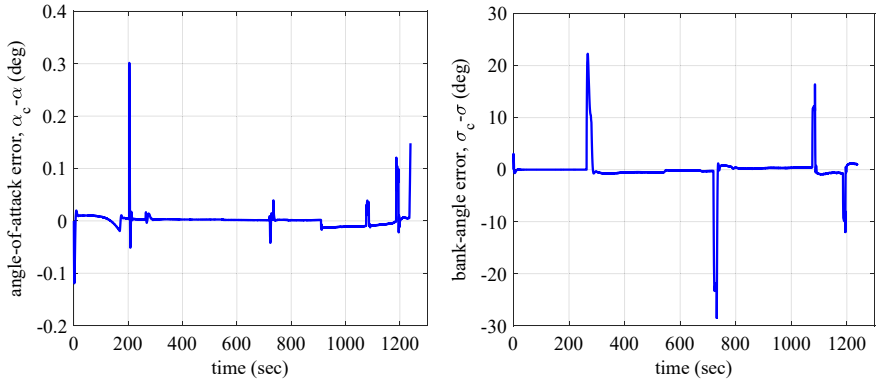


Fig. 7.69 Control errors for the nominal trajectory

The moment the body flap is activated ($t \approx 194$ s), α is stabilised at the nominal value. The diverging α is only partly compensated for by the pitch jets, but apparently due to the design assumptions and simplifications, this offset is not properly controlled. On the other hand, it should be noted that the gains are computed allowing a 2° overshoot of α, so the offset is very well within range. A revision of the trim law (trim with pitch jets, when the body flap is not active) or a gain computation for the first 200 s with a smaller allowable overshoot is not required, considering the small divergence.

Returning to the time history of α, from $t \approx 200$ s the controlled α has a persistent difference compared to the nominal one, which can be related to either the guidance system or the attitude controller. Figure 7.69, the left plot, can be consulted in finding the answer, since this figure indicates how well the attitude controller performs. As we see, the differences are very small at all times, from which it can be concluded that the "problem" is not related to the attitude controller, but a result of guiding the vehicle towards the target along the reference trajectory.

In the same figure, four peaks in the right half of the graph need more explanation. During bank reversals, the guidance system keeps the commanded variation in α at a constant value, which means that during the reversal no correction other than a change in the reference value takes place. It was discussed above, that σ is smaller than it should be during a reversal, resulting in a larger vertical lift. Any resulting error can only be compensated for after the reversal has been completed. This shows as an abrupt change in the commanded angle of attack at $t \approx 730$ s, $t \approx 1{,}080$ s, $t \approx 1{,}200$ s and $t \approx 1{,}250$ s.

In Fig. 7.69 (right plot), the difference between σ_c and σ has been plotted, again indicating how well the attitude controller performs. The first difference between actual and nominal σ is found just before $t = 100$ s, which is due to a relatively sudden activation of the guidance system. Since during the first 100 s, the actual α is not identical to the nominal value, a small altitude difference will be built up. Altitude control is activated after some 80 s to avoid excessive commands to the attitude controller in this low-dynamic pressure regime. Once altitude control is active, it is faced with a discrete altitude error resulting in a sudden change in σ_c. However, only a small peak value of 2° is observed, well within the gain overshoot-value of 5°. The limitations of the attitude control system in case of sudden changes that are relatively large is shown at $t \approx 270$ s, where σ_c is linearly increased at a rather large rate. An adjustment of the guidance scheme seems to be required; on the other hand, neither the ailerons nor the roll jets have been used at their maximum capacity, so there is still a margin in the attitude controller.

The three remaining peaks are all related to the bank reversals. It suffices to say that the change in σ_c, as generated by the guidance system based on a predicted bank rate, is too large for the attitude control system to achieve. Moreover, the reversals induce a significant altitude error, and each time a rather large corrective σ is required to compensate for this (Fig. 7.68, right plot).

The last attitude angle to be discussed is β (Fig. 7.70). While designing the control system, it became clear that there is a (strong) coupling between β and σ, the already mentioned roll-yaw coupling. Therefore, an induced β can be expected in case there is a sudden change in σ. Indeed, a number of peaks is observed during the bank reversals, although they damp out relatively quickly. At $t \approx 1{,}000$ s, however, β diverges slowly, without coming back to zero again. It is assumed that this diversion is due to the simplifications made during the control-system design, e.g., neglecting terms due the rotation of both the Earth and the local horizontal plane. This indicates that the translational motion has at least some influence on the rotational motion, which the LQR in its current design cannot compensate for completely.

Figures 7.71, 7.72, and 7.73 show the deflections of the control surfaces and the moments generated by the reaction-control thrusters (the roll and pitch jets have been included as well, despite the marginal activity). Most activity takes place during the bank reversals, or when a sudden change in commanded attitude arises. The ailerons (Fig. 7.71, left) show peak values when bank reversals are initiated; smaller deflections are required when $\bar{q}$ increases towards the end of the flight. The peaks in the elevator deflection (Fig. 7.72, left) are also due to the bank reversals, although only indirectly. During the bank reversal the commanded variation of α is kept

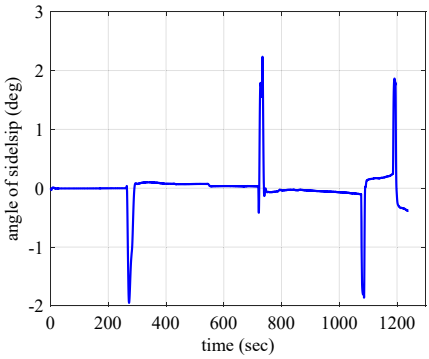


Fig. 7.70 Angle of sideslip for the nominal trajectory ($\beta_c = 0°$), flown in closed loop

constant, which means that the guidance system does not execute its energy control. To compensate for induced energy errors a sudden change in α_c (and hence elevator deflection) right after completing the reversal is the result.

Note that the elevator and aileron deflection should be combined to get the left and right elevon deflection, and since the related deflections are due to corrective control only, also the trimmed deflection of the elevator should be included. The trimmed deflection of the body flap has been depicted in Fig. 7.74. After a constant value of $\delta_{b,trim} = 15°$, active body-flap trim is started at $\bar{q} = 100$ N/m^2. Mach effects result in a slowly decreasing deflection. When α starts decreasing (at $M \approx 12$), $\delta_{b,trim}$ increases because of a larger nose-up moment (the maximum pitch moment in that Mach range lies around $\alpha = 25°$). Below $M = 3$, the pitch moment changes to nose down, resulting in negative trim deflections. Towards the end of the flight, there is a strong nose-down moment that results in a maximum upward deflection of the body flap ($\delta_{b,trim} = -20°$). As a result, the trimmed elevator deflection is about $-10°$ in the very last phase of the mission, which gives an elevon deflection of about $-20°$. It remains to be studied in more detail, whether the margins in manoeuvrability capability are sufficient.

The rudder (Fig. 7.73, left) is twice deflected at its maximum value to start a bank reversal and to compensate for the induced β. Unfortunately, the two rudders are only outward deflectable, which means that only half the available control moment can be used. Therefore, it is advised, since not even uncertainties in the vehicle design have been included, that both rudders should be used for yaw control or that the effective rudder area is increased. Increasing equilibrium deflection angles are found after $t \approx 900$ s, to compensate for the slowly diverging β. Large control signals are also found for the yaw jets (Fig. 7.73, right), similarly to those of the rudder, when a strong variation of σ is commanded. The total fuel mass that is used by the RCS is 10 kg.

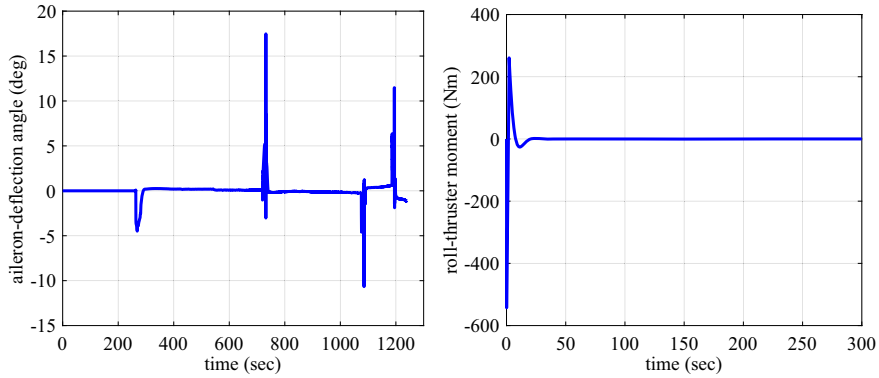


Fig. 7.71 Roll control: aileron deflection and roll-thruster moment

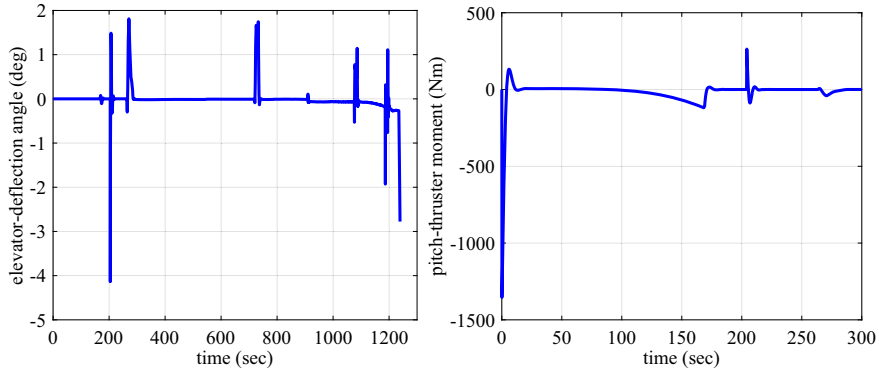


Fig. 7.72 Pitch control: elevator deflection and pitch-thruster moment

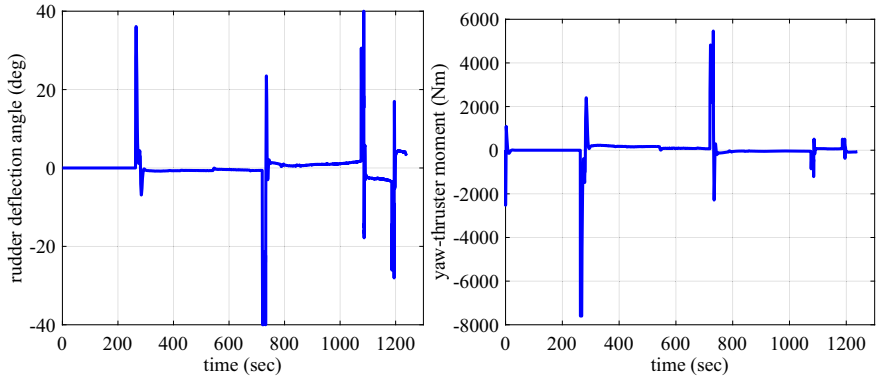


Fig. 7.73 Yaw control: rudder deflection and yaw-thruster moment

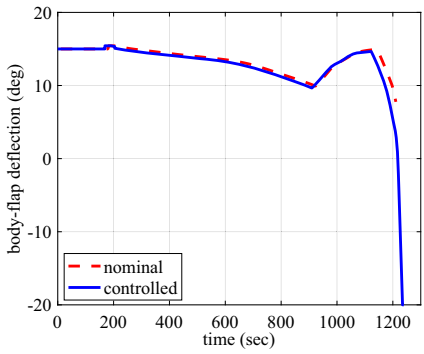


Fig. 7.74 Body-flap deflection used for trim, for the nominal trajectory (solid line), and controlled trajectory (dashed line)

7.6 Guidance

7.6.1 Introduction

Re-entry guidance methods have been extensively described in literature. In the fifties and sixties, many studies were conducted within the framework of the American space programme, and, more specific, the Apollo project. Wingrove (1963) gives a survey of atmospheric re-entry guidance concepts that were most commonly used in those days, some of which still form the basis of nowadays concepts. Because of an extensive reference list with almost 100 entries, this paper is a very good starting point for understanding re-entry dynamics and guidance for moderate lift-to-drag vehicles returning from space with either sub- or super-circular velocities. In principle, the guidance was based on lift and drag control, which also forms the basis of the guidance of the high lift-to-drag ratio Space-Shuttle, but we will discuss this later.

Wingrove presents the guidance methods under two general classifications: guidance using predicting capabilities and guidance using a nominal trajectory. The former group contains those methods that predict possible future trajectories and do not use a stored nominal trajectory. These predictions are computed with a fast, inner loop, as is also described by Braun and Powell (1992), and Powell and Braun (1992). They present a predictor-corrector guidance algorithm for a Mars aerocapture that is also applied to determine the pull-up conditions for an SSTO space plane (Powell et al. 1991). It is also possible that all or part of the possible trajectories can be considered to have an approximate explicit solution, that can directly be applied. The nominal trajectory variables of guidance methods from the second classification are stored on-board and are used as a reference. Variations in the corresponding measured variables are used by the guidance, either to bring back the vehicle onto the nominal trajectory (path controller) or to establish a new trajectory to reach the destination (terminal controller). Three methods are described, i.e., path control using either fixed or optimised time-varying feedback gains, and terminal control using

influence coefficients. For more details, the reader is referred to the excellent paper by Wingrove (1963).

ACRI (1992) does an analysis of several guidance and control techniques for moderate lift-to-drag atmospheric transfer vehicles. Focussing on the guidance techniques, they describe a number of guidance concepts belonging to two essentially different families, i.e., guidance based on the tracking of a reference and guidance based on the tracking of an objective, a difference that was already established by Wingrove. The first guidance concept of the first family includes altitude-, energy-, density- and dynamic-pressure trajectory tracking. By means of state feedback, the error between the related output and its reference is constrained to satisfy a linear differential equation (exact linearisation), and a standard pole-placement technique is used to guarantee stability and an adequate time response. The second concept consists of altitude, density and drag sliding modes. This kind of guidance aims at delivering a robust scheme based on Lyapunov theory. By definition of a sliding surface and a Lyapunov stability criterion, the error will be driven to zero once the sliding-surface equation is forced to zero. The resulting control consists of an equivalent and a normalised part, for which analytic expressions were derived.

In case significant deviations from a reference trajectory arise, excessive guidance corrections and effort may be the result that could potentially risk the integrity of the mission. In that case, one should question the choice to keep tracking the original optimal trajectory, to accept a sub-optimal solution, or to re-optimise the remaining part of the trajectory. The latter choice may, in fact, be more optimal, depending on the deviation from the original trajectory. To support this, the third generation guidance algorithms uses modern computational capabilities to generate real-time optimal trajectories and integrate this with tracking into a robust guidance and control system. Examples of such new guidance algorithms are methods using acceleration guidance and closed-form analytical solutions combined with predictor-corrector techniques (Leavitt and Mease 2007; Shen and Lu 2003), convex optimisation (Sagliano and Mooij 2018) and successive convexification (Acikmese and Ploen 2007; Wang and Grant 2017; Liu et al. 2016), interpolation of optimum trajectories (Sagliano 2016; Gonzalez Puerta et al. 2018), predictor-corrector methods (Zimmerman et al. 2003; Lu 2014), and neural networks (Schierman et al. 2004), to name but a few.

Recent advances in computational power and efficient optimisation techniques has stimulated the progress in the field of on-board optimisation, and as a result this lead to a paradigm shift in the application of guidance algorithms, which can potentially fully exploit the capabilities of the re-entry vehicle (AIAA 2017; Simplicio et al. 2019). On one hand, pseudo-spectral optimisation could provide real-time optimisation capabilities due to its fast convergence properties by solving the optimal control problem as a non-linear programming problem (NLP) on-board and correcting for errors produced by model uncertainties and other variables (Rao and Clarke 2002; Bollino et al. 2006; Jorris 2007; Han et al. 2012). However, these methods require accurate initial estimates and remain computationally intense, with the additional disadvantage that no upper bound on the computational time can be given (D'Souza and Sarigul-Klijn 2014; Sagliano 2019).

Alternatively, convex optimisation has seen considerable developments, and has been applied in highly dynamic entry descent and landing operations showing promising results, specially for reusable launch vehicles, which require a real-time capable guidance system (Acikmese and Ploen 2007; Blackmore 2016; Scharf et al. 2017).

These concepts, no matter how relevant they could be to the discussion, are too too complex for fundamental understanding and cannot be discussed in just a few pages. Unfortunately, we have to restrict ourselves to providing a few of the references, and focus on a more "physical" discussion, rather than a mathematical one.

Re-entry vehicles, like HORUS and the Space Shuttle, are winged, high lift-to-drag vehicles that employ combined bank-angle and angle-of-attack control to reach a landing site on the Earth's surface. The earlier introduced guidance models, however, were developed for a moderate lift-to-drag transfer vehicle with constant angle of attack and only bank-angle control, and are therefore not directly applicable to just any vehicle and mission. For that reason, it is also interesting to look at guidance of high lift-to-drag vehicles, although we should not forget that the Shuttle guidance was developed from the Apollo guidance. Even the recently launched Mars Science Laboratory employs a guidance scheme, the so-called entry terminal point controller guidance algorithm that has been derived from the final phase Apollo Command Module guidance. Like Apollo, MSL modulates the bank angle to control range based on deviations in range, altitude rate, and drag acceleration from a reference trajectory (Mendeck and Craig 2011).

The design principle of the Shuttle guidance is that the heat load into the structure must be minimised to achieve minimum thickness and weight of the thermal protection system. This is achieved by flying the maximum angle of attack of 38° during the period of high heating rate and then reducing this to about 28° to achieve good cross-range characteristics. Furthermore, to minimise the Shuttle's internal structure weight, the aerodynamic loads are limited during the re-entry. The range to be flown during entry is a unique function of the drag acceleration profile maintained throughout the entry flight. The Space Shuttle entry guidance is designed on the principle of analytically defining a desired drag acceleration profile, which is possible because the local flight-path angle is close to zero, and commanding the Shuttle attitudes to achieve the desired profile. It uses a combination of bank-angle and angle-of-attack modulation for this trajectory control. The bank angle is in this case used to control both the total entry range and the cross-range component of the entry range. The magnitude of the bank angle controls total range, and the direction (or sign) of the bank angle controls the Shuttle's heading relative to the path required to initiate the turn onto the runway final approach. Because trajectory response to bank-angle modulation is relatively slow, the angle of attack is modulated to achieve the reference drag acceleration on a short-period basis. In addition, the angle of attack is selected to minimise the aerodynamic heating environment while achieving the required cross range. For more details, the reader is referred to the publication by Harpold and Graves (1979). The performance of the Space-Shuttle entry guidance system proved to be excellent during all of the Space-Shuttle flights. A detailed discussion on the flight results of the first three flights is given by Harpold and Gavert (1983).

The MBB (1988) study on HORUS also gives a complete description of a guidance system, tailor-made for HORUS and its particular nominal mission. The separation of the guidance controller into a horizontal and a vertical one is the same as for the Space Shuttle, but the vertical guidance is not based on a velocity-referenced drag profile, but on a number of other parameters. The main reason for this different design was, that the nominal trajectories of HORUS were not necessarily composed of sections with constant flight parameters.

On the other hand, Jänsch and Markl (1991) modified the Shuttle guidance for the Hermes nominal mission from a 28.5° inclined orbit to a landing site in Istres, to meet with the larger cross-range demands. Also Serrano-Martínez and Parra (1987) applied with success a variation of the Space-Shuttle entry guidance to the Hermes spaceplane, for both an Istres landing (high cross-range capability) and a Kourou landing (low cross-range capability). Last but not least, Roenneke and Markl (1994) modified the Shuttle guidance laws to apply it to European missions, since the original design did not prove to be accurate enough. As the Shuttle guidance is based on a drag function constructed from pieces of parabola functions, they used a more general approach to model the drag function as a cubic spline over a number of break points. To achieve the required range, the drag level of each break point is computed with numerical trajectory optimisation. Moreover, the drag is scheduled versus the total energy, compared with Earth relative velocity in case of the Space Shuttle. The above mentioned method has three advantages (Roenneke and Markl 1994): an accurate analytic approximation of the range to go is possible regardless of the local flight-path angle; no distinction of different trajectory segments is necessary; the Terminal-Area Energy Management (TAEM) interface is a natural terminal state.

Terminal-Area Energy Management will be discussed in Chap. 8. Here, we will focus on the Space-Shuttle guidance (Sect. 7.6.2), since it is a well-proven, well-documented design that can be used for many different missions and vehicles. In Sect. 7.6.3 we will discuss two other guidance principles (trajectory tracking and lateral guidance), whereas Sect. 7.6.4 concludes this section on guidance systems with a case study about heat-flux tracking.

7.6.2 Guidance System Space Shuttle

The Shuttle entry guidance controlled the trajectory from the moment it entered the atmosphere down to the TAEM phase. At the end of the entry phase, and consequently the beginning of the TAEM phase, the Shuttle had a velocity of 762 m/s (2,500 ft/s) and the altitude was about 24.4 km (80,000 ft) (Harpold and Graves 1979). The mission profile comprised returns from orbit with an inclination varying between 28.5° and 104° with an altitude up to 500 nmi. Thermal loads were minimised by flying an angle-of-attack profile of about 40° during the high-heating period and decreasing it thereafter to have sufficient control on the lateral range. The bank angle rate and acceleration were limited to 5°/s and 1.7°/s^2 to minimise the demands on the flight control system. In addition, the dynamic pressure was constrained to limit the

control-surface hinge moments. A limit of 2.5 g for the normal aerodynamic load was imposed to minimise structure weight and to safeguard the crew. These constraints put a limit on the minimum Orbiter altitude at a given velocity during entry. If the Orbiter would be lower, too much energy was dissipated and one of the constraints would have been violated. The maximum altitude is limited by the equilibrium flight conditions. The purpose of the entry guidance system of the Shuttle can be stated as (Harpold and Graves 1979):

1. Guide the Orbiter along a trajectory that minimises the demands on the Orbiter's design
2. Deliver the Orbiter to an acceptable energy state and attitude at the initiation of the TAEM guidance

Below, we will first discuss the operating principle and theoretical basis of the guidance system, and conclude with that part of the theory that is specific to the Shuttle.

7.6.2.1 Guidance Principle

The basic concept on which the guidance system operates is the analytical definition of a desired drag acceleration profile that should be flown during entry. To achieve this it commands the attitude of the Orbiter. The range-to-go during entry is a unique function of the drag acceleration profile. This range can be predicted by analytical techniques for geometric drag acceleration functions of the Earth-relative velocity if the flight path angle is close to zero. For low velocities, the flight-path angle cannot be kept small and the range can be predicted analytically if the independent variable is changed from the Earth-relative velocity to energy with respect to Earth. The flight through the entry corridor is then achieved by linking these two geometric functions together. The drag acceleration profile is adjusted to account for constraints and designed to minimise the accumulated heat load. The resulting range is predicted by an analytical solution to the equations of motion and compared to the actual range. The profile is iteratively adjusted until these two converge to the same value. The resulting drag acceleration profile is then defined to be the reference drag-acceleration profile. The altitude rate and the L/D in the vertical plane that are required by the trajectory controller are also computed analytically from this reference drag profile.

The control law computes the control commands by comparing the actual flown profile to the reference profile (as a function of range-to-go). This law, based on a linearisation of the flight dynamics, is a homogeneous second-order differential equation with constant coefficients. Bank-angle and angle-of-attack modulation can be used to achieve the commanded L/D profile. The Orbiter uses σ-modulation as the primary control variable, α is used secondary to minimise the aerodynamic heating, while achieving the required cross-range. Hence, bank-angle modulation is used to control the total entry range and cross-range. The magnitude of the bank angle determines the total entry range and the direction determines the cross-range, the latter of which will be discussed later in Sect. 7.6.3. The angle-of-attack is modulated to achieve the reference profile on a short-period basis, because the response to bank

angle modulation is relatively slow. This reduces the phugoid motion during a bank reversal and provides additional manoeuvrability to compensate for deviations.

The Shuttle entry comprises four different phases to achieve a successful entry, thus also the drag-acceleration profile consists of four different segments. Each segment has its own analytical representation of the range prediction and reference drag acceleration profiles. The four segments are:

1. Temperature Control (quadratic in drag)
2. Equilibrium Glide
3. Constant Drag
4. Transition (linear in drag).

7.6.2.2 Theoretical Foundation

The equations of motion are formulated in spherical components under the assumption of a non-rotating Earth such that the Coriolis and centripetal accelerations can be neglected. The three dynamic equations are in that case given by (see Sect. 6.1 for the complete set):

$$\dot{V} = -\frac{D}{m} - g\sin\gamma = -\tilde{D} - g\sin\gamma \tag{7.236}$$

$$V\dot{\gamma} = \left(\frac{V^2}{R} - g\right)\cos\gamma + \frac{L}{m}\cos\sigma = \left(\frac{V^2}{R} - g\right)\cos\gamma + \tilde{L}\cos\sigma \tag{7.237}$$

$$V\cos\gamma\dot{\chi} = \frac{V^2}{R}\cos^2\gamma\sin\chi\tan\delta + \frac{L}{m}\sin\sigma = \frac{V^2}{R}\cos^2\gamma\sin\chi\tan\delta + \tilde{L}\sin\sigma \tag{7.238}$$

In the above equations, $\tilde{D}$ and $\tilde{L}$ are the drag and lift force per unit mass (or, equivalently, the drag and lift acceleration).

7.6.2.3 Range Prediction

The time rate of change of the range is calculated by:

$$\dot{s} = V\cos\gamma \tag{7.239}$$

With $\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}s}{\mathrm{d}V}\frac{\mathrm{d}V}{\mathrm{d}t}$ and $\frac{\mathrm{d}V}{\mathrm{d}t}$ given by Eq. (7.236), we get after integration:

$$s = -\int \frac{V\cos\gamma}{\tilde{D} + g\sin\gamma}\mathrm{d}V \tag{7.240}$$

For small angles $\cos\gamma \approx 1$ and $\sin\gamma \approx \gamma$, which is a realistic assumption for gliding flight, so, with $D >> g\gamma$:

$$s = -\int \frac{V}{\tilde{D}} \mathrm{d}V \tag{7.241}$$

The previous equation forms the basis for all flight segments. For each segment the drag is formulated as a function of the velocity and Eq. (7.240) is integrated. In the final part, the flight-path angle is not sufficiently small anymore and the independent variable is changed to energy. The energy per unit mass of the vehicle at any moment is:

$$E = gh + \frac{1}{2}V^2 \tag{7.242}$$

To express the change of range with altitude we start with the kinematic equation $\dot{h} = V \sin\gamma$ and get

$$\dot{h} = \frac{\mathrm{d}h}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}s} V \cos\gamma \tag{7.243}$$

yielding $\frac{\mathrm{d}h}{\mathrm{d}s} = \tan\gamma$. Together with Eq. (7.242) we obtain the equation for the range:

$$s = -\int \frac{\cos\gamma}{\tilde{D}} \mathrm{d}E. \tag{7.244}$$

7.6.2.4 The Reference L/D

The reference L/D is constructed from the altitude rate and drag acceleration. For small flight-path angles the altitude rate can be approximated by $\dot{h} = V\gamma$. Differentiating the altitude rate with respect to time, using small-angle approximations and substituting Eqs. (7.236) and (7.237) results in:

$$\ddot{h} = -\tilde{D}\frac{\dot{h}}{V} + \left(\frac{V^2}{r} - g\right) + (L/D)\tilde{D} \tag{7.245}$$

To find an alternative expression for $\ddot{h}$ we will start with the time derivative of the drag acceleration, defined by $\tilde{D} = \frac{C_D S}{m}\frac{1}{2}\rho V^2$:

$$\dot{\tilde{D}} = \frac{\dot{C}_D S}{m}\frac{1}{2}\rho V^2 + \frac{C_D S}{m}\frac{1}{2}\dot{\rho}V^2 + \frac{C_D S}{m}\rho V \dot{V} \tag{7.246}$$

Both sides of Eq. (7.246) divided by $\tilde{D}$ gives

$$\frac{\dot{\tilde{D}}}{\tilde{D}} = \frac{\dot{C}_D}{C_D} + \frac{\dot{\rho}}{\rho} + 2\frac{\dot{V}}{V} \tag{7.247}$$

The time derivative of the density follows from the expression of the exponential atmosphere, $\rho = \rho_0 e^{-\beta h}$, and is

$$\dot{\rho} = -\beta \rho \dot{h} \tag{7.248}$$

For small values of the flight-path angle we can ignore the weight term with respect to the drag, so Eq. (7.236) reduces to $\dot{V} = -\tilde{D}$. Both derivatives $\dot{\rho}$ and $\dot{V}$ substituted in Eq. (7.247) yields

$$\frac{\dot{\tilde{D}}}{\tilde{D}} = \frac{\dot{C}_D}{C_D} - \beta \dot{h} - 2\frac{\tilde{D}}{V} \tag{7.249}$$

or

$$\dot{h} = \frac{1}{\beta}\left(\frac{\dot{C}_D}{C_D} - \frac{\dot{\tilde{D}}}{\tilde{D}} - 2\frac{\tilde{D}}{V}\right) \tag{7.250}$$

Taking the time derivative of $\dot{h}$ in Eq. (7.250) and equating it to Eq. (7.245) gives after some manipulation:

$$\begin{aligned}\ddot{\tilde{D}} - \dot{\tilde{D}}\left(\frac{\dot{\tilde{D}}}{\tilde{D}} - \frac{3\tilde{D}}{V}\right) + \frac{4\tilde{D}^3}{V^2} = &- \tilde{D}\beta\left(\frac{V^2}{R} - g\right) - \tilde{D}^2\beta\frac{L}{D} - \\ &- \frac{\dot{C}_D\tilde{D}}{C_D}\left(\frac{\dot{C}_D}{C_D} - \frac{\tilde{D}}{V}\right) + \frac{\ddot{C}_D\tilde{D}}{C_D}\end{aligned} \tag{7.251}$$

One can isolate L/D in Eq. (7.251), so it is possible to construct the L/D reference profile once a function for the drag is assumed. This profile is a generalisation of the drag-acceleration profile. The foregoing two equations are used by the control law to steer the Orbiter to the desired drag-acceleration profile.

7.6.2.5 Control Law

The linear relations for small deviations from the reference profile are written as:

$$\begin{aligned}\delta_{(L/D)} &= (L/D)_{cmd} - (L/D)_0 \\ \delta_{\ddot{\tilde{D}}} &= \ddot{\tilde{D}} - \ddot{\tilde{D}}_0 \\ \delta_{\dot{\tilde{D}}} &= \dot{\tilde{D}} - \dot{\tilde{D}}_0 \\ \delta_{\tilde{D}} &= \tilde{D} - \tilde{D}_0 \\ \delta_V &= V - V_0\end{aligned}$$

The linear feedback function $\delta(L/D) = k_1\delta\tilde{D} + k_2\delta\dot{\tilde{D}} + k_3\delta V$ is used to achieve a desired dynamic response in the control to the reference profile, the gains are expressed by k with an index. By taking first-order difference equations of Eq. (7.251) and the linear feedback function, a second-order homogeneous equation with constant coefficients is found in the form of a damped oscillator. The general form is: $\delta\ddot{\tilde{D}} + 2\zeta\omega\delta\dot{\tilde{D}} + \omega^2\delta\tilde{D}$. Here, ζ is the damping coefficient and ω is the undamped natural frequency. The term $\delta\dot{\tilde{D}}$ is eliminated by applying a first order difference to Eq. (7.250), because it has to be obtained numerically otherwise. This changes the control law to $\delta(L/D) = k_1\delta\tilde{D} + k_2\delta\dot{h} + k_3\delta V$. One can thus obtain for the commanded L/D:

$$(L/D)_{cmd} = (L/D)_0 + k_1\delta\tilde{D} + k_2\delta\dot{h} + k_3\delta V \tag{7.252}$$

With the velocity used as an independent variable, one consequently has $\delta V = 0$, since any velocity deviation just shifts the reference profile in the on-board tables. A steady state drag acceleration error is introduced when the altitude rate is determined from the inertial navigation system. The resulting control law, with an integral controller to correct for the steady-state error, is:

$$(L/D)_{cmd} = (L/D)_0 + k_1(\tilde{D} - \tilde{D}_0) + k_2(\dot{h} - \dot{h}_0) + k_4\int_0^t(\tilde{D} - \tilde{D}_0)dt. \tag{7.253}$$

7.6.2.6 Guidance Logic

The theory presented in the previous subsection is generally applicable and the selection of the shape and drag acceleration profiles depends on the vehicle and mission flown. The entry trajectory of the Orbiter consists of five segments, the quadratic segment is flown twice. A figure of the entry corridor for a 40° angle-of-attack profile can be seen in Fig. 7.75.

The drag coefficient and its derivatives are a function of α, and for the Shuttle the following empirical relations are used in the calculations (Harpold and Graves 1979):

$$C_D = K_1 + K_2\alpha + K_3\alpha^2 + K_4 e^{-\left(\frac{V-K_5}{K_6}\right)} \tag{7.254}$$

$$\dot{C}_D = K_7(\tilde{D} + g\dot{h}/V)e^{-\left(\frac{V-K_5}{K_6}\right)} + \dot{\alpha}_0(K_8\alpha_0 + K_9) \tag{7.255}$$

The constants K_1–K_9 depend on the vehicle aerodynamics. The terms $\dot{C}_D$ are relatively small and can be deleted to save computation time. The control-law gains are complex expressions and empirical equations were again used:

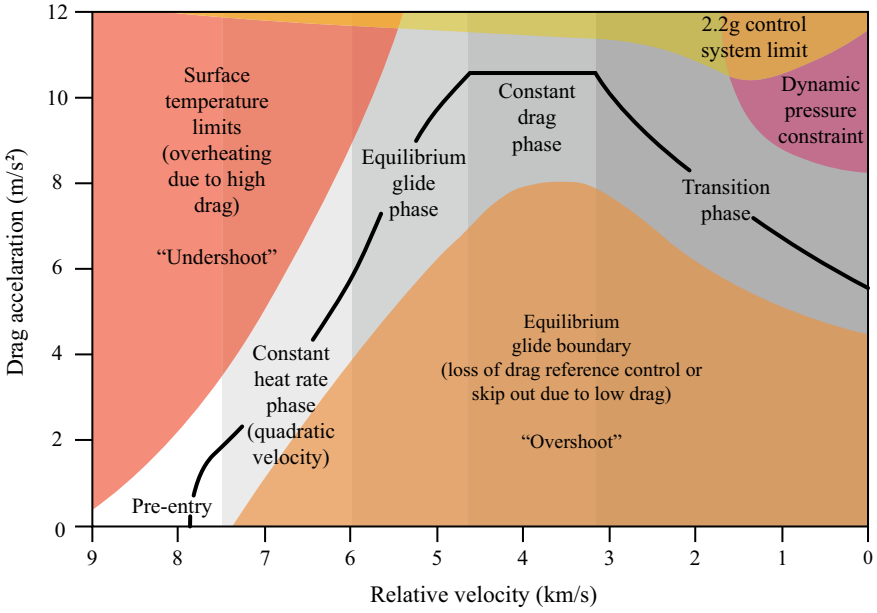


Fig. 7.75 Entry corridor showing different flight segments

$$k_1 = K_{10} D^{K_{11}} + K_{12}(\tilde{D} - \tilde{D}_0) \tag{7.256}$$

$$k_2 = K_{13} \tilde{D}^{K_{14}} \tag{7.257}$$

The bank angle modulation is computed using:

$$\sigma_{cmd} = \cos^{-1}\left(\frac{(L/D)_{cmd}}{L/D}\right) + k_5(\alpha - \alpha_0) \tag{7.258}$$

Stable platform acceleration data in the navigation system were used to estimate L/D in flight. The angle-of-attack modulation is given by:

$$\Delta\alpha = C_D(\tilde{D}_0 - \tilde{D})/k_6 \tag{7.259}$$

The concept of planning a drag-acceleration profile was originally developed for the Apollo entry capsule and modified for the Space Shuttle. Both missions have proven that this is a successful guidance strategy. It is also a strategy that can be generally applied with a fine tuning of the shape of the profile depending on mission and vehicle. An advantage of drag acceleration planning and tracking is that the method is robust to errors introduced by aerodynamic force modelling. However, the

Shuttle's guidance system generates a drag acceleration profile for the longitudinal motion. By combining it with lateral guidance, the cross range can be increased and extreme lateral points can be reached more accurately (Saraf et al. 2004).

A disadvantage, though, is that the reference profile is not calculated on-board. Abort scenarios for the Shuttle are all designed separately and this requires a significant effort.[14] On-board mission planning, on the other hand, does not have this drawback. Furthermore, it can also take off-nominal conditions into account and hence, allows for the generation of a more accurate trajectory. On-board mission planning would increase safety, by taking into account off-nominal conditions, and reduce cost by lowering the pre-mission design effort. According to Mease and Kramer (1994), the Shuttle's trajectory tracker is designed for Shuttle mission conditions and therefore has a limited operating domain.

7.6.3 Other Guidance Principles

7.6.3.1 Trajectory Tracking

In Sect. 7.5.4.2 we discussed the theory of optimal control, which we subsequently applied to the design of the HORUS attitude controller for longitudinal and lateral motion. Of course, this form of perturbation control can also be used for the design of a reference-trajectory tracking guidance, in the form of a linear state feedback controller. The (linear) system of the error dynamics to be controlled is in state-space form, i.e.,

$$\Delta\dot{\mathbf{x}} = \mathbf{A}\Delta\mathbf{x} + \mathbf{B}\Delta\mathbf{u} \qquad \mathbf{y} = \mathbf{C}\Delta\mathbf{x} \tag{7.260}$$

In Chap. 6 the linearised model for a winged re-entry vehicle has been derived. The states representing translational motion are ΔV, $\Delta\gamma$ and ΔR, which are, in fact, the states for vertical motion. As (corrective) guidance parameters, $\Delta\mathbf{u}$, we only have the option to select the angle of attack, $\Delta\alpha$ and bank angle, $\Delta\sigma$, which are the two variables that define the magnitude and orientation of the aerodynamic force components (drag D and lift L). Retaining the related state derivatives we get for $\mathbf{A}$:

[14] Potential solutions for abort guidance are discussed by, for instance, Gonzalez Puerta et al. (2018) and Dominguez Calabuig and Mooij (2021), who consider both pseudo-spectral methods and convex optimisation, respectively, for (real-time) on-board application.

$$\mathbf{A} = \begin{bmatrix} -\frac{1}{mV_0}\left(M_0 \frac{\partial C_D}{\partial M} \bar{q}_0 S_{ref} + 2D_0\right) & -g_0 \cos\gamma_0 & \cdots \\ \frac{1}{V_0}\left(-\dot{\gamma}_0 + \frac{2V_0}{R_0}\cos\gamma_0\right) + \frac{\cos\sigma_0}{mV_0^2}\left(M_0 \frac{\partial C_L}{\partial M} \bar{q}_0 S_{ref} + 2L_0\right) & -\left(\frac{V_0}{R_0} - \frac{g_0}{V_0}\right)\sin\gamma_0 & \cdots \\ \sin\gamma_0 & V_0 \cos\gamma_0 & \cdots \end{bmatrix}$$

$$\begin{bmatrix} \cdots & 2\frac{g_0}{R_0}\sin\gamma_0 \\ \cdots & \left(\frac{2g_0}{V_0} - \frac{V_0}{R_0}\right)\frac{\cos\gamma_0}{R_0} \\ \cdots & 0 \end{bmatrix} \tag{7.261}$$

The corresponding partials for $\Delta\alpha$ and $\Delta\sigma$ make up the control matrix $\mathbf{B}$:

$$\mathbf{B} = \begin{bmatrix} -\frac{1}{m}\frac{\partial C_D}{\partial \alpha}\bar{q}_0 S_{ref} & 0 \\ \frac{\cos\sigma_0}{mV_0}\frac{\partial C_L}{\partial \alpha}\bar{q}_0 S_{ref} & -\frac{L_0}{mV_0}\sin\sigma_0 \\ 0 & 0 \end{bmatrix} \tag{7.262}$$

The non-zero coefficients in $\mathbf{B}$ are the same as $a_{V\alpha}$, $a_{\gamma\alpha}$ and $a_{\gamma\sigma}$, as given in Sect. 6.4.1. Note that a principal disadvantage of this tracking method is that in case the nominal bank angle is zero, the corresponding coefficient, $a_{\gamma\sigma}$, is also zero, and there can be no bank-angle corrections.

For state feedback, the guidance law is defined to be

$$\Delta\mathbf{u} = -\mathbf{K}\Delta\mathbf{x} \tag{7.263}$$

However, because the altitude is a kinematic variable, it can only be indirectly controlled by changing γ (note the zero coefficients of $\mathbf{B}$). Altitude and flight-path angle are combined into a so-called pseudo altitude, h^*:

$$h^* = h + K_\gamma \gamma \tag{7.264}$$

where K_γ is a velocity-dependent gain to account for the large velocity regime, i.e.,

$$K_\gamma = 2\frac{c_{h,2}}{c_{h,1}} V_0 \tag{7.265}$$

with V_0 being the reference (=nominal) velocity. The altitude-control parameters $c_{h,1}$ and $c_{h,2}$ can be found from a linear analysis of the altitude-control dynamics by specifying a preferred eigenfrequency and damping. Typical numerical values are $c_{h,1} = 0.005$ and $c_{h,2} = 0.05$.

This means that our output matrix $\mathbf{C}$ becomes

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2\frac{c_{h,2}}{c_{h,1}} V_0 & 1 \end{bmatrix} \tag{7.266}$$

and we will use output feedback to guide the vehicle:

$$\Delta \mathbf{u} = -\mathbf{K}\Delta \mathbf{y} \tag{7.267}$$

With the definition of the two weighing matrices $\mathbf{Q}$ and $\mathbf{R}$ as

$$\mathbf{Q} = \text{diag}\left\{ \frac{1}{\Delta V_{max}^2} \quad \frac{1}{\Delta h_{max}^{*2}} \right\} \tag{7.268}$$

$$\mathbf{R} = \text{diag}\left\{ \frac{1}{\Delta \alpha_{max}^2} \quad \frac{1}{\Delta \sigma_{max}^2} \right\} \tag{7.269}$$

we can solve the Riccati equation to obtain $\mathbf{K}$ with the theory discussed in Sect. 7.5.4.2.

7.6.3.2 Lateral Guidance

It is the task of the lateral entry guidance (or *horizontal* guidance in contrast with the vertical guidance discussed in, for instance, Sect. 7.6.2) to bring the vehicle in the vicinity of the landing site, so that the Terminal Area Energy Management (TAEM) guidance can take over and steer the vehicle towards a safe landing. The layout of the targeting strategy is shown in Fig. 7.76. It is noted that the following discussion is partially based on MBB (1988), but is generally applicable to any re-entry vehicle with sufficient lift.

The targeting point of the horizontal guidance is located at a distance d_{rw} from the centre of the runway, in opposite direction to the landing direction. The Terminal Area is reached, when the vehicle reaches the so-called TAEM interface is reached. This is a circular area, with a radius of R_T and the centre located in the targeting point T. Typical values for d_{rw} and R_T are 12 km and 0.75° (≈83 km), respectively. After intersecting the TAEM interface, typically at an altitude of 25 km and a velocity equal to Mach 3, a dedicated TAEM guidance system takes over, and the vehicle is guided towards a so-called Heading Alignment Cylinder (HAC). Depending on initial position and flight direction, either a right (HAC-R) or left (HAC-L) turn can be flown. Finally, the vehicle glides towards the runway in a straight descending flight and lands. The guided flight in the Terminal Area is discussed in some more detail in Chap. 8.

As has become clear from the above discussion, the objective of the horizontal guidance is to reach the TAEM interface in such a way that the TAEM guidance has no problem of steering the vehicle to the runway. In principle, this corresponds to a heading error that is within prescribed limits. Note that the energy state with which the vehicle reaches the TAEM interface, i.e., velocity and altitude, is controlled by the vertical guidance. So the first task of the horizontal guidance is to determine the sign of the initial bank angle (when it differs from zero), since the vertical guidance gives only the absolute value. Because there is a direct relation between the sign of the bank angle and the heading error, we can use the latter one for this purpose.

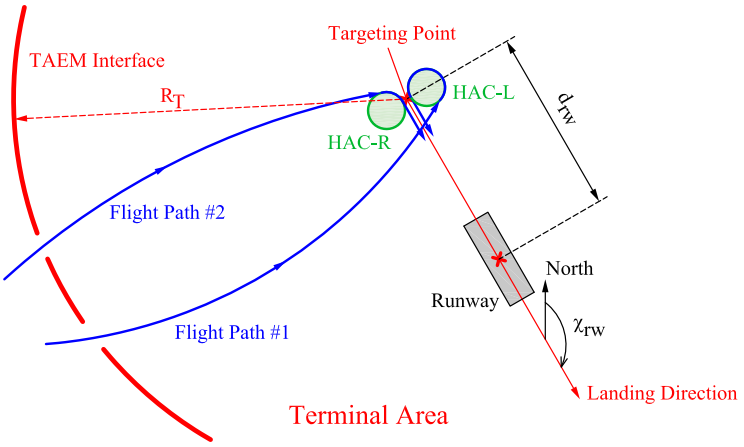


Fig. 7.76 Guidance towards TAEM interface

To restrict the heading error, we will define a so-called heading-error dead band that gives us the maximum allowable error (see Fig. 7.77 for an example of what the dead band looks like). The upper boundary of this dead band is driven by two notions. In the first place, the variation of the length of the actual travelled and the nominal trajectory should be kept (relatively) small. This tends to decrease the allowable heading error. In the second place, intersection with the TAEM interface must be guaranteed, which imposes restrictions on the dead band closer to the targeting point. The lower boundary is dictated by the fact that a narrow dead band increases the number of bank reversals, which is not very favourable. Confinement of this number is therefore desirable. Partly based on experience gained with the Space Shuttle, dead-band values between 10° and 30° give a reasonable compromise.

When the vehicle is still far away from the TAEM interface, the definition of the heading-error dead band is rather uncritical. We will call this the coarse dead band. However, in close proximity of the interface a too large dead band might result in the vehicle passing outside the interface. On the other hand, when the dead band is too narrow, there will be an increasing number of bank reversals. To overcome this, the so-called low-distance heading-error dead band, $\chi_{d,low}$, is derived from the requirement that if at a certain distance d, the heading error χ_e amounts to $\chi_{d,low}(d)$, then it will be reduced to zero when reaching the TAEM interface provided that the proper sign of the bank angle is applied. The heading error is defined as $\chi_e = \chi_{nav} - \chi_T$, with χ_{nav} being the actual heading given by the navigation system and χ_T the heading to the targeting point from current position. Numerical values of the low-distance heading-error dead band can be found by doing a backward integration for the selected vehicle and mission.

The geometry to find an expression for χ_T is shown in Fig. 7.78 (see Appendix C, Sect. C.1, for more details on spherical trigonometry). For the current location (τ,δ), or $(\tau_{nav}, \delta_{nav})$ when navigation errors are included, and the target location

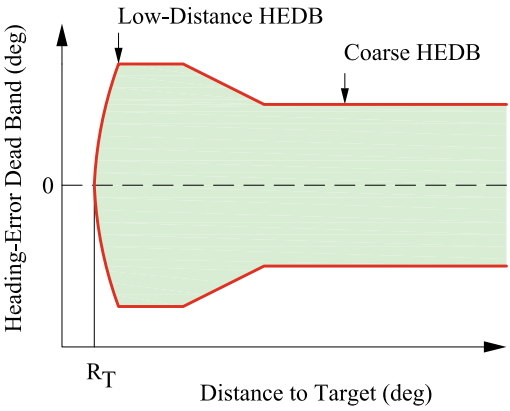


Fig. 7.77 Heading-error dead band

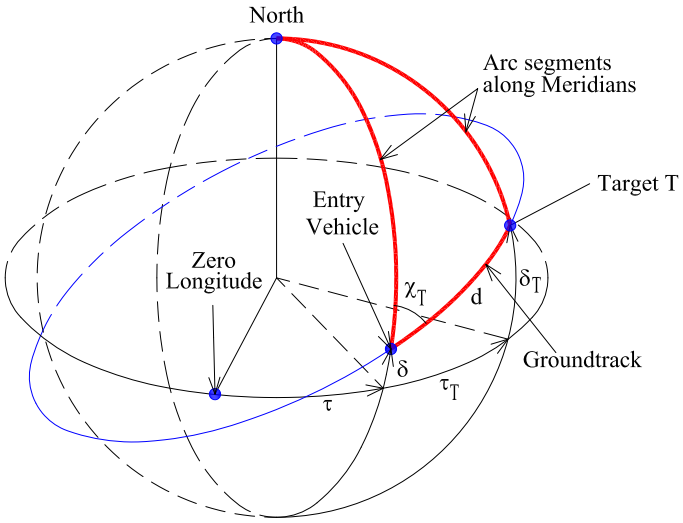


Fig. 7.78 Spherical geometry for deriving the heading to the targeting point T

(τ_T, δ_T), the targeting heading is written as:

$$\sin(\tau_T - \tau)\cot\chi_T = \cot\left(\frac{\pi}{2} - \delta_T\right)\sin\left(\frac{\pi}{2} - \delta\right) - \cos(\tau_T - \tau)\cos\left(\frac{\pi}{2} - \delta\right) \tag{7.270}$$

or

$$\chi_T = \arctan\left[\frac{\sin(\tau_T - \tau)\sin\left(\frac{\pi}{2} - \delta_T\right)}{\cos\left(\frac{\pi}{2} - \delta_T\right)\sin\left(\frac{\pi}{2} - \delta\right) - \cos(\tau_T - \tau)\cos\left(\frac{\pi}{2} - \delta\right)\sin\left(\frac{\pi}{2} - \delta_T\right)}\right] \tag{7.271}$$

As we mentioned above, a bank reversal will be initiated if

$$|\chi_e| > |\chi_d| \tag{7.272}$$

but only when the bank angle does not have the proper sign already, or, in mathematical terms, when:

$$\chi_e \text{sign}(\sigma) < 0 \tag{7.273}$$

During the bank reversal, the commanded bank angle changes with a constant theoretical bank rate $\dot{\sigma}_c$ from an initial value σ_0 to the final value $-\sigma_0$. Assume the guidance algorithm operates on the guidance interval Δt_g. Once a bank reversal is initiated by the guidance algorithm, the bank angle will be increased or decreased according to

$$\sigma_{i+1} = \sigma_i + \dot{\sigma}_c \Delta t_g \tag{7.274}$$

with the first value of σ_i being equal to σ_0. The initial value σ_0 is the last value generated by the vertical guidance. The commanded bank rate may be defined as an input function of the dynamic pressure, done so because the dynamic pressure defines the banking capabilities due to deflecting the aerodynamic control surfaces. For higher dynamic pressures, a higher bank rate is applied. It should be noted that during the bank reversal, no (or limited) variations in commanded angle of attack and absolute bank angle are computed by the vertical guidance logic to prevent oscillating command signals. When the commanded variations are kept constant, they are adjusted again after finishing the bank reversal.

It is noted, though, that particular attention should be paid to attitude control during the reversal. For instance, if the vehicle rotates faster than commanded, but is going "in the right direction" it would do no good if the attitude controller would try to slow down the vehicle to be in line with the guidance command. Therefore, the (absolute) largest of either previous guidance command or the current bank angle estimated by the navigation system should be used. However, one has to safeguard that the vehicle will also not rotate excessively fast.

The heading-error dead band is an input function of the distance to target d, or equivalently d_{nav}, if we assume the distance to be provided by the navigation system as a processed value. The distance to target is given by the angle between the Earth radius vectors to both the vehicle and targeting point, and can be expressed by

$$\cos d = \sin\delta \sin\delta_T + \cos\delta \cos\delta_T \cos(\tau - \tau_T) \tag{7.275}$$

Here, τ_T and δ_T are the longitude and latitude of the targeting point T, respectively. The location of the targeting point (h_T, τ_T, δ_T) can be derived from the known location of the runway $(h_{rw}, \tau_{rw}, \delta_{rw})$, the landing direction χ_{rw} and the distance d_{rw} of the targeting point to the centre of the runway (Fig. 7.79). Given these parameters, the location of T is given by:

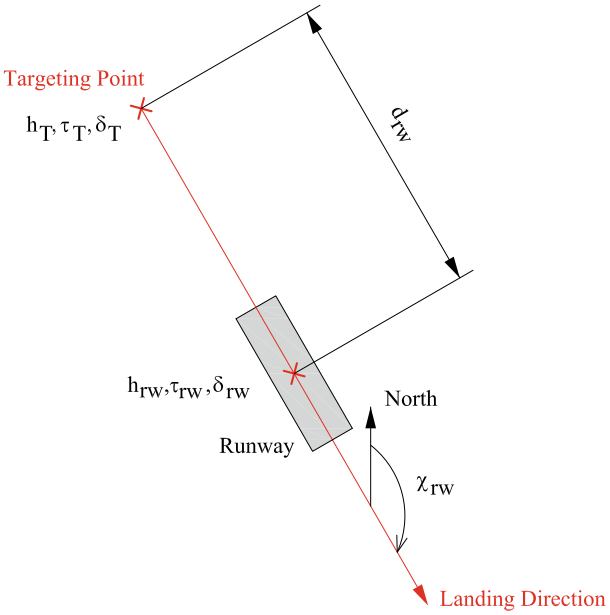


Fig. 7.79 Runway location and related parameters

$$\tau_T = \tau_{rw} - \frac{d_{rw}}{R_e} \sin \chi_{rw} \tag{7.276}$$

$$\delta_T = \delta_{rw} - \frac{d_{rw}}{R_e} \cos \chi_{rw} \tag{7.277}$$

$$h_T = h_{rw} \tag{7.278}$$

However, rather than using the actual (measured) distance d_{nav}, applying the final, low-distance dead band is based on the so-called effective distance d_{eff}, related to d_{nav} by

$$d_{eff} = d_{nav} - d_{bank} \tag{7.279}$$

where d_{bank} is the distance that will be travelled during the bank reversal. It is assumed that during the reversal no net change in χ_e will be generated. The equation to calculate d_{bank} is:

$$d_{bank} = \Delta t_{bank} \frac{V_{nav} \cos \gamma_{nav}}{R_s} \tag{7.280}$$

with

$$\Delta t_{bank} = 2 \left| \frac{\sigma_0}{\dot{\sigma}_c} \right| \tag{7.281}$$

The low-distance dead-band check is confined to small distances $d_{eff} < d_{comp}$, where the comparing distance d_{comp} is defined by

$$d_{comp} = R_T + d_{lim} \tag{7.282}$$

The limiting distance d_{lim} is an input parameter and has a typical value of 0.5°. Due to the impact of this last bank reversal and due to the deviations from the nominal flight conditions the requirement of zero χ_e will not be exactly met at the TAEM interface.

The Influence of Bank Reversals on the Flight-Path Angle

During a bank reversal there will be an increase in flight-path angle. The extent of this increment is proportional to the inverse of the bank rate. In other words, the faster the vehicle can bank the smaller the error in flight-path angle will be. It is obvious that an error in flight-path angle will have its influence on the altitude control, and therefore the corrective bank-angle control. Furthermore, the energy dissipation will differ from the nominal dissipation, which might lead to corrective angle-of-attack control. Fast execution of the bank reversal would in that case be favourable.

The variation of the flight-path angle with time can be approximated by

$$\dot{\gamma} \approx \frac{1}{m}\left(\frac{V}{R} - \frac{g}{V}\right)\cos\gamma + \frac{L}{mV} \tag{7.283}$$

by considering the motion in a vertical plane (no sideslip and no banking) and assuming a non-rotating Earth. The increase of γ can be estimated from the variation of the lift force with time during the bank reversal, since it is reasonable to assume that the variation of velocity and altitude are much smaller during the manoeuvre, so:

$$\Delta\gamma = \int_0^{\Delta t} \Delta\dot{\gamma}\mathrm{d}t = \int_0^{\Delta t} \frac{1}{mV}\Delta L(t)\mathrm{d}t \tag{7.284}$$

Assuming a constant bank rate $\dot{\sigma}$ the duration of the bank reversal Δt from $\sigma = -\sigma_0$ to $\sigma = \sigma_0$ becomes

$$\Delta t = 2\frac{\sigma_0}{\dot{\sigma}} \tag{7.285}$$

The variation of the vertical lift force with time is

$$\Delta L(t) = L(t) - L(t = t_0) = C_L \bar{q} S_{ref}\left[\cos(-\sigma_0 + \dot{\sigma}t) - \cos(-\sigma_0)\right] \tag{7.286}$$

where we have assumed that the magnitude of the lift is constant and that only the orientation changes due to the banking manoeuvre. It is noted, though, that a bank

reversal *will* result in a variation of the angle of attack, so the absolute lift force is in that sense not constant. Equation (7.286) substituted into Eq. (7.284) yields

$$\begin{aligned}\Delta\gamma &= \frac{L}{mV}\int_0^{\Delta t}\left[\cos(-\sigma_0+\dot{\sigma}t)-\cos(-\sigma_0)\right]\mathrm{d}t \\ &= \frac{L}{mV}\left[\frac{\sin(-\sigma_0+\dot{\sigma}t)}{\dot{\sigma}}-\cos(-\sigma_0)t\right]_0^{\Delta t}\end{aligned} \tag{7.287}$$

Solving Eq. (7.287) yields

$$\Delta\gamma = \frac{L}{mV}\left[\frac{\sin(-\sigma_0+\dot{\sigma}\Delta t)-\sin(-\sigma_0)}{\dot{\sigma}}-\cos(-\sigma_0)\Delta t\right] \tag{7.288}$$

Rearranging terms and substituting Eq. (7.285) finally gives us

$$\Delta\gamma = 2\frac{L}{mV\dot{\sigma}}\left[\sin\sigma_0-\sigma_0\cos\sigma_0\right] \tag{7.289}$$

where we have used Eq. (7.285). Since the error in γ is inversely proportional to the bank rate, it must therefore be chosen as large as possible to minimise the flight-path angle error, while on the other hand as small as possible to minimise the demands on the attitude controller. In other words, these requirements have to be balanced.

To finalise this mission analysis, in Sect. 7.5.7 we have discussed the design of an attitude controller for HORUS, given the reference mission shown in Fig. 7.67. For the bank manoeuvres flown (Fig. 7.68, right), the induced flight-path angle error is shown in Fig. 7.80. During each of the bank reversals there is an increase in $\Delta\gamma$. As we discussed before, the error is proportional to the inverse of the bank rate, so the faster the vehicle can bank, the smaller the error will be. However, the commanded bank rate is almost constant for each reversal. Since the dynamic pressure is increasing with time, then the lift force increases with time, and, according to Eq. (7.289), $\Delta\gamma$ increases with L. It is obvious that an error in γ will have its influence on the altitude control, and the corrective σ-control. Furthermore, the energy dissipation will differ from the nominal dissipation, which might lead to corrective α-control. In conclusion, especially during bank reversals significant attitude errors may occur. The case study discussed in Sect. 7.5.7 confirms this.

7.6.4 Case Study: Heat-Flux Tracking

A small and low-cost re-entry vehicle can be a good means for testing new heat-resistant materials and qualifying newly developed sub-systems in a realistic flight environment, to complement ground tests and numerical experiments. A disadvantage of a ballistic re-entry test vehicle, however, is the limited manoeuvrability, so

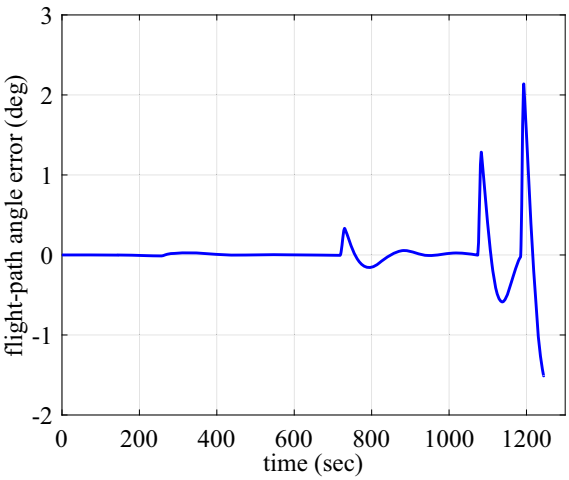


Fig. 7.80 Flight-path angle error as a function of time

that it becomes difficult to predict the landing site, and the small flexibility to cope with flight-path constraints like the heat flux or total load. These difficulties can only be overcome by actively controlling the aerodynamic forces acting on the re-entry vehicle, like the COLIBRI (COncepts of LIfting Bodies for Re-entry Investigations), a vehicle studied by Schöttle and Burkhardt (1995). The feasibility study presented by Mooij et al. (1995) concerns the aerodynamic controllability of such a small re-entry test vehicle, designated *Hyperion-1*. The vehicle has a triangular cross section with rounded corners, a spherical nose and three control flaps mounted at the base, as is shown in Fig. 7.81. The three flaps can be independently deflected to allow for roll, pitch, and yaw control. The vehicle will fly in a bank-to-turn mode where the bottom flap is used for trim-angle control and both upper flaps are used for yaw and roll control.

The current case study is focussed on detailing one of the test requirements, namely to test new, heat-resistant materials under realistic re-entry conditions. At the time this study was initiated, the Russian submarine-launched Volna was the designated launcher, because of its low launch cost. Even though the Volna is not available anymore, we stick to this launcher (and its associated requirements), because it provides representative separation conditions. Because of size limitations of the Volna nosecone, the original *Hyperion-1* is too large (the shape and dimensions of the original configuration are shown in Fig. 7.82). Therefore, this case study gives an overview of the results of a preliminary mission analysis for two scaled-down versions of *Hyperion-1*.

Summarised, the objective of this evaluation is to show that a reduced-size *Hyperion-1* that fits in the Volna nose cone, can fulfill the mission requirement of TPS testing, i.e., to fly for a minimum of 600 s with a heat flux of 300 kW/m^2 at a downstream location on the vehicle. Due to the Volna nose-cone limitations, the maximum vehicle length is set at 1.6 m with a maximum width of 1.2 m. The vehicle mass is 400 kg and the separation conditions from the Volna are such that

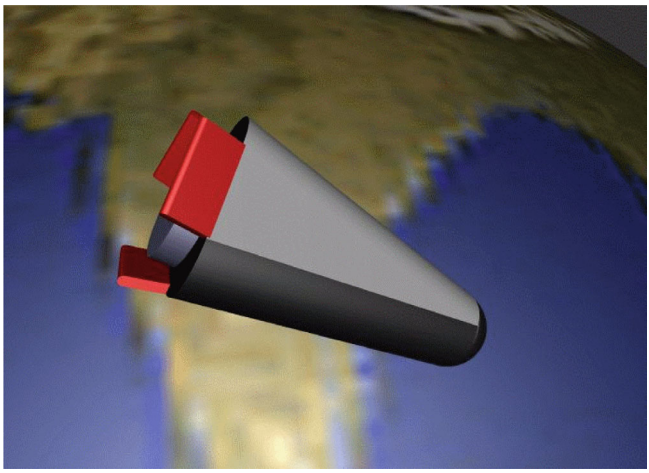


Fig. 7.81 Hyperion-1

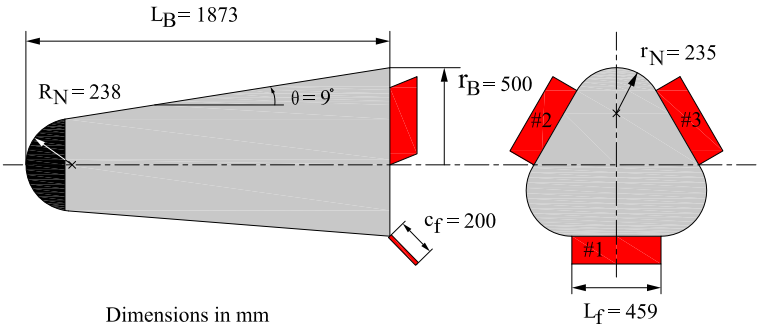


Fig. 7.82 Hyperion-1 geometry

the entry velocity and flight-path angle at 100 km altitude are $V_E = 6$ km/s and $\gamma_E = -5.5°$. Below, we will first introduce the vehicle configurations, then discuss the guidance-system design, and finally show some results.

7.6.4.1 Vehicle Configurations

The two *Hyperion* variants are shown in Figs. 7.83, 7.84, and 7.85 (*Hyperion-1a* is the left one, with the larger nose radius, and *Hyperion-1b* is the right one with the smaller nose radius). The dimensions as listed in Table 7.5 apply. Note that the configurations under study do not have control surfaces. Any trim stability and active control should either be done by moving masses (centre-of-mass control) or RCS thrusters. This is not considered in this case study.

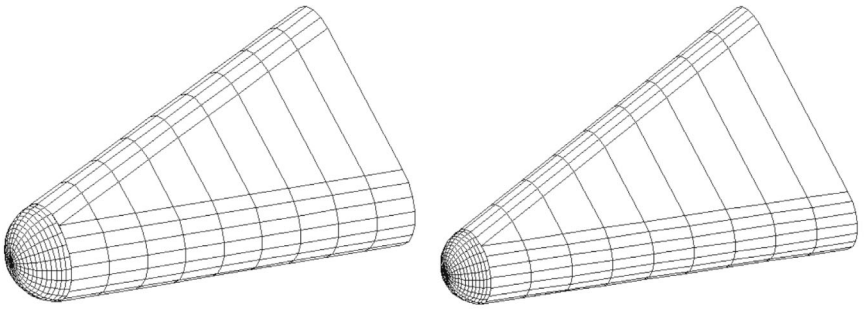


Fig. 7.83 Dimetric view of the two Hyperion-1 variants

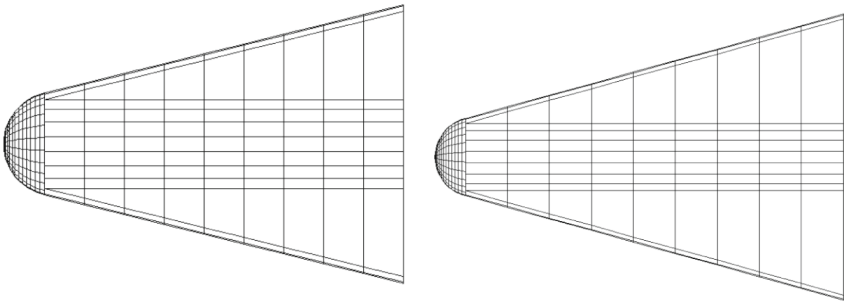


Fig. 7.84 Top view of the two Hyperion-1 variants

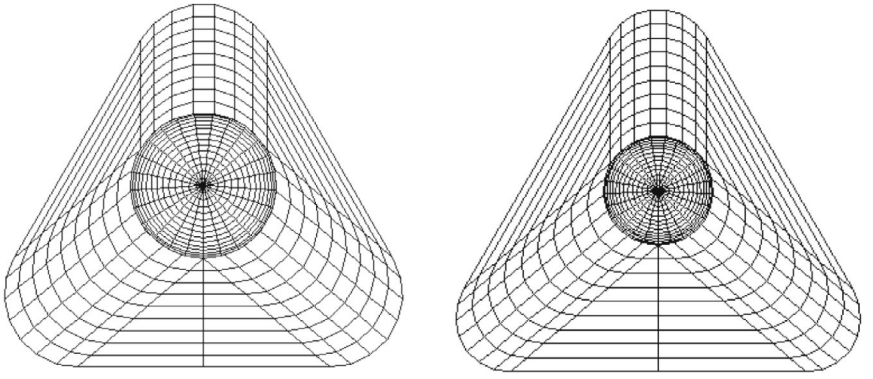


Fig. 7.85 Front view of the two Hyperion-1 variants

Table 7.5 Two reduced-size Hyperion-1 configurations

Name	Unit	Hyp1a	Hyp1b
Upper body angle	[deg]	12.00000	13.50000
Side angle upper panels	[deg]	60.00000	60.00000
Rounded radius base	[m]	0.20000	0.15000
Radius top of nosecone	[m]	0.20000	0.15000
Angle of dihedral bottom	[deg]	0.00000	0.00000
Body width at the base	[m]	1.09282	1.10829
Radius base of nose cone	[m]	0.20000	0.15000
Body height at the base	[m]	1.00000	1.00000
Nose radius	[m]	0.20447	0.15426
Nose length	[m]	0.16196	0.11825
Triangular body length	[m]	1.41139	1.45785
Body length	[m]	1.57335	1.57611

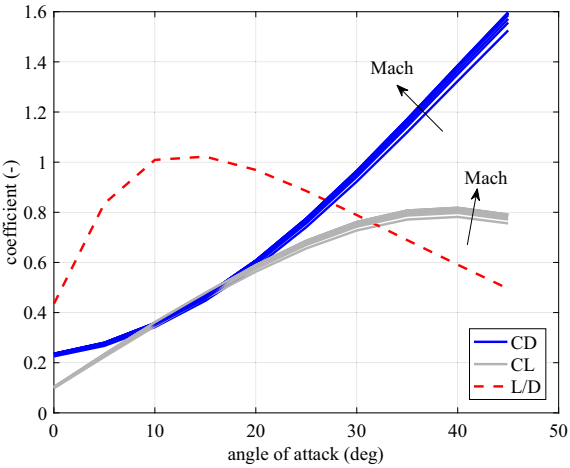


Fig. 7.86 Aerodynamic properties of the *Hyperion-1a* configuration

In Figs. 7.86 and 7.87, the aerodynamic properties are shown, computed with a modified Newtonian panel method. It is clear that the Mach effect is small. For configuration *Hyperion-1a*, the maximum L/D is about 1 at $\alpha = 15°$, whereas the maximum L/D for configuration *Hyperion-1b* is 1.2 at $\alpha = 10°$. Note that for $\alpha = 0°$ the lift is non-zero, due to the geometric asymmetry.

7.6.4.2 Guidance-System Design

To design a guidance system for heat-flux tracking we need to make some assumptions first. The test requirement is that the vehicle needs to fly a trajectory such that

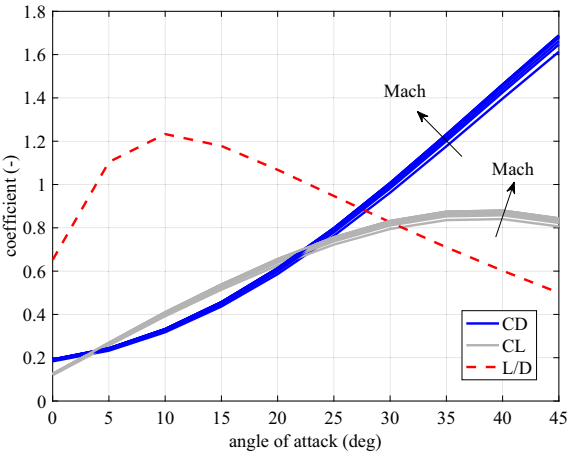


Fig. 7.87 Aerodynamic properties of the *Hyperion-1b* configuration

at a downwards location, i.e., at some distance from the nose where the material sample can be mounted, the maximum heat flux is $q_c = 300$ kW/m^2. To model the heat flux at this location we would like to have an analytical model available, which can be used to design the guidance system. But, as it turns out, there are no real analytical models that accurately describe these locations, so the only solution would be to derive an approximation from, for instance, CFD results. However, this could mean that if we change the vehicle configuration also this approximation may have to be adjusted. If we would look at the problem from a more global perspective, and studying the CFD results for several vehicle configurations then we can of course always assume that there exists a location where the flux scales with the stagnation heat flux. It goes without saying that the actual location of the material sample needs to be verified by detailed CFD analysis, but for the sake of this study we can assume that such a location exists. This makes the design of the guidance system a lot easier, because for the stagnation point we do have a simplified model available.

So the original requirement of $q_c = 300$ kW/m^2 is translated into a stagnation-point heat flux. It is assumed that the downstream location has a 2.5 times lower flux, which means that $q_{c,nose} = q_{c,setpoint} = 750$ kW/m^2. Assuming a laminar flow, we found in Chap. 3 that the stagnation-point convective heat flux is given by

$$q_{c_{lam}} \propto \frac{1}{\sqrt{R_N}} \sqrt{\frac{\rho_\infty}{\rho_0}} \left(\frac{V_\infty}{V_c} \right)^{c_2} \left[1 - \frac{T_w}{T_{w_{ad}}} \right] = c_1^* \frac{1}{\sqrt{R_N}} \rho_\infty^{0.5} V_\infty^{c_2} \left[1 - \frac{T_w}{T_{w_{ad}}} \right] \quad (7.290)$$

The above model assumes that the vehicle wall is in thermal (radiation) equilibrium. As a result we need to calculate the wall temperature T_w by equating the incoming convective flux with the outgoing radiative flux q_{rad} for an assumed adiabatic wall temperature, T_{aw} (here: $T_{aw} = 100$ K):

$$q_{rad} = \varepsilon \sigma T_w^4 \tag{7.291}$$

with ε being the emissivity of the applied material and σ being the Stefan–Bolzmann constant ($\sigma = 5.670400 \cdot 10^{-8}$ J s^{-1}m^{-2}K^{-4}.). So, for thermal equilibrium $T_w = \sqrt[4]{\frac{q_c}{\varepsilon\sigma}}$. Calculating q_c becomes an iterative procedure, which usually converges in a few steps. Alternatively, instead of using the so-called *hot-wall* model, we could assume that the wall does not radiate, effectively putting the wall temperature to zero. As a result we obtain the following *cold-wall* model:

$$q_c = c_1 \frac{1}{\sqrt{R_N}} \sqrt{\rho} V^{c_2} \tag{7.292}$$

This model is also known as Chapman's Equation. Note that for convenience we have dropped the ∞ subscript; also, ρ is the atmospheric density and relates to the altitude where the vehicle is flying with velocity V. A final remark: in literature the velocity exponent, c_2, is in the range 3–3.15, and can have different values from one source to the other. Be aware that always the correct combination of c_1 and c_2 must be used. In this study, the density at sea level $\rho_0 = 1.225$ kg/m^3 and the local circular velocity is assumed to be constant, $V_c = 7{,}905$ m/s. Then, $c_1 = 5.28137 \cdot 10^{-11}$ for a calculated flux in MW/m^2 with c_2 taken as 3.15.

Based on the specified set point, let us define the tracking error, $q_{c,err}$:

$$q_{c,err} = q_c - q_{c,setpoint} \tag{7.293}$$

To reduce (and keep) this error to zero, we will employ a so-called PI-regulator. As we discussed in Sect. 7.5.3 a P(roportional)-I(ntegral) regulator (or *controller*) takes the error as input, and the combination of a proportional scaling and the integrated error added to this results will give us the required steering (or *guidance* command). We will link the heat-flux error with the mechanism that causes the heat in the first place, i.e., the drag. By changing the drag more or less heat can be dissipated. So, the PI-regulator is given by

$$\Delta D_{cmd} = -K_P q_{c,err} - K_I \int_0^t q_{c,err} \mathrm{d}t \tag{7.294}$$

with K_P and K_I being the proportional and integral gain, respectively. These gains can either be determined by a linearised system analysis (in combination with pole placement), or by trial and error, i.e., by selecting gain values and tune them until a reasonable response is obtained.

The above commanded drag increment is only a correction for the situation that the heat-flux error is not zero. Given the set point for the heat flux, this corresponds with a nominal value of the drag, D_0. The nominal drag can be derived from the condition that for a stable (i.e., constant) heat flux, the first derivative of q_c should

be zero. But let us start with the (simplified, vertical-plane) equation of motion for a non-rotating Earth:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \;\Rightarrow\; D_0 = -mg\sin\gamma - m\left.\frac{\mathrm{d}V}{\mathrm{d}t}\right|_0 \tag{7.295}$$

So, if we can find an alternative expression for $\left.\frac{\mathrm{d}V}{\mathrm{d}t}\right|_0$ and substitute that in Eq. (7.295) we find the corresponding value of D_0. From the condition $\dot{q}_c = 0$, from Eq. (7.292) it follows that:

$$\dot{q}_c = \frac{c_1}{\sqrt{R_N}}\left(\frac{1}{2\sqrt{\rho}}\frac{\mathrm{d}\rho}{\mathrm{d}t}V^{c_2} + c_2\sqrt{\rho}V^{c_2-1}\frac{\mathrm{d}V}{\mathrm{d}t}\right) = 0 \tag{7.296}$$

or

$$\frac{1}{2}\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t}V + c_2\frac{\mathrm{d}V}{\mathrm{d}t} = 0 \;\Rightarrow\; \frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{1}{2c_2}\frac{V}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t} \tag{7.297}$$

The partial derivative $\frac{\mathrm{d}\rho}{\mathrm{d}t}$ can easily be obtained from the exponential-atmosphere model $\rho = \rho_0 e^{-\beta h}$, with $\beta = \frac{1}{H_s}$ being the inverted scale height. So:

$$\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}h}\frac{\mathrm{d}h}{\mathrm{d}t} = -\beta\frac{\mathrm{d}h}{\mathrm{d}t} \tag{7.298}$$

Substituting the kinematic altitude relation

$$\frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma \tag{7.299}$$

in Eq. (7.298) yields

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\beta V\sin\gamma \tag{7.300}$$

The above equation combined with Eq. (7.297) results in

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{1}{2c_2}\beta V^2\sin\gamma \tag{7.301}$$

and thus

$$D_0 = -mg\sin\gamma - m\frac{1}{2c_2}\beta V^2\sin\gamma \tag{7.302}$$

Note that a commonly used value for c_2 is 3, which changes Eq. (7.302) into

$$D_0 = -mg\sin\gamma - \beta\frac{mV^2}{6}\sin\gamma \tag{7.303}$$

Finally:

$$D_{cmd} = D_0 + \Delta D_{cmd} \tag{7.304}$$

By definition of the drag ($D = C_D \frac{1}{2}\rho V^2 S$) and assuming a drag-coefficient, C_D, dependency on the Mach number and angle of attack the commanded drag easily translates into a commanded angle of attack, α_{cmd}, which is the actual guidance command:

$$D_{cmd} = C_{D,cmd} \frac{1}{2}\rho V^2 S \Rightarrow C_{D,cmd} = C_{D,cmd}(\alpha_{cmd}, M) = \frac{D_{cmd}}{\frac{1}{2}\rho V^2 S} \tag{7.305}$$

Usually the (theoretical) drag coefficient is stored in the on-board computer as a polynomial or a multi-dimensional data table. By an inverse procedure, α_{cmd} is retrieved for a given flight condition (i.e., combination of h and V). Starting with the initial conditions provided in the introduction, the entry is started with an angle of attack of $\alpha = 40°$, corresponding with the maximum lift configuration. From literature it is known that this will result in the lowest maximum heat flux. The moment the maximum heat flux is encountered the applied guidance algorithm checks whether the set point of the heat flux to be tracked is reached. Once this is the case, the above described angle-of-attack modulation is applied to track this set point.

To prevent skipping flight that will remove kinetic energy from the system that could otherwise be used to prolong the flight, bank-angle modulation is applied to fulfill the equilibrium-glide condition. This condition assumes a (small) flight-path angle that remains constant. The simplified equation of motion that describes this motion is given by:

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) = 0 \tag{7.306}$$

To be able to manipulate this motion, as a control force we have the lift force, L. Of course, L has a similar dependency on M and α as the drag force, but we cannot use the angle-of-attack dependency since we are already using this to regulate the heat flux. If we realise, however, that in fact the *vertical* component of the lift determines the flight-path angle we can rewrite Eq. (7.306) to

$$L\cos\sigma_{cmd} - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) = 0 \tag{7.307}$$

or

$$\cos\sigma_{cmd} = \frac{mg}{L}\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{7.308}$$

For a given flight condition and α_{cmd} we can evaluate L and then the complete right-hand side of Eq. (7.308). By subsequently taking the arccos we get σ_{cmd}. As a consequence of this guidance approach, the occurring *g-load* is quite low, and very well acceptable.

7.6.4.3 Results

The objective is, as mentioned in the introduction, to test TPS systems with an input flux of 300 kW/m^2, which is about the limit value for metallic TPS systems. However, the location of the TPS test sample is not the stagnation point (the nose), but somewhere downstream. As a rule of thumb we can state that the stagnation heat flux is about 2.5 times higher in the stagnation point. This means that the heat-flux set point is $\dot{q}_c = 750$ kW/m^2. It is also important to note that the selected heat-flux model will influence the outcome of the simulations. Commonly, a cold-wall model is used, but this gives an overestimate of the heat flux of about 20%. This means that the input for TPS design is conservative and will lead to a heavier thermal-protection system. Therefore, we will assume radiation equilibrium, which means a hot radiating wall, with an emissivity of the applied material of $\varepsilon = 0.85$. The nose radii are $R_N = 0.20447$ m and 0.15426 m for the two *Hyperion* configurations.

In Figs. 7.88 and 7.89, the results are shown for *Hyperion-1a*. Two observations:

1. the maximum heat flux is 1.86 MW/m^2, which can only be lowered by either changing the entry conditions (shallower entry, or lower velocity) or by increasing the lift (e.g., by adding a body flap that can also be used for trim),
2. the duration for a heat flux larger than 300 kW/m^2 (stagnation heat flux 750 kW/m^2) is about 120 s.

Similarly, the results for *Hyperion-1b* are shown in Figs. 7.90 and 7.91, from which we conclude:

1. the maximum heat flux is 2.25 MW/m^2 (the smaller nose radius increases the stagnation heat flux).
2. the duration for a heat flux larger than 300 kW/m^2 (stagnation heat flux 750 kW/m^2) is about 160 s. Note that by using bank angle modulation towards the end of the flight, the duration can be extended, because of the non-zero lift at $\alpha = 0°$. Alternatively, the drag will be a bit lower at a negative angle of attack, which decreases the minimum-drag configuration. This will also extend the flight time at the heat-flux set point.

In the remainder of this chapter, a brief sensitivity analysis is executed to see the influence on the results. Three additional simulations are done for concept 1b:

1. Nominal entry conditions, but now with a cold-wall heat flux model,
2. Cold-wall, but with adjusted flight-path angle at entry: $\gamma_E = -0.5°$,
3. Cold wall, shallow entry, but with adjusted entry velocity: $V_E = 6.8$ km/s.

Note: a fourth simulation could be the decrease of mass to 250 kg, which will reduce the peak heating.

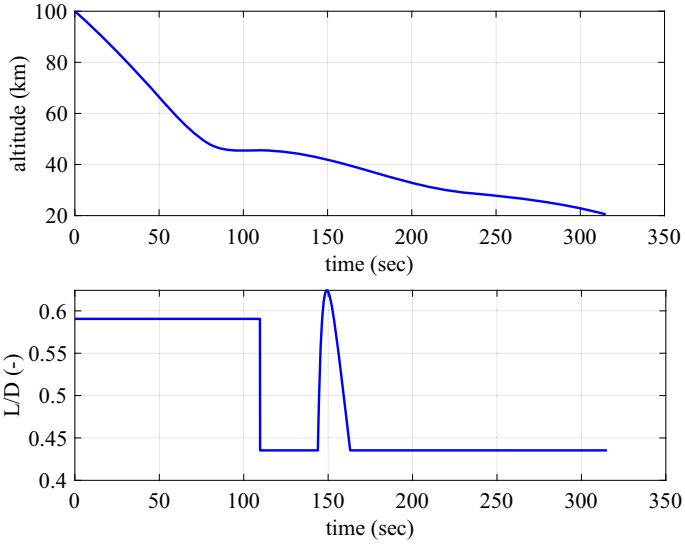


Fig. 7.88 *Hyperion-1a*: nominal entry, mass = 400 kg, radiation equilibrium, $R_N = 0.20447$ m, $V_E = 6$ km/s and $\gamma_e = -5.5°$

It can be seen in Figs. 7.92 and 7.93 how each of the changes affect the heat-flux profile. Even just using the (principally wrong) cold-wall heat-flux model increases the heat-flux set-point duration with almost 70 s. A shallower entry and a larger entry velocity will also increase the duration. Moreover, the shallow entry will decrease the maximum heat flux, whereas the larger entry velocity will increase it again.

To see the effect of adding a body flap, the following simulation has been executed: assuming a body flap, mounted at the back of the vehicle (*Hyperion-1b*), flush with the bottom, the lift and drag coefficient will scale proportional to the increased projected surface area of the bottom of the vehicle. Assuming that this bottom surface can be approximated by a triangle with base 1.0 m and height 1.5 m, the surface area is $S = 0.75$ m^2. Defining a body flap with a width of 0.8 m and a chord length of 0.2 m, 0.25 m, and 0.3 m, the additional surface area becomes $\Delta S_1 = 0.16$ m^2, $\Delta S_2 = 0.20$ m^2, and $\Delta S_3 = 0.24$ m^2. This will increase the aerodynamic coefficients by $\Delta C_D = 0.2133 \sin\alpha$ and $\Delta C_L = 0.2133 \cos\alpha$, a change in lift of 21.3% for $\alpha = 0°$. The second and third flap configuration change by 26.7% and 32%, respectively. An overview of the flap configuration is shown in Fig. 7.94. Results of the three simulations, compared with the nominal heat flux are shown in Figs. 7.95 and 7.96. The flight time increases, not only for the platform but also the total flight time. This means that with additional bank modulation the heat-flux platform can be extended even more. Due to the increased lift the peak heat flux and *g-load* decrease. It is interesting to note that for flap configuration 3, a second skip-control manoeuvre has to be executed. Apparently, the additional lift is too large for equilibrium flight.

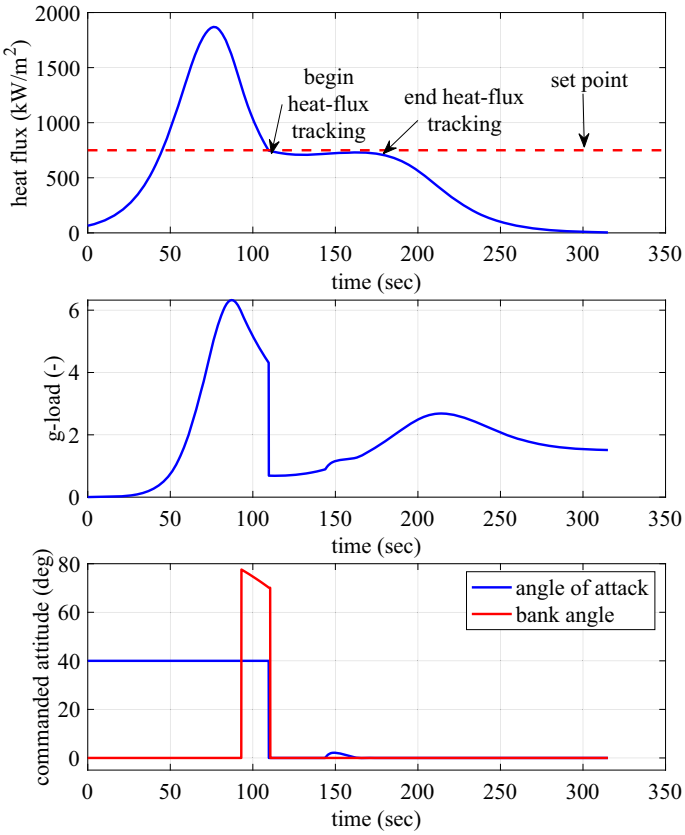


Fig. 7.89 *Hyperion-1a*: nominal entry, mass = 400 kg, radiation equilibrium, $R_N = 0.20447$ m, $V_E = 6$ km/s and $\gamma_e = -5.5°$—concluded

Remark: the results can be further improved by changing the initial conditions. The impact of this has already been presented for the nominal configuration. Future optimisation of the results may aim at:

1. Flying the trajectory with an optimiser,
2. Locally adjusting the nose radius to reduce the peak heating (e.g., conform the EXPERT and PRE-X re-entry test vehicles),
3. Verification of the aero-thermodynamics results with CFD analysis, and
4. Iteration of the trajectory simulation with updated aerodynamic coefficients (from CFD).

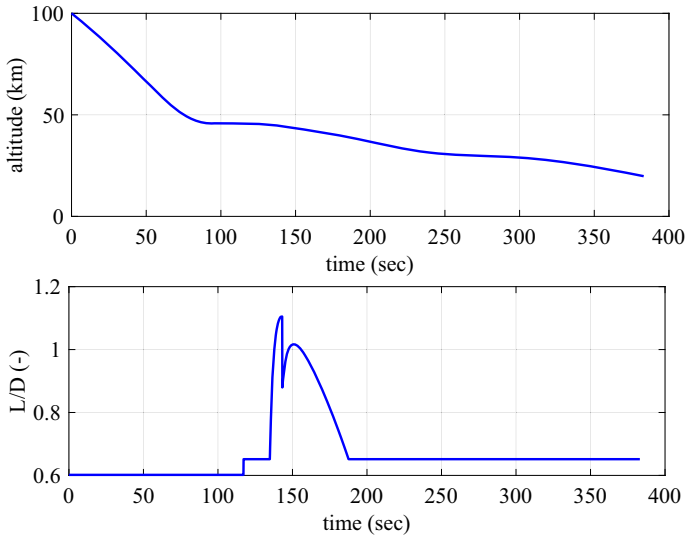


Fig. 7.90 *Hyperion-1b*: nominal entry, mass = 400 kg, radiation equilibrium, $R_N = 0.15426$ m, $V_E = 6$ km/s and $\gamma_e = -5.5°$

7.6.4.4 Concluding Remarks

For the two adapted *Hyperion-1* configurations it can be stated that the default Volna separation conditions are not suitable to allow for a 600-s time frame of TPS testing. By allowing for a shallower entry and a larger entry velocity, the duration of the test period can be significantly increased. However, the maximum stagnation heat flux is such that particular attention should be paid to the initial flight phase and a possible reconfiguration of the nose. One solution maybe to increase the lift at high angle of attack, and a second solution could be to adjust (i.e., increase) the nose radius in the stagnation point at the high entry angle of attack.

It is advised to optimise the trajectory of the proposed *Hyperion-1* configurations before the launcher conditions and vehicle configuration are reviewed and possibly changed.

The addition of a flap, which increases the lift coefficient by about 20–30%, decreases the maximum heat flux, but before going into great detail it is better to first review the applied engineering heat-flux model, which is on the conservative side. Rather than changing the vehicle it is perhaps a better option to change the trajectory, i.e., lower the flight-path angle for a shallower entry. Note that a skipping flight is not an option to decrease the maximum heat flux, since the skip occurs after passing the peak value. Therefore, a review of the Volna separation conditions is required before optimising the vehicle shape.

Furthermore, the following aspects should be studied in more detail:

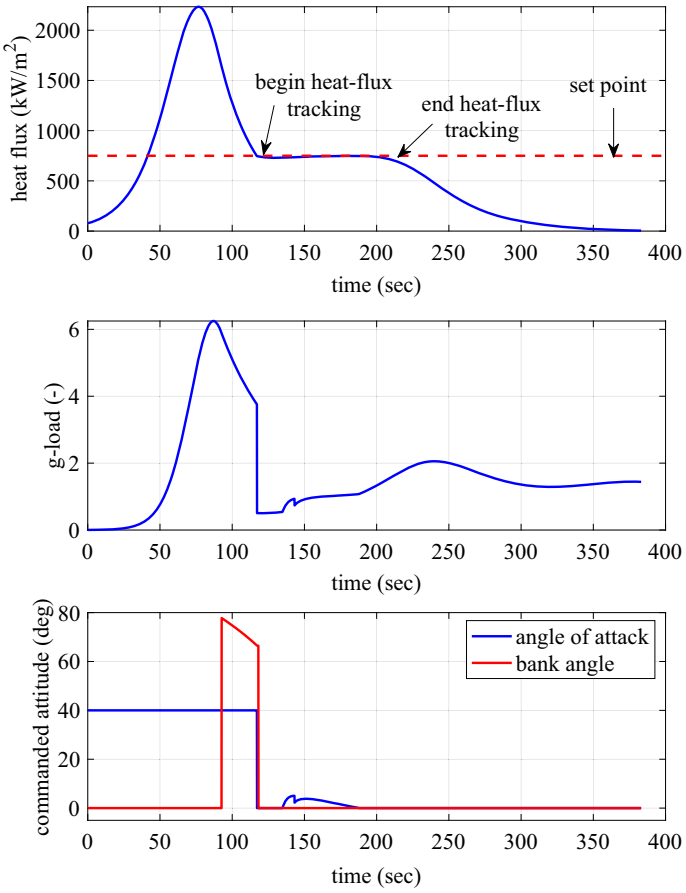


Fig. 7.91 *Hyperion-1b*: nominal entry, mass = 400 kg, radiation equilibrium, $R_N = 0.15426$ m, $V_E = 6$ km/s and $\gamma_e = -5.5°$—concluded

1. The heat-flux model applied in this study is still conservative. It should be reviewed by comparison with CFD results and catalycity effects should be added in the engineering model.
2. The applied aerodynamic database should be verified and tuned with CFD results; the updated database should be used to re-generate the trajectory, and iteratively the CFD results should be updated.
3. To lower the drag without affecting the lift, the vehicle design should be flattened on the top part. Of course, this should be done taking the volume requirements into account.
4. At zero angle of attack, the vehicle has a non-zero lift coefficient. To further extend the constant heat-flux platform a bank manoeuvre could be executed.

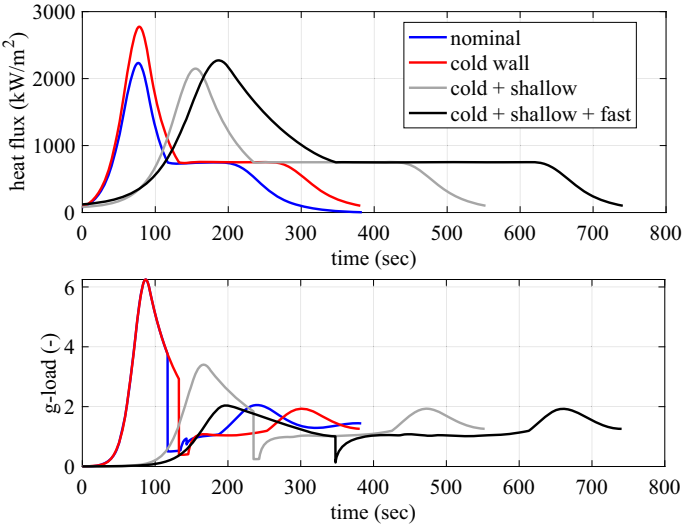


Fig. 7.92 *Hyperion-1b*, with a mass of m = 400 kg and a nose radius of $R_N = 0.15426$ m; influence of heat-flux model and entry conditions. Shown here are the heat flux and *g-load* versus time

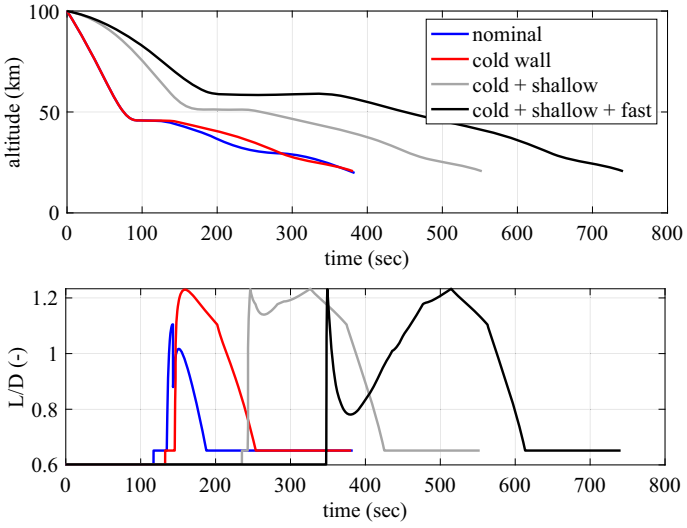


Fig. 7.93 *Hyperion-1b*, with a mass of m = 400 kg and a nose radius of $R_N = 0.15426$ m; influence of heat-flux model and entry conditions. Shown are the altitude and L/D versus time

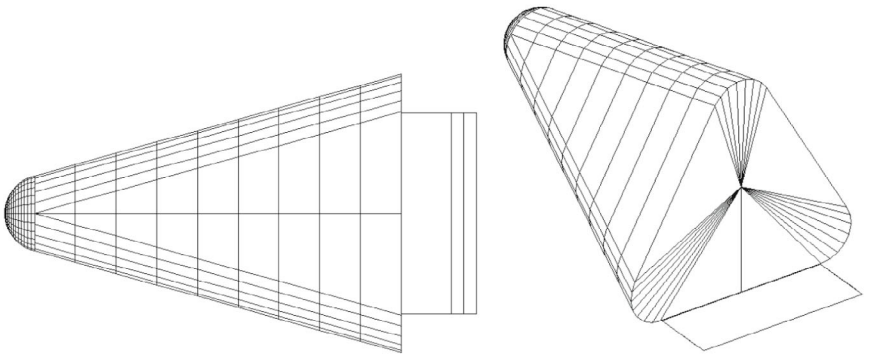


Fig. 7.94 Bottom view *Hyperion-1b* with three flap configurations: chord length $c = 0.2, 0.25,$ and 0.3 m (left), and dimetric view with largest flap configuration (right)

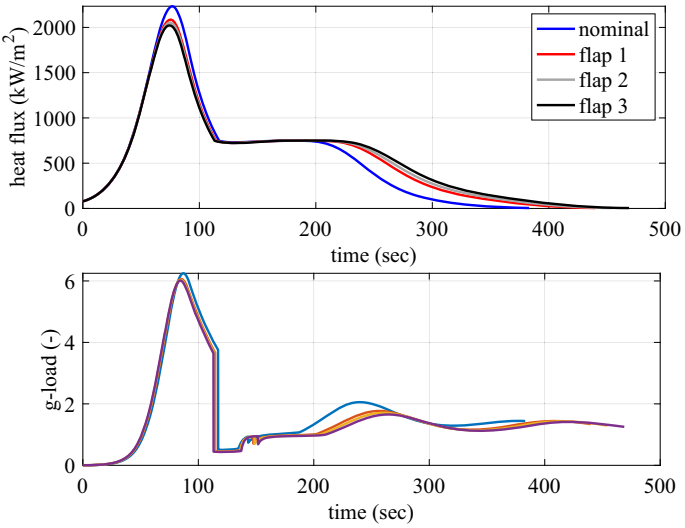


Fig. 7.95 *Hyperion-1b* with three flap geometries; m = 400 kg and $R_N = 0.15426$ m; heat flux and *g-load*

However, it is also possible to first decrease the angle of attack to the minimum drag configuration and after reaching that configuration the aforementioned bank manoeuvre can be executed (of course after increasing the angle of attack again to have sufficient lift).

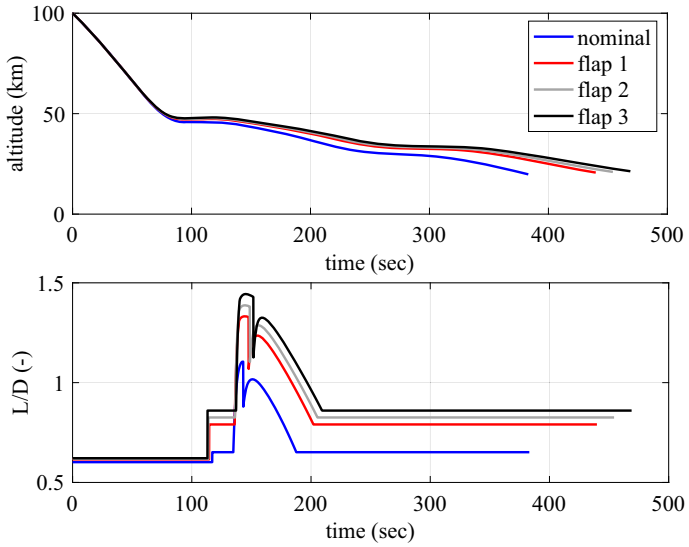


Fig. 7.96 *Hyperion-1b* with three flap geometries; m = 400 kg and $R_N = 0.15426$ m; altitude and L/D

Problems

Judge whether the following statements are TRUE or FALSE (Problems 7.1–7.15).

7.1. In an accelerometer, the bias error increases with increasing accelerations.
7.2. A single bank reversal may be enough to reach the terminal area with the right flight conditions.
7.3. Between accelerometers and gyroscopes, only accelerometers have (constant) bias errors.
7.4. The navigation system can estimate more than just the state of the vehicle.
7.5. Decreasing the integral gain of a PID controller will lower the maximum overshoot.
7.6. Hypersonic vehicles primarily rely on ground-based command and control centres for real-time guidance and control during flight.
7.7. Navigation systems of entry vehicles often incorporate redundant sensors and fusion algorithms to enhance reliability and accuracy, especially during critical phases of flight.
7.8. Entry vehicles typically rely on precomputed trajectories for guidance, with minimal real-time adjustments.
7.9. Hypersonic vehicles rely solely on passive stabilisation systems to maintain stability during flight.
7.10. Adaptive control algorithms can be used in hypersonic vehicles to adjust control strategies in real-time based on changing flight conditions.

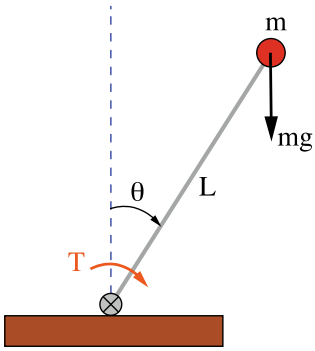


Fig. 7.97 The inverted pendulum

7.11. The elevator is the primary control surface to execute a bank reversal.
7.12. Between guidance and control, it is the guidance system that commands the actuators.
7.13. The key parameter that drives the lateral guidance of a gliding vehicle is the heading error.
7.14. Hypersonic vehicles can benefit from Reinforcement Learning (RL, a subclass of machine learning) algorithms to autonomously learn control policies and adapt to uncertain environments.
7.15. To determine the 3D position from GPS signals only three GPS satellites are required.

Navigation system: IMU measurement errors (Problems 7.16–7.17).

An entry vehicle has a constant acceleration of $\mathbf{a} = (-11.8, 1, 4.3)^T$ m/s^2; its navigation sensor, an IMU, measures this acceleration to calculate the velocity and position. However, the IMU has several (constant) instrument errors, i.e., a bias of 0.0025 m/s^2 for the X- and Z-axis, and -0.018 m/s^2 for the Y-axis, scale factors of $(0.02, -0.01, 0.04)^T$ for the X-, Y- and Z-axis, and a general mounting misalignment for all axes of $1.3 \cdot 10^{-1}$.

7.16. State the measurement equation.
7.17. Calculate the errors in acceleration, velocity, and position after 65 s along the three axes.

Control-system design—Inverted pendulum (Problems 7.18–7.25).

This set of problems deals with a classic control problem, i.e., that of the inverted pendulum (Fig. 7.97 shows a simplified version of this problem). A rigid bar of length L, with a mass m on top, is mounted vertically on a flat surface. Gravity is the only force acting on m, and enforces a downward motion on the mass. This motion can be countered with a torque, T, acting on the hinge connecting the bar to the surface. The deviation from the local vertical is given by the angle θ. The nominal pendulum configuration is $L = 1.5$ m, and $m = 1.2$ kg. The gravitational acceleration is $g = 9.81$ m/s^2.

The second-order differential equation describing the motion is given by:

$$mL^2\ddot{\theta}(t) = mgL\sin\theta(t) + T(t) \tag{7.309}$$

You are tasked with designing a PID controller to keep the inverted pendulum in an upright position. It is advised to create a small computer program that allows you to design this controller, and analyse its performance. Hint: a simple Euler integrator is sufficient to gain understanding of this problem.

7.18. Convert Eq. (7.309) into two first-order differential equations, linearise those by assuming small deviations θ, and write the answer in state-space form.
7.19. Determine the eigenvalues of the (open-loop) system. What can you conclude? Refer to Chap. 6, Sect. 6.2 for reference.
7.20. As input to the design process, which criteria could you use to select the gains of your controller?
7.21. Design a proportional controller, taking the error on θ as input. What is your conclusion?
7.22. Add a derivative component to your controller, and analyse the performance.
7.23. Is it necessary to add an integral component to your controller? Motivate your answer.
7.24. So far, the control torque was unconstrained. What happens when $|T_{min}| = T_{max} = 5$ Nm? What if this reduced to 1 Nm? Show the results.
7.25. Finally, what can you say about the stability and controllability of the system when either L or m is changed?

(*) *Optimal control-system design—bank-reversal control (Problems 7.26–7.29).*

Having established the basics of control-system design, we will now consider the problem of controlling the bank reversal of a winged re-entry vehicle. It is assumed that the lateral motion of the vehicle is decoupled from the longitudinal one. The error dynamics of the lateral system can be described by the following linear system (see also Sect. 7.5.7, describing the design of an optimal controller for HORUS)[15]:

$$\begin{pmatrix} \Delta\dot{p} \\ \Delta\dot{r} \\ \Delta\dot{\beta} \\ \Delta\dot{\sigma} \end{pmatrix} = \mathbf{A}\begin{pmatrix} \Delta p \\ \Delta r \\ \Delta\beta \\ \Delta\sigma \end{pmatrix} + \mathbf{B}\begin{pmatrix} \Delta\delta_a \\ \Delta\delta_r \\ \Delta M_{T,z} \end{pmatrix} \tag{7.310}$$

with

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & -3.5453 & 0 \\ 0 & 0 & -0.2809 & 0 \\ 0.6428 & -0.7660 & -5.83\cdot 10^{-4} & -1.03\cdot 10^{-3} \\ -0.7660 & -0.6428 & -1.21\cdot 10^{-3} & 0 \end{bmatrix} \tag{7.311}$$

[15] This series of questions deals with designing a Linear Quadratic Regulator (LQR) for lateral control. To do so, one could use Matlab® with the command [K, ,] = lqr(A,B,Q,R); to calculate the gain matrix **K**. To use another programming language, one may have to search for the appropriate toolbox/library that provides an equivalent to the lqr command.

and

$$\mathbf{B} = \begin{bmatrix} 2.8710 & 0 & 0 \\ -0.2162 & -0.0337 & 1.2407 \cdot 10^{-6} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{7.312}$$

The lateral state is defined by roll and yaw rate (p and r), and angle of sideslip and bank angle (β and σ). The controls consist of the aileron and rudder deflection, δ_a and δ_r, and the yaw moment, $M_{T,z}$.

7.26. Design a baseline controller for lateral motion, for a step command of $\sigma_c = 1°$. Evaluate the performance.

7.27. Repeat this for $\sigma_c = 5°$ and $\sigma_c = 10°$. Draw conclusions. In case you have to change the controller, designed under Problem 7.26, explain why.

7.28. Without changing the controller as designed under Problem 7.26, initialise your system with $\Delta\beta = 5°$, whereas the other states are zero. Run the sim ulation, and discuss what you find.

7.29. With a ramp function, simulate a bank reversal from $-65°$ to $+65°$, with a commanded bank rate of 10°/s. Use your baseline lateral controller from the previous question. Analyse the performance and draw conclusions. If you do not have a stable response, adapt your controller.

Guidance-system design—g-load tracking (Problems 7.30–7.34*).*

A flight through the entry corridor may encounter two load constraints, i.e., one for the thermal load, i.e., the heat-flux constraint, and one for the mechanical load, defined by the *g-load*. Flying along the heat-flux constraint as long as possible, will minimise the integrated heat load and thus the mass of the thermal-protection system. When that constraint can no longer be followed and the *g-load* constraint is encountered, it can be reasoned that to track that constraint will shorten the flight time, and will also contribute to lowering the integrated heat load. The current series of questions will deal with the development of a *g-load* tracking law, following similar steps as discussed in the case study covered in Sect. 7.6.4, where a heat-flux tracking law has been developed.

7.30. Starting from the expression for flight duration, Eq. (5.137):

$$t_{flight} = \frac{1}{2}\frac{V_c}{g}\frac{L}{D}\ln\left(\frac{1 + V_E/V_c}{1 - V_E/V_c}\right) \tag{7.313}$$

and the definition of *g-load*,

$$n_g = \frac{\sqrt{D^2 + L^2}}{mg_0} \tag{7.314}$$

motivate that tracking n_g will shorten the flight. The expression for t_{flight} has been derived under the assumption that the lift-to-drag ratio is constant. Extending this assumption, this also holds for C_D and C_L individually.

7.31. For *g-load* tracking, n_g should remain constant during a major portion of the flight, i.e., $\dot{n}_g = 0$. Starting from its definition, derive an expression for the nominal drag force, $D_{cmd,0}$, that fulfils this requirement. An exponential atmosphere can be assumed.
7.32. Given the definition of the *g-load* error, $n_{g,err} = n_g - n_{g,setpoint}$, define the three components of a PID controller.
7.33. As $D_{cmd,0}$ is only the nominal drag, define a PI-controller to compensate for perturbations in the *g-load*. Also, give the complete expression for the commanded drag, D_{cmd}.
7.34. Given that the drag coefficient is a function of angle of attack and Mach number, how does one obtain the commanded angle of attack, α_{cmd} from D_{cmd}?

References

Acikmese B, Ploen SR (2007) Convex programming approach to powered descent guidance for Mars landing. J Guid Control Dyn 30(5):1353–1366

ACRI/LAN (1992) Guidance and control for moderate lift/drag reentry - final report. ESA Contract no. 9359/91/NL/JG (SC), ACRI, Valbonne

Acquatella P, Falkena W, Van Kampen E-J, Chu QP (2012) Robust nonlinear spacecraft attitude control using incremental nonlinear dynamic inversion. AIAA 2012-4623, In: AIAA guidance, navigation, and control conference, Minneapolis, Minnesota, 13–16 Aug

AIAA (2017) Special issue on computational guidance and control. J Guid Control Dyn 40(2)

Amzajerdian F, Roback VE, Brewster PF, Hines GD, Bulyshev A (2016) Imaging flash Lidar for autonomous safe landing and spacecraft proximity operation. AIAA-2016-5591, In: AIAA space 2016, Long Beach, CA, 13–16 Sept

Atkins BM, Queen EM (2015) Internal moving mass actuator control for Mars entry guidance. J Spacecr Rock 52(5):1294–1310

Balaram J (2005) Sherpa moving mass entry descent landing system. DETC2005-85283; In: Proceedings of IDETC/CIE 2005 ASME 2005 international design engineering technical conferences computers and information engineering conference, Long Beach, CA, 24–28 Sep

Barkana I, Kaufman H (1993) Simple adaptive control of large flexible space structures. IEEE Aerosp Electr Syst Mag 29(4):1137–1149

Barkana I, Guez A (1990) Simple adaptive control for a class of non-linear systems with application to robotics. Int J Control 52(1)

Bayle GP (1984) Space Shuttle entry flight control off-nominal design considerations. J Guid Control Dyn 7(1):9–14

Betts JT (2010) Practical methods for optimal control and estimation using nonlinear programming, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia

Blackmore L (2016) Autonomous precision landing of space rockets. Bridge: Link Eng Soc 46(4):15–20, National Academy of Engineering

Bollino K, Ross M, Doman D (2006) Optimal nonlinear feedback guidance for re-entry vehicles. AIAA-2006-6074, In: AIAA guidance, navigation, and control conference and exhibit, Keystone, CO, 21–24 Aug

Boskovich B, Kaufmann RE (1966) Evolution of the Honeywell first-generation adaptive autopilot and its applications to F-94, F-101, X-15, and X-20 vehicles. J Aircraft 3(4):296–304

Braun RD, Powell RW (1992) Predictor-corrector guidance algorithm for use in high-energy aerobraking system studies. J Guid Control Dyn 15(3):672–678

Bryson AE Jr (1985) New concepts in control theory, 1959–1984. J Guid Navig Control 8(4):417–425

Bryson AE Jr, Ho YC (1975) Applied optimal control. Wiley

Byrne RH, Robinett RD, Sturgis BR (1996) Moving mass trim control system design. AIAA-1996-3826-CP

Chu QP, Van Woerkom PThLM (1997) GPS for low cost attitude determination: a review of concepts, in-flight experiences and current development. Acta Astronautica 41(4–10):421–433

Cobleigh BR, Whitmore SA, Haering Jr EA, Borrer J, Roback VE (1999) Flush Airdata Sensing (FADS) system calibration procedures and results for blunt forebodies. In: NASA/TP-1999-209012

Conway BA (ed) (2010) Spacecraft trajectory optimization. Cambridge University Press, Cambridge, UK

Cooke DR (1982) Space Shuttle stability and control test plan. AIAA-1982-1315, In: From: AIAA 9th atmospheric flight mechanics conference, San Diego, CA, 9–11 Aug

Dominguez Calabuig GJ, Mooij E (2021) Optimum on-board abort guidance based on successive convexification for atmospheric re-entry. In: SciTech forum - AIAA guidance, navigation, and control conference. On-Line Event

Doyle J, Lenz K, Packard A (1986) Design examples using μ-synthesis: Space Shuttle lateral axis FCS during reentry. In: Proceedings of the 25th conference on decision and control, pp 2218–2223

D'Souza AF (1988) Design of control systems. Prentice Hall Inc, Englewood Cliffs, New Jersey

D'Souza SN, Sarigul-Klijn N (2014) Survey of planetary entry guidance algorithms. Progr Aerosp Sci 68:64–74

Epple RGE, Altenbach RE (1983) Dynamic stability testing of the orbiter flight control system/flexible body interaction. J Guid Control Dyn 6(6):456–460

Farrell JA, Barth M (1999) The global positioning system & inertial navigation. McGraw-Hill, New York

Frangos C, Yavin Y (1992) Design methodology for linear optimal control systems. J Guid Control Dyn 15(5):1302–1304

Gonzalez Puerta A, Mooij E, Yabar Vallesz C (2018) Reachability analysis to design zero-wait entry guidance. AIAA-2018-1316, In: SciTech forum - AIAA guidance, navigation, and control conference, Kissimmee, FL, pp 8–12

Goodman JL (2003) The Space Shuttle and GPS - a safety-critical navigation upgrade. In: Proceedings of the 2nd international conference on COTS-based software systems, Ottawa, Canada, 10–12 Feb

Gopal M (1989) Modern control system theory. Wiley Eastern Ltd., New Delhi, Second Reprint

Grewal MS, Weill LR, Andrews AP (2001) Global positioning systems, inertial navigation, and integration. Wiley

Groves PD (2008) Principles of GNSS, inertial, and multisensor integrated navigation systems. Artech House

Hamilton PN (1982) Flexible body stabilization for Space Shuttle aerosurface control loops. AIAA-1982-1532, In: From: AIAA guidance and control conference, San Diego, CA, 9–11 Aug

Han P, Shan J, Meng X (2012) Re-entry trajectory optimization using an hp-adaptive radau pseudospectral method. Proc Inst Mech Eng Part G: J Aerosp Eng 227(10):1623–1636

Harpold JC, Gavert DE (1983) Space Shuttle entry guidance performance results. J Guid Navig Control 6(6):442–447

Harpold JC, Graves CA (1979) Shuttle entry guidance. AAS-78-147, In: From: Advances in the astronautical sciences, vol 38, Part I: the future United States space program, pp 99–132

Holmberg NA, Faust RP, Holt HM (1980) Viking 75 spacecraft design and test summary. Volume I - lander design, NASA Reference Publication 1027

Jänsch C, Markl A (1991) Trajectory optimization and guidance for a Hermes-type reentry vehicle. AIAA-1991-2659, In: From: AIAA guidance, navigation, and control conference, New Orleans, LA, 12–14 Aug

Jenkins DR (2008) Space Shuttle. The history of the national space transportation system. The first 100 missions, 3rd edn, 5th printing
Jet Propulsion Laboratory (1966) Surveyor I mission report. Part I - mission description and performance. NASA CR-77795/JPL Technical Report No. 32-1023, 31 Aug
Jet Propulsion Laboratory (2011) Mars Science Laboratory. http://mars.jpl.nasa.gov/msl/. Accessed 21 Nov 2011
Jorris TR (2007) Common aero vehicle autonomous re-entry trajectory optimization satisfying waypoint and no-fly zone constraints. PhD thesis, Air Force Institute Of Technology, Wright-Patterson Air Force Base, Ohio
Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME (Series D) J Basic Eng 82:35–45
Kaufman H, Barkana I, Sobel K (1998) Direct adaptive control algorithms: theory and applications, 2nd edn. Springer, New York
Klinar WJ, Saldana RL, Kubiak ET, Smith Jr EE, Peters WH, Stegall HW (1975) Space Shuttle flight control system, vol 4, paper 6.2. In: Proceedings of the IFAC 6th World Congress, Boston/Cambridge, MS, 24–30 Aug
Kokotovic PV (1985) Recent trends in feedback design: an overview. Automatica 21(3):225–236
Kuo BC (1987) Automatic control systems. Prentice Hall Inc, Englewood Cliffs, New Jersey
Leavitt JA, Mease KD (2007) Feasible trajectory generation for atmospheric entry guidance. J Guid Control Dyn 20(2):473–481
Lewis FL, Vrabie DL, Syrmos V (2012) Optimal control, 3rd edn. Wiley
Liu X, Shen Z, Lu P (2016) Entry trajectory optimization by second-order cone programming. J Guid Control Dyn 39(2):227–241
Lorenz RD (2006) Spin of planetary probes in atmospheric flight. J Br Interplanet Soc 59(8):273–282
Lu P (2014) Entry guidance: a unified method. J Guid Control Dyn 37(3):713–728
Luo J, Lan CE (1995) Determination of weighting matrices of a linear quadratic regulator. J Guid Navig Control 18(6):1462–1463
Martins JRRA, Kroo IM, Alonso JJ (2000) An automated method for sensitivity analysis using complex variables, AIAA-2000-689. In: 38th aerospace sciences meeting, Reno, NV, 10–13 Jan
MBB Space Communication and Propulsion Systems Division (1988) Study on re-entry guidance and control. Final report, ESA report reference: ESA CR (P) 2652, MBB, Munich
McDermott AM, Makowski K (1982) Space Shuttle linearized guidance, navigation, and control stability analysis, AIAA-1982-1555. In: AIAA guidance and control conference, San Diego, CA, 9–11 Aug
McHenry RL, Brand TJ, Long AD, Cockrell BF, Thibodeau JR III (1979) Space Shuttle ascent guidance, navigation, and control. J Astronaut Sci XXVI I(1):1–38
Mease KD, Kremer JP (1994) Shuttle entry guidance revisited using nonlinear geometric methods. J Guid Control Dyn 17(6):1350–1356
Mehiel EA, Balas MJ (2004) Adaptive control for a deployable optical telescope, AIAA-2004-5222. In: Proceedings AIAA guidance, navigation, and control conference, Providence, RI, 16–19 Aug
Mendeck GF, Craig LE (2011) Entry guidance for the 2011 Mars Science Laboratory mission, AIAA 2011-6639. In: AIAA atmospheric flight mechanics conference, Portland, Oregon, 08–11 Aug
Misra P, Enge P (2006) Global positioning system signals, measurements, and performance. Ganga-Jamuna Press
Mooij E (2001) Numerical investigation of model reference adaptive control for hypersonic aircraft. J Guid Control Dyn 24(2):315–323
Mooij E (2014) Passivity analysis for non-linear, non-stationary entry capsules. Int J Adapt Control Signal Proc 28:708–731
Mooij E (2018) Simple adaptive bank-reversal control for non-linear winged re-entry vehicles. Math Eng Sc Aerosp 9(1):85–110

Mooij E, Maree AGM, Sudmeijer KJ (1995) Aerodynamic controllability of a selected re-entry test vehicle, IAF-95-V.4.04. In: 46th International astronautical congress, Oslo, Norway, 2–6 Oct
Mooij E, Chu QP (2001) IMU/GPS integrated navigation system for a winged re-entry vehicle, AIAA paper 2001-1469. In: AIAA guidance, navigation, and control conference, Montreal, Canada, 6–9 Aug
Mooij E, Chu QP (2002) Tightly-coupled IMU/GPS re-entry navigation system, AIAA paper 2002-5005. In: AIAA guidance, navigation, and control conference, Monterey, CA, 5–8 Aug
Mulder JA, Chu QP, Sridha JK, Breeman JH, Laban M (1999) Non-linear aircraft flight path reconstruction review and new advances. Prog Aerosp Sci Elsevier 35:673–726
Myers TT, Johnston DE, McRuer DT (1987) Space Shuttle flying qualities and criteria assessment. NASA contractor Report 4049. NASA, Washington, D.C
NASA (2011) Phoenix Mars lander. http://www.nasa.gov/mission_pages/phoenix/main/index.html. Accessed 16 Dec 2011
Nguyen VH, Hishimi JT, Payne TH, Woosley EW (1990) Space Shuttle descent flight verification by simulation: a challenge in implementing flight control system robustness, AGARD-CP-489. In: Space vehicle flight mechanics conference (paper 9). Neuilly sur Seine, France
NOAA National Geodetic Survey (2011) GPS orbit data. http://www.ngs.noaa.gov/orbits/orbit_data.shtml. Accessed 16 Dec 2011
Ogata K (2010) Modern control engineering, 5th edn. Prentice Hall
Paraskevopoulos PN (2002) Modern control engineering. Marcel Dekker, Inc
Petsopoulos T, Regan FJ, Barlow J (1996) Moving-mass roll control system for fixed-trim re-entry vehicle. J Spacecr Rockets 33(1):54–60
Pontryagin LS, Boltyanskii VG, Gramkreledze QV, Mishchenko EF (1962) The mathematical theory of optimal processes. Intersciences, New York
Powell RW, Braun RD (1992) A six-degree-of-freedom guidance and control analysis of Mars aerocapture, AIAA-1992-0736. In: 30th aerospace sciences meeting & exhibit, Reno, NV, 6–9 Jan
Powell RW, Shaugnessy JD, Cruz CI, Naftel JC (1991) Ascent performance of an air-breathing horizontal-takeoff launch vehicle. J Guid Control Dyn 14(4):834–839
Rao A, Clarke K (2002) Performance optimization of a maneuvering re-entry vehicle using a legendre pseudospectral method, AIAA-2002-4885. In: AIAA atmospheric flight mechanics conference and exhibit, Monterey, CA, 5–8 Aug
Roenneke AJ, Markl A (1994) Re-entry control to a drag-vs-energy profile. J Guid Control Dyn 17(5):916–920
Rosello AD (1995) A vehicle health monitoring system for the Space Shuttle reaction control system during reentry. MSc Thesis MIT; also: NASA-CR-188370
Sagliano M (2016) Development of a novel algorithm for high performance reentry guidance. PhD thesis, University of Bremen
Sagliano M, Mooij E (2018) Optimal drag-energy entry guidance via pseudospectral convex optimization, AIAA-2018-1315. In: SciTech Forum - AIAA guidance, navigation, and control Conference, Kissimmee, FL, 8–12 Jan
Sagliano M (2019) Generalized hp pseudospectral convex programming for powered descent and landing. J Guid Control Dyn 42(7):1562–1570
San Martin M (2010) The MSL SkyCrane landing architecture - a GN&C perspective. In: International planetary probe conference, Barcelona, Spain, June 14–18
Saraf A, Leavitt J, Ferch M, Mease KD (2004) Landing footprint computation for entry vehicles, AIAA-2004-4774. In: AIAA guidance, navigation, and control conference and exhibit, Providence, Rhode Island, 16–19 Aug
Scharf DP, Acikmese B, Dueri D, Benito J, Casoliva J (2017) Implementation and experimental demonstration of onboard powered-descent guidance. J Guid Control Dyn 40(2):213–229
Schaub H, Junkins JL (2014) Analytical mechanics of space systems, Sect. 6.2. AIAA Education Series

Schierman JD, Ward DG, Hull JR, Gandhi N, Oppenheimer MW, Doman DR (2004) Integrated adaptive guidance and control for re-entry vehicles with flight-test results. J Guid Control Dyn 27(6):975–988

Schleich WT (1982) The Space Shuttle ascent guidance and control, AIAA-1982-1497. In: AIAA guidance and control conference, San Diego, CA, 9–11 Aug

Schletz B (1982) Use of quaternions in Shuttle guidance, navigation, and control, AIAA-1982-1557. In: AIAA guidance and control conference, San Diego, CA, 9–11 Aug

Schöttle UM, Burkhardt J (1995) Entwurf einer kleinen semiballistischen Rückkehrkapsel. Space Course, Lehrgang über Raumtransportsysteme, Stuttgart

Serrano-Martínez JB, Parra JA (1987) Hermes hypervelocity re-entry trajectory study. ESA contract no. 6932/86/F/CG, Final report GMVSA 212/87, Grupo de Mecanica del Vuelo, Madrid

Shahian B, Hassul M (1993) Control system design using MATLAB®. Prentice Hall, Englewood Cliffs, New Jersey

Shen Z, Lu P (2003) On-board generation of three-dimensional constrained entry trajectories. J Guid Control Dyn 26(1):111–121

Sieberling S, Chu QP, Mulder JA (2010) Robust flight control using incremental nonlinear dynamic inversion and angular acceleration prediction. J Guid Control Dyn 33(6):1732–1742

Simplicio P, Marcos A, Bennani S (2019) Guidance of reusable launchers: improving descent and landing performance. J Guid Control Dyn 42(10):2206–2219

Space-X (2011) SpaceX's dragoneye navigation sensor successfully demonstrated on Space Shuttle. http://www.spacex.com/press.php?page=20090925. Accessed 3 Dec 2011

Stone HW, Powell RW (1976) Space Shuttle orbiter entry guidance and control system sensitivity analysis, AIAA CP-76-03. In: AIAA 3rd atmospheric flight mechanics conference, Arlington, TX, June 7–9, pp 169–175

Wang J, Sundararajan N (1995) A nonlinear flight controller design for aircraft. Control Eng Pract Nr 3(6):813–825

Wang Z, Grant MJ (2017) Constrained trajectory optimization for planetary entry via sequential convex programming. J Guid Control Dyn 40(10):2603–2615

Welch G, Bishop G (2001) An introduction to the Kalman filter. In: SIGGRAPH 2001 Course 8. University of North Carolina at Chapel Hill

Wingrove RC (1963) Survey of atmosphere re-entry guidance and control methods. AIAA J 1(9):2019–2029

Zarchan P, Musoff H (2009) Fundamentals of Kalman filtering - a practical approach, 2nd edn. American Institute of Aeronautics and Astronautics

Zimmerman C, Dukeman G, Ganson J (2003) Automated method to compute orbital reentry trajectories with heating constraints. J Guid Control Dyn 26(4):523–529

Chapter 8
Terminal Area Energy Management



8.1 Introduction

A spacecraft in Earth orbit has a large potential and kinetic energy: for each kg in a low Earth orbit the total energy amounts to about 30 MJ. To guarantee a safe return to Earth, it has to reduce this large energy to the level of potential energy that the vehicle has after landing on the surface, and, what is even more constraining, with zero relative velocity. The only mechanism to dissipate the energy is by aerodynamic friction with the vehicle wall during its flight through the atmosphere. This energy dissipation has to take place in a time on the order of minutes in case of ballistic vehicles. For gliding vehicles, this time can be as long as 20 min, depending on the mission profile. Moreover, gliding vehicles can control the energy dissipation by actively changing the aerodynamic force acting on the vehicle, as part of the guidance problem. The energy-dissipation problem forms a fundamental part of the guidance of one of the last phases of atmospheric entry, notably the terminal area energy management (TAEM) phase.

Generally, it is assumed that the atmosphere starts at an altitude of 120 km, although the official definition for the beginning of space (or the boundary of the Earth's atmosphere) is an altitude of 100 km, the so-called Kármán line, as defined by the *Fédération Aéronautique Internationale* (FAI). When an entry vehicle returns to Earth, it will usually start its return flight with a deorbit-burn manoeuvre. After a coasting phase the vehicle re-enters into the atmosphere. A typical flight of a *gliding* entry vehicle consists of three atmospheric phases:

- **Hypersonic re-entry**
  The hypersonic atmospheric re-entry starts at an altitude of 120 km and continues down to an altitude of about 25 km. During this phase, the vehicle, with a typical blunt shape, flies at high angle of attack, and decelerates from an initial velocity of Mach 24 down to approximately Mach 2.5. This phase is characterised by a high heating rate and temperatures can be as high as 1500 °C. The highest temperatures

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_8

are reached on the nose of the vehicle, at the bottom of the vehicle, and on the leading edges of the wings. These are all known stagnation points of aerodynamic flow. Due to their much steeper entries ballistic vehicles will encounter even far higher heat loads.

- **Terminal area energy management**
  Following the hypersonic descent is the terminal-area energy-management phase. During this phase the vehicle is brought to the correct interface position with the subsequent approach and landing phase, which is on the extension of the centreline of the runway with the correct amount of energy (velocity and altitude), as well as the correct heading.
- **Approach and landing**
  The last phase includes the final part of the runway approach and the actual landing. The interface between the TAEM and landing phases is called the approach and landing interface (ALI).

The hypersonic atmospheric re-entry phase has been discussed extensively in Chaps. 5 and 7, covering the flight mechanics and guidance, navigation, and control, respectively. Most notably, Sect. 7.6 provides details on the guidance system of the Space Shuttle for the hypersonic entry and descent. The subsequent phase, i.e., the flight through the terminal area, will be the topic of this chapter.

The terminal area is used to manage the vehicle's energy while aligning the vehicle with the runway threshold. During the TAEM phase, the flight is unpowered, the vehicle flies through the transonic flight regime, and the flight characteristics of the vehicle change significantly. The most characteristic feature of the TAEM phase is a heading alignment cylinder (HAC), see also Fig. 8.1 (Fig. 7.76, Sect. 7.6.3.2, gives an alternative view of the terminal area). This imaginary cylinder touches the extension of the centreline of the runway and the vehicle will perform a turn around the wall of this cylinder to get aligned with this centreline. Two of these cylinders may exist, one on each side of the runway centreline, to allow for left or right turns, depending on the energy state of the vehicle and local wind conditions.

During the TAEM phase, the vehicle's energy level is regulated such that the vehicle will arrive at the HAC and ALI with the correct amount of energy. To dissipate more energy, the flight path can be enlarged by making so-called S-turns, although in principle any turn (in a banked flight situation) will contribute to an increase of the energy dissipation. Other ways to achieve this are to fly faster by flying at a steeper descent angle or to use speed brakes.

Following the nominal entry flight for gliding vehicles is not the only situation when terminal-area guidance is used. In the event of one or more engine failures during the ascent of a space plane, the flight must be aborted, necessitating a return to a designated landing site. Throughout this return phase, careful management of the vehicle's energy is imperative to ensure it reaches the runway with optimal energy levels. Consequently, a terminal-area guidance phase becomes essential in such scenarios.

Under normal circumstances for re-entry flights, meticulous analysis and pre-flight simulations determine a singular or relatively limited starting point for the

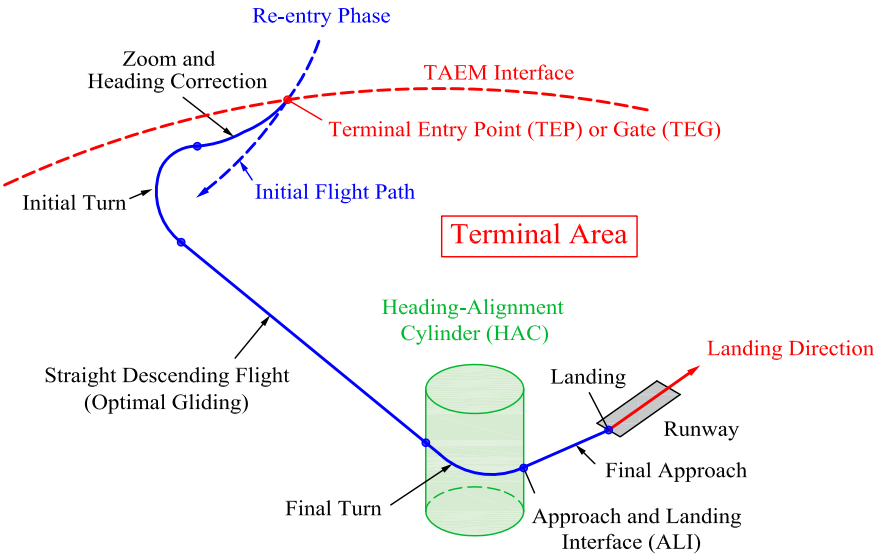


Fig. 8.1 Overview of the terminal area energy management phase

TAEM phase. The hypersonic re-entry guidance system is responsible for guiding the vehicle to this specific location, termed the terminal entry point (TEP) or terminal entry gate (TEG). However, complications arise if a malfunction occurs during the ascent phase, leading to a flight abort. In such cases, it is uncertain whether the vehicle will align with the previously identified TEG. Additionally, variability in runway orientations among different landing sites prevents the specification of a universal TEG across all scenarios. Consequently, the terminal area may initiate from any position along an entry circle surrounding the landing site, presenting significant challenges in flight planning and TAEM execution.

In terminal-area studies, the flight trajectories are commonly starting from nominal conditions. In a real-life scenario, though, the operational conditions will not be nominal and several variations can be present. These off-nominal conditions must be considered through extensive sensitivity studies, and include both variations in the initial conditions (energy level, heading angle, and flight-path angle) and variations in the models (density, wind, vehicle characteristics). In this chapter, though, we will focus on the variations in the initial conditions to study the effect on the vehicle's energy level and required energy dissipation. Hence, besides the position (on the entry circle) at the start of the terminal area, the TEG is also determined by the variations in initial conditions for which the vehicle is able to reach the runway.

The analysis of the terminal-area flight phase or the development of TAEM guidance systems allows us to use simplified flight-mechanics models. Therefore, we will begin the discussion with an overview of all flight-mechanics related aspects, beginning from the non-linear equations of motion (Sect. 8.2) and introducing several simplifications, i.e., the non-rotating flat-Earth model, the steady-state flight,

and the energy approximation. Before we put this theory into practice, in Sect. 8.3 the most important existing guidance algorithms are discussed, notably those of the Space Shuttle and the Russian shuttle Buran. Finally, the case study in Sect. 8.4 will present many details about optimal trajectories through the terminal area for a winged re-entry vehicle. As before, we will use HORUS as our reference vehicle; details on this vehicle can be found in Appendix B.2.

It is noted that for planetary missions, such as those for landing on Mars, also some form of terminal-area guidance is required. This kind of guidance is usually referred to as terminal-point guidance. Due to the (limited) scope of the current discussion, we will focus on the Earth-related missions, most notably on unpowered, winged re-entry vehicles. Powered, terminal descent flight will be discussed elsewhere (Chap. 9).

8.2 Flight Mechanics

The non-linear equations that describe the translational motion of a re-entry vehicle (see Chap. 4), can be used to study optimal trajectories and the performance of the vehicle. By propagating the complete trajectory we can determine the path or *integral* performance, and, due to the non-linear character, it requires numerical methods. This results in the most accurate solution, but is also computationally the most complex and expensive in terms of CPU time. In addressing the computational challenges inherent in optimal-control problems or performance analysis, significant efforts have been devoted to generating solutions that are near-optimal. Furthermore, empirical observations from actual flight experiences, alongside comparisons between different optimal solutions and suboptimal approximations, frequently demonstrate minimal disparities in performance. With proper execution, a straightforward analysis often yields solutions closely resembling the optimal solution.

For instance, one common metric for evaluating performance is the integral-performance value, such as maximum range. Conversely, point performance focuses on performance parameters at specific points along the flight path. These parameters may pertain to instantaneous quantities like minimum drag, minimum flight-path angle, or minimum radius of turn. In this chapter, we explore the potential of using point performance parameters to analyse or approximate integral performance. Two commonly employed approximations for this purpose are the steady-state approximation and the energy-state approximation. These two approximations will be discussed later in this section, notably in Sects. 8.2.3 and 8.2.5. In between we will also look at the maximum dive (Sect. 8.2.4), as this will give the minimum range opposed the maximum-range flight given by the two mentioned approximations. Before we come to that, the non-linear equations of translational motion will be reviewed (Sect. 8.2.1). The basis for the two mentioned approximations is driven by two assumptions, i.e., a flat and non-rotating Earth. This will be presented in Sect. 8.2.2.

Finally, to conclude this section, the out-of-plane motion due to steady, coordinated turns will be discussed in Sect. 8.2.6.

8.2.1 Non-linear Equations of Motion

For performance prediction of the longitudinal range, a point-mass model for the motion in the vertical plane would be adequate. However, since we will also study the influence of a turn on the optimal trajectories (Sect. 8.2.6), we need a set of equations that describes the full 3D motion. These equations have been extensively discussed in previous chapters, but for the sake of convenience they are restated. In the current chapter, the equations of motion for a rotating Earth (with a shape described by the reference ellipsoid) with a central gravity field are used to simulate and calculate an optimal TAEM trajectory. To facilitate the subsequent approximations the position and velocity of the vehicle are defined with spherical coordinates.

So, for an Earth with a central gravity field, the dynamic and kinematic equations read:

$$\dot{V} = -\frac{D}{m} + g\sin\gamma + \omega_{cb}^2 R\cos\delta(\sin\gamma\cos\delta - \cos\gamma\sin\delta\cos\chi) \tag{8.1}$$

$$\begin{aligned} V\dot{\gamma} =& \frac{L\cos\sigma}{m} - g\cos\gamma + 2\omega_{cb}V\cos\delta\sin\chi + \frac{V^2}{R}\cos\gamma + \\ &+\omega_{cb}^2 R\cos\delta(\cos\delta\cos\gamma + \sin\gamma\sin\delta\cos\chi) \end{aligned} \tag{8.2}$$

$$\begin{aligned} V\cos\gamma\dot{\chi} =& \frac{L\sin\sigma}{m} + 2\omega_{cb}V(\sin\delta\cos\gamma - \cos\delta\sin\gamma\cos\chi) + \\ &+\frac{V^2}{R}\cos^2\gamma\tan\delta\sin\chi + \omega_{cb}^2 R\cos\delta\sin\delta\sin\chi \end{aligned} \tag{8.3}$$

$$\dot{R} = \dot{h} = V\sin\gamma \tag{8.4}$$

$$\dot{\tau} = \frac{V\sin\chi\cos\gamma}{R\cos\delta} \tag{8.5}$$

$$\dot{\delta} = \frac{V\cos\chi\cos\gamma}{R} \tag{8.6}$$

It is noted that in the above formulation, the bank angle is defined such that a positive bank will lead to a positive turn. Hence, in Eq. (8.3) the term $\frac{L\sin\sigma}{m}$ appears as a positive term.

The aerodynamic forces (lift L and drag D) are defined as:

$$L = C_L\bar{q}S_{ref} \tag{8.7}$$

$$D = C_D\bar{q}S_{ref} \tag{8.8}$$

with the dynamic pressure defined as $\bar{q} = \frac{1}{2}\rho V^2$. The aerodynamic lift and drag coefficients C_L and C_D are usually not constant, but depend on the vehicle configuration,

control-surface deflections, Mach number, Reynolds number, attitude of the vehicle (the angles of attack and sideslip), and the altitude. The Mach number is defined as $M = \frac{V}{a}$, where a is the local speed of sound. The atmospheric properties can be calculated using the US Standard Atmosphere 1976, which provides an accurate representation for most applications. To avoid making the subsequent discussion overly complicated, an atmosphere without wind is assumed.

To anticipate subsequent discussions on steady-state approximation and the energy-state approximation, a few definitions are introduced here.

The total mechanical energy, $\mathcal{E}_{tot}$, is the sum of the potential and the kinetic energy, $\mathcal{E}_p$ and $\mathcal{E}_k$[1]:

$$\mathcal{E}_{tot} = \mathcal{E}_p + \mathcal{E}_k = mgh + \frac{1}{2}mV^2 \tag{8.9}$$

In the energy-state approximation, a constant gravity field is assumed. Using this constant g_0, the total mechanical energy as well as the individual terms on the right-hand side in Eq. (8.9) can be divided by the weight, and the so-called energy height is obtained:

$$\mathcal{E}_h = \frac{\mathcal{E}_{tot}}{mg_0} = \frac{\mathcal{E}_{tot}}{W} = h + \frac{1}{2\,g}V^2 \tag{8.10}$$

where we have introduced the vehicle weight $W = mg_0$. The driving parameters in $\mathcal{E}_h$ are, of course, the velocity and altitude. In the altitude-velocity plane contours with constant energy can be plotted (Fig. 8.2). Flights with either constant or varying energy height play an important role in the energy-state approximation.

The flight range of a vehicle flying over a spherical Earth, R_f, is defined to be the projection of the trajectory on the surface of the Earth. This corresponds with the integral value of the horizontal velocity components, yielding R_{f_x} and R_{f_y}:

$$R_{f_x} = \int \dot{x}\mathrm{d}t = \int V \sin\chi \cos\gamma \mathrm{d}t \tag{8.11}$$

$$R_{f_y} = \int \dot{y}\mathrm{d}t = \int V \cos\chi \cos\gamma \mathrm{d}t \tag{8.12}$$

However, it is common practice to relate the flight range to the motion in a fixed vertical plane, which means that the heading angle is no longer relevant. The flight range becomes the integral value of the absolute horizontal velocity, i.e.,

$$R_f = \int V \cos\gamma \mathrm{d}t \tag{8.13}$$

[1] The general definition of the potential energy is $\mathcal{E}_p = -m\frac{\mu}{R}$, which puts the zero reference at infinity. However, if referenced to a flat surface and taking this surface as the zero reference, the definition becomes $\mathcal{E}_p = mgh$. Since we will be using the flat-Earth approximation throughout this chapter, the latter definition is the one that will be used.

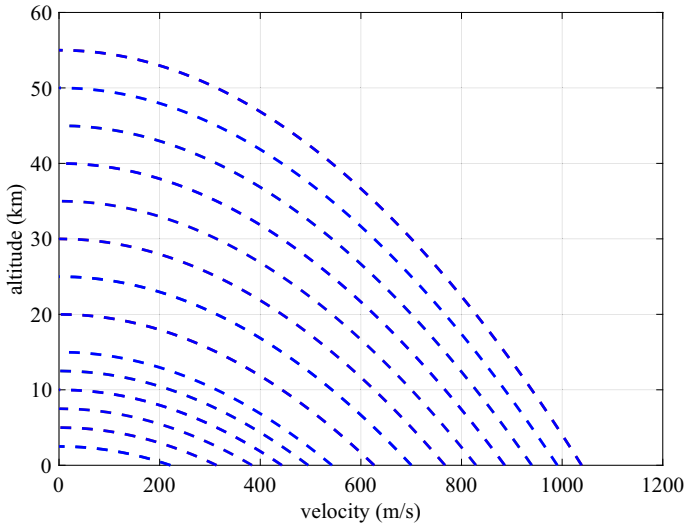


Fig. 8.2 Lines with constant energy height in the altitude-velocity plane

8.2.2 Flat, Non-rotating Earth

In some studies related to the terminal area, e.g., the one by Kluever and Horneman (2005), a simplified version of the equations as stated above is used. Moreover, these simplified equations could also be used when deriving approximation theorems for the point performance of unpowered, winged re-entry vehicles. A commonly used simplification is the use of a flat, non-rotating Earth.

The rotation of the Earth appears in two extra terms in the force equation: the centripetal acceleration and the Coriolis acceleration. The centripetal acceleration depends on the direction of the position vector and hence, the latitude. The acceleration is zero at the poles and maximum at the equator. Assuming an altitude of 25 km above the equator, the maximum magnitude of the centripetal acceleration, $\omega_{cb}^2 R$, is equal to $0.034\,\mathrm{m/s^2}$. The Coriolis acceleration depends on the magnitude and the direction of the velocity of the vehicle with respect to the rotating Earth. The magnitude of this Coriolis acceleration has an important effect in a high-speed, long-range flight. The acceleration is zero when the flight path is parallel to a meridian (flying north or south) and is maximum when flying parallel to the equator. Assuming a flight situation at an altitude of 25 km with a velocity of $V = 750\,\mathrm{m/s}$ (a Mach number of about $M = 2.5$) parallel to the equator, the maximum Coriolis acceleration $2\omega_{cb}V$ is equal to $0.109\,\mathrm{m/s^2}$. For an accurate result, both accelerations should be taken into account. But because a terminal area flight is not a long-range flight and because of the small values of the accelerations compared to the gravitational acceleration, these accelerations are often ignored in terminal-area studies by considering a non-rotating Earth, $\omega_{cb} = 0$.

Assuming the Earth is a perfect sphere with equal mass distribution, the radial component of the gravity is inversely proportional to the square of the distance from the centre of the Earth. In a flat Earth assumption, it is appropriate to consider g as constant. Since atmospheric flights are conducted in a relatively thin layer of the atmosphere as compared to the Earth's radius, g_0 can indeed considered to be a constant. However, also here it applies that when the flight duration is long and accurate results are required, g as a function of altitude should be considered. For an altitude of 25 km, the error induced by a constant g is around 8% of the Coriolis acceleration and 2.5 times as large as the centripetal acceleration.

Next, also the term $\frac{V^2}{R}$ will disappear from the force equation. For a vehicle flying at twice the speed of sound at sea level, the ratio is in the same order as the Coriolis acceleration and based on that it could already be neglected. However, for a flat Earth R reaches the limit value of infinity, which means that the term is actually zero.

For a flat Earth it does not make sense anymore to express the position in spherical coordinates, because latitude and longitude are no longer defined. Switching to Cartesian x, y, and z coordinates, and realising that the time derivative of the position is simply the velocity in corresponding directions (see Fig. 8.3) we get the following set of equations for a non-rotating flat Earth:

$$\dot{x} = V \sin\chi \cos\gamma \tag{8.14}$$

$$\dot{y} = V \cos\chi \cos\gamma \tag{8.15}$$

$$\dot{z} = \dot{h} = V \sin\gamma \tag{8.16}$$

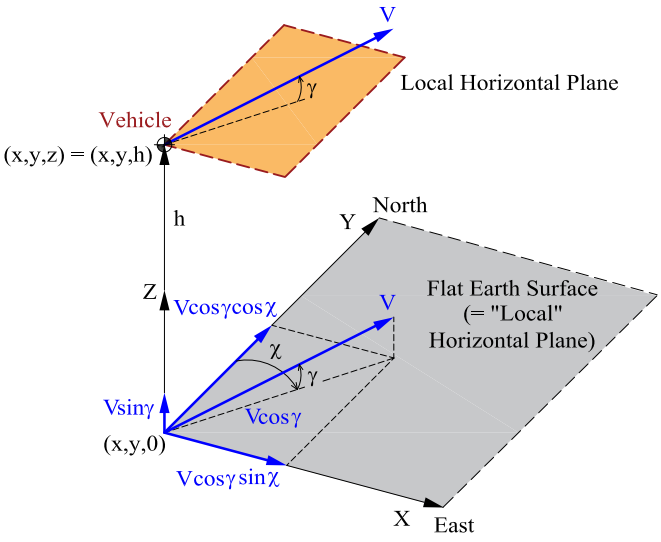


Fig. 8.3 Cartesian position and velocity for a flat-Earth approximation

$$\dot{V} = -\frac{D}{m} - g_0 \sin\gamma \tag{8.17}$$

$$V\dot{\gamma} = \frac{L\cos\sigma}{m} - g_0 \cos\gamma \tag{8.18}$$

$$V\dot{\chi} = \frac{L\sin\sigma}{m\cos\gamma} \tag{8.19}$$

These equations are used in the many studies of the terminal area phase of winged re-entry vehicles found in literature. However, one should realise that each assumption introduced above has possibly only a minor effect on its own resulting in only small inaccuracies. But stacking assumption on assumption, the accumulated effect may be significant and for an accurate result, the equations of motion for the flight over a rotating, spherical Earth must be used, Eqs. (8.1)–(8.6).

8.2.3 Steady-State Approximation

The vehicle's capabilities are constrained by two factors: the maximum range achievable from an initial state and the minimum distance attainable through a maximum-dive strategy. These boundaries effectively define the vehicle's operational limits. Consequently, any point falling within these boundaries is accessible by the vehicle. To enhance the flexibility in selecting a terminal-area trajectory to the landing site under off-nominal conditions, determining the maximum range reachable from the terminal-area entry point becomes crucial, as well as the remaining range from each state during the TAEM phase. The maximum-range capability using a steady-state approximation will be discussed in detail here; the minimum range obtained in a diving flight will be presented in the next Sect. 8.2.4.

The fundamental tool used in the steady-state glide theory is the performance curve, also known as the velocity polar or velocity hodograph. An example of a performance curve is given in Fig. 8.4. This curve provides insight into the point performance of the vehicle. Specifically, it illustrates the rate of descent, or vertical velocity, of the vehicle at each horizontal velocity. The length of a vector extending from the origin to a point on the performance curve indicates the magnitude of the airspeed along the corresponding flight path. Meanwhile, the angle formed between the horizontal axis and the velocity vector represents the flight-path angle. Notably, the validity of the performance curve is subject to specific altitude and vehicle weight parameters.

The performance curve is constructed by using the force equations for a stationary glide. Assuming a constant magnitude of the velocity ($\dot{V} = 0$) with a constant direction ($\dot{\gamma} = 0$), and zero bank angle, the dynamic equations Eqs. (8.17) and (8.18) are reduced to:

$$-D - W\sin\gamma = 0 \tag{8.20}$$

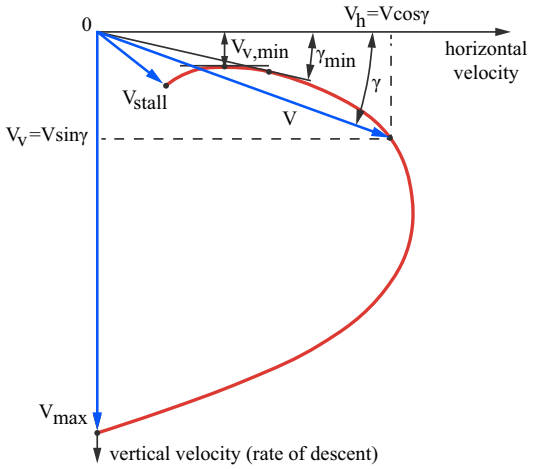


Fig. 8.4 Performance curve for a gliding flight

$$L - W\cos\gamma = 0 \tag{8.21}$$

These equations represent a force equilibrium both in parallel and normal direction with respect to the flight path, see Fig. 8.5. The resulting aerodynamic force (vector sum of lift force and drag force) is in equilibrium with the weight. Substituting the expressions for the lift and drag force gives:

$$-C_D\frac{1}{2}\rho V^2 S = W\sin\gamma \tag{8.22}$$

$$C_L\frac{1}{2}\rho V^2 S = W\cos\gamma \tag{8.23}$$

The airspeed in a stationary glide is obtained by re-writing Eq. (8.23):

$$V = \sqrt{\frac{W}{S}\frac{2}{\rho}\frac{1}{C_L}\cos\gamma} \tag{8.24}$$

with $\frac{W}{S}$ being the wing load. Dividing Eq. (8.23) by Eq. (8.22) gives an expression for the flight-path angle:

$$-\tan\gamma = \frac{C_D}{C_L} \tag{8.25}$$

It is easily verified that a flight with minimum flight-path angle is flown with a maximum lift-to-drag ratio $\frac{C_L}{C_{D\,max}}$.

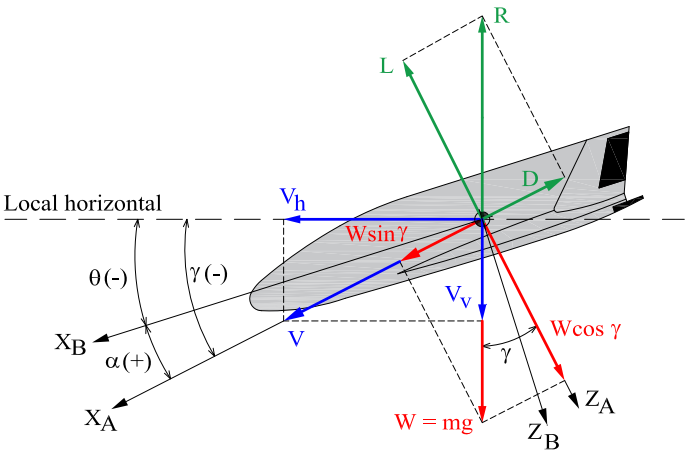


Fig. 8.5 Steady glide conditions

The rate of descent is equal to the vertical component of the flight velocity. Using the velocity and the flight-path angle, the rate of descent, RD, is defined as (see also Eq. (8.16)):

$$RD = -V_v = \dot{h} = V \sin\gamma \tag{8.26}$$

Substituting Eqs. (8.24) and (8.25) into Eq. (8.26) yields

$$RD = V \sin\gamma = V \frac{C_D}{C_L} \cos\gamma = \sqrt{\frac{W}{S} \frac{2}{\rho} \frac{C_D^2}{C_L^3} \cos^3\gamma} \tag{8.27}$$

In case of a gliding flight the flight-path angle is often small ($\cos\gamma \approx 1$), so the expressions for velocity and rate of descent simplify to

$$V = \sqrt{\frac{W}{S} \frac{2}{\rho} \frac{1}{C_L}} \tag{8.28}$$

$$RD = V \frac{C_D}{C_L} = \sqrt{\frac{W}{S} \frac{2}{\rho} \frac{C_D^2}{C_L^3}} \tag{8.29}$$

At low subsonic velocities and ignoring the Reynolds number effect, the quantities V, γ, and RD are fully determined by the angle of attack through the lift and drag coefficient. Inspecting Eq. (8.28) (or, equivalently, Eq. (8.24) for a constant flight-path angle) we see that for a given vehicle weight a flight with minimum velocity is flown at $C_{L_{max}}$, corresponding with a large angle of attack. A flight with minimum descent

rate, on the other hand, is flown with $\frac{C_L^3}{C_{D\,max}^2}$. Both $\frac{C_L}{C_{D\,max}}$ and $\frac{C_L^3}{C_{D\,max}^2}$ are vehicle characteristics, also known as the glide number and climb number, respectively.

For subsonic flight, the performance curve for a specific altitude can be calculated by a simple process:

1. Take a specific value for the angle of attack,
2. calculate the corresponding values for the lift and drag coefficient,
3. compute the flight-path angle and velocity,
4. evaluate the rate of descent and the horizontal velocity, and
5. repeat this process for the whole angle of attack interval.

It is crucial to acknowledge that the construction of the performance curve hinges on the feasibility of attaining a steady-state situation, meaning the aerodynamic coefficients and characteristics must permit the vehicle to achieve this stationary state.

Developing the performance curve for a supersonic re-entry vehicle like the Space Shuttle or HORUS presents greater complexity compared to a simple subsonic glider. The aerodynamic properties of such vehicles cannot be adequately described by a single lift and drag curve due to compressibility effects at high velocities. Both lift and drag coefficients vary not only with the angle of attack but also with the Mach number. As velocity changes, so does the Mach number, necessitating the use of different lift and drag curves to determine the corresponding coefficients.

Consequently, the calculation of performance curves becomes more intricate, requiring an iterative process to pinpoint a single point on the performance curve for a specific altitude. The steps required to find a single point on the performance curve, i.e., the horizontal velocity and the rate of descent, corresponding to a specific velocity V_{req} between the stall speed and maximum velocity, are:

1. Specify the velocity, V_{req}, for which it is required to calculate the horizontal component and the vertical component (rate of descent) during steady state,
2. calculate corresponding Mach number at the specified altitude,
3. specify an initial guess for the angle of attack,
4. calculate the lift and drag coefficient using the aerodynamic model. C_D and C_L depend on the angle of attack and the Mach number,
5. calculate γ using Eq. (8.25),
6. calculate V using Eq. (8.24),
7. compare V with V_{req}, and
8. update the guess for the angle of attack and repeat steps 4, 5, 6 and 7 until the difference $\left|V - V_{req}\right|$ is small enough.
9. When V $\approx V_{req}$, the converged values for α and γ are found, corresponding to a steady state at V_{req} and a specified altitude. Using the obtained γ, the horizontal and vertical velocity components are computed.

Repeating this process for multiple velocities between the stall speed and the maximum velocity provides the complete performance curve for a specific altitude. The maximum velocity follows from the $\bar{q}_{max}$ constraint, evaluated at the specified altitude. In Sect. 8.4 the performance curves for HORUS will be calculated.

Two important points can be extracted from the performance curve:

- **Point for maximum endurance**
  The maximum endurance (maximum flight time) occurs when the rate of descent is the smallest and can be found at the point where the performance curve has a horizontal tangent. This corresponds with $V_{v,min}$ in Fig. 8.4.
- **Point for maximum range**
  In a steady glide, achieving maximum range entails maintaining the optimal ratio between horizontal velocity and the negative rate of descent (Reichmann 1976). This optimal ratio can be determined by drawing a line from the origin tangent to the performance curve, as illustrated in Fig. 8.4. By reading the values on the horizontal and vertical axes, the corresponding horizontal velocity and rate of descent can be identified. It is noteworthy that a flight aiming for maximum range operates at a higher velocity compared to one aiming for maximum endurance. As depicted in Fig. 8.4, a maximum-range glide corresponds to the minimal flight-path angle. Using Eqs. (8.20) and (8.25), this minimal flight-path angle requirement for maximum-range flight translates into a minimum drag criterion and a maximum lift-to-drag ratio, denoted as $(C_L/C_D)_{max}$.

Because of this last ratio, a common interpretation of the steady-state approximation is to construct an angle-of-attack profile for a maximum lift-to-drag ratio, as a function of velocity or Mach number, and use this profile to calculate the maximum range.

In a steady gliding flight, the rate of energy dissipation due to the drag, must be equal to the rate of loss of potential energy. Normally, the total rate of energy dissipation consists of two terms, one accounting for the rate of potential energy and another one accounting for the rate of kinetic energy. But in a steady gliding flight, where neither the magnitude nor the direction of the velocity changes, the rate of kinetic energy is equal to zero. Therefore, the total rate of energy dissipation is equal to the rate of potential energy.

The performance curve as given in Fig. 8.4 indicates that the rate of descent during a maximum-range flight is larger than RD_{min}. Hence, it can be concluded that a maximum-range flight is conducted at a velocity for which the rate of total energy height dissipation $\frac{\mathrm{d}\mathcal{E}_h}{\mathrm{d}t}$ is larger than the minimum value. The energy height $\mathcal{E}_h$ is defined by Eq. (8.10).

Because of the changing density, altitude has an effect on the performance curve. A performance curve is only valid for a specific altitude and hence, several performance curves can be constructed for different altitudes. The lower the density, the faster the vehicle flies and sinks. But the ratio between the velocity and the sink rate stays the same (Reichmann 1976). This means that the performance curve shifts downward and to the right along a straight line through the origin when the altitude increases and the density decreases. The performance curves for different altitudes have a joint tangent and hence the same minimum flight-path angle for maximum range.

The ratio of the steady-glide velocity for a maximum-range glide at an altitude h_1 and at sea level h_0 is calculated by using Eq. (8.24), inserting the constant flight-path angle and assuming a constant lift coefficient:

$$\frac{V_{h1}}{V_{h0}} = \frac{\sqrt{\frac{W}{S}\frac{2}{\rho_1}\frac{1}{C_L}\cos\gamma}}{\sqrt{\frac{W}{S}\frac{2}{\rho_0}\frac{1}{C_L}\cos\gamma}} = \sqrt{\frac{\rho_0}{\rho_1}} \tag{8.30}$$

This equation indicates that a maximum-range flight requires that the vehicle is controlled such the dynamic pressure or equivalent airspeed[2] is kept constant throughout the glide, so:

$$\frac{V_{h1}}{V_{h0}} = \sqrt{\frac{\rho_0}{\rho_1}} \Rightarrow \sqrt{\rho_1}V_{h1} = \sqrt{\rho_0}V_{h0} \Rightarrow \bar{q}_1 = \frac{1}{2}\rho_1 V_{h1}^2 = \frac{1}{2}\rho_0 V_{h0}^2 = \bar{q}_0 \tag{8.31}$$

Thus, the gliding vehicle effectively maintains a quasi-steady flight regime, where the equivalent airspeed remains constant while the true airspeed decreases with decreasing altitude. Consequently, an enhanced approach to solving the maximum-range problem involves maintaining the dynamic pressure or equivalent airspeed at a constant value, leveraging a steady-state approximation. This concept of constant equivalent airspeed is employed, for instance, in the terminal-area study conducted by MBB (1988).

It can be inferred that by adopting a quasi-steady state assumption, the maximum-range problem transitions from a dynamic optimisation problem to a static optimisation problem. This transformation simplifies the global path optimisation problem into a series of localised point-performance problems. Through optimising these point-performance problems across various altitudes, an estimate of the overall quasi-steady path performance can be derived.

8.2.4 Diving Flight in the Vertical Plane

Besides the maximum range, the capabilities of the vehicle are also limited by the minimum attainable distance. This minimum-range capability corresponds to a maximum dive.

In Sect. 8.2.3, it was stated that the maximum range is obtained by flying at the smallest flight-path angle and with the lowest amount of drag. Hence, it is logical that a diving flight is executed at a larger flight-path angle and drag level. Looking at Fig. 8.4, it can be seen that a larger-flight path angle can be obtained by flying either at a lower velocity or at a higher velocity, with respect to the velocity required for a maximum-range flight. Of course, the most natural instinct in a diving flight is to lower the nose and hence to dive towards the ground with a larger velocity. In fact, Büchner (2003) used this strategy and the maximum dive path was calculated using a diving flight at maximum dynamic pressure or maximum velocity.

[2] The equivalent airspeed is the airspeed at sea level in the International Standard Atmosphere at which the dynamic pressure is the same as the dynamic pressure at the true airspeed and altitude at which the aircraft is flying.

Table 8.1 Flight-path angle, drag and angle of attack for a stationary point with an altitude of 5 km and a velocity of 150 m/s, as a function of speed-brake deflection

Speed-brake deflection [deg]	γ [deg]	Drag [N]	α [deg]
0.0	–14.81	65,298	7.52
40.0	–20.05	87,576	7.30

To dive more rapidly, speed brakes can be used. A speed-brake deflection may be modelled as an increase of the drag coefficient. Using Eq. (8.25), this increase in drag results in a steeper trajectory. As an example, Table 8.1 shows the related parameters for HORUS. Subsequently, the flight-path angle, the drag level, and the required angle-of-attack for two speed-brake settings at a stationary point at an altitude of 5.0 km with a velocity of 150 m/s are shown. Following an augmentation of speed-brake deflection, achieving force equilibrium at the same velocity necessitates a greater flight-path angle. Consequently, the required angle-of-attack experiences a slight decrease. Moreover, a larger speed-brake deflection directly amplifies orbiter drag and trajectory steepness, both of which escalate energy dissipation relative to range. In Sect. 8.4, a complete diving flight both with and without speed brakes is presented to indicate the difference in obtained range.

8.2.5 Energy-State Approximation

An improved approximation is the energy-state approximation, where it is assumed that kinetic and potential energy can be exchanged back and forth in zero time without loss of total energy (Bryson et al. 1969). The total energy is now considered as the state variable of the system and is continuously decreasing for a gliding vehicle. The total energy per unit of weight, or energy height $\mathcal{E}_h$, was already defined in Sect. 8.2.1, Eq. (8.10), and is repeated here for convenience:

$$\mathcal{E}_h = h + \frac{1}{2g} V^2 \tag{8.32}$$

The most important derivative used in the energy-state approximation is the rate of energy dissipation concerning glide range:

$$\frac{\mathrm{d}\mathcal{E}_h}{\mathrm{d}R_f} = \frac{\mathrm{d}\mathcal{E}_h}{\mathrm{d}t} \frac{\mathrm{d}t}{\mathrm{d}R_f} \tag{8.33}$$

The time derivative of the energy height, $\frac{\mathrm{d}\mathcal{E}_h}{\mathrm{d}t}$, is expressed as:

$$\frac{\mathrm{d}\mathcal{E}_h}{\mathrm{d}t} = \dot{h} + \frac{V\dot{V}}{g_0} \tag{8.34}$$

Substituting Eqs. (8.16) and (8.17) into Eq. (8.34) yields

$$\frac{\mathrm{d}\mathcal{E}_h}{\mathrm{d}t} = \frac{-VD}{mg_0} \tag{8.35}$$

$\frac{\mathrm{d}R_f}{\mathrm{d}t}$ is equal to the time rate of change of the flight range R_f, or the horizontal component of the velocity, $V\cos\gamma$. Inserting the expressions for $\frac{\mathrm{d}\mathcal{E}_h}{\mathrm{d}t}$ and $\frac{\mathrm{d}R_f}{\mathrm{d}t}$ in Eq. (8.33) results in:

$$\frac{\mathrm{d}\mathcal{E}_h}{\mathrm{d}R_f} = \frac{-VD}{mg_0}\frac{1}{V\cos\gamma} = -\frac{D}{mg_0}\frac{1}{\cos\gamma} \tag{8.36}$$

The energy of a vehicle consists of a potential (or altitude) component and a kinetic (or velocity) component. Plotting a trajectory in terms of the altitude as a function of the velocity results in a trajectory in the energy space or the (V, h) space. In Fig. 8.2 we have shown this energy space with curves of constant energy height.

Within the energy-state approximation, with the total energy being a state variable, the maximum range is obtained by minimising the magnitude of $\mathrm{d}\mathcal{E}_h/\mathrm{d}R_f$. This means that, assuming a small influence of the cosine of the flight-path angle, the drag has to be constantly minimised for a specific energy height to maximise the range.

The drag level depends on the angle of attack (expressed through the drag coefficient), altitude (dictating density), and velocity, as indicated in Eq. (8.8). Therefore, to represent the drag level within the (V, h) space, it is necessary to define a particular angle-of-attack value, ensuring that drag becomes solely dependent on altitude and velocity. A common approach is to specify the angle of attack necessary to achieve equilibrium between lift and weight.

By employing this method to specify α, the drag level can be computed for each combination of altitude and velocity. For a given energy level, there exists a unique combination of altitude and velocity at which the vehicle encounters minimal drag. Consequently, a trajectory in the energy space can be delineated where drag is minimised at each energy level. If the initial conditions do not align with this minimum drag or maximum-range glide path, the vehicle should either execute a zoom-climb or zoom-dive with constant energy to converge with the curve (Bryson et al. 1969; Vinh 1981). Towards the conclusion of the flight, enhancing the range can be achieved by executing a final flare, during which kinetic energy is converted into potential energy until reaching stall speed. Figure 8.6 depicts a schematic of a typical maximum range glide path in the energy space.

In reality, executing a zoom-climb or zoom-dive without energy loss is not possible for any vehicle, particularly an unpowered glider. Consequently, these discontinuities in altitude and velocity are unrealistic. The total energy of a gliding vehicle invariably decreases, resulting in a curved trajectory in the energy space.

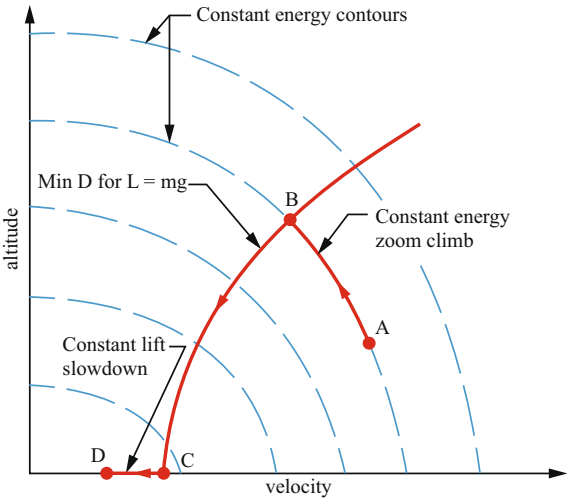


Fig. 8.6 Sketch of a typical maximum range glide. Data taken from Bryson et al. (1969)

Example 8.1 (*Energy Space of a Supersonic Aircraft*) Bryson et al. (1969) used the energy-state approximation to calculate the maximum glide path for a supersonic aircraft. In this example we will show the 2D- and 3D representation of the drag in the energy space and compare our calculated plots with the one available from literature. In that way we can verify our models and approach that will be used for the case study of the terminal-area flight of an unpowered, winged re-entry vehicle in Sect. 8.4.

The aerodynamic characteristics of the supersonic aircraft used in the maximum glide-range problem are calculated using:

$$C_L = C_{L\alpha}\alpha \tag{8.37}$$

$$C_D = C_{D0} + C_{L\alpha}\alpha^2 \tag{8.38}$$

where the parameters $C_{L\alpha}$ and C_{D0} are Mach dependent. Table 8.2 lists the numerical values. The reference surface is equal to 500 ft^2 (46.45 m^2) and the mass of the aircraft is equal to 34,200 lb (15,513 kg). With the given data one can show that the aircraft has relatively good gliding capabilities with a maximum lift-to-drag ratio of 10.3 and 7.6 in the subsonic and supersonic speed regime, respectively.

Figure 8.7a shows the contours of constant drag in the energy space, as shown by Bryson et al. (1969). The contours of constant drag are calculated for an equilibrium situation, where the lift equals the weight of 34,200 lb (15,513 kg).

As stated earlier, our aim is to compute the constant-drag contours and to confirm the values and shape as indicated in Fig. 8.7a. With the aerodynamic model given by Eqs. (8.37)–(8.38) and the assumption of lift being equal to weight, the drag was calculated for the complete (V, h) space. In Fig. 8.7b the calculated contours of constant drag for the supersonic aircraft are given, whereas Fig. 8.8 shows a 3D-view

Table 8.2 Mach-dependent aerodynamic parameters of a supersonic aircraft (Bryson et al. 1969)

M	$C_{L\alpha}$	C_{D0}
0.0	2.240	0.0065
0.4	2.325	0.0055
0.8	2.350	0.0060
1.2	2.290	0.0118
1.6	2.160	0.0110
2.0	1.950	0.0086
2.4	1.700	0.0074
2.8	1.435	0.0069
3.2	1.250	0.0068

of the drag levels. The drag valley in which the maximum range path is situated, is clearly visible. The calculated drag levels and the shape of the drag valley correspond with Fig. 8.7a.

The lift-equal-to-weight assumption is actually not correct, because a gliding vehicle does not perform a level flight in a steady situation, but a descending flight. So the real equilibrium in a zero-bank situation is described by:

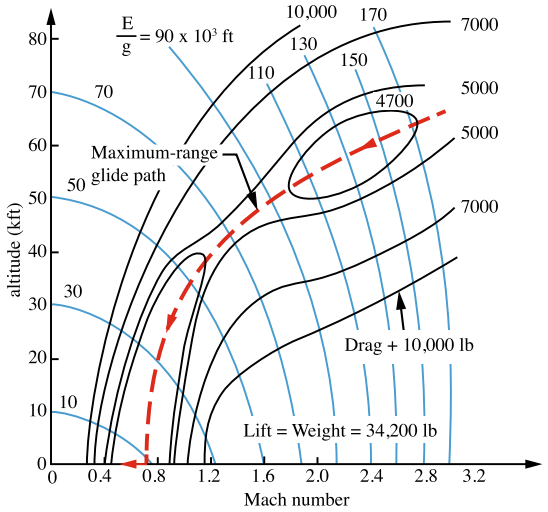
$$L = W \cos \gamma \tag{8.39}$$

With the incorporation of the flight-path angle, determining the required angle of attack makes an iterative process akin to the one outlined in Sect. 8.2.3 for calculating the performance curve necessary. Once the angle of attack is established, the corresponding drag level can be ascertained at the specific altitude and velocity. However, findings suggest that the inclusion of the flight-path angle does not significantly alter the shape of the drag valley. Only minor adjustments in the drag levels are observed, attributed to the low values of the flight-path angle and consequently, the near-unity value of the cosine term. When the flight-path angle is more substantial, such as in the case of a winged re-entry vehicle, the influence of the flight-path angle becomes more pronounced.

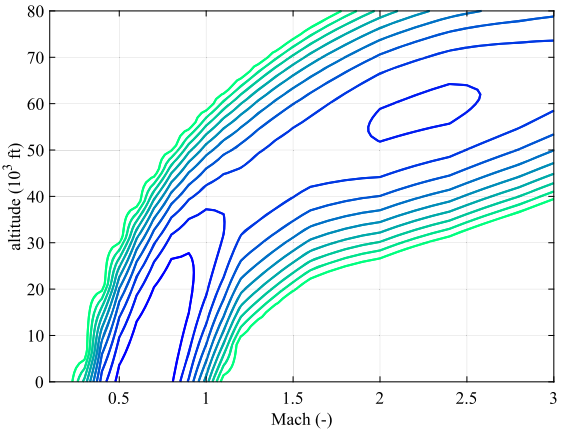
Note that contrary to the common practice of using SI-units, Figs. 8.7b and 8.8 were calculated in the foot-pound-second system to be able to compare the calculated results with those from Bryson et al. (1969). ■

8.2.6 Steady Coordinated Turn

The basic manoeuvre to change the flight-path heading is a coordinated turn without slip. Such a manoeuvre is required, for instance, when following the "mantle" of the heading alignment cylinder, HAC. Figure 8.9 shows a vehicle in a steady level



(a) Maximum range glide path and contours of constant drag for $L{=}W$. Data taken from Bryson *et al.* (1969).



(b) Contours of constant drag

Fig. 8.7 Energy space of a supersonic aircraft

coordinated banked-turn. A coordinated turn is a turn where the inward centripetal force[3] required to pull the airplane toward the centre of the turn is accomplished by the horizontal component of the lift and, where both the resultant aerodynamic force and the vector sum of the weight and the outward centrifugal force[4] are in the plane of symmetry of the vehicle. In a coordinated turn there will not be any sideslip, such that the side force is zero. In a banked flight situation, the direction of the lift vector is changed and is no longer parallel to the weight vector. That

[3] The centripetal force is the external force required to make a body follow a curved path.

[4] The centrifugal force is an outward force associated with rotation.

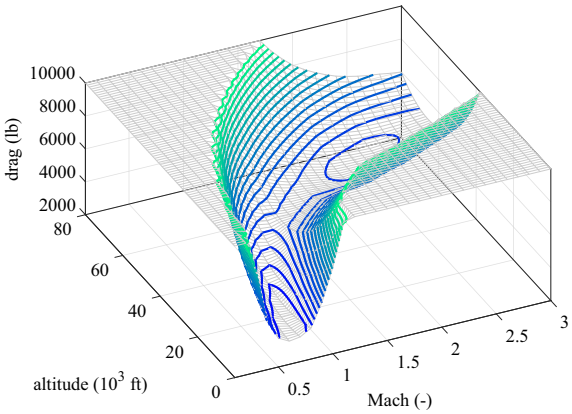


Fig. 8.8 3D-view of the contours of constant drag for a supersonic aircraft

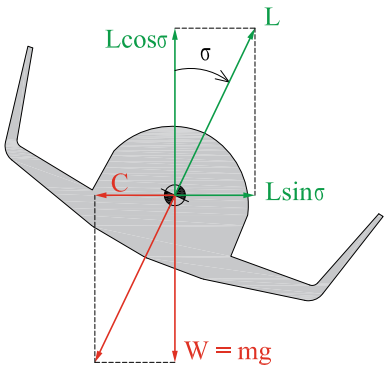


Fig. 8.9 Vehicle in a steady level banked situation

means that to achieve equilibrium between the vertical component of the resulting aerodynamic force vector and the weight vector, the lift must be larger than in a wings-level situation. If the velocity remains unchanged, this additional component is obtained by increasing the angle-of-attack.

In general, a gliding re-entry vehicle makes a descending turn. The equilibrium equations for the instantaneous conditions on the descending spiral flight path are obtained by assuming that the velocity does not change, nor in magnitude nor in direction ($\dot{V} = \dot{\gamma} = 0$). Using these assumptions, the dynamic equations as given by Eqs. (8.17) through (8.19), reduce to:

$$-D - mg_0 \sin\gamma = 0 \tag{8.40}$$

$$L\cos\sigma - mg_0 \cos\gamma = 0 \tag{8.41}$$

$$\frac{L\sin\sigma}{m\cos\gamma} = V\dot{\chi} \tag{8.42}$$

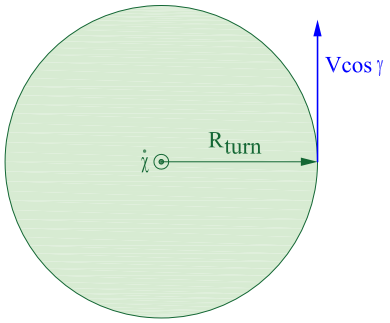


Fig. 8.10 Relation between turn radius and heading rate

The rate of turn is defined as the degrees per second the vehicle changes heading and is defined by the term $\dot{\chi}$. For a gliding vehicle, starting from the notion that only the horizontal component of the velocity ($V \cos\gamma$) drives the rate of change of the heading, one can write the rate of turn as (Fig. 8.10):

$$\dot{\chi} = \frac{V \cos\gamma}{R_{turn}} \tag{8.43}$$

which easily follows from the notion that the tangential velocity equals the product of angular rate with the radial distance. Inserting the above equation into Eq. (8.42) gives, after rearranging terms:

$$\frac{V^2}{R_{turn}} m \cos^2\gamma = L \sin\sigma \tag{8.44}$$

where the term on the left defines the horizontal centrifugal force, while the term on the right represents the horizontal component of the lift. Therefore, Eqs. (8.42) and (8.44) affirm that in a coordinated turn, the centrifugal force equals the centripetal force. Consequently, the vehicle lacks any inclination to move either inward or outward and follows a circular path. In such a scenario, the vehicle only experiences an additional force perpendicular to the plane of the wings, without the tendency for lateral sliding (no sideslip).

The additional force in a direction perpendicular to the plane of the wings is expressed by the load factor. The load factor is a parameter, which describes the ratio of the lift force to the weight of the vehicle. Using Eq. (8.41), the load factor during a gliding turning flight is expressed by:

$$n = \frac{L}{mg_0} = \frac{\cos\gamma}{\cos\sigma} \tag{8.45}$$

It is obvious that the larger the bank angle, the larger the required additional lift force, which will increase the load factor.

The airspeed in a gliding turn is obtained by inserting the expression for the lift force in Eq. (8.41) and rearranging the terms:

$$V = \sqrt{\frac{mg_0}{S}\frac{2}{\rho}\frac{1}{C_L}\frac{\cos\gamma}{\cos\sigma}} \tag{8.46}$$

Dividing Eq. (8.40) by Eq. (8.41) results in an expression for the flight-path angle:

$$-\tan\gamma = \frac{C_D}{C_L}\frac{1}{\cos\sigma} \tag{8.47}$$

A relation for the turn radius can be obtained by substituting the expression for the lift force into Eq. (8.44), using Eq. (8.46) and rearranging the terms:

$$R_{turn} = \frac{V^2\cos\gamma}{g_0\tan\sigma} \tag{8.48}$$

This equation confirms that in a turn, a larger bank angle and lower airspeed yields a smaller turn radius.

8.3 Terminal Area Guidance Algorithms

8.3.1 Introduction

Now that we have covered the flight mechanics aspects of terminal-area flight, including some common approximations that usually find their way into guidance-system design, we will look at some actual guidance algorithms. The goal of the terminal-area guidance is to properly manage the energy of the vehicle while aligning the vehicle with the runway. This is achieved by using a guidance algorithm that calculates commands that are used to influence the trajectory of the vehicle, in a similar fashion to the re-entry guidance systems discussed in Sect. 7.6. The main focus of this section will be on two guidance systems that have actually been used in flight. Section 8.3.2 describes the algorithm of the Space Shuttle, while the algorithm of the Russian Shuttle Buran is discussed in Sect. 8.3.3. Even though in the last few years, several other terminal-area studies have been performed, we will not discuss these due to space limitations. The interested reader is referred to the corresponding literature, e.g., Büchner (2003), Da Costa (2003), Kluever and Horneman (2005), Hull et al. (2005), and Kluever (2007).

Before we delve deeper into the specifics of terminal-area guidance systems, let us focus on the problem at hand, in fact, let us recap what we discussed earlier in Sect. 7.6. To guide the vehicle along a nominal trajectory, the path of the vehicle must be influenced. This can be done by changing the magnitude and direction of the

control force which results in an acceleration or deceleration. For unpowered flight, the only force available to deflect the trajectory is the aerodynamic force, with its main components the lift and drag force. The aerodynamic force is a function of the aerodynamic shape of the vehicle, its velocity and attitude with respect to incoming air flow, and atmospheric properties. An efficient way to change the magnitude of the aerodynamic force is to change the attitude using the angle of attack and to change the aerodynamic shape of the vehicle by using speed brakes. The direction of the lift vector can be changed by rolling the vehicle around its velocity vector. The angle over which the vehicle is rolled, is called the bank angle. By using a bank angle, the lift force is split into a vertical and a horizontal part. For a zero side-slip flight, the angle of attack in combination with the bank angle fully determine the attitude of the vehicle. Summarising, the guidance system can use three guidance variables to control the aerodynamic force and to guide the vehicle: (i) the angle of attack, α, the bank angle, σ, and the speed-brake deflection, δ_{sb}.

There are two important classes of guidance systems, based on how the reference input is obtained. The first class is based on guidance using nominal trajectories, while the second class is based on guidance using predicting capabilities.

- **Guidance using nominal trajectories**. When employing nominal trajectories, an explicit reference trajectory is used. This reference trajectory takes the shape of nominal profiles, for example, the nominal altitude, dynamic pressure, or energy as a function of the flight range. The actual flight state is compared with the nominal reference trajectory, and control laws are used to generate commands that steer the current trajectory towards the reference trajectory. Within this category, two approaches are possible. Either the reference trajectory is calculated prior to the flight and the nominal profiles are stored in the flight computer, or the reference trajectory is calculated at the beginning or during the TAEM phase onboard of the vehicle.
- **Guidance using predicting capabilities**. A guidance system employing predictive capabilities operates without necessitating a reference trajectory or nominal profile. Instead, it determines commanded control variables by computing an implicit nominal trajectory at regular intervals and comparing it to the actual state of the vehicle. For instance, a system might calculate the necessary energy to reach the ALI, compare it with the vehicle's current energy, and derive the requisite commands accordingly. This implicit trajectory can be computed either through closed-form approximations (analytical formulas) or fast-time predictions achieved by numerically solving the equations. By leveraging this method of predictive capabilities, the system gains significant flexibility in accommodating variations in the initial state and flight disturbances.

8.3.2 Space Shuttle

The American Space Shuttle flew for many years and each flight, its TAEM guidance performed well. So thanks to its age, the Space-Shuttle guidance may serve as a convenient basis for technology extrapolation. Moreover, some guidance algorithms for new-generation vehicles are based on the Space-Shuttle guidance, see, for example, the paper by Kluever and Horneman (2005). Therefore, it is essential to understand the guidance logic that started it all.

During the early 1970s, many organisations were involved in the development of the TAEM guidance of the Shuttle and many different options were considered. In 1975, a study contract was given to McDonnell Douglas Astronautics Company in St. Louis to make the final guidance consisting of all the good qualities from several developed algorithms (Moore 1991). After some fundamental modifications, this guidance algorithm was used for the Shuttle flights. In this section, both the original guidance and the guidance as used during the actual flights are described.

Terminal Area Flight Phases

The terminal area of a Space Shuttle started at Mach 2.5, at about 85,000 ft ($\approx$ 26 km) altitude and finished at the approach and landing interface at an altitude of 3 km and Mach 0.4 (Moore 1991). The TAEM phase was divided into four flight phases. These flight phases are illustrated in Fig. 8.11 and described below.

(0) **S-turn phase** The S-turn phase (Phase 0) was implemented when the Shuttle possessed an excess of energy that could not be sufficiently reduced to meet the desired final interface conditions solely through the use of speed brakes and dynamic-pressure modulation. During this phase, the Shuttle was instructed to deviate from its trajectory, moving away from the tangency point of the heading-alignment cylinder. Once a significant amount of energy had been dissipated, the acquisition phase commenced, guiding the Shuttle to an appropriate tangential alignment with the HAC. Throughout the S-turn manoeuvre, the Shuttle operated at maximum dynamic pressure and maximum speed-brake deflection, within the confines of flight conditions, aiming to achieve the maximum rate of energy dissipation possible.

(1) **Acquisition phase** In the acquisition phase (Phase 1), the Shuttle was directed towards an appropriate point of tangency with the HAC. In the original TAEM guidance of the Shuttle, the HAC had a radius of 20,000 feet. The acquisition phase could be split into two segments. Initially, a turn was executed to align with a heading tangent to the HAC. Subsequently, the Shuttle maintained level flight until it aligned with and captured the HAC (Camara 2003).
During phase 1, energy was regulated by adjusting the dynamic pressure and, when the Shuttle was flying at subsonic speeds, by adjusting the speed brake. Design requirements prohibited speed-brake adjustments in the supersonic speed range. The acquisition phase concluded when the Shuttle was within 4,000 feet (approximately 1,220 m) of the HAC to prevent significant overshooting.

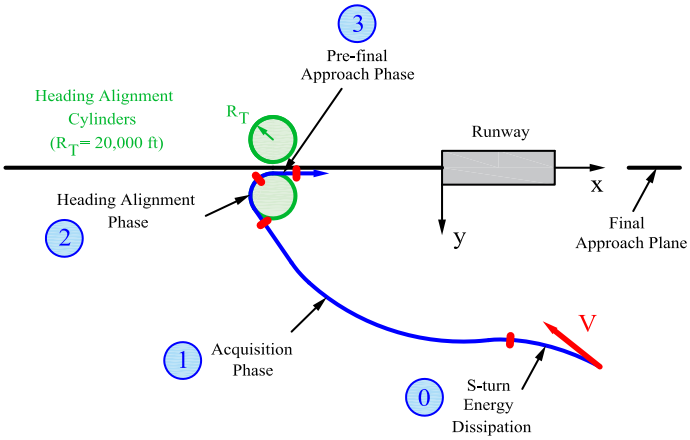


Fig. 8.11 TAEM guidance phases of the Space Shuttle

(2) **Heading-alignment guidance phase** In the heading-alignment guidance phase (Phase 2), the Shuttle was controlled to directionally track the HAC. Again, the energy-dissipation rate was controlled by dynamic-pressure modulation and the speed brake. This phase was continued until the heading was within 20° of the final-approach plane, at which moment a switch was made to the pre-final guidance phase. The final-approach plane is a vertical plane through the extended runway centreline.

(3) **Pre-final guidance phase** During the pre-final guidance phase (Phase 3), the lateral guidance directed the Shuttle towards the final approach plane. The longitudinal guidance regulated the speed brake to adjust the dynamic pressure and attain an equivalent air speed of 290 knots (approximately 149 m/s), as stipulated for the auto-land interface. Additionally, the required glide slope was maintained.

The ground-track geometry as depicted in Fig. 8.11 is used in both the original guidance algorithm from McDonnell Douglas and the algorithm during actual Space Shuttle flights.

Energy Control

The original Shuttle TAEM guidance was developed to do its primary function of energy management with an energy controller (Kraemer and Ehlers 1975). A ground-track range predictor estimates the range-to-go for the different guidance phases and adds the appropriate components to obtain the total ground track. This ground track was used by the Shuttle TAEM guidance system to steer the unpowered vehicle to an energy-versus-range profile. The energy is normalised with respect to the weight of the vehicle, and is thus given by Eq. (8.10).

To achieve the required energy state, the guidance system must modify the energy dissipation to be above or below the nominal rate. The rate of energy dissipation with glide range R_f, Eq. (8.13), has been derived in Sect. 8.2.5 and is given by

Eq. (8.36). The longitudinal guidance system compared the current energy gradient with a desired gradient, one among several stored onboard, using the disparity to navigate the Shuttle towards the nominal energy state. This current energy gradient was computed by comparing the Shuttle's energy at successive sampling intervals and dividing it by the computed ground-track distance within each interval. Since the onboard profiles were established for a wings-level flight condition (zero bank angle), while the current energy gradient accounted for energy loss due to both wings-level gliding and banking manoeuvres, an adjustment was made to align the two gradients for comparison.

The guidance system employed three primary methods to adjust the energy dissipation rate relative to the remaining range: initiating a turn away from the runway (S-turn) to extend the range-to-go and the duration for energy dissipation, modulating the speed-brake deflection, and adjusting the dynamic pressure. The angle of attack was determined through dynamic-pressure control. The energy level dictated the necessary dynamic pressure, subsequently converted into a required angle of attack. Full-deflection range modulation of the speed brake was only feasible in the subsonic speed regime due to design limitations, effectively used for modulating energy rate. In the supersonic region, the speed-brake deflection was set to 55°, while in the subsonic regime, it could vary between 0° and 87.2°.

During the S-turn phase, all three methods of increasing energy dissipation were employed to achieve maximum dissipation. In the acquisition and heading-alignment phases, the speed brake was used at subsonic speeds to rectify minor energy errors. In case the energy error escalated, dynamic pressure was then used for energy control. In the pre-final phase, the primary objective was to capture the auto-land glide slope and to regulate the airspeed for auto-land requirements. Here, the two longitudinal control channels (normal acceleration and speed brakes) were independently used (Kraemer and Ehlers 1975). Specifically, the normal-acceleration command (utilising the elevator) was employed to manage altitude, while the speed brakes were used to regulate dynamic pressure.[5] Consequently, the guidance depicted in Fig. 8.11 was not applicable in the pre-final phase.

Lateral control adjusted the commanded bank angle to ensure the vehicle adhered to a predetermined ground track. Throughout the S-turn phase, the bank angle was set to 50°. In the acquisition phase, lateral guidance was tailored to maintain the Shuttle's heading tangentially aligned with the HAC, while during the heading-alignment phase, it facilitated a turn around the HAC. In the pre-final phase, lateral guidance steered the vehicle to align with the extension of the runway centreline.

Performance of the Guidance Algorithm

From the preceding discussion, it can be concluded that the energy-range profiles served as the primary reference for computing commands. These profiles, obtained through simulations, remained fixed and were used for various initial conditions. Notably, banked-flight scenarios were not factored into these profiles, necessitating

[5] It is noteworthy that longitudinal control in this phase was contrary to that used during gliding, where altitude was controlled with speed brakes and velocity with normal acceleration (or elevator).

a correction factor for those cases involving banking. Additionally, the HAC, with a fixed radius, could be positioned at two downrange locations relative to the runway threshold.

A closed-loop investigation was conducted to assess this guidance principle under different headings, initial positions, winds, and turbulence (Kraemer and Ehlers 1975). Overall, the performance of the TAEM system was satisfactory, although in severe wind conditions, the dynamic pressure occasionally exceeded tolerance limits. It was anticipated that fine-tuning the system for tighter energy control and slight modifications to the energy control algorithms would bring results within tolerance, even under severe winds.

The algorithm employed during the initial Shuttle flight is detailed by NASA (1980), with subsequent modifications outlined by Moore (1991). Originally, energy control was managed by adjusting dynamic pressure and speed brakes. However, after additional simulations, two major issues emerged. Firstly, severe winds caused deviations beyond tolerances, and secondly, controlling dynamic pressure solely with normal acceleration resulted in oscillatory response characteristics due to dynamic pressure's dependency on drag as well as lift (Moore 1991).

Addressing these issues, a modification was made wherein the altitude component of energy was regulated independently with normal acceleration. This separation of altitude and dynamic-pressure control resolved both problems. Thus, all Shuttle missions thereafter employed this altitude controller, with reference altitudes serving as the primary input for computing commanded normal acceleration, while dynamic-pressure references determined commanded speed-brake deflection. Notably, all altitude, dynamic-pressure, and energy reference profiles continued to be calculated pre-flight.

Similarly, as in the original guidance, two options existed for the placement of the HAC downrange from the runway threshold. Typically, the Shuttle approached in the plane through the extension of the runway centreline, at a nominal entry point (NEP). However, in cases of low energy, where it was determined that the altitude at the approach and landing interface would not be achieved, the HAC centre could be shifted to a modified entry point (MEP).

During the inaugural Shuttle flight in April 1981, a circular HAC with a fixed radius of 20,000 ft was used. Subsequent missions adopted a spiral HAC, forming a heading-alignment cone in three dimensions. The radius at the intersection with the final approach plane or the final radius of the HAC had a nominal value, but could be adjusted to a smaller value if the altitude fell below a specified reference value. Another modification introduced was the optional TAEM targeting (OTT), which allowed for an overhead approach (OH) to the runway alongside a straight-in approach (SI), conserving energy for subsonic dissipation near the runway. The altitude difference between SI and OH increased the likelihood of successfully navigating weather disturbances along the landing-site approach. OTT debuted on STS-5 in 1982. The combination of NEP/MEP and SI/OH led to four possible HAC positions, with the actual choice contingent on pre-flight energy boundaries.

A notable drawback of the Shuttle algorithm was its limited flexibility to accommodate off-nominal conditions, and was heavily reliant on extensive pre-mission planning.

8.3.3 Buran

The first (and only) flight of the Buran orbiter (November 15, 1988) has proved the efficiency of the trajectory-design principle and guidance algorithm. Despite the strong gusty wind, which was present during the first flight, the guidance system performed its task perfectly and with high accuracy. Therefore, a similar approach can be used in the development of guidance systems for other winged re-entry vehicles. This section briefly describes the guidance algorithm of the Buran and is based on the overview given by Kirpischikov (2009).

Re-entry

The descent and landing phase commenced within the area of allowable entry conditions (AAEC), situated at an altitude of 100 km. This expansive area encompassed numerous daily orbit circuits, ensuring safety throughout the orbital flight.

The cornerstone of the Buran guidance system revolved around trajectory tubes. Descent trajectories were delineated by various constraints, necessitating consideration of a wide array of disturbances during design, including atmospheric parameter variations and vehicle uncertainties. These disturbances exhibit random characteristics, fluctuating not only between flights, but also within the same flight. Consequently, establishing a fixed descent trajectory was unfeasible. Instead, trajectory design-tubes were adopted, and during a flight the actual trajectory was expected to fall inside these tubes with a certain probability. The tubes were determined with a probability of 0.997.

The descent and landing phase comprised three stages, each governed by its own guidance algorithm, delineated by the vehicle's altitude:

1. Descent stage: altitude = 100–20 km
2. Pre-landing manoeuvring stage (PM): altitude = 20–4 km
3. Pre-landing approach and landing stage: altitude = 4–0 km

The TAEM phase for both the HORUS and the Space Shuttle starts approximately at an altitude of 25 km. But the pre-landing manoeuvring stage for Buran began at 20 km altitude. This is the reason why in addition to this pre-landing manoeuvring stage, also the final part of the descent stage will be discussed.

Final Part of the Descent Stage

Figure 8.12 illustrates the tube trajectories and cross-sections of the tube during the final segment of the descent phase, along with the cylinder of energy dispersion (CED) indicated by the red circle. The HACs are denoted by number 2, the runway by number 3, the flight trajectory flown during the first Buran flight by number 4,

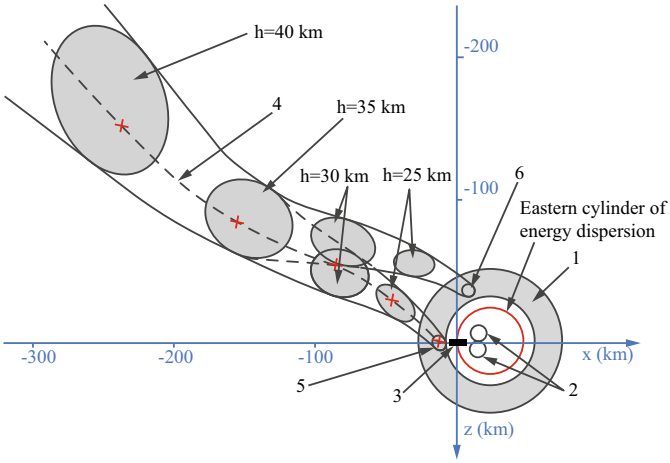


Fig. 8.12 Final leg of Descent Phase (inspired by Kirpischikov (2009)). Numbers are explained in the text

and numbers 5 and 6 represent the areas where the trajectory tubes terminate at the interface with the pre-landing manoeuvring stage.

The guidance algorithm steers the vehicle towards a specified region of trajectory parameters at an altitude of 20 km, termed the final area, depicted in Fig. 8.12 by the grey region (labelled as 1). Within this area, the velocity vector should align tangentially to the CED, with a permissible deviation of no more than 15°. Additionally, the velocity magnitude must fall within a designated range, and the distance along a great circle arc from the Orbiter's centre of mass to the CED tangency point must lie between 19 and 45 km. These trajectory parameters define the end conditions of the descent stage and the initial conditions for the pre-landing manoeuvring stage.

The lateral trajectory control algorithm guides the vehicle towards a direction aligned with one of the two tangents to the CED. As a result, the trajectory tube is split into two smaller tubes at an altitude of 35 km. Following numerous simulations, it was determined that the probability of the Buran entering the left terminal area (relative to the Orbiter's flight direction) was approximately 0.3. Hence, there is a 30% likelihood that the vehicle enters the left tube. Consequently, the probability of the vehicle entering the right tube is 0.7.

Pre-landing Manoeuvring Stage

The main task of the guidance algorithm at PM is to guide the vehicle from the final area of the descent stage to the vicinity of the key point. This key point is located at an altitude of 4 km above the runway centreline on a distance of approximately 14.5 km from the centre of the runway. Figure 8.13 shows the trajectory tubes for the pre-landing manoeuvring stage. Two CEPs and two key points are shown, which are needed to accommodate for an eastern or western wind. All trajectories are generated assuming that the vehicle is approaching from the west.

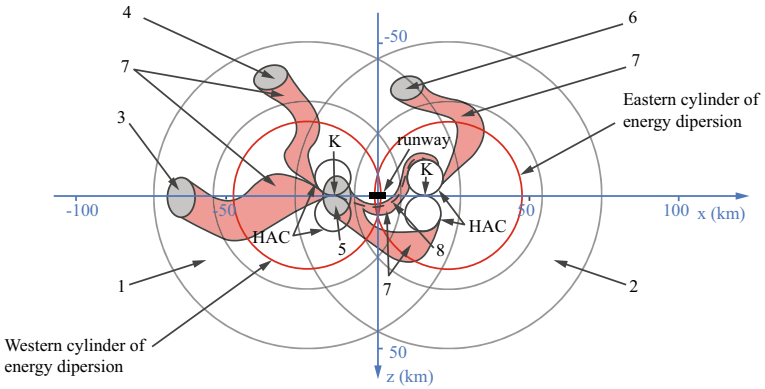


Fig. 8.13 Pre-landing manoeuvring stage (inspired by Kirpischikov (2009)). Numbers are explained in the text

At the pre-landing manoeuvring stage, dispersion of excessive energy is done. Generally, this can be done by three methods:

1. Change of trajectory by turns for matching the current energy versus range state with the designed state,
2. change of dynamic pressure, or
3. change of lift-to-drag ratio

For Buran, a hybrid approach was adopted. The fundamental concept involved altering the trajectory path, achieved through spiral turns, coupled with programmed adjustments to the lift-to-drag ratio and dynamic pressure levels. Alterations in the lift-to-drag ratio were enacted by deploying the air brakes. When facing a tailwind, the vehicle would be directed to maintain the maximum dynamic pressure possible at each altitude level to facilitate thorough energy dissipation. Conversely, in the presence of a headwind, minimal dynamic pressure was sought to enable flight with the maximum lift-to-drag ratio (Kirpischikov 2009). It is essential to highlight that this applies solely to a segment of the pre-landing manoeuvring stage. Towards the conclusion of this phase, when progressing in a straight trajectory towards the pivotal point along the extension of the runway centreline, it becomes imperative for the dynamic pressure to surpass the minimum threshold, particularly in the event of a headwind. Drawing from gliding expertise, it is observed that wind intensity diminishes closer to the ground compared to higher altitudes. Should one maintain minimum speed while encountering a reduction in wind velocity, the vehicle's speed may decrease, potentially leading to a stall.

At the onset of the PM stage, the guidance system constructed a three-dimensional reference trajectory contingent upon the energy level and the Orbiter's status. This reference trajectory linked the initial position with the key point. Typically, the projection of this reference trajectory onto the horizontal plane comprised five distinct geometric trajectory segments:

1. A turn-away spiral,
2. a corrective turn spiral,
3. a tangent to the HAC,
4. an arc of the HAC, and
5. a straight line to the key point.

The trajectory for which the energy difference was within an acceptable limit, was accepted as the nominal reference trajectory.

A turn-away manoeuvre was performed in case the best trajectory without a turn-away was insufficient to dissipate the energy: the longer the flight the more energy will be dissipated. The reference trajectory would be redefined when the deviation from the reference trajectory is too large, and at the end of a turn manoeuvre (end of turn-away and corrective turn).

During the flight of the Buran, the eastern CEP was used, a decision influenced by prevailing wind conditions. As previously mentioned, following extensive simulations, it was determined that the probability of a trajectory leading to the right terminal area (in relation to the flight direction) was 0.7. This area was indeed employed during the maiden flight. Commencing from this terminal region, two potential trajectory tubes were feasible. One trajectory tube circumnavigated the southern HAC, while the other encircled the northern HAC. Analysis indicated a probability of 0.97 for the southern HAC and 0.03 for the northern HAC to be selected. Despite the higher probability favouring the southern HAC, the conditions during the Buran flight demanded the adoption of the northern HAC, as depicted in Fig. 8.14.

8.4 Case Study: Terminal-Area Flight of HORUS

Following the discussion of the TAEM guidance algorithms of the Space Shuttle and the Buran in the previous section, we will now have a closer look at the minimum- and maximum-range capabilities of an unpowered, winged re-entry vehicle. The current case study aims at characterising optimal longitudinal trajectories and longitudinal guidance strategies in the terminal area of such a vehicle. The following questions will be addressed:

1. What are the optimal trajectories to perform a maximum-range and a maximum-dive flight?
2. What is the influence of variations in the initial altitude and velocity in these trajectories?
3. What is the influence of a turning flight on these trajectories?
4. What is the minimum energy required to fly a specific distance?
5. What is the influence of a change in initial position and velocity on this minimum?

It is noted that currently only changes in the initial state are considered. Other dispersions, such as atmospheric or aerodynamic uncertainties, should ultimately be

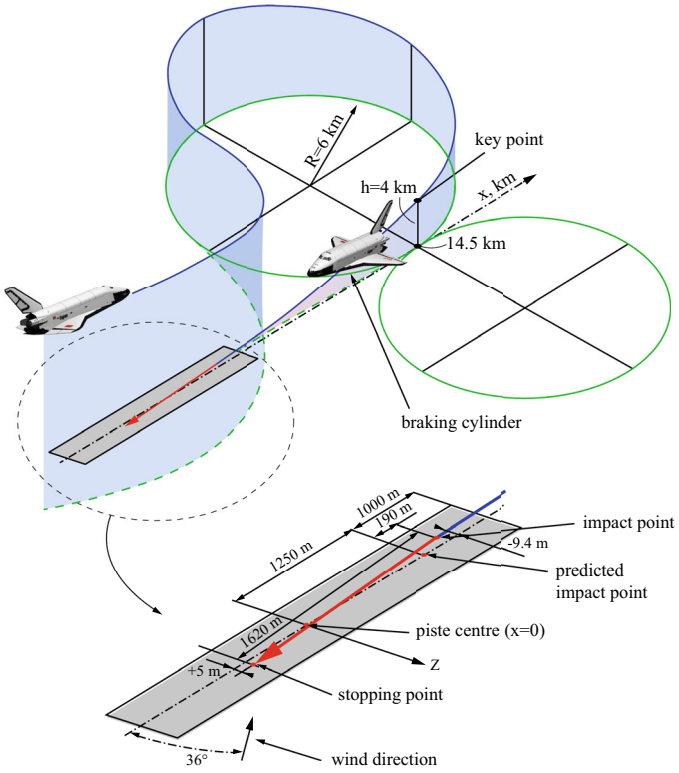


Fig. 8.14 Flight of Buran around the HAC during first flight. Inspired by Kirpischikov (2009)

investigated, but are currently not included. Note that this case study is largely based on two previously published papers (De Ridder and Mooij 2011a, b).

To answer the above research questions we will employ a numerical approach. The problem at hand will be optimised by means of a global optimisation tool. This tool is based on an evolutionary algorithm, for the simple reason that it was available to us. The numerical results will be compared with the simplified models discussed before (steady-state and energy approximations, see Sects. 8.2.3 and 8.2.5), such that conclusions from both can be drawn.

To this end, Sect. 8.4.1 will introduce the specific vehicle characteristics applicable to the terminal area. We will use the HORUS-2B reference model as we did before. The two subsequent Sects. 8.4.2 and 8.4.3 will briefly describe the results obtained with the steady-state and energy-state approximations. The numerical results of the maximum-range trajectories obtained with the global optimiser are presented in Sect. 8.4.4. These optimal trajectories are, of course, also compared with those obtained with the steady-state and energy-state approximations. Section 8.4.5 describes the optimal maximum-dive trajectories, yielding the minimum range the vehicle can fly without overshooting its target. Together with the maximum range,

the so-called energy-tube concept can be conceived (Sect. 8.4.6) that can be used to plan the terminal-area flight. Last but not least, in Sect. 8.4.7 the influence of a turn on the optimal trajectories is investigated. A summary of the major findings concludes this case study in Sect. 8.4.8.

8.4.1 Vehicle Characteristics

The design and analysis of terminal-area missions is driven by the vehicle requirements. These requirements are derived from the vehicle structural limitations, and from the trimming and manoeuvrability restrictions imposed by flight dynamics. The constraints encountered during the terminal area can be summarised as:

1. Flight envelope,
2. maximum load factor, n_z,
3. speed-brake deflection, δ_{sb},
4. angle-of-attack corridor, and
5. maximum bank angle, σ_{max}.

The above constraints will be further detailed below.

Flight Envelope

The flight envelope delineates the region of altitude and airspeed within which a vehicle is allowed to operate, establishing the confines of its altitude-speed operational range. For aircraft, the flight envelope encompasses various boundaries, including the stall limit, maximum design dive speed, maximum design diving Mach number, and maximum flight altitude. However, for a winged re-entry vehicle, the latter two boundaries hold little relevance. The maximum speed is characterised by the maximum dynamic pressure, rather than by stipulating a maximum Mach number. Given that a re-entry vehicle originates from outer space, the altitude boundary becomes irrelevant. Thus, the remaining two boundaries pertinent to a winged re-entry vehicle within the terminal area are the stalling speed and the maximum velocity or maximum dynamic pressure.

Before we move on, we will calculate the performance curves according to the iterative scheme discussed in Sect. 8.2.3. Figure 8.15 shows these performance curves for several altitudes, i.e., 0 km (sea level), 5 km, 10 km and 12 km. Each point on a performance curve corresponds to equilibrium-glide conditions for a constant angle of attack at the specified altitude. Along the performance curve, the angle of attack is changing. As mentioned earlier, the maximum velocity follows from the $\bar{q}_{max}$ constraint, evaluated at the specified altitude. The characteristics of the stationary maximum-range points are give in Table 8.3. Note that only a part of the performance curves can be used, because of the maximum velocity boundary. All the stationary maximum-range points have Mach numbers below 0.8. The table indicates that in a subsonic glide, the equivalent airspeed has approximately a constant value of 116 m/s

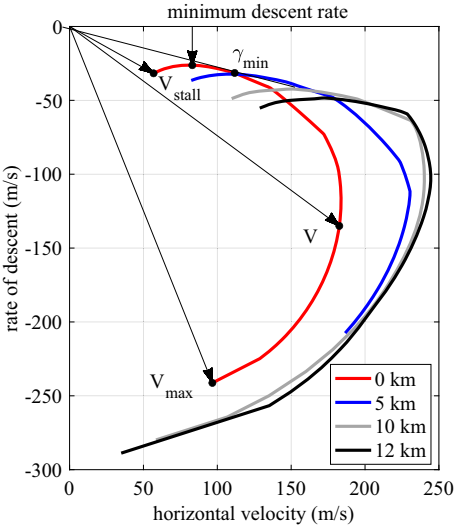


Fig. 8.15 Performance curves of HORUS for different altitudes

Table 8.3 Stationary maximum range points for 0–12 km altitude

Altitude	0 km	5 km	10 km	12 km
Ground velocity (m/s)	116.09	150.10	200.10	231.09
Horizontal velocity (m/s)	111.75	145.43	194.02	223.88
Rate of descent (m/s)	31.47	37.14	48.96	57.28
Angle of attack (deg)	7.52	7.54	7.60	7.50
Flight-path angle (deg)	–15.73	–14.33	–14.16	–14.35
Equivalent airspeed (m/s)	116.09	116.35	114.93	116.63
Mach	0.34	0.47	0.67	0.78

while the true velocity varies. This is consistent with the definition of a quasi-steady flight. The corresponding angle of attack for a maximum-range glide is also approximately constant, and is equal to 7.5°.

The maximum dynamic pressure is associated with the structural safety of the vehicle. No value for HORUS was found in the literature. In a terminal-area study of the Hopper vehicle, a conservative value of 10 kPa was used (Büchner 2003). Based on this number, a somewhat less conservative number of 15 kPa is used for HORUS in this study to be able to investigate a larger number of possible trajectories.

The stalling speed is defined as the minimum speed at which the lift equals the weight or the minimum steady speed in free air under 1-g conditions. Hence, it is the lowest speed at which a force equilibrium is possible. The stalling speed V_{ST} can be calculated from the aerodynamic data of the vehicle with

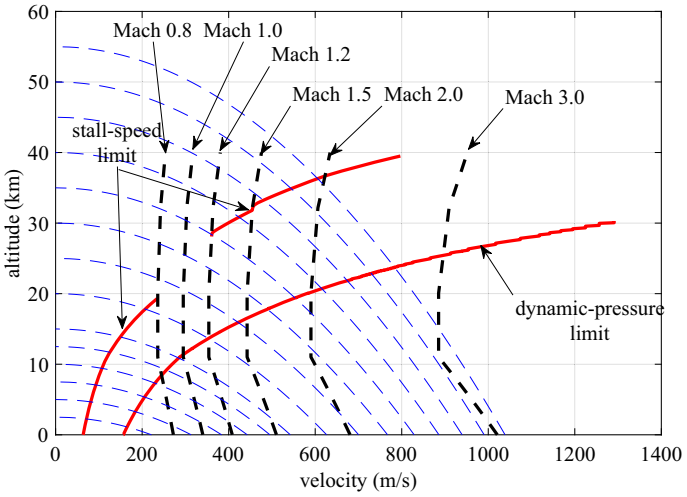


Fig. 8.16 Flight envelope of HORUS, in between the stall limit and the maximum dynamic pressure. The dashed blue lines indicate curves of constant energy height

$$V_{ST} = \sqrt{\frac{mg}{S}\frac{2}{\rho}\frac{1}{C_{Lmax}}\frac{\cos\gamma}{\cos\sigma}} \tag{8.49}$$

This equation indicates that in a banked flight situation, the stalling speed is larger than in a wings-level flight situation.

Based on the aerodynamic data presented by MBB (1988), it is evident that the maximum lift coefficient varies with Mach number. Determining the stalling speed involves initiating with a suitably high velocity at a particular altitude and progressively reducing this velocity until the required lift coefficient for steady-glide conditions matches the maximum lift coefficient. Once reached, further reduction in velocity becomes unattainable, as the required lift coefficient surpasses the maximum available lift coefficient.

Figure 8.16 illustrates the computed boundaries of the flight envelope in the (V, h) space, or the energy space. Additionally, the figure includes various Mach lines denoting distinct Mach regimes. The stall-speed limit depicted in this figure corresponds to the stall speed during wings-level flight. However, during a turn with a specified bank angle, this limit shifts to the right (higher velocities). Furthermore, the stall speed is calculated based on a vehicle mass of 26,029 kg; any alteration in mass necessitates a recalculated stall-speed limit. It is important to acknowledge that theoretically, HORUS has the capability to operate beyond the stall limit. Nonetheless, the stall limit signifies the lowest velocity achievable in a steady-state flight scenario, implying that transient or dynamic manoeuvres with velocities lower than the steady-state stall speed are feasible, albeit improbable to be used by HORUS.

Load Factor

In describing the aerodynamic loads on a vehicle, the so-called load factor is used. This parameter refers to the ratio of the resultant aerodynamic force to the weight of the airplane, for a specific velocity. The increase in lift is of prime significance and hence, the load factor is usually obtained by dividing the lift by the weight of the vehicle:

$$n = \frac{L}{mg} \tag{8.50}$$

Loads exerted on an airborne vehicle stem from two primary sources: manoeuvring loads, induced by the vehicle's acceleration due to control actions, and gust loads, incurred from encountering atmospheric turbulence. As the present study does not incorporate wind or gusts, gust loads are not considered. Manoeuvring loads occur due to alterations in the angle of attack, with increased angles of attack typically observed during turns to maintain force equilibrium in banked flight situations, consequently elevating the load factor. Ordinarily, the permissible manoeuvring load varies with velocity, delineating a manoeuvring load envelope. However, since such an envelope is unavailable for HORUS, a constant value of 2.5 for the maximum load factor is employed, a value comparable to that of the Space Shuttle (Kraemer and Ehlers 1975).

Speed-Brake Deflection

The deflection of the speed brakes is also connected with the structural safety and with allowed hinge moments. The speed brakes of HORUS are only used at subsonic velocities, $M < 0.95$. The maximum (outward) deflection angle is $\delta_{sb_{max}} = 40°$.

Angle-of-Attack Corridor

The elevon deflection used for trim purposes is limited, because a sufficient margin in elevon deflection should remain to control the vehicle. Therefore, the allowable angle-of-attack interval is constrained due to limitations in trim range. MBB (1988) states that these restrictions are a function of Mach number. Figure 8.17 presents the restrictions on the angle-of-attack with a linear interpolation between the data points. In the transonic regime, a narrow angle-of-attack corridor exists: the angle of attack cannot be larger than 10°.

Bank-Angle Limit

There are no predefined bank-angle limits available for HORUS in existing literature. However, for the purpose of this study, realistic bank-angle limits are adopted following a scheme akin to that of the Space Shuttle. When the Mach number exceeds 0.95, the maximum allowable bank angle is capped at 30°. Conversely, below a Mach number of 0.75, the maximum bank angle varies based on the phase of operation for HORUS. Refer to Table 8.4 for phase-dependent values of the maximum bank angle (and Sect. 8.3.2 for an explanation of the various phases). For Mach numbers between 0.75 and 0.95, a linear interpolation approach is employed. Additionally, alongside the maximum bank angle, restrictions are imposed on the rate of change of the bank angle. This limit has been set to $\dot{\sigma} = 15°/s$.

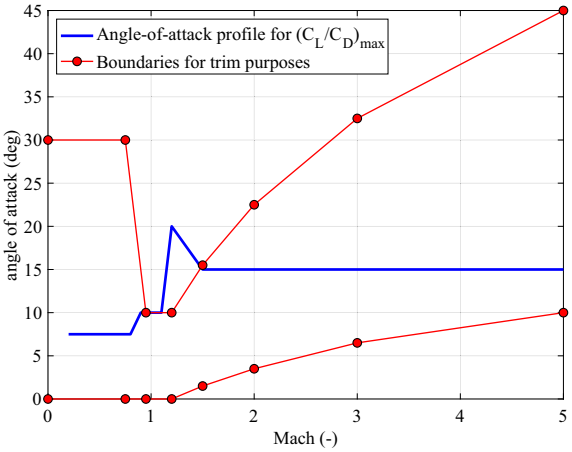


Fig. 8.17 Angle-of-attack boundaries due to trim considerations, based on data from MBB (1988)

Table 8.4 Phase dependent bank angle limits for $M < 0.75$

Phase	Bank-angle limit [deg]
Acquisition	50.0
Heading alignment	60.0
Pre-final	30.0
2nd acquisition	60.0

8.4.2 Steady-State Approximation

From the theoretical description of the steady-state approximation (Sect. 8.2.3) we saw that in a steady glide, the maximum range is flown with a maximum ratio between the horizontal velocity and the negative of the rate of descent, yielding a minimum flight-path angle. The requirement of minimum flight-path angle was subsequently translated into a maximum lift-to-drag ratio, L/D. Hence, a common interpretation of the steady-state approximation is to establish an angle-of-attack profile for a maximum L/D, as a function of velocity or Mach number, and use this profile to calculate the maximum range.

Using this angle-of-attack profile, the range has been calculated for three subsonic and three supersonic initial conditions by numerically propagating the trajectory. The three subsonic initial conditions all have an identical initial energy, but with a different distribution of kinetic and potential energy, i.e., a different combination of altitude and velocity. The same holds for the three supersonic points. The initial flight-path angle for subsonic initial conditions is the flight-path angle required for a steady-state glide at the initial velocity, while the initial flight-path angle for supersonic initial conditions is set to –10°, to match the flight-path angle at the end of the

Table 8.5 Initial conditions and obtained range using $\alpha_{(L/D)max}$

	V [m/s]	h [km]	γ [°]	$\mathcal{E}_h$ [km]	M [–]	Range[km]
Initial state 1	108.50	9.40	–29.24	10.00	0.36	32.78
Initial state 2	171.60	8.50	–14.55	10.00	0.56	34.88
Initial state 3	228.00	7.35	–17.65	10.00	0.73	33.76
Initial state 4	583.10	35.00	–10.00	52.51	1.89	134.63
Initial state 5	661.64	30.00	–10.00	52.51	2.20	136.35
Initial state 6	732.00	25.00	–10.00	52.51	2.45	130.79

hypersonic descent (Sect. 7.6). The initial conditions and the corresponding ranges using $\alpha_{(L/D)max}$ are given in Table 8.5.

The results for subsonic flight indicate that although the initial energy height is equal for all three simulation cases, the obtained maximum range is different for the three trajectories. It seems that the distribution of the total energy in kinetic and potential energy is important in determining the maximum range, so one should, in fact, be able to find an optimal value for the distribution of the total energy. Hence, it is not possible to only specify the minimum required energy height for a specific target range, but it is also necessary to specify for which combination of altitude and velocity this is valid. This is an important conclusion for studying the maximum range and required energy during off-nominal conditions.

For the supersonic initial conditions, similar findings are observed: depending on the distribution of the total energy in potential and kinetic energy the maximum attained range differs. Moreover, inspecting the velocity and flight-path angle profiles, using the angle-of-attack profile for a maximum lift-to-drag ratio, indicates that these are not constant. This is in violation with the original steady-state assumption. Although it is acceptable to use a steady-state approximation for relatively low velocities, apparently one cannot extend the use of an angle-of-attack profile for maximum lift-to-drag ratio from subsonic to supersonic velocities without any proof that this really results in a maximum-range flight.

Because a winged re-entry vehicle enters the TAEM phase at about Mach 2.5 and needs to dissipate the kinetic energy excess to meet with the required final TAEM conditions at about Mach 0.5, the vehicle performs a decelerating gliding flight and hence, it is a dynamic process. Due to the presence of this dynamic term, it is not always possible to solve the maximum-range flight as a sequence of static problems, and one is forced to resort to dynamic optimisation techniques.

Based on these conclusions, some questions can be raised:

- Is a flight with an angle-of-attack profile for a maximum lift-to-drag ratio the correct strategy that results in a maximum range?
- If not, can another strategy be found, which will allow HORUS to fly a larger distance than the one obtained using the angle-of-attack profile for maximum lift-to-drag ratio?

To address these questions, in Sect. 8.4.4 the maximum range of HORUS for the same six starting points from Table 8.5 will be determined by means of trajectory optimisation. This will provide the optimal control vector, in terms of angle of attack, that yield the maximum range.

8.4.3 Energy-State Approximation

In Sect. 8.2.5, where we discussed the energy-state approximation it was shown that for maximum range the vehicle has to fly with minimum drag. In particular, Fig. 8.6 shows a typical maximum-range trajectory for a supersonic aircraft. The contours of constant drag are obtained by calculating the drag level at each combination of altitude and velocity throughout the (V, h)-space. Due to the Mach dependency of lift and drag coefficients, this is an iterative process. As an example, in Fig. 8.18 the contours of constant drag for HORUS in the subsonic velocity regime are shown.

This figure indicates that for a specific energy level, there is one combination of altitude and velocity for which the drag encountered by the vehicle is minimal. Hence, at each energy level in the energy space a trajectory for which the drag is minimal can be defined. Note that Fig. 8.8 shows a 3D-plot of the contours of constant drag for the supersonic aircraft that was used by Bryson et al. (1969) to derive the energy-state approximation theory.

Figure 8.19 shows the subsonic and supersonic trajectories, obtained in the previous section, in the energy space (V, h space). The contours of constant drag in the

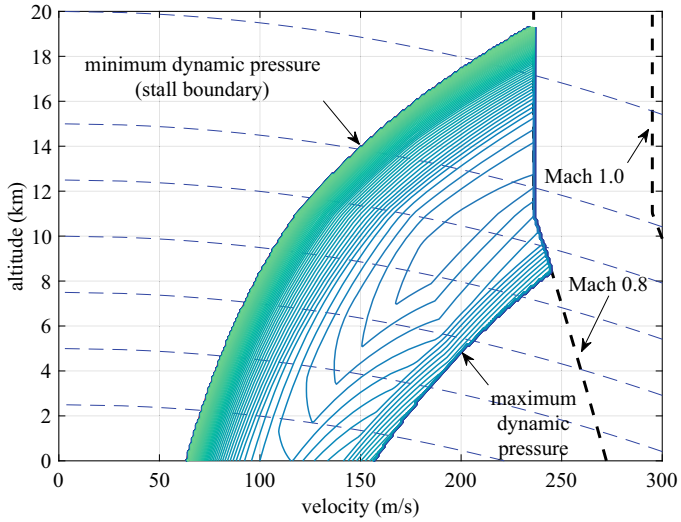
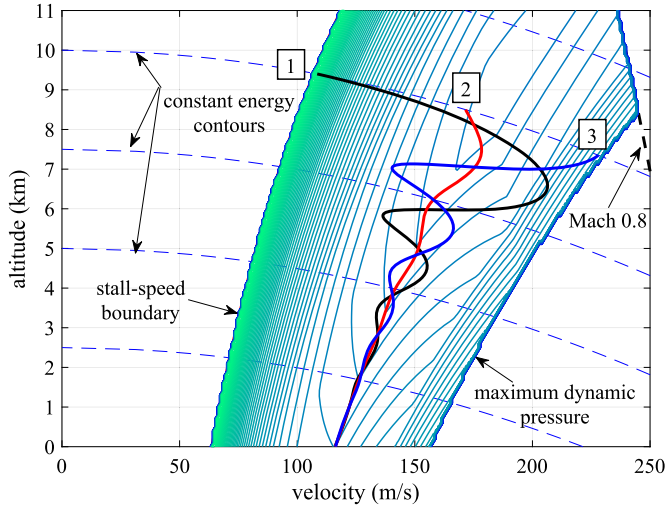
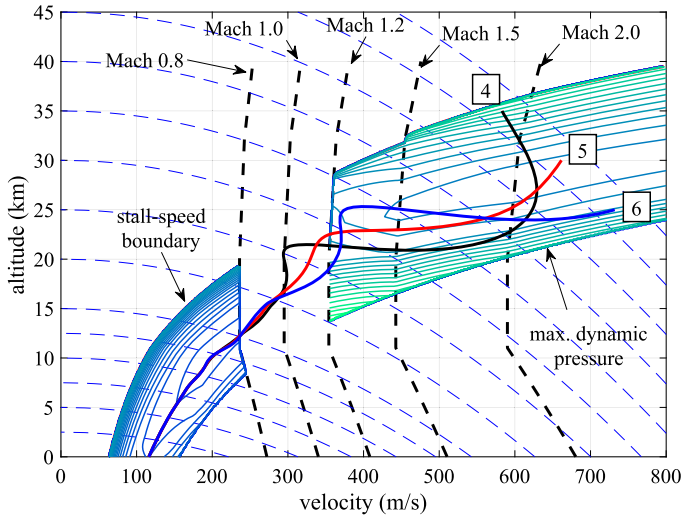


Fig. 8.18 Contours of constant drag for HORUS in the subsonic velocity regime



(a) Subsonic



(b) Supersonic

Fig. 8.19 Trajectories flown with $\alpha_{(L/D)max}$-initial conditions according to Table 8.5

energy space for HORUS are calculated between the maximum dynamic-pressure boundary and the steady-state stall speed. No contours of constant drag are shown in the transonic speed regime, because the vehicle capabilities in this region are insufficient to have an equilibrium flight situation. One could just say that a “falling” motion is performed in the transonic regime.

The trajectories using $\alpha_{(L/D)max}$ for the initial subsonic points fluctuate around the drag valley and are damped out near the end of the flight. These oscillations are induced when starting at an initial state (altitude and velocity) away from the drag valley and flying with a constant angle of attack corresponding with a maximum L/D (= 7.5° in the subsonic phase). A time-varying angle of attack could improve the trajectory response and prevent the oscillatory behaviour. In case of supersonic initial points, there are no fluctuations when entering the subsonic region, as they have already damped out. Also in the supersonic region the fluctuations are less, because of the lower sensitivity to an off-nominal angle of attack. The main cause of the oscillations is the low lift-to-drag ratio, which prevents a quick transient response, and therefore induces an overshoot of the equilibrium state and only a slow recovery. A similar simulation with the supersonic aircraft used by Bryson et al. (1969) shows far less oscillations due to the much larger L/D and thus a much better damped transient response.

8.4.4 Optimal Trajectories in the Vertical Plane

The steady-state approximation and energy-state approximation are used to approximate the "real" optimal solutions. To compare the approximations with the optimal solution, an evolutionary algorithm (EA) is used as an optimiser to find the optimal control vector (angle of attack) for optimal trajectories in the vertical plane, using the equations of motion given by Eqs. (8.14)–(8.19).

An EA is a global optimiser which mimics nature's own search system, i.e., evolution, to find an optimum (Michalewicz 1996), and is an easy-to-use tool to explore a design region of interest. The working principle of the EA is based on the principle of survival of the fittest. An EA starts with generating an initial population. Using several operators, such as crossover and mutation, the fittest individuals of the population are chosen to produce offspring. After several generations, the population has evolved into fitter individuals and the evolution is stopped when a pre-defined stop criterion is met. For this case study, the EA is used to optimise a control vector, which represents discrete values of the angle of attack at fixed discrete values of the energy height (the number of nodes was chosen such to balance accuracy and computational load):

$$\mathcal{E}_{h(subsonic)} = [0.0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0]\ \text{km}$$

$$\mathcal{E}_{h(supersonic)} = [0.0, 1.0, 10.0, 15.0, 20.0, 25.0, 30.0, 35.0, 40.0, 45.0, 50.0, 55.0]\ \text{km}$$

The first grid is used for subsonic initial states, while the second grid is used for supersonic initial states. Between these discrete values, linear interpolation is used. The energy height was chosen as the independent variable, because for a gliding re-entry vehicle it is continuously decreasing, while both altitude and velocity can

Table 8.6 Obtained maximum optimal range for HORUS and improvement with respect to steady-state solution

	Range $\alpha_{optimal}$ [km]	% improvement
Initial state 1	34.87	6.38
Initial state 2	35.97	3.13
Initial state 3	35.18	4.21
Initial state 4	138.93	3.19
Initial state 5	138.80	1.80
Initial state 6	132.95	1.65

either decrease or increase. Also time was not used, because the time at the end of the TAEM is unknown before the simulation, whereas the energy is supposed to reach a specified value at the ALI. The approach and landing phase is outside the scope of the current discussion. Therefore, for simplicity, it is assumed that the final energy is zero (zero altitude and zero relative velocity), and the trajectory does not stop at ALI, but continues until the vehicle touches the ground (touchdown). Although energy height is chosen as the independent value to specify the angle-of-attack profile, time is used as the independent value during trajectory simulations.

The optimal trajectory problem is subjected to several control-variable and path constraints: α (see Fig. 8.17), $\dot{\alpha}$ (3°/s), stall speed and maximum dynamic pressure (e.g., Fig. 8.18), and maximum load factor ($n = 2.5$).

The optimal maximum ranges, obtained using the EA, for the same initial points in the (V, h) space from Table 8.5, are given in Table 8.6. These results show that the range can be improved by 3–6% and 1–3% with respect to the subsonic and supersonic $\alpha_{(L/D)max}$ trajectories, respectively.

The resulting profiles of subsonic optimal angle-of-attack are depicted in Fig. 8.20, alongside $\alpha_{(L/D)max}$ (= 7.5°). Throughout the majority of the optimal subsonic maximum-range glide, the angle of attack closely approximates $\alpha_{(L/D)max}$, except during the initial and final phases of the flight. Corresponding optimised trajectories in the energy space are illustrated in Fig. 8.21, revealing that optimal solutions involve initial manoeuvres to align with the drag valley, followed by adherence to this valley until executing a final flare manoeuvre. Contrasting these trajectories with those obtained using $\alpha_{(L/D)max}$ (Fig. 8.19), it is evident that optimal trajectories exhibit fewer fluctuations around the drag valley. Nonetheless, minor fluctuations persist due to the EA's global optimisation approach, which often yields solutions close to optimal with less-than-ideal convergence properties.

Performing a local optimisation that uses the EA result as an initial estimate would aid in mitigating the fluctuations (De Ridder and Mooij 2009). This is demonstrated in Fig. 8.22, showcasing two maximum-range glide trajectories for the supersonic aircraft. To mitigate the fluctuations inherent in the optimal solution obtained via the EA, a local Monte-Carlo analysis is conducted on the optimal α-profile. In this anal-

ysis, 1000 new angle-of-attack profiles are randomly generated, the corresponding trajectories propagated, and the profile yielding the maximum range is chosen as the new optimal one. Each of the 1000 α-profiles is generated by extracting the α values at the grid points–obtained using the EA–and additional intermediate points, followed by the addition of a small random $\Delta\alpha$ value. All α values are adjusted simultaneously, with a distinct random $\Delta\alpha$ applied at each grid point. Figure 8.22 demonstrates that fluctuations are mitigated through local optimisation post the EA solution. Only during the initial flight phase do fluctuations persist. While local optimisation can enhance the solution, the optimal EA solutions for HORUS, as depicted in Fig. 8.21, already clearly indicate their alignment with the drag valley. Despite minor fluctuations, it can be inferred that the energy-state approximation effectively addresses the subsonic maximum-range problem of a winged re-entry vehicle, representing an improvement over the steady-state approximation.

As depicted in Fig. 8.23, the optimal solutions obtained for initial supersonic conditions deviate from the drag valley. This behaviour can be attributed to two factors. Firstly, trajectories commencing above the drag valley best illustrate this deviation, which stems from the interplay between vehicle characteristics and high velocities. Optimal trajectories for the aforementioned supersonic aircraft demonstrate a slight initial overshoot of the drag valley, followed by alignment with it until the final flare manoeuvre. Notably, this aircraft has relatively good gliding capabilities, with maximum L/D values of 10.3 and 7.6 for subsonic and supersonic regimes, respectively. However, when adapting the original supersonic aircraft aerodynamics to exhibit maximum L/D values of 4.3 and 3.2 for subsonic and supersonic conditions, respectively, optimal trajectories for this modified supersonic aircraft with

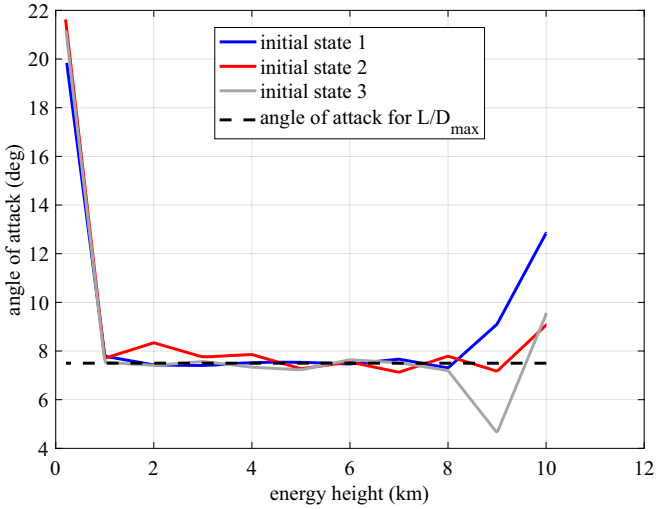


Fig. 8.20 Optimised subsonic angle-of-attack profiles compared to $\alpha_{(L/D)max}$ (= 7.5°). Data taken from De Ridder and Mooij (2011b)

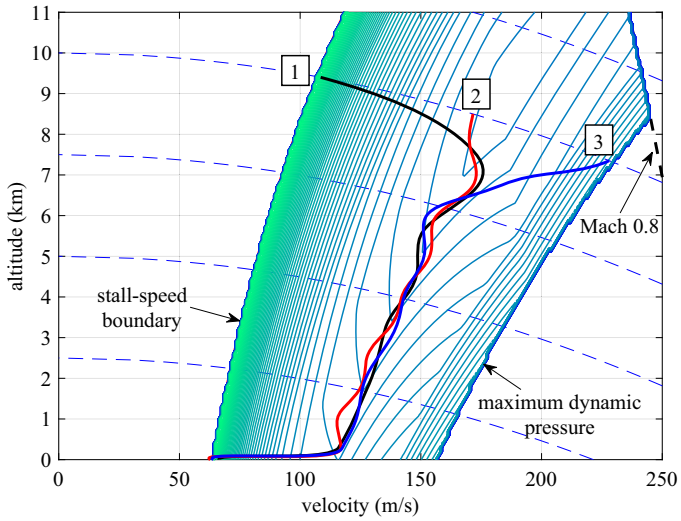


Fig. 8.21 Subsubsonic optimal maximum-range solutions. Data taken from De Ridder and Mooij (2011b)

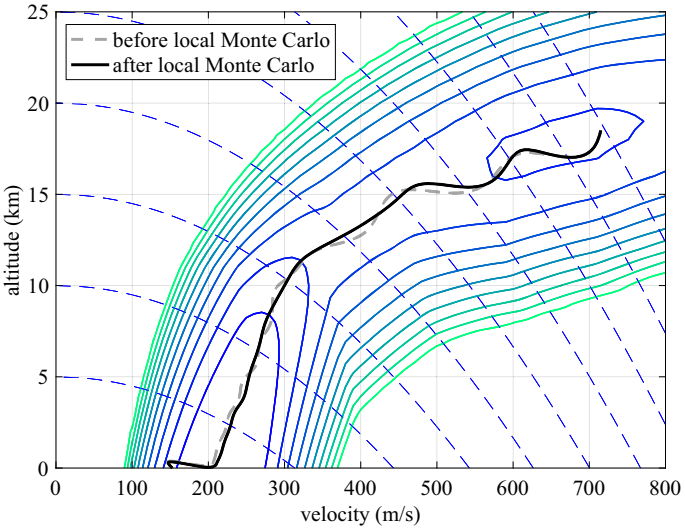


Fig. 8.22 Maximum range glide trajectory of supersonic aircraft, before and after local Monte Carlo. Data taken from De Ridder and Mooij (2011b)

diminished gliding capabilities significantly overshoot the drag valley before converging towards the maximum range path. This observation aligns with the optimal supersonic trajectories of HORUS, which initially overshoot the drag valley due to excessive energy expenditure required for manoeuvring and subsequent reduction of

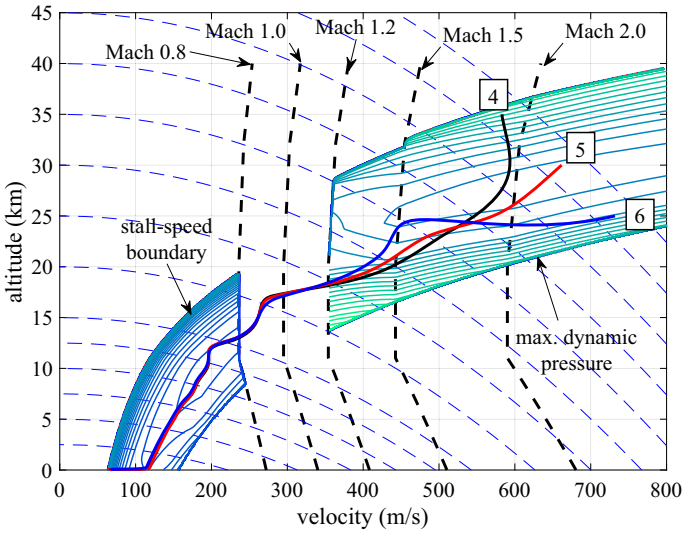


Fig. 8.23 Supersonic optimal maximum-range solutions. Data taken from De Ridder and Mooij (2011b)

the overshoot. For further insights, the reader is referred to the paper by De Ridder and Mooij (2009).

For supersonic aircraft with relatively robust gliding capabilities, initial overshoots were negligible in the context of the overall trajectory. Similarly, for vehicles exhibiting poor gliding capabilities at subsonic speeds, these overshoots remained minor. However, when operating at supersonic velocities, transient manoeuvres and overshoots become significant factors. During the requisite pull-up manoeuvre to track the drag valley without overshooting, the encountered drag far exceeds that experienced in the optimal trajectory. Consequently, the drag contours computed using the angle of attack necessary for equilibrium become invalid. Nevertheless, the overarching principle of minimising drag for maximising range remains valid, albeit these manoeuvres are executed at varying angles of attack.

A second rationale for deviating from the drag valley pertains to the optimal trajectory within the transonic velocity regime. As depicted in Fig. 8.23, all three optimal trajectories commence the transonic phase with identical state conditions in terms of altitude and velocity. Throughout the transonic and subsonic phases, these trajectories exhibit near-identical behaviour. Hence, it can be concluded that the path delineated through the transonic regime represents the optimal solution, with the trajectory in the supersonic segment tailored to attain these optimal conditions at the transonic entry point.

8.4.5 *Maximum-Dive Trajectories in the Vertical Plane*

Achieving maximum range entails flying at the shallowest flight-path angle while minimising drag. Consequently, flight configurations involving energy-intensive or diving manoeuvres necessitate larger flight-path angles or higher drag levels compared to those during maximum-range flight. Examining the performance curves of HORUS (Fig. 8.15) alongside contours of constant drag in the energy space reveals that a larger flight-path angle and drag level can be attained by flying either at a lower or higher velocity. This is contingent upon whether the angle of attack surpasses or falls short of the maximum L/D angle. However, in the context of executing a maximum dive, it is customary to lower the vehicle's nose, prompting many maximum-dive strategies to advocate flight at maximum dynamic pressure (Da Costa 2003; Büchner 2003; Hull et al. 2005).

Using the EA, optimal maximum-dive trajectories (aiming to minimise range) are computed for two subsonic initial points (initial points 1 and 2), as shown in Table 8.5. The EA seeks the trajectory with the smallest range. Figure 8.24 depicts the resulting optimal-dive trajectories when no speed brakes are utilised. Achieving the maximum dive involves flying at the minimum dynamic pressure or stall-speed boundary and subsequently accelerating (by lowering the nose of the vehicle) towards the end of the gliding phase. Some fluctuations are noticeable in the optimal diving trajectory for initial point 2. This can be explained by the fact that the global solution has not yet fully converged, leaving room for potential improvements.

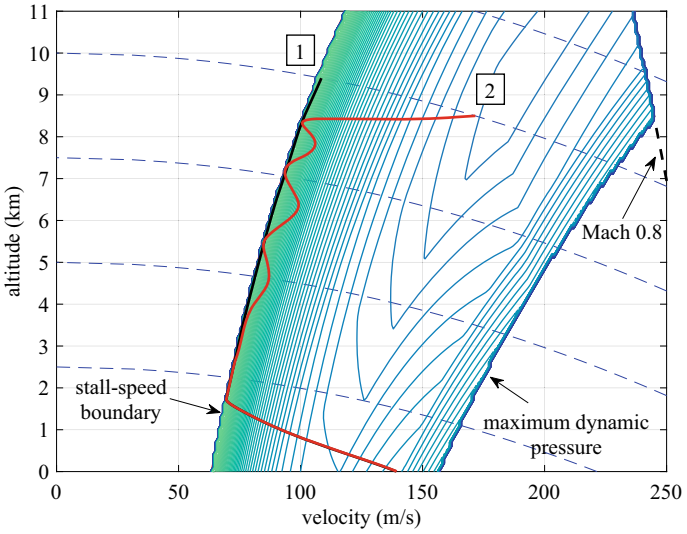
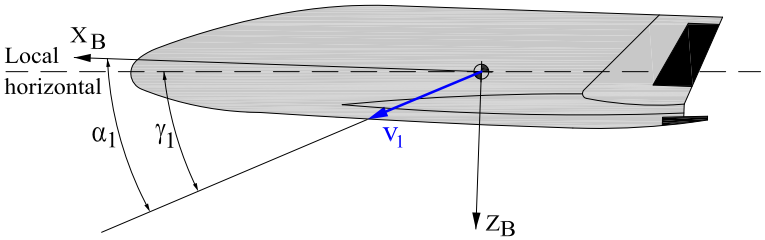
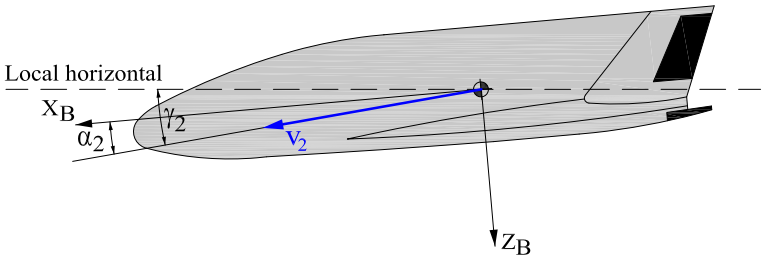


Fig. 8.24 Optimal maximum-dive trajectory without the use of speed brakes. Data taken from De Ridder and Mooij (2011b)



(a) Dive at minimum dynamic pressure



(a) Dive at maximum dynamic pressure

Fig. 8.25 Difference between the steady-state dive situation at minimum and maximum dynamic pressure

To explain why the optimal maximum-dive solution involves flying at a minimum dynamic pressure, the drag contours for a steady-state flight situation are also depicted in Fig. 8.24. At the minimum dynamic pressure boundary, the encountered drag is higher compared to when flying at the maximum dynamic pressure boundary due to the large angle of attack (resulting in large lift and drag coefficients). Figure 8.25 compares the steady-flight situations at minimum and high dynamic pressures. While at minimum dynamic pressure the velocity is lower, the flight-path angle is larger, and the nose is raised above the horizon.

Although the optimal trajectories for a maximum dive indicate that a minimum range is obtained by flying at a minimum dynamic pressure, we would still use a maximum dive at maximum dynamic pressure for three reasons:

1. It is the natural instinct to lower the nose of the vehicle and to perform a dive at a high velocity,
2. in the high-nose situation, astronauts do not have a clear view on the ground and the runway, and they may need this as a reference, and
3. a disadvantage of flying at a low dynamic pressure is the inherent reduction in control authority, which makes it difficult to reject wind disturbances.

It is therefore safer to perform a maximum dive at a high dynamic pressure. However, the solution of flying at minimum dynamic pressure might still be used as a back-up in case a dive at maximum dynamic pressure is insufficient.

Kluever (2007) applied a minimum dynamic pressure as the reference dynamic pressure to generate optimal TAEM trajectories in case of effector failures. It is observed that reducing the velocity can decrease the turn radius, thereby mitigating the impact of limited bank capabilities in case of these failures. Additionally, it is asserted that achieving the minimum dynamic pressure involves flying at a relatively shallow flight-path angle, which typically extends the vehicle's overall range (Kluever 2007). However, our investigation suggests that for HORUS, flying at minimum dynamic pressure occurs at a larger flight-path angle, which, in fact, diminishes the range.

8.4.6 Energy-Tube Concept

In the previous sections, questions related to the maximum range that can be achieved, given a specific initial state, were answered. But for the purpose of energy management and estimation, it is more valuable to know the minimum amount of energy that is required to fly a specific ground track (range).

It is customary to employ a singular value for the minimum energy height. If the vehicle's energy state drops below this threshold, reaching the targeted range becomes unattainable (Büchner 2003). However, both steady-state and optimised solutions reveal that despite identical initial energy heights, different maximum ranges are achieved, as depicted in Table 8.5. This phenomenon is also evident in Fig. 8.26, which illustrates all initial points (shaded area) capable of achieving at least 35.0 km, signifying the minimum required energy, i.e., the maximum-range boundary. Should an initial state lie below this boundary (indicating lower energy), reaching the target range of 35.0 km, even with a maximum-range strategy, becomes unfeasible. The figure highlights that not all combinations of initial altitude and velocity, with an initial energy height of 10.0 km, will yield a maximum range of at least 35.0 km, as the maximum-range boundary does not align with lines of constant energy. This observation is crucial, since it implies that in scenarios where a guidance system must handle off-nominal conditions, a single minimum required energy value for a particular range cannot be specified for all initial states. Therefore, besides specifying the minimum required energy for a specific range, it is imperative to define the altitude and velocity combinations for which this holds true. The same principle applies to the maximum-energy boundary.

In general, these capability parameters of the vehicle can be expressed by the so-called energy-tube concept. All the combinations for which it is possible to fly a desired range, without overshoot or undershoot, form an energy area in the energy space. This energy area is the cross-section of the energy tube corresponding with the desired range. For a more general situation, Fig. 8.27 shows the energy tube cross-

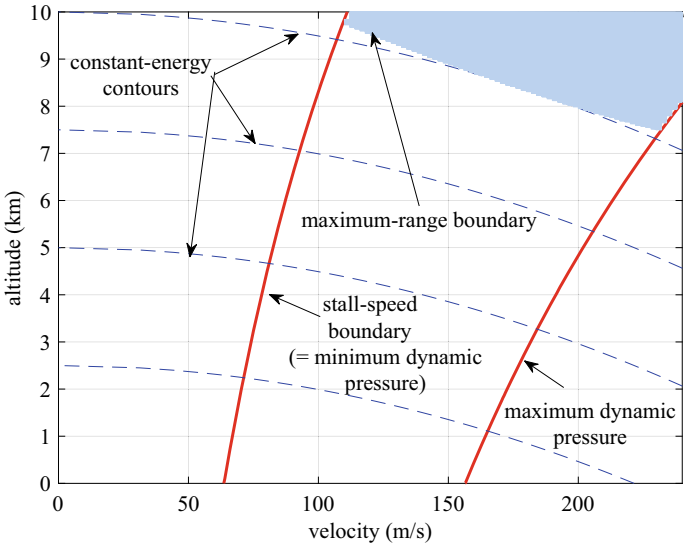


Fig. 8.26 Combinations of initial altitude and velocity which yield a maximum range of at least 35.0 km (shaded area), indicating the maximum-range boundary. Data taken from De Ridder and Mooij (2011b)

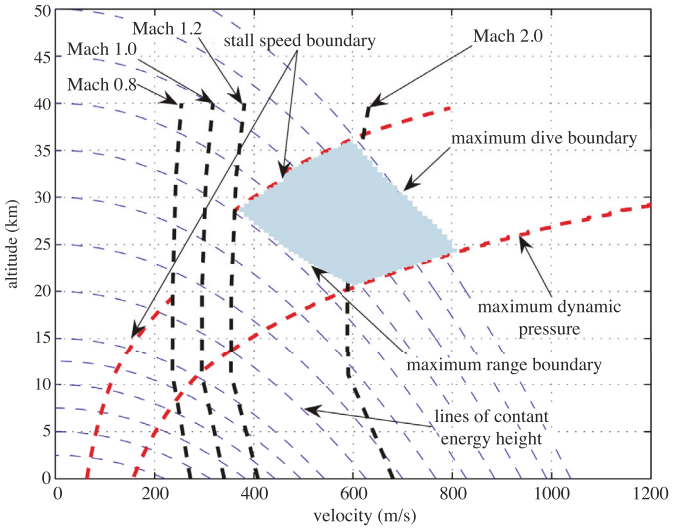


Fig. 8.27 Flight envelope of HORUS and cross-section of energy tube (shaded area) for a target range of 100 km. Data taken from De Ridder and Mooij (2011b)

section for a target range of 100 km, together with the corresponding maximum-range and maximum-dive boundaries.

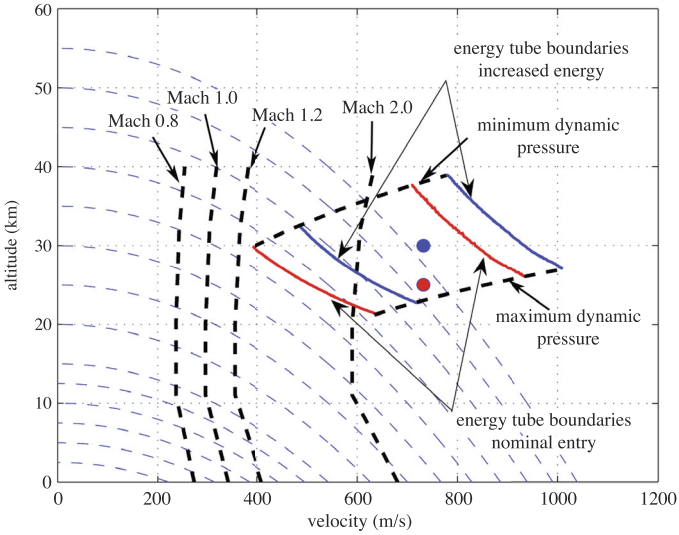


Fig. 8.28 Energy tube cross-section for the nominal TAEM entry point and energy high situation. The RED dot indicates the nominal entry point, while the BLUE dot indicates an entry point with a higher energy (altitude = 30 km). The minimum and maximum dynamic-pressure boundaries do not change. Data taken from De Ridder and Mooij (2011b)

For an initial state located on the maximum-range boundary, achieving the 100 km target range necessitates employing a maximum-range strategy, while an initial state on the maximum-dive boundary mandates a maximum-dive strategy to avoid overshooting the target range. Between these extremes, different specific strategies are required to precisely attain a range of 100 km, contingent on the position within the cross-section. If a point is selected above the maximum-dive boundary, HORUS will exceed the target range, even with the maximum-dive strategy in effect. Conversely, choosing a point below the maximum-range boundary renders HORUS incapable of reaching the target due to insufficient maximum-range capabilities, resulting in undershooting.

The concept of the energy-tube cross-section holds significance for vehicle design applications. It defines the vehicle's capability to safely reach a specific target under various off-nominal conditions. A wider cross-section implies a broader range of off-nominal conditions the vehicle can accommodate. Modifying the vehicle's aerodynamic characteristics, such as gliding capabilities, will influence these cross-sections, allowing for tailored performance. Beyond its use as a pre-flight analysis tool, this concept can inform on-board guidance algorithms by estimating the requisite energy for reaching the target, assessing the current position within the energy tube vis-a-vis boundary capabilities, and facilitating the calculation of an optimal HAC location using a dedicated HAC-planning algorithm.

A typical illustration of HAC planning involves considering the following scenario: Fig. 8.28 depicts the energy-tube cross-section for a case where the vehicle possesses higher potential and/or kinetic energy, observed at an altitude of 30 km and the nominal TAEM entry point. The blue dot represents this energy-high case. If the optimal HAC position for the nominal entry is used, denoted by the red line delineating the energy-tube cross-section boundary, the energy-high case (blue dot) aligns closely with the maximum-dive capabilities. However, employing the original optimal HAC position for the nominal entry point becomes problematic for the high-energy case. Doing so would necessitate deviating from the nominal longitudinal strategy towards a strategy requiring greater energy dissipation, closer to the maximum dive strategy, to avoid overshooting the runway threshold. This poses an issue as HORUS would then have reduced capability to respond to off-nominal conditions during subsequent phases of TAEM, which might demand strategies involving varying degrees of energy dissipation. Hence, a HAC-planning algorithm becomes essential to adjust the HAC position until the new initial state once again falls between the maximum-dive and maximum-range capabilities.

By dynamically adjusting the HAC, the planning algorithm ensures that HORUS commences the TAEM phase within the energy tube's "sweet spot", thereby maximising its ability to respond to situations requiring either more or less energy dissipation. Furthermore, employing a variable HAC configuration expands the effective total energy tube cross-section compared to a single fixed HAC position. Further insights on the energy-tube concept and its use in trajectory-planning algorithms can be gained from the paper by De Ridder and Mooij (2011a).

8.4.7 Influence of a Turn on the Optimal Trajectories

In several papers the same reference velocity or dynamic pressure is used for both a wings-level and a turning flight (Büchner 2003; Hull et al. 2005; Kluever 2007). In this section, the influence of a turn on the optimal trajectories and reference dynamic pressure is investigated.

In the current study, the optimal turning problem for a winged re-entry vehicle during the terminal area is formulated as follows: given a HAC with a specific constant radius, what is the optimal longitudinal strategy (angle of attack), such that minimal energy is dissipated during the turn around that specific HAC?

Using the EA, only the angle of attack is optimised for a specific HAC radius, while the required bank angle is calculated by a lateral tracking algorithm. The initial starting point of the flight is situated on the HAC surface and after turning 270° around the HAC, a straight flight is executed. During this straight flight, a maximum-range flight is performed. Executing a turning flight with minimal energy dissipation entails preserving the maximum available energy post-turn. Subsequently, if a wings-level maximum-range flight is pursued following the turn, the criterion shifts from minimum-energy dissipation during the turn to a maximum-range criterion post-turn. While the initial energy at the turn's outset is predetermined, forecast-

ing the final energy value at the turn's conclusion is unfeasible. Therefore, energy cannot serve as the independent variable during the optimal turn problem. Instead, the heading is used.

After the optimal trajectories are obtained by executing the EA, an explanation for the shape of the resulting optimal trajectories can be obtained from the energy-state approximation. Bryson et al. (1969) and Vinh (1981) exclusively employed the energy-state approximation to approximate the maximum-range problem in the vertical plane. This section extends the energy-state approximation to address the challenge of optimising a turn around a circular HAC. During such a turn, with a fixed radius, the ground-track range covered remains constant. Consequently, the optimisation criterion of minimising energy dissipation during the HAC can be reformulated as the minimisation of energy dissipation concerning range around the HAC: $(\frac{\mathrm{d}\mathcal{E}h}{\mathrm{d}R_f})_{min}$. This criterion mirrors the maximum-range criterion in the energy-state approximation for wings-level maximum-range flight. Thus, the same methodology is applicable to the optimal-turn problem around a HAC.

Using Eq. (8.36) and assuming a small influence of the flight-path angle, the drag has to be minimised during the turn to dissipate the least amount of energy. The banked flight has to be taken into account during the drag calculation and hence, the contours of constant drag must be recalculated for a specific radius of the HAC. The bank angle now has an influence on the required angle of attack for an equilibrium stationary-flight situation, see Eq. (8.22). The larger the bank angle, the larger the required angle of attack (lift) to maintain vertical equilibrium at an equal velocity. Also, throughout the energy space, different bank angles are required to have a constant HAC radius. From Eq. (8.48) we get an expression for the required steady-state bank angle to fly a specific turn radius, which changes with velocity and flight-path angle:

$$\tan\sigma = \frac{V^2\cos\gamma}{g_0 R_{turn}} \tag{8.51}$$

The optimal turn around a HAC with a radius of 5 km is evaluated for three initial states, see Table 8.5. For ease of comparison, the three initial states all have an initial flight-path angle of –10°. Figure 8.29 shows an example of an optimal turn around the 5 km radius HAC and the straight flight afterward.

Figure 8.30 shows both the resulting optimal turning trajectories starting at three different initial states (altitude and velocity) for a HAC radius of 5 km, obtained using the EA, and the drag contours (5 km HAC radius) in the energy space. The optimal trajectories are located in the new drag valley for a banked situation and initial manoeuvres are performed to intercept the new drag valley. The figure also indicates the equivalent airspeed both in the drag valley for a 5 km radius HAC and during a maximum-range wings-level flight. The drag valley for the banked situation is shifted to a lower equivalent airspeed with respect to the maximum-range path. Hence, the optimal turn is performed at a lower velocity with respect to the optimal maximum-range velocity. At the end of the turn the strategy changes to one with the

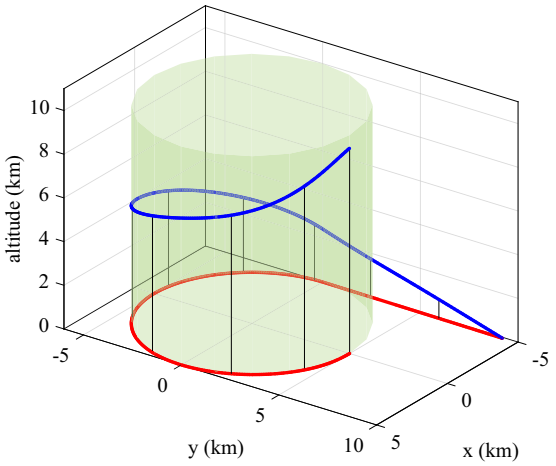


Fig. 8.29 3D view of an optimal turn around a HAC with radius 5 km for initial state 1. Data taken from De Ridder and Mooij (2011a)

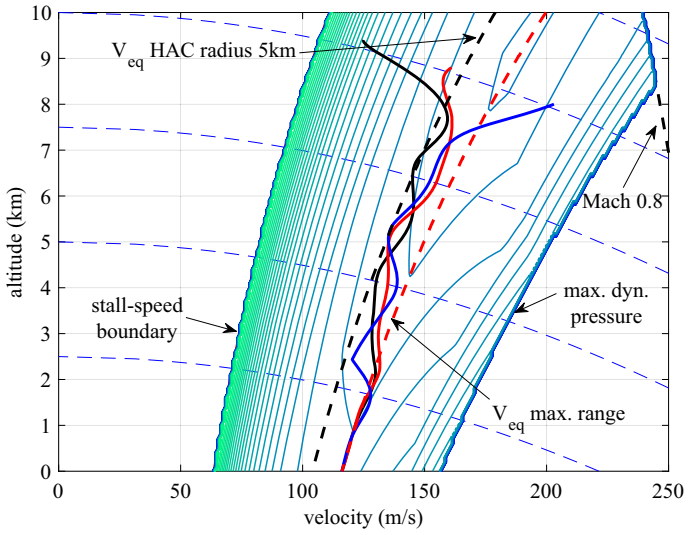


Fig. 8.30 Contours of constant drag for a HAC with a radius of 5 km and optimal EA trajectories. Data taken from De Ridder and Mooij (2011b)

equivalent airspeed of the maximum-range glide. This switch is clearly visible in Fig. 8.30. A lower velocity during the turn requires a higher angle of attack. Hence, it can be concluded that the optimal turn is performed at a higher angle of attack and a lower velocity.

The radius of the HAC turn has an influence on the optimal trajectories. The smaller the turn radius, the larger the required bank angle. When the radius is

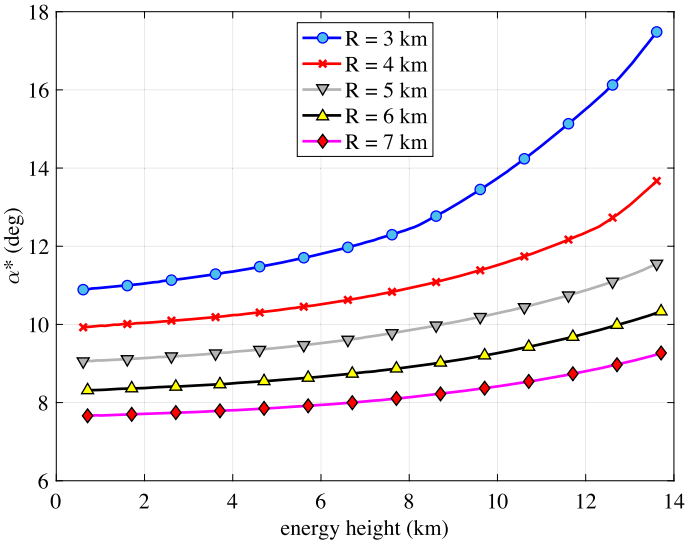


Fig. 8.31 Open-loop angle of attack as a function of energy height and turn radius for a turn with minimum energy dissipation. Data taken from De Ridder and Mooij (2011b)

decreased, the drag valley is shifted to lower equivalent airspeeds. For a 5 km radius optimal turn, the optimal equivalent airspeed (equivalent airspeed in the drag valley) is 108.0 m/s, while for a turn radius of 4 and 6 km, the optimal values are 104.0 m/s and 112.0 m/s, respectively. The lower the radius, the lower the optimal velocity and the higher the required angle of attack to fly at these lower velocities. Figure 8.31 shows the steady-state angles of attack in the drag valley as a function of the turn radius and the energy height. The figure confirms that the smaller the turn radius, the higher the optimal angle of attack.

The important conclusion is that the optimal airspeed during a turn is not equal to the optimal airspeed in a straight flight and changes as a function of the turn radius. The smaller the radius, the lower this velocity.

8.4.8 Conclusions

Although it is common to use the angle-of-attack profile for a maximum lift-to-drag ratio as the maximum-range strategy, here we have shown that a larger range can be obtained by using a different strategy. The energy-state approximation states that the optimal strategy to perform a subsonic maximum-range flight is to fly with minimum drag, which is confirmed by numerical optimisation using a genetic algorithm, despite the vehicle's poor gliding capabilities. If the vehicle starts at an initial state, which is not located on this path, transient manoeuvres are performed to intercept the

maximum-range glide path. However, in case of a supersonic maximum-range flight, the energy-state approximation (using an equilibrium condition) is no longer valid, because the drag encountered during the transient manoeuvres is too large. The general principles of dissipating the least amount of energy with respect to range and flying at minimum drag is still valid, though.

Significant improvements (3%–6%) in range can be obtained for a subsonic maximum-range flight using the optimal strategies, comparing with the trajectories using $\alpha_{(L/D)max}$. Only minor improvements are obtained during a supersonic flight (1%–3%). Hence, the angle-of-attack profile for L/D_{max} is used for a supersonic maximum range flight, while for a subsonic maximum-range flight the longitudinal strategy is to track the reference dynamic pressure in the drag valley. This leads to acceptable deviations from the optimal ranges.

An optimal subsonic turn (one that minimises energy loss) is performed at a larger angle of attack and at a lower velocity than during a subsonic wings-level maximum-range flight. The smaller the radius, the lower the optimal velocity and the larger the required angle of attack. The energy-state approximation is valid for the optimal-turning problem. The drag valley is different for different turn radii; when the turn radius decreases the valley is shifted to lower dynamic pressures.

The optimal dive (one that minimises range and maximises energy loss) is obtained by flying along a minimum dynamic-pressure profile close to the stall boundary, because this yields the largest encountered drag. The energy-state approximation indicates that the drag encountered at the minimum dynamic-pressure boundary is larger than at the maximum dynamic-pressure boundary. When the inherent reduction in control capabilities is too large at this minimum dynamic pressure, a dive near maximum dynamic pressure should be considered.

An important conclusion from all the optimal trajectories is that the optimal dynamic-pressure profiles, especially the initial transient parts, change for different initial states. A different initial state requires a different optimal dynamic-pressure profile, if the complete flight is considered. Hence, it is not possible to use a single fixed dynamic-pressure profile in case of off-nominal conditions.

Although it is common to specify a single value for the minimum required energy that is needed to reach a specific ground-track range using the maximum-range strategy, simulations indicate that the maximum range differs for initial states with the same initial energy. Therefore, in addition to the minimum energy height, it is also necessary to specify for which combination of kinetic and potential energy this is valid.

Problems

Judge Whether the Following Statements are TRUE or FALSE (Problems 8.1 to 8.10).

8.1 The terminal area energy management (TAEM) guidance system of the Space Shuttle continuously adjusts the vehicle's trajectory and descent rate based on

real-time sensor feedback to account for changing environmental conditions and landing constraints.

8.2 TAEM guidance of the Space Shuttle is primarily concerned with minimising the vehicle's descent rate to achieve a soft touchdown on the runway.

8.3 The terminal area for winged entry vehicles is defined approximately 300 km around the runway.

8.4 If an entry vehicle has an excess energy, the only option to remove that energy is to fly out of the TAEM area again.

8.5 Targeting the terminal area usually begins a few hundred kilometres away from the runway.

8.6 The key parameter in targeting the terminal area is the heading error, i.e., the difference between the current flight heading and the nominal one aimed at the terminal-area targeting point.

8.7 The performance curve for a gliding flight relates the flight velocity to the distance-to-runway.

8.8 In the steady-state approximation the minimum glide angle corresponds to a flight at maximum lift-to-drag ratio.

8.9 The unit of energy height is Joule per meter

8.10 TAEM guidance of the Space Shuttle primarily focuses on managing the vehicle's energy to ensure a safe and precise landing.

Theoretical Aspects of Terminal-Area Flight (Problems 8.11 to 8.15).

At the end of the hypersonic descent, a winged re-entry vehicle enters the so-called Terminal Area. Here, a dedicated guidance system, the so-called Terminal Area Energy Management system takes over to control the flight range to-go. As a first approximation to design such a system we assume that the vehicle is flying in a stationary glide.

8.11 Starting from the equations of planar motion–see Problem 5.10–set-up the dynamic equations of motion for a stationary glide. State your assumptions. Note that the flight-path angle is negative when the velocity is below the local horizon.

8.12 From the dynamic equations as found under Problem 8.11, derive expressions for (i) the (equilibrium) velocity, (ii) the flight-path angle, (iii) the rate of descent, and (iv) the horizontal range.

8.13 Sketch the performance curve of the vehicle, and indicate (i) velocity, (ii) flight-path angle, (iii) minimum flight-path angle, (iv) minimum rate of descent, and (v) the stall speed. Pay attention to axis labels and positive axis directions.

8.14 Given the stationary-flight condition, when will the vehicle obtain a maximum range?

8.15 Suppose that during the glide, the vehicle would fly a turn with constant bank angle. How would the expressions for equilibrium velocity and flight-path angle change?

Performance Calculations for Terminal-Area Flight (Problems 8.16 to 8.21).

Given an unpowered, winged entry vehicle in a steady-state gliding flight, starting from $h = 10\,\text{km}$ with $V_{eq} = 114.93\,\text{m/s}$. Assume constant aerodynamics, $C_D = 0.068$ and $C_L = 0.272$, referenced to $S_{ref} = 110\,\text{m}^2$. The mass of the vehicle is 19,000 kg. The atmosphere can be modelled with an exponential model ($\rho_0 = 1.225\,\text{kg/m}^3$ and $H_s = 6{,}950\,\text{m}$).

8.16 What is the groundspeed of the vehicle?
8.17 Calculate the equilibrium glide angle and the rate of descent.
8.18 At 5 km altitude, the vehicle enters a coordinated turn. For a turn rate of 4°/s, what will be the corresponding turn radius and the required bank angle? State any assumption that you make. What can you conclude when the turn rate has to be 5°/s?
8.19 Assuming the bank angle from the previous question, how does this affect the equilibrium glide angle in the turn? What does this mean for your assumption(s) in the previous question?
8.20 Using the new equilibrium glide angle, recalculate the turn radius.
8.21 What is the load factor experienced during the turn?

References

Bryson A, Desai M, Hoffman W (1969) The energy-state approximation in performance optimization of supersonic aircraft. J Aircr 6(6):481–488

Büchner T (2003) Trajectory generation strategy for the terminal area of a reusable launch vehicle. IAC-03-W.1.06

Camara FMC (2003) Development and design of the terminal area energy management guidance for a reusable launch vehicle. MSc thesis, Delft University of Technology

da Costa RR Studies for terminal area GNC of reusable launch vehicles, AIAA-2003-5438. In: AIAA guidance, navigation and control conference and exhibit, Austin, TX, 11–14 August 2003

De Ridder S, Mooij E (2009) Optimal terminal-area strategies and energy-tube concept for a winged re-entry vehicle, AIAA-2009-5769. In: AIAA guidance, navigation and control conference and exhibit

De Ridder S, Mooij E (2011) Optimal terminal-area strategies and energy-tube concept for reusable launch vehicles. Acta Astronaut 68(7–8):915–930

De Ridder S, Mooij E (2011) Optimal longitudinal trajectories for reusable space vehicles in the terminal area. J Spacecraft and Rockets 48(4):642–653

Hull JR, Gandhi N, Schierman JD (2005) In flight TAEM/final approach trajectory generation for reusable launch vehicles, AIAA 2005-7114. In: Infotech aerospace, Arlington, Virginia, 26–29 September

Kirpischikov VP Trajectories of buran orbiter's descent and landing, algorithms of the automatic guidance and control. http://www.buran-energia.com/documentation/documentation-akc-guidance-control.php. Accessed 26 Sep 2020

Kluever CA (2007) Terminal guidance for an unpowered reusable launch vehicle with bank constraints. J Guid Control Dyn 30(1):162–168

Kluever CA, Horneman KR Terminal trajectory planning and optimization for an unpowered reusable launch vehicle, AIAA 2005-6058. In: AIAA guidance, navigation and control conference and exhibit, San Francisco, California, 15–18 August 2005

Kraemer JW, Ehlers HL (1975) Shuttle orbiter guidance system for the terminal flight phase. Instrument Society of America, pp 6.5 1–6.5 10
MBB Space Communication and Propulsion Systems Division (1988) Study on re-entry guidance and control. Final report, ESA report reference: ESA CR (P) 2652, MBB, Munich
Michalewicz Z (1996) Genetic algorithms+data structures=evolution programs. Springer
Mooij E (1998) Space plane flight dynamics. Analysis of guidance and control concepts. PhD thesis, Delft, The Netherlands
Moore TE (1991) Space Shuttle entry terminal area energy management. NASA-TM-104744
NASA (1980) MCC level C formulation requirements. Shuttle TAEM guidance and flight control (STS-1 baseline), NASA-TM-81100
Reichmann H Zum Problem der Fahroptimierung im Streckensegelflug. PhD thesis (in German), Universität Karlshruhe, Institut für Praktische Mathematik
Vinh NX (1981) Optimal trajectories in atmospheric flight. Elsevier Scientific Publishing Company
Welch A, Welch L, Irving F (1970) New soaring pilot. Butler and Tanner Ltd

Chapter 9
Entry, Descent, and Landing Systems



Entry, descent, and landing systems (EDLS) should guarantee the safe delivery of either manned or unmanned payloads to the surface of a planet or a moon, with a certain specified terminal state. This could relate to the soft landing of a rover on Mars, astronauts on the Moon, or a sample-return mission to Earth. They all have in common that the payload should be "operational" after the landing in the broadest sense of the word: the rover should be able to begin its mission, the astronauts should not be injured in any way, and the samples should not be contaminated, because of a too hard landing. This means that the thermo-mechanical load occurring during the complete mission towards the surface should be well below the specified maximum level. For the EDLS this usually means that the final (or *impact*) velocity should be below a certain value.

However, we should distinguish between the unpowered and unaided flight of entry capsules and winged entry vehicles, and systems that are assisted by subsystems that provide extra deceleration in addition to that induced by the vehicle itself. The "unaided" flight of entry vehicles has been extensively discussed in Chap. 5, covering the flight mechanics of three entry mechanisms, i.e., ballistic, gliding, and skipping flight. To have control over the occurring thermo-mechanical loads, some vehicles are equipped with a guidance and control system, a topic covered in Chaps. 7 (guidance, navigation and control of the hypersonic entry and descent) and 8 (the guided flight of winged vehicles in the so-called terminal area).

The term "unaided" may give rise to some confusion, because by changing their attitude, these vehicles can exert control over the trajectory with a change of the aerodynamic force by changing its attitude. This could be achieved by means of, for instance, control surfaces, and not from drag-inducing devices like parachutes. So affecting the state lies within the capabilities of the vehicle itself, hence we call this unaided. When the aerodynamic characteristics of the vehicle do not suffice to guarantee a certain terminal state, additional external forces are necessary: the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_9

vehicle has to be "aided" by additional means that are only operational for specific mission phases.

When in the atmosphere, a parachute is a simple way to increase the drag area of the vehicle and hence to decrease the terminal velocity. When the additional deceleration is not enough—as is the case in the thin atmosphere of Mars—retro rockets have to provide the helping hand. In an atmosphereless environment guidance by means of aerodynamic forces is, of course, non-existent. In that case, the propulsion system becomes the only way to brake and steer to the right terminal conditions. On the other end of the spectrum, when the atmosphere is so thick that one does not actually need a (large) parachute, other deceleration strategies should be considered: changing to a smaller parachute at lower altitudes, or even removing the parachute all together. Identifying the possibilities for the EDLS and finding out about the characteristics of different types of decelerators will be the topic of this chapter.

To better understand the development of descent and landing systems, Sect. 9.1 gives a brief overview of the historical development of in particular parachute systems, because these systems have the longest history to begin with and they are a major subsystem in most planetary landers. Propulsion systems as well, of course, but since the application of these sub-systems only began with the initiation of the lunar programs of the USA and the Soviet Union, there is actually not so much "old" history to discuss.

The two main components of the descent and landing system, i.e., the parachute and the propulsion system will be discussed separately, because of their fundamentally different operating principle and design. In Sect. 9.2 the parachute system will be introduced by means of the principal design requirements, following by a description of the individual hardware elements, different types of parachutes and operational aspects. A brief discussion on simulation models will conclude that section. To distinguish between the propulsion systems needed for the powered descent and, for instance, the ascent engine to lift off from the surface of the Moon, the next Sect. 9.3 will only discuss the powered descent, thereby focussing on the different guidance systems that are available to guarantee a soft and/or pinpoint landing. Although usually parachutes and/or engines are used, there are also some other means to decelerate a vehicle. A brief summary of these means is given in Sect. 9.4.

Landing close to areas of scientific interest is often complicated, because of the accessibility of the terrain. In many areas there are craters, rocks, and slopes, and also the lighting conditions are not always ideal. To increase operational flexibility or even to avoid mission failure, planetary landers can benefit from hazard detection and avoidance systems, which—in combination with advanced guidance laws—can aid in steering the vehicle to an alternative landing site. This kind of system also increases vehicle autonomy at large distance from Earth, when communication between ground control and the lander takes too long. Section 9.5 addresses these advanced descent and landing systems.

Having covered the descent and terminal phase just prior to landing, Sect. 9.6 will discuss the actual landing in terms of which systems are available and how they can absorb the remaining energy during the landing shock (=mechanical load). Since this section concludes the overview of elements of the descent and landing system, what

is left to show are some real-life mission examples. Section 9.7 reviews a number of missions, focussing on the design of the EDLS. Included missions are the Surveyor and Apollo lunar landers, the two Viking landers on Mars, the Galileo mission to Jupiter, the Earth sample-return mission Stardust, the two Mars Exploration Rovers MER-A and MER-B, and finally the success story of the Mars Science Laboratory with the *Curiosity* rover. The Huygens mission to Titan will be elaborated upon in a dedicated case study (Sect. 9.8) that includes simulation results and analyses supporting the EDLS design.

9.1 History

The history of parachutes is a long one and—as far as the Europeans are concerned—dates back to as early as the end of medieval days.[1] Famous are Leonardo da Vinci's drawings of a pyramid-shaped parachute, conceived around 1485 and commonly assumed to be the first western design of a parachute. However, Leonardo himself was inspired by the work of the *Quattrocento* Sienese engineers, who dealt with man's movement through air, proposing a parachute and two very unlikely wings, as can be seen in Fig. 9.1. These drawings have been dated to the 1470s or not much later (White 1968), which means that this parachute is at least 10 years older than Leonardo's. The Sienese parachute was intended as an escape device to allow people to jump from burning buildings, but there is no evidence that it was ever used.

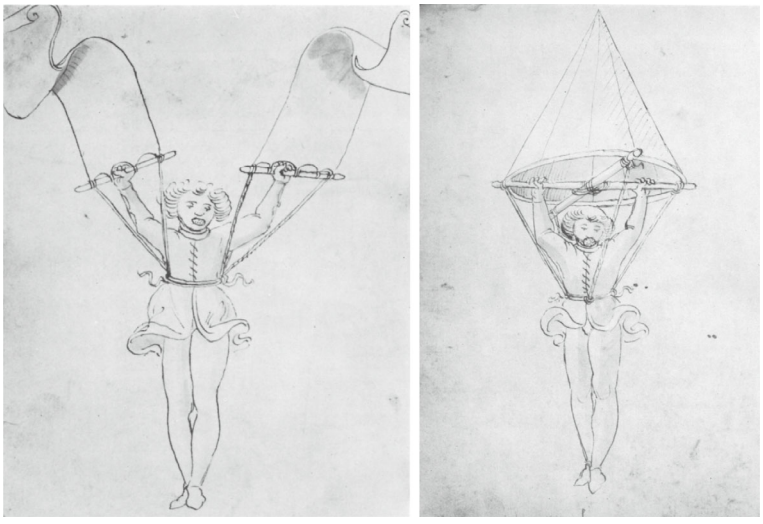


Fig. 9.1 Wing and parachute design of Sienese engineers, around the 1470s (Images: public domain)

[1] Chapter 1 gives a more global overview of the history of, in particular, launch and re-entry systems.

But parachute history has more surprises. Evidence found in the historical archives of Peking, China, and translated by the French monk, Vasson, indicates that parachute-like devices were used in China as early as the 12th century. Still, if we include legends besides facts, we can trace the use of parachutes even further back, although it was still happening in China. In 90 B.C. the Chinese historian Si Ma Chian wrote in historical records about Shun, a 23rd–22nd century B.C. legendary leader of ancient China, whose half-century of rule was said to be one of the longest in Chinese history. When he was young the then current emperor Yao was testing him to make sure he would make a better emperor than any of his nine sons. Shun was given a district to rule and prove he could do that. Shun's stepmother and stepbrother became extremely jealous and conspired to kill Shun (some claim it was actually Shun's own father).

Legend says that at one time Shun's stepbrother Xiang lit a barn on fire, and convinced Shun to climb onto the roof to put the fire out. Once up the roof Xiang took away the ladder, trapping Shun on the burning roof. Shun skilfully made a parachute out of his hat and cloth to increase air resistance to reduce the falling speed (other sources mention that Shun may have used two large bamboo hats). He jumped down from the roof and landed down in safety (a similar story came up around 90 B.C. when a man trapped in a burning tower used two conical straw-hats as a drag or descent-delay parachute).

In 200 B.C. acrobats were performing stunts of great falls using something similar to parachutes during their acrobatics in the palaces of the early Han Emperors. That this actually may have been true is confirmed by other historical accounts from 1308 A.D., which indicate that the acrobatics of the parachute activities also appeared in the palace of the Yuan Emperor. Although it seems that the invention of the parachute was for performances to entertain the emperor or to avoid a murder attempt, there were also some more serious developments.

Derived from the Chinese invention of the kite (called a "flying sail") in the 5th century B.C., from *ca.* 290 B.C. until 250 A.D. many instances of 'man-lifting' kites bearing an observer aloft for military surveillance have been reported.[2] Lastly, according to the Chinese archives there were several winged-flight experiments imitative of birds around *ca.* 550–577. Parachute development was stimulated in a more bizarre way, in the sense that Chinese officials experimented with parachute designs by compelling condemned prisoners to jump from towers and cliffs—and this was, in fact, the part of the archives that was translated by the French monk Vasson.

Abbas Ibn Firnas (muslim poet, astronomer, musician and engineer) made several attempts to build a flying machine. In 852 he jumped from the minaret of the Grand Mosque in Cordoba using a loose cloak stiffened with wooden struts, more resembling a parachute than a flying machine. He hoped to glide like a bird, but in fact, he did not. Fortunately, the cloak slowed his fall sufficiently to sustain only minor injuries. In 875—he was already 70—he had perfected his machine of silk

[2] The (detailed) timeline on which the current overview is based can be found on http://www.combat.ws/S4/MILTERMS/PARAHIST.HTM (last access: August 14, 2021). This timeline is also discussed by Poynter (2007), and some of the events are mentioned in the book by Harding (2006).

and eagles' feathers, and tried it again. This time he jumped from a mountain, flew to a significant height and stayed up in the air for ten minutes. Alas, he crashed on landing, survived and correctly concluded that it was because he had not given his device a tail so it would stall on landing. Baghdad international airport and a crater on the Moon are named after him.

Talking about minarets, there was a very high mosque in Guangzhou (Canton), China, which was the centre of a story written down by Yue Ke (1183-*ca.* 1242). One day in 1192, people found that one leg of the golden rooster at the top of the mosque was stolen. Of course they wondered how the thief could manage to escape with such a big piece of gold. The mystery was solved when the thief was caught. He confessed that he had used two umbrellas to descend from the top of the mosque. The umbrellas served as a parachute for the thief! Let us hope for the sake of his ingenuity, that he was not forced to parachute-jump from another tower or cliff to further parachute development!

Finally, we have reached Europe. As said earlier, around 1485 Leonardo da Vinci visualised the drag principle to slow down people by means of a parachute. In the *Codex Atlanticus*, the device pictured was a pyramid-shaped structure by means of which, the sketch implied, a man might leap from a tower or burning building without greatly endangering his life (Fig. 9.2). Moreover, da Vinci also conjectured that it could be used for military deployment if soldiers were launched into the air by cannon. However, as far as is known, he did not reduce his sketch to practice.

In 1595, Fausto Veranzio, a Hungarian mathematician living in Italy, published the idea of a parachute being used as a "fall breaker": one of the illustrations in his *Machinae Novae* is a sketch of a parachute dubbed Homo Volans ("The Flying Man"), see Fig. 9.3. Having examined Leonardo da Vinci's rough sketches of his parachute, Veranzio designed a parachute of his own. He modified the original parachute sketch by keeping the square frame, but replacing the canopy with a bulging sail-like piece of cloth. In a number of in particular parachuting publications, there are assertions that in 1617, Veranzio, then aged 65 (and seriously ill), had tested his parachute by jumping from St Mark's Campanile in Venice, or according to other sources, from St. Martin in Bratislava or even from a bridge in Venice. There are some pieces of the puzzle not fitting, though. The story does not explain how anybody could have achieved to get the (heavy) wooden frame out of the tower into a horizontal position, and at the same time move himself under the parachute without the apparatus tilting and collapsing. That would already be hard for a healthy and strong man, let alone for someone ill and close to his death. The story about Fausto Veranzio testing his parachute thus appears to be an early urban legend. No evidence has been found that anybody ever tested Veranzio's parachute.

In 2000, though, a Mr. Adrian Nicholas decided to demonstrate the viability of Leonardo's design (Harding 2006). He constructed a parachute according to the available drawings, using only tools and and materials that would have been available to the old master. On June 25, 2000, Nicholas prepared to jump from a hot-air balloon in Mpumalanga, South Africa, despite the warning of many experts that the design would not work. At 10,000 ft, surrounded by two helicopters and two parachutists, Nicholas parachuted for 5 min, Fig. 9.4a. Despite the predictions that the parachute



Fig. 9.2 Leonardo da Vinci's pyramid-shaped parachute, published in the *Codex Atlanticus* (1495) (Image: public domain)

would tip over, fall apart or spin uncontrollably, Leonardo's parachute made such a smooth and slow descent that the two parachutists had to brake twice to remain at the same altitude as Nicholas. Nicholas, however, feared the wooden frame might crush him upon landing (it weighed 80 kg), so he cut himself free at 3,000 ft and used a modern parachute to reach the ground.

Nicholas was not the only one to test Da Vinci's design, though. A Swiss amateur parachutist, Oliver Vietti-Teppa, made a successful jump in 2008. Although following the guidelines of the more than 500 year-old design, he just used the canopy without the original wooden frame to make the design a lot lighter. The parachute consisted of four equilateral triangles, 7 m on each side, and made from modern parachute fabric. Vietti-Teppa jumped from a helicopter, and the parachute opened at 650 m. He became the first person to have made it safely to the ground with the Leonardo model, Fig. 9.4b.

But now the question arises: who, actually, was the first to test Leonardo's design? Was it Adrian Nicholas, who, despite using all the original materials, was not confident enough to descend all the way down to the ground, or was Oliver Vietti-Teppa's jump the first successful test? He made it all the way down, but used modern materials to construct the parachute and had removed the wooden frame from the design. We leave it to the readers to make up their own mind about who they think is right...

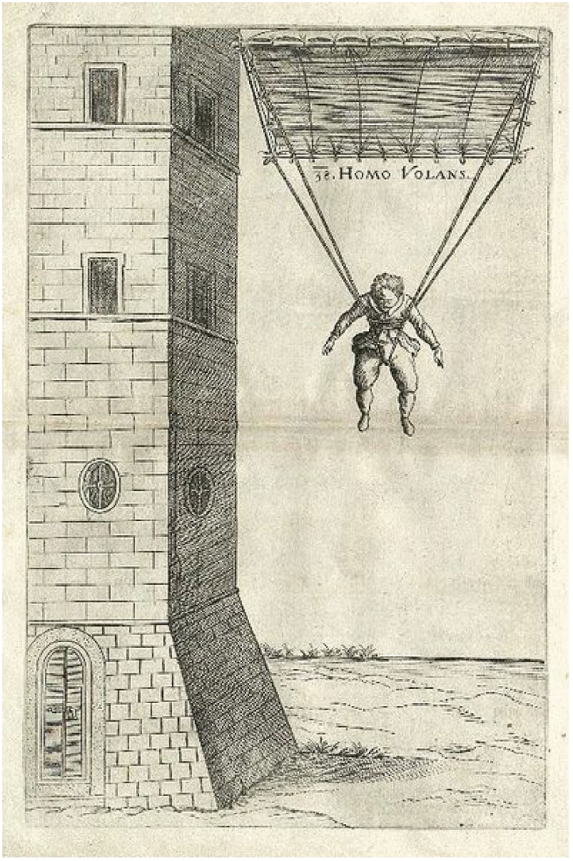


Fig. 9.3 Fausto Veranzio's sketch of Homo Volans ("The Flying Man") in *Machinae Novae* (Image: public domain)

Credit for the invention of the first practical parachute frequently goes to the Frenchman Sebastien Lenormand (1757–1837) who demonstrated the parachute principle in 1783.[3] He became the first human to make a witnessed descent with a parachute by jumping from a tree in a private garden with the help of two modified umbrellas (White 1968). He was also the first to come up with the term parachute (from the Greek *para* meaning "against", and the French *chute*, "fall"), although the Chinese had been using the same principle for more than a thousand years before him. Lenormand continued to refine his invention and later that same year (26 December, 1783), he jumped from the tower of the Montpellier observatory in front of a crowd that included Joseph Montgolfier, using a 14-ft parachute with a rigid wooden frame.

[3] As we mentioned in Chap. 1, in 1633 Lagari Hasan Çelebi launched in a 7-winged rocket, lifted off and rose high in the sky, before landing in the sea using a simple parachute. Maybe his parachute did not look like one, we should also not forget this pioneer.



(a) Adrian Nicholas (2000) (Image: Heathcliff O'malley)

(b) Oliver Vietti-Teppa (2008) (Image: AFP)

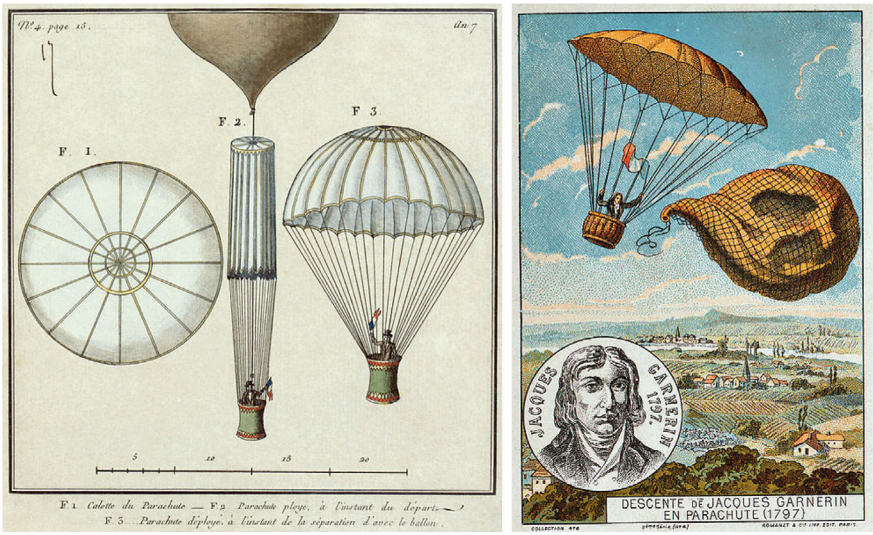
Fig. 9.4 Two modern-day tests of Leonardo da Vinci's pyramid-shaped parachute

He shared the idea with many before him that the parachute could be used to help people escaping from burning buildings.

The Montgolfier brothers Joseph-Michel (1740–1810) and Jacques-Étienne (1745–1799), were the inventors of the Montgolfière-style hot air balloon. Their first experimental flight took place on 19 September 1783, with a sheep, a duck and a rooster as "passengers", followed by the first manned flight on 21 November of the same year. Concurrent with his experiments on hot-air balloons, Joseph Montgolfier became interested in the concept of parachute-like devices. He experimented with the design by dropping a number of animals from towers. Subsequently, he himself tested his invention by jumping from the roof of his house at Annonay and later from "greater heights", which could have been from one of his balloons.

Another French aeronaut (i.e., a pilot of a balloon or lighter-than-air aircraft), Jean Pierre Blanchard (1753–1809), claimed the invention of the parachute in 1785, but we know now that he had not the right to do so. Blanchard made his first successful balloon flight in Paris on 2 March 1784, in a hydrogen gas balloon launched from the Champ de Mars. Later, in 1785, he made the first aerial crossing of the English Channel. The number of balloon accidents, which occurred during this period, led to the consideration of the parachute as a possible means of saving lives. In 1785 Blanchard did the first successful parachute experiment with an animal from a balloon, and became the first to introduce the emergency use of parachutes. After more successful experiments with animals, including his own dog, he actually used the concept himself, when in 1793 his hot-air balloon ruptured and he used a parachute to escape. Subsequent development of the parachute focused on it becoming more compact. In the late 1790s, Blanchard began making parachutes from folded silk, taking advantage of silk's strength and light weight, as compared to the earlier concept of linen stretched over a wooden frame.

Although Blanchard may have designed the first frameless folding-cloth parachute, he should at least share that honour with André-Jacques Garnerin (1769–1823), who is officially credited as the inventor of the frameless parachute. He made



(a) Schematic depiction (b) Proof of concept

Fig. 9.5 Garnerin's first parachute used as emergency parachute in 1797 (Images: public domain)

the first jump from high altitude with the help of a non-rigid parachute on October 22, 1797 (Fig. 9.5). At an altitude of around 900 m he cut the ropes of the balloon and unfolded a ribbed parachute canopy, which folded like a sunshade-parasol (7 m diameter, with 36 suspension lines). Due to the lack of vents the canopy started to oscillate violently, which drove the notion to insert vents in the canopy (Garnerin did that after his jump in England in 1802). Although his landing was bumpy, Garnerin emerged uninjured.

In 1837, the umbrella-shaped parachute had become standard. One of its defects, though, was the violent swinging during the descent. Earlier, around 1810, the English pioneer of aeronautical engineering and sometimes referred to as "the father of aerodynamics", Sir George Cayley (1773–1857), had proposed to remedy the oscillatory problem with a cone-shaped parachute which would be used with the apex pointing downwards. Inspired by the works of Cayley, the English artist Robert Cocking (1776–1837) came to a similar conclusion after his calculations that this oscillation could be avoided if the parachute were made conical (Fig. 9.6). If the surface area was large enough, the descent would not be too rapid.

Cocking, 61 at the time, had constructed a conical parachute with a basket below in which he could sit. Although unexperienced in parachute jumping he insisted to make the jump himself from a balloon. Immediately after the release it became obvious that Cocking had neglected to take the mass of the parachute into account. With a weight of 10 times that of a modern parachute (just over 100 kg), the parachute descended far too quickly and the canopy took the shape of an umbrella inverted by a strong wind. As a consequence, the drag area was too small and Cocking fell to

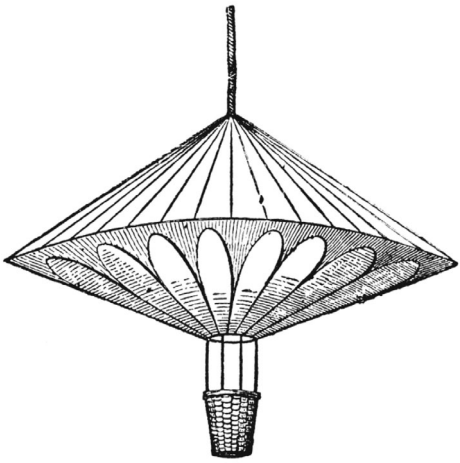


Fig. 9.6 Robert Cocking's ill-fated conical parachute (Image: public domain)

his death. However, his notion of a conical parachute being more stable than the standard shape in those days proved to be correct, as nowadays almost all parachutes for planetary landing have a conical shape.

The use of parachutes as an escape system for persons aboard balloons unable to land safely drove the improvements in parachute design. In 1887, Captain Thomas Baldwin invented the first parachute harness. Gleb Kotelnikov invented he knapsack parachute, further popularised by the German circus performers Katchlen Paulus and Paul Letterman. In 1890, Letterman and Paulus jumped from a balloon wearing a "bag packed" tandem parachute so they could deploy a second canopy in series after discarding the first. This two-parachute backpack system and "cut-away" technique is still used when a reserve parachute replaces a defective/malfunctioning main canopy. Further improvements included the inclusion of a pilot chute (1910), which deployed the primary or main canopy from its container; a similar system is still in effect for planetary descent systems.

In 1903 the Wright brothers made their first powered flight, and with this the development of the parachute picked up speed. Military interest in parachutes began during World War 1 (1914–1918). An Australian pilot made the first escape by parachute from a disabled plane, on the Russian front in 1916, five years after the first civilian did so (in 1911, Grant Morton made the first free-fall jump from an airplane). Three months later, another Australian pilot made a safe escape from another disabled plane. By 1917 the parachute had proven itself, and both Germans and English were equipping their air forces with these life-saving devices (Fig. 9.7). By the Summer of 1918, parachutes were in wide use on all fronts.

In 1913, Captain M. Douade designed an emergency parachute system to lower an entire airplane; however, his plans were cancelled by the war. In 1919, though, the idea had picked up and the "Guardian Angel" emergency parachute system was manufactured by E.R. Cathrop Ltd. (established in London). The system was credited with saving many lives. In 1929, Captain Roscoe Turner landed his airplane with a

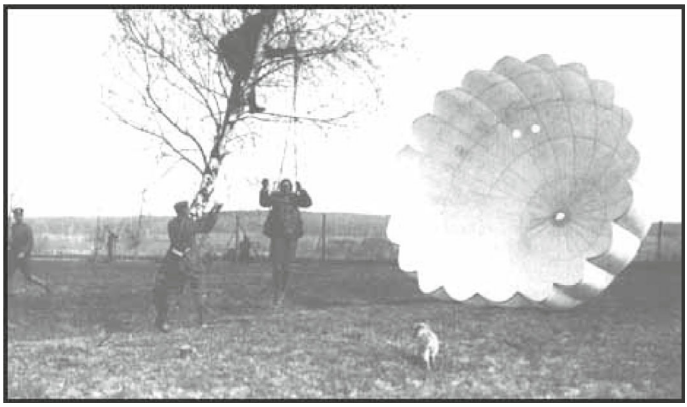


Fig. 9.7 Landing with a parachute remains difficult with trees around: a German pilot and his parachute are being disentangled from a tree (1918) (Image: public domain)

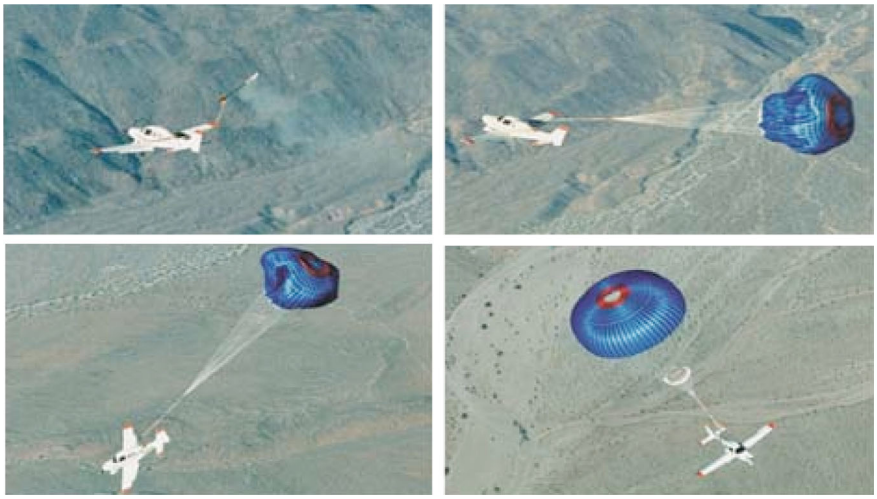


Fig. 9.8 The Cirrus SR20 is the first aircraft with a parachute as safety feature (Image: NASA)

parachute, before 15,000 spectators in an air-show at Santa Ana, California. This public display of vehicle recovery stimulated NASA in later years (1961) to do more research on this type of recovery, and the agency adopted parachutes for the final phase of the re-entry in the Gemini space-capsule program. And, in a 1996 prototype made by Ballistic Recovery Systems Inc. (now commonly known as BRS Aerospace) for Cessna light airplanes, the FAA approved aircraft descent parachutes. The Cirrus light plane became the first production model to have a built-in ballistic parachute for emergency descent (Fig. 9.8).

The first mass military jump by five US Marines and four USN sailors took place on 24 September, 1927, at Anacostia Flats, Washington DC. The group jumped from an altitude of about 500 m. Later that year, on 6 November, the Italian Air Force experimented by parachuting nine aviators from an altitude slightly above 500 m. Brigadier General William C. "Billy" Mitchell directed that six soldiers would exit a Martin bomber airplane by parachute with full-field equipment and thus becoming the first "paratroopers" in the world. Other countries followed, and by 1930 the parachute delivery of troops and equipment had become a standard procedure in the German and Russian army.

Of course the military requirements dominated the period in between the two world wars, and in particular during World War II. Also, from 1930 onwards the operational speed regime increased and so-called high-speed parachutes were developed, in particular useful for entry applications. Discoverer 14 (also known as KH-1 9009) was the first successful low-resolution photo-surveillance spacecraft (a "spy" satellite) launched by the US Air Force within the Corona program co-managed by DARPA (NSSDC 2012). On 19 August, 1960, usable photographic film images taken by the satellite were recovered by a C-119 recovery aircraft. On the C-119's third pass over the parachute, the recovery gear trailing behind the aircraft successfully snagged the parachute canopy (Fig. 9.9a).

This was the first successful recovery of film from an orbiting satellite and the first mid-air recovery of an object returning from Earth orbit. This method of retrieval is a risky technique, and is only used when other forms of landing are infeasible. In case of the KH-1 this was done, because of the delicate nature of the films to be retrieved. Another example was the planned recovery of the Genesis entry capsule (2004) carrying a sample of the solar wind that was so delicate that it would have been damaged by a parachute landing (Fig. 9.9b). Unfortunately the recovery parachute failed to open and the capsule crashed in the Utah dessert.

The first official parachute used for manned Earth entry took place on 12 April 1961, when Juri Gagarin—the first man to be launched into space in the Vostok-1—returned to Earth in his entry capsule after orbiting the Earth for 1 h and 48 min. At an altitude of 7 km from the ground, the hatch of the spacecraft was released, and two seconds later Gagarin was ejected. This was done as a safety measure in case the Vostok main parachute would not deploy (and also to protect the cosmonauts from the large landing shocks that occurred), an operation that was repeated for all successive 5 manned Vostok flights. At 2.5 km altitude, the main parachute was deployed from the Vostok spacecraft. It landed safely, although it bounced up once before it came to rest and it created an impressive furrow, because the parachute dragged the capsule along (Fig. 9.10). Gagarin's parachute opened almost right away, and about ten minutes later he landed some 280 km to the west of the planned landing site (near Baikonur).

The notion of being able to perform a precision landing, (mostly) independent from the wind conditions, was one of more to stimulate the design of steerable parachutes, also known as parafoils. A parafoil is a non-rigid (textile) airfoil with an aerodynamic cell structure, which is inflated by the wind. Once inflated by so-called ram-air inflation the shape of the parafoil is forced into a classic wing cross-section. Domina Jalbert (1904–1991) developed the device in 1964 and patented it in 1966.



(a) Recovery of the KH-1 film canister by a C-119 (1960) (Image: USAF).



(b) A practice run for the planned retrieval of Genesis, using helicopters flown by Hollywood stunt pilots contracted by NASA (Image: NASA).

Fig. 9.9 Mid-air retrieval



Fig. 9.10 The parachuted landing of Vostok-1, without Gagarin: left, officially released landing photo, without a trace of the impact (reconstruction photo); right, the actual landing furrow in the wheat field (Images: RKK Energia)

Compared to a simple round canopy, a parafoil has greater steerability, will glide further and allows greater control of the rate of descent. A potential space example is found in the descent system of NASA's Crew Return Vehicle and corresponding experimental vehicle X-38 (cancelled in 2002 due to budget cuts). In Fig. 9.11 the first free flight and successful parafoil landing of X-38 is shown.

The world record for the longest skydive, from a height greater than 31 km, was achieved by Colonel Joseph William Kittinger II (born 27 July, 1928) in 1960. For



Fig. 9.11 X-38, first free flight on March 12, 1998 (Image: NASA)

over 50 years his altitude record held, until on 14 October 2012 Austrian skydiver Felix Baumgartner went up to the stratosphere in a capsule attached to a helium balloon as part of the Red Bull Stratos project. At an altitude of 39 km he jumped down, and after a free fall of 4 min and 19 s (yet still 17 s short of Kittinger's jump), reaching a top speed of 1,342 km/h (Mach 1.24), he finally landed safely in Roswell, New Mexico. Not only did Baumgartner break the altitude record, he also set the altitude record for a manned balloon flight, and the greatest free fall velocity.

Although parachute systems are not the only landing systems to be used in entry and descent missions, they have always been a very important and mostly essential part of both manned and unmanned spaceflight. Any mission aiming to land on another celestial body, being it a planet with an atmosphere or the atmosphereless Moon, either a parachute system or a propulsion system for a powered descent—or a combination of both—is part of the design. In this chapter, apart from the fundamental aspects in descent-system design, flight-mechanics issues and required guidance, many missions will be discussed that may serve as realistic examples. This means that details on specific entry and descent missions will be discussed in Sect. 9.7. Therefore, we will conclude this historical overview with only some "firsts".

The Luna programme was a series of robotic spacecraft missions sent to the Moon by the Soviet Union between 1959 and 1976, as either an orbiter or lander. Fifteen of those missions were successful, and on 3 February 1966, Luna-9 made the first

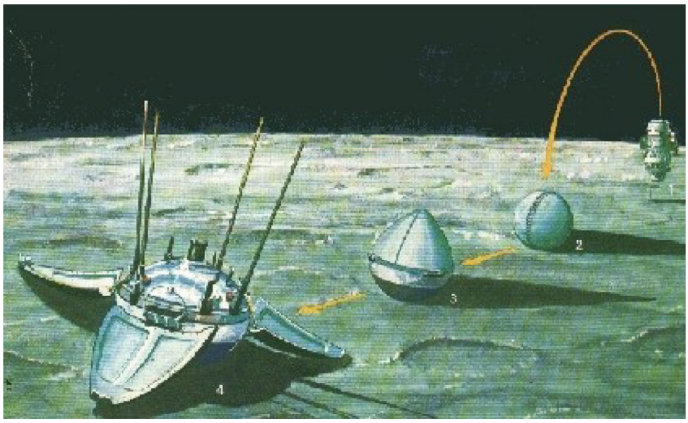
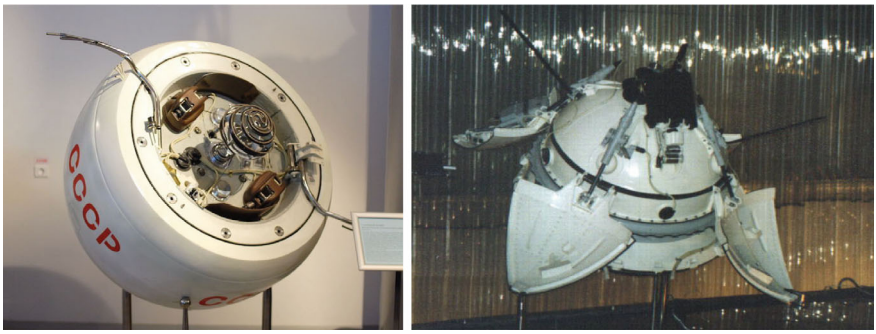


Fig. 9.12 Luna-9 landing with landing bag (Image: RKK Energia)

successful soft landing on the Moon. Its descent and landing sequence was as follows. At 75 km altitude, 48 s before landing, the descent vehicle fired its retro rockets to begin reducing its velocity of 2.6 km/s (Harvey 2007b). At that moment, also the landing bags that would protect the landing vehicle at impact were inflated. At 250 m from the surface the main retrorocket was turned off and four smaller engines were used to slow down the spacecraft, until at a height of about 5 m—measured with a contact sensor that touched the ground—the engines were shut down. The landing capsule was ejected and impacted the surface at 22 km/h, bouncing several times and coming to rest in *Oceanus Procellarum*, see Fig. 9.12.

The first manned landing on the Moon, indicating successful powered descent and soft landing, took place on 20 July 1969 by Apollo-11. It also included the first successful ascent from another celestial body. The astronauts returned to Earth in the Apollo Command Module, and landed in the sea with a cluster of three large parachutes. In this chapter, the parachute system of Apollo will be touched upon on several occasions; the guided descent of the Lunar Module will be treated in more detail in Sect. 9.7.1.

In 1970 the Russian Venera-7 became the first successful lander to reach the surface of Venus, although in the end it did not use a parachute to do so. The probe was equipped with a parachute that deployed at an altitude of about 60 km. Unfortunately the parachute failed shortly before landing, but aiding to Venera's survival was that it was very close to the surface. It impacted at 17 m/s and toppled over, but survived. It continued to operate for 23 min before succumbing to the heat and pressure. Venera-7 achieved the first broadcast from another planet's surface. Later versions of Venera did not have a parachute at all, but were equipped with a so-called drag skirt; this will be discussed in more detail later (Sect. 9.4). Finally, on 2 December 1971, the Russian Mars-3 lander made the first soft landing on Mars, using a combination of parachutes and retro rockets. In Fig. 9.13 a Venera (nr. 4; Venera-7 had a similar configuration) and Mars-3 are shown. Clearly visible is the entry probe at the bottom of the Venera;



(a) Venera-4 (Image: WikiCommons/User Rave).

(b) Mars-3 (Image: NASA).

Fig. 9.13 Two successful Soviet planetary landers

note the resemblance of Mars-3 with Luna-9. To conclude, the Russians have been credited with many firsts, which was the situation in the early days of the Space Race.

Two noteworthy American and European missions are Galileo and Huygens. On 7 December 1995 the Galileo atmospheric probe (Fig. 1.51) made a parachuted descent into Jupiter's atmosphere, not to land (as this is impossible for a gas planet), but to take atmospheric measurements. The very turbulent atmosphere and very high atmospheric pressure at lower altitudes made this a very challenging mission. The design of the entry and descent system is further detailed in Sect. 9.7.3. Finally, the joint NASA/ESA mission Cassini/Huygens set out to study the Saturnian system.

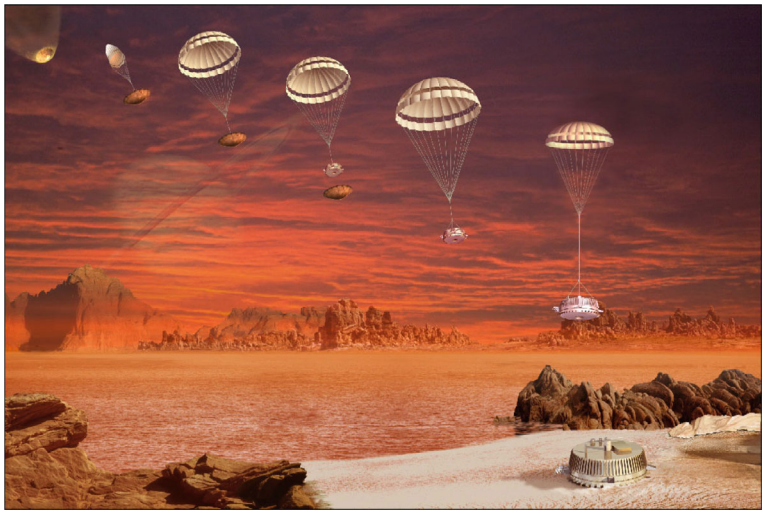


Fig. 9.14 Huygens descent sequence (Image: ESA)

The ESA built Huygens probe was sent into the atmosphere of Titan, Saturn's largest moon, see Fig. 9.14. On 14 January 2005, the Huygens probe made a soft landing on Titan's surface. Until this date, it remains the most distant landing of any craft launched from Earth. Huygens will be the topic of a case study in Sect. 9.8.

9.2 Parachute Systems

9.2.1 Introduction

In aerospace context a parachute is a folding umbrella-shaped device of light fabric used especially for making a safe descent from an airplane. In the space field, the *parachute decelerator system* (PDS) is an integral part of the entry, descent, and landing system. In essence, it is a flexible elastic body whose inflated shape is dependent on the flow conditions. It provides an aerodynamic drag suitable to decelerate an entry vehicle from supersonic to subsonic speeds until the landing, or until the start of the final propulsive descent to the surface. The need for parachutes stems basically from one or two requirements:

1. To limit the impact speed to a range that is acceptable to the landing system (mechanical load constraint)
2. To control the descent duration to maximise the scientific return in a given altitude range (communication constraint).

The connection between parachute and payload may put additional constraints on the design; in relation to the second requirement the payload may be required to spin, e.g., when a camera is mounted on the lander and supposed to scan the horizon. In case of the Huygens probe, for instance, there was a strict requirement on the range of allowable spin rates, both in terms of peak rate and a minimum rate towards the end of the descent. Apart from increasing the drag of a system, the parachute may also provide a horizontal velocity component (called *drive*) to the system, or even to be the means whereby this system acquires a lift force. Another function of a parachute system is to give specified degrees of stability to the payloads to which it is attached.

The performance of the parachute is driven by aspects arising from the environment as well as the vehicle. Obviously, when for a given parachute a payload is heavier the decelerating effect will be smaller, and the velocity at touch down will consequently be larger. The drag force is not only depending on the size and type of parachute, but also on the dynamic pressure. A denser atmosphere will create a larger drag and thus a slower descent through the atmosphere, especially when the

gravitational pull is low. In that respect, Venus and Titan are "easy", because of the high ρ/g. Mars, on the other hand, is most demanding with respect to parachute design, because of the low ρ/g.[4]

At any time, a trade off has to be made between the capability to control the descent (through the drag area $C_D S$) and the energy-absorption capacity of landing the landing system (e.g., the landing legs, landing bags, or a crushable structure). It may be clear that the total payload (or landing) mass will drive the design of both descent and landing system. A further complication may be the presence of surface winds. In case a low landing velocity is required to alleviate the landing loads, a large(r) horizontal surface wind will have a large impact on the incidence angle with which the system will touch down. This might induce asymmetric landing loads, concentrated on a small portion of the landing system. Therefore, since wind is a factor that is largely beyond engineering control, margins in the load-carrying structure as well as the descent velocity have to be taken into account.

A parachute system can consist of more than one (type of) parachute. The *main* parachute is responsible for the major part of the deceleration and stability of the parachute-payload system. By itself, the main parachute can be split into multiple, clustered parachutes in case the payload mass is (very) large or when the risk of partial or no deployment has to be reduced by including redundancy in the design. A *pilot* parachute is a small parachute, which is attached to a deployment bag or the vent of a larger parachute, used to provide the force required to deploy this larger parachute. A *drogue* parachute, finally, is a parachute attached to the payload and used to provide stabilisation or initial deceleration or both. Having a drogue parachute usually implies that a larger parachute will be deployed later in the event sequence. In that case it could also be used as the pilot chute for main-parachute deployment.

In Fig. 9.15 a typical parachute-system sequence is depicted, related to the future manned Orion mission—similar to the one of Apollo. Once the proper deployment conditions are reached, the back cover is jettisoned such that the drogue chutes can be deployed. In this particular example, the two drogue chutes are relatively large, so they are deployed in stages, a process called *reefing*. The drogue chutes inflate, and start decelerating and stabilising the payload. Once stable and a certain velocity reduction has been obtained, the drogue chutes are released. It is now time to deploy the main parachutes. However, for such a heavy payload as the Orion, large main parachutes are required. To guarantee release from the capsule, smaller pilot chutes are deployed, one for each of the three main parachutes. At line stretch, the main chutes start inflating, again in a reefed manner to avoid excessive opening forces. Once the main parachutes are fully deployed the system will reach an equilibrium velocity that allows for a safe landing with an acceptable impact load. Before landing the forward heat shield is jettisoned, such that the additional landing systems can be deployed. This could be landing bags or landing legs (Earth or Mars landing),

[4] The surface density of Venus is $\rho_0 = 65\ \text{kg/m}^3$, at a gravitational acceleration of $g_0 = 8.87\ \text{m/s}^2$. Corresponding values for Titan are $\rho_0 = 5.5\ \text{kg/m}^3$ and $g_0 = 1.352\ \text{m/s}^2$, for Mars $\rho_0 = 0.020\ \text{kg/m}^3$ and $g_0 = 3.71\ \text{m/s}^2$, and for Earth $\rho_0 = 1.225\ \text{kg/m}^3$ and $g_0 = 9.81\ \text{m/s}^2$. Consequently, the ratios of ρ_0/g_0 for Venus, Titan, Mars and Earth are 7.33, 4.07, $5.4{\cdot}10^{-3}$ and 0.12 kg s^2/m^4, respectively.

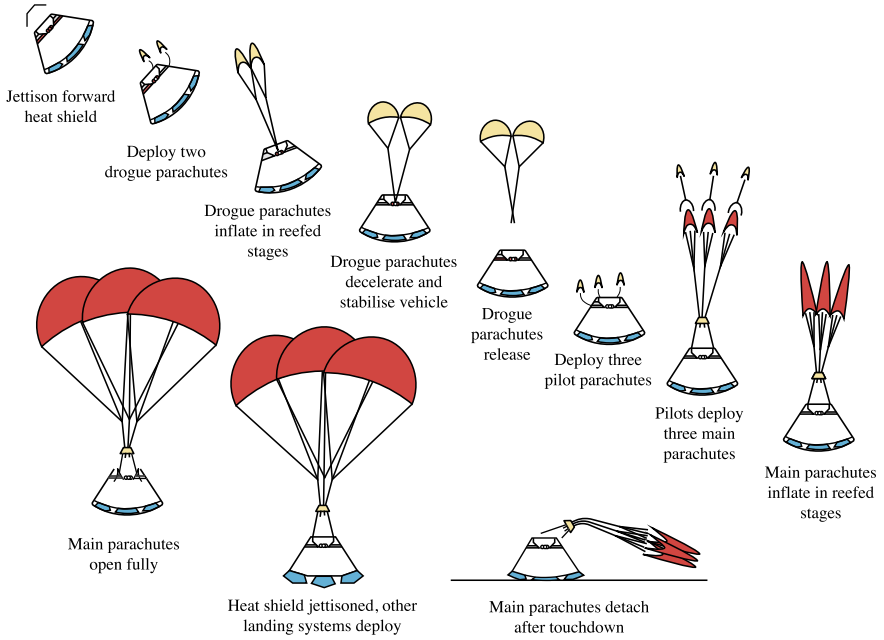


Fig. 9.15 Sequence of events for descent and landing of the Orion capsule

or even the retro rockets (typical for Mars landing). Once the system has landed, the parachutes are cut off. In case retro rockets need to be fired, the parachutes are usually released just before activation. Many of the above aspects will be discussed in more detail in later sections.

Roughly speaking, parachutes can be divided into three categories. The first category consists of the *conventional* parachutes, whose sole purpose is generating drag. This type of parachute is non-gliding and non-rotating. *Rotating* parachutes form the second category. An aerodynamic torque results in autorotation of the canopy and, as a consequence of the centrifugal forces, in an increase of the drag area. In the third category we find *gliding* parachutes, which resemble low aspect wings and are also known as *parafoils*. They are capable of imparting a horizontal velocity component to the parachute and its payload. Due to space limitations, the type of parachute considered in this book is the conventional one, as it is by far the most used for space missions.

In the upcoming sections many design and operational aspects of parachutes will be discussed. The standard works by Ewing et al. (1978) and Knacke (1987) have been the main source of information, and are gratefully acknowledged. We begin the discussion by looking at the hardware elements making up the parachute system, Sect. 9.2.2. As parachutes come in different types, depending on application and performance specifications, in Sect. 9.2.3 the most commonly used parachute types for space missions are introduced.

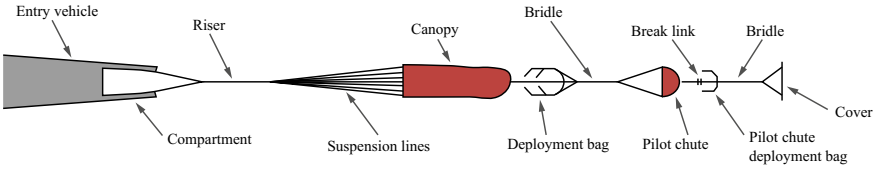


Fig. 9.16 Main elements of a generic parachute system

9.2.2 Hardware Elements

Figures 9.16 and 9.17 show the main elements of a parachute system, both at the moment of jettisoning and when the parachute is fully inflated. Before being used, the parachute is folded and contained in a canister and/or *deployment bag*, which is a textile container from which the parachute deploys. Initiation of the parachute-deployment process begins with releasing the cover that closes the vehicle and protects the folded parachute. This release could, for instance, be done with a dedicated

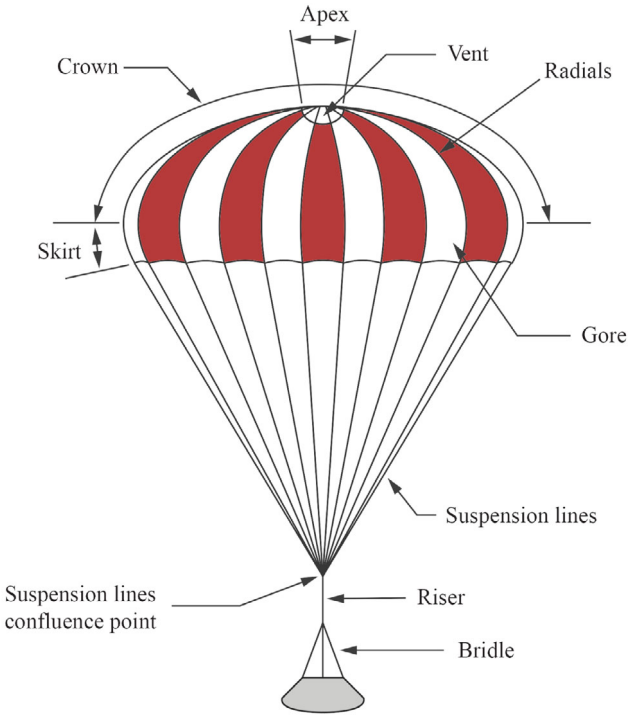


Fig. 9.17 Deployed parachute elements

separation mechanism with pyro bolts/nuts, activated by an electrical current, as was the case for the Huygens-probe backshell separation.

The (back) cover may be directly connected to a (pilot) chute with a *bridle*—a means of providing a (multi-point) connection to a deployment bag or vehicle from a parachute or riser—and includes a break link to disconnect the cover once the (pilot) chute is inflated. A more active means to deploy the parachute can be a drogue gun, a tractor rocket, or a mortar. A *drogue gun* is a deployment device used to fire a projectile at high velocities away from the payload. A riser connecting the projectile to a parachute forces the deployment to begin. Similarly, a tractor rocket is a deployment device with a small rocket at the front, used to pull the parachute system out of the payload and deploy it into the airstream. Finally, a mortar is a more powerful deployment device, which is typically used in applications where performance of drogue or pilot chute may be questionable, and in applications where deployment time has to be minimised (Pleasants 1974).

The pilot chute is connected with the drogue or main parachute with another bridle. The deployment bag should guarantee a controlled and incremental, orderly deployment of riser, suspension lines and canopy, in that order, as the bag moves away from the body in the process.

The *canopy* is a cloth surface that inflates to a developed shape, and is the major drag producing element of the parachute. The open region at the apex of the canopy is called the *vent*, which allows turbulent trapped air escape from its centre. The decreased turbulence and thus smoother airflow has a positive effect on the parachute's stability, and, as a consequence, increases the induced drag force. The canopy has load-bearing members running from the suspension lines at the skirt to the vent lines. The section of the canopy in between two radials is called a *gore*. The *suspension lines* are load-bearing members extending from the canopy skirt to the payload. They transmit the retarding force from the canopy to the body, usually through a riser at the towing end. The *riser* is a single load-carrying member below the convergence of the suspension lines that increases the trailing distance between parachute and payload. It can use a single-point or multi-point attachment scheme. The *harness* (or alternatively, the bridle) is a suspension for the payload. Using a multiple point harness, it is possible to introduce an arbitrary orientation of the payload. This was used in the case of Apollo to reduce the loads at sea impact, by having the (sharper) corner hitting the water first, rather than the blunt heat shield.

9.2.3 Parachute Types

There are quite a few parachute geometries available with different inflation performance, drag coefficient, stability, manufacturing cost, and so on. Lowest in cost is the cruciform, as it is easily manufactured as two strips of fabric sewn at an orthogonal intersection. This kind of parachute is used a lot by the military, but has not seen much use on planetary probes, because of the generally poorer stability. The four

more commonly used parachutes in that application bracket are the ribless guide-surface, the conical-ribbon, the disk-gap-band, and ring-sail parachute.

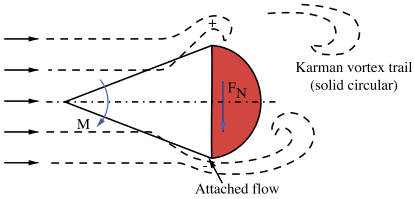
A key feature of a parachute is its porosity, at both the macroscopic (gaps in a ribbon parachute or ring-sail, or in a disk-gap-band chute), and microscopic (porosity in the fabric) scales. If the canopy is porous, a part of the flow can go through, rather than around. This improves the inflation characteristics and in the operational phase its stability. Without sufficient porosity, the parachute will exhibit undesirable oscillations. Porosity and its effect on stability can be measured in wind tunnels or during drop tests, but because the microscopic porosity is particularly sensitive to variations in Reynolds number, additional attention is required to apply test data to different flight conditions.

In this section we will briefly describe the above mentioned four most commonly used parachute types for planetary missions, but before we start a brief introduction to a parachute's primary design space is given. It is noted that the description of the four parachute types is a summary only, and more information can be found in the standard work by Ewing et al. (1978). This design guide also includes detailed mathematical models of how to design a parachute system. Due to limited space, these design models will not be repeated here, but the reader is referred to the quoted reference instead.

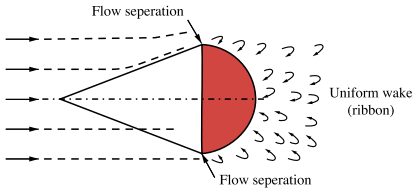
Drag Versus Stability Space

Two key design criteria for the parachute system are the parachute's drag-generating capability and stability characteristics. Obviously, one would like to maximise both, and reduce the oscillations as much as possible. The stability of parachute-descent systems is one of the necessary conditions for successful planetary landings, most notably those on Mars due to the very small performance margins as a result of the thin atmosphere, or those in a strong and/or turbulent wind environment, such as Titan. In other words, it is important to consider the factors, which influence the attitude of the lander during the crucial terminal descent phase. From an engineering point of view, the environmental effects due to wind and wind gusts are the major sources of uncertainty while the lander descends under the parachute (Martin and Zurek 1993). The necessity to somehow absorb the large uncertainties associated with environmental factors into the design results in safety margins of 20–100% and thus incur associated mass penalties. Additionally, these uncertainties of mission design may rule out scientifically attractive landing sites and severely limit available options. It is meaningful to undertake a study of the effects of wind (gusts) on the dynamic stability of parachute systems, since it eliminates one of the uncertainties for a successful landing on Mars.

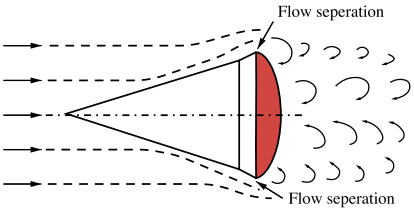
A parachute's stability, or in other words, limiting its oscillations, originates from the flow field around the canopy. In Fig. 9.18, this has been schematised for several canopy configurations. Consider the first configuration in Fig. 9.18a. The canopy is an imporous hemisphere, which means no air can pass through the canopy. Consequently, it has to flow around it and at the edges it will separate and form alternating



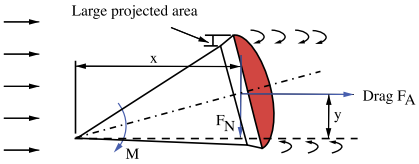
(a) Imporous canopy, destabilising moment $M = F_N \cdot x$



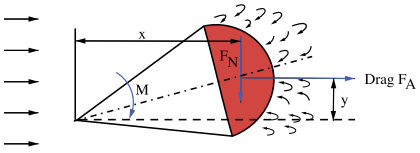
(b) Porous canopy, $F_N = 0 \Rightarrow M = 0$



(c) Canopy with separation edge (guide surface) $F_N = 0 \Rightarrow M = 0$



(d) Canopy with separation edge, stabilising moment $M = F_A \cdot y + F_N \cdot x$



(e) Porous canopy, with stabilising moment $M = F_A \cdot y + F_N \cdot x$

Fig. 9.18 Parachute stability

vortices, forming a so-called Kàrmàn vortex trail. The separation results in alternate pressure areas on opposite sides of the canopy, and these produce the parachute's oscillations.

Suppose now that the canopy cloth is porous, or that it has (venting) openings of another sort (Fig. 9.18b). Part of the flow can now pass through, creating a uniform wake flow with small vortices. The remaining air flows around and (uniformly) separates around the canopy's leading edge. This situation avoids forming a (destabilising) Kàrmàn vortex trail. This type of flow field underlies the ribbon parachute and all slotted parachute canopies. The guide surface parachute, on the other hand, has a sharp separation edge around the skirt of the canopy, which generates a strong, uniform airflow separation around the leading edge, see also Fig. 9.18c. Moreover, as shown in Fig. 9.18d, a large stabilising normal force originates from the inverted leading edge (guide surface). Whenever the parachute deviates from its zero angle-of-attack position, a stabilising moment is induced, with contributions from both the normal and axial (=drag) force, F_N and F_A: $M = F_A \cdot y + F_N \cdot x$. A similar effect occurs for the porous, hemispherical parachute in Fig. 9.18e, although the stabilising moment is smaller, because of the (much) smaller F_N.

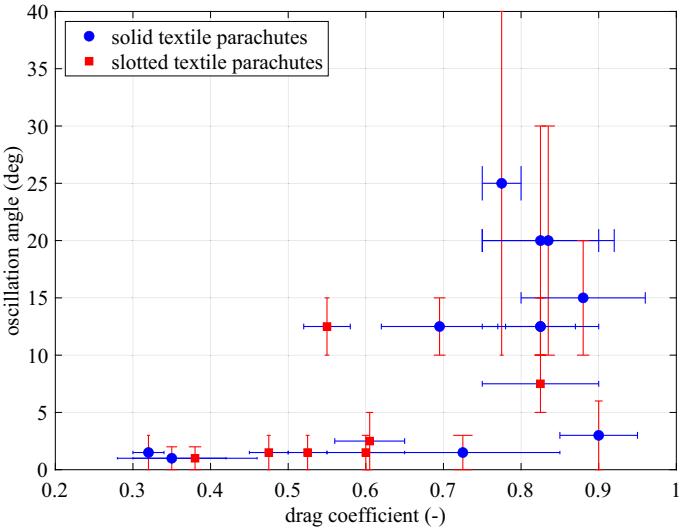


Fig. 9.19 Drag versus stability space, with the range per parachute indicated by error bars. Data: Knacke (1987)

In summary, the known main design features for creating stable parachutes are uniform airflow separation around the leading edge of the canopy, a certain amount of air flowing through the canopy, and a large restoring moment, such as created by a guide surface.

In case the payload itself is inherently stable, the used parachutes must also be stable and not interfere with the stability of the vehicle. Parachutes with good stability and high drag are used, such as the ribbon, ring-slot, or cross parachutes. On the other hand, if a fast-falling, unstable payload must be stabilised, but at the same time keep a high descent rate, a parachute with a strong negative $\mathrm{d}C_m/\mathrm{d}\alpha$, such as a guide surface parachute, could be the parachute of choice.

In Fig. 9.19, the oscillation angle as a function of the drag coefficient has been plotted, for two main types of parachute, being the solid and slotted textile parachutes. Data for this plot have been taken from Knacke (1987), who listed a range for both the drag coefficient and the oscillation angle. These ranges have been indicated as error bars. From the figure we see that in general the solid textile parachutes have larger drag coefficients, but are also less stable. When stability is an issue, the slotted textile parachutes provide a good design option. The commonly used disk-gap-band parachute for planetary missions score reasonably well in both objectives.

The final selection of the parachute system is a compromise between the required stability and induced drag force. Even though these criteria are very important ones, they are not the only design criteria. as also the following are to be considered:

1. Heritage. It may save a lot of time and money if one can use the experience (and design data) of previous projects. The more compatible the system and mission requirements are, the more beneficial this is, of course.
2. Drag area versus mass trade. A smaller drag area because of a larger drag coefficient will lead to mass savings that may either help meeting the design requirements, or will allow for more payload mass.
3. Robustness. No matter how many sensitivity analyses are done, there may always be uncertainties that could either not be assessed or their magnitude properly predicted. Therefore, it is a useful trade to have a more robust design that allows for just that.
4. Deployment and inflation reliability. The inflation process is a critical part of the operational phase. A higher reliability will therefore increase mission success.
5. Cost and ease of fabrication. Reducing cost while meeting the requirements is an obvious benefit. Ease of fabrication may actually contribute to that, but may also lead to more reliable hardware.

Guide Surface (Ribless) Parachutes

The guide-surface parachute (Fig. 9.20) has seen few space applications. One known example is that it was used by Pioneer Venus as a pilot chute. A few of the parachute's characteristics are listed below:

- Low drag ($C_{D,0} \approx 0.3$) with good stability ($-3° \leq \alpha \leq 3°$)
- Used in situations where stability is principal consideration (drogue, pilot)
- Abrupt transition at maximum projected diameter and subsequent flow separation is reason for stability characteristics
- Appropriate for supersonic and subsonic applications
- Difficult to manufacture.

Conical Ribbon Parachutes

Circular or "flat" canopies usually have wider oscillations than conical types and are therefore rarely used. Conical parachutes have triangular gores (see Fig. 9.21), and the parachute has thus a conical shape. However, once fully inflated, this cone tends to be rather rounded. The conical-ribbon parachute has been used by Pioneer Venus and the Galileo Jupiter entry probe. As a subsonic parachute, it was used, amongst others, for the Space Shuttle's Solid Rocket Boosters, and the Russian equivalent for both the Soyuz and the Ariane-5 Booster Recovery System. Some of its more important characteristics are:

- Ribbon-type parachutes can have high porosities
- Moderate drag ($C_{D,0} \approx 0.5$) with good stability ($-3° \leq \alpha \leq 3°$)
- Appropriate for subsonic and supersonic applications

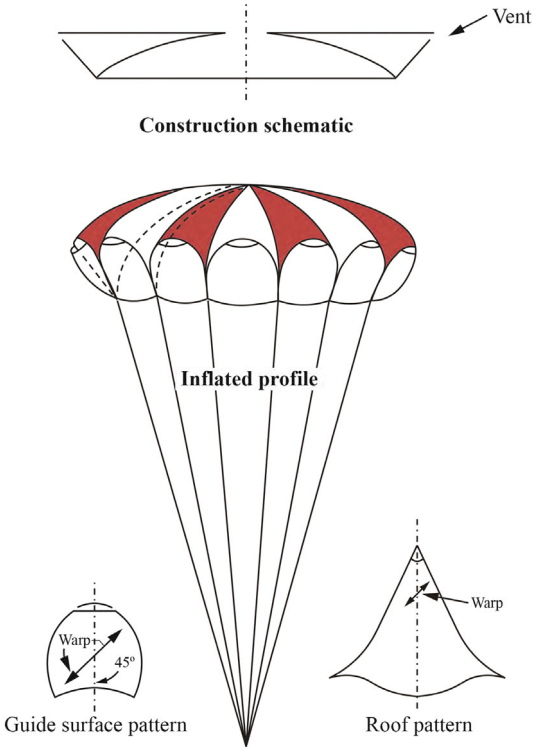


Fig. 9.20 Guide surface (ribless) parachute

- Can be made very strong (especially if manufactured from Kevlar), because materials can be made stronger as ribbons than as broadloom fabric, so they can be deployed at high dynamic pressure
- Relatively high weight per unit drag area.

Disk-Gap Band Parachute

Disk-gap-band (DGB) parachutes (Fig. 9.22) form a variant of circular canopies that have better stability characteristics: the gap allows a through-flow that stays better attached to the canopy. Therefore, it has no asymmetric flow separation, which is a known cause to induce oscillations. Each gore is approximately triangular, with a rectangular segment to form the band. The DGB parachute is the parachute system most commonly used for planetary missions: Viking, Mars Pathfinder (MPF), Mars Polar Lander (MPL), Phoenix, Beagle-2, Mars Exploration Rover (MER), Huygens, Genesis, Stardust, Mars Science Laboratory (MSL), and InSight. Due to the extensive mission heritage, built up since the days of the Viking programme, and running through all of NASA's Mars missions, this type of parachute is the starting point when designing a parachute system for planetary missions.

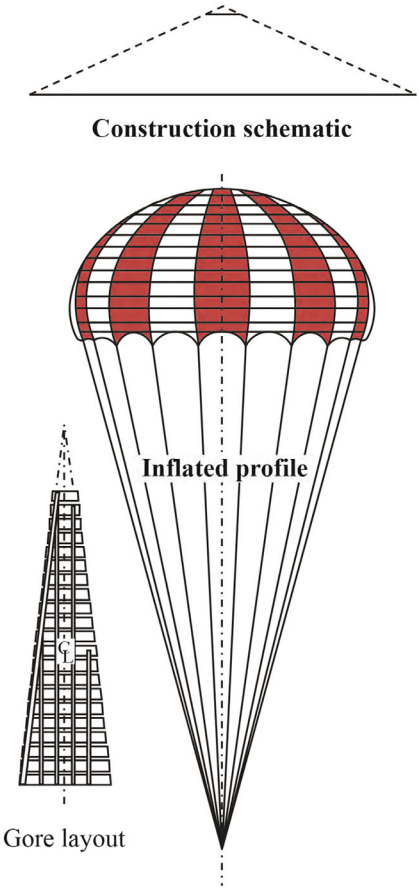


Fig. 9.21 Conical ribbon parachute

Some of its key characteristics are:

- Low-to-moderate drag ($C_{D,0} \approx 0.4$ to 0.7) with good-to-moderate stability (α: $\pm 5°$ to $\pm 15°$)
- Drag can be traded for stability by changing the gap and band heights
- Appropriate for subsonic and supersonic applications
- Strong heritage data at supersonic speeds in low density atmospheres is key to its continued use.

Ring-Sail Parachutes

Ring-sail parachutes are something of an intermediate between a ribbon chute and a hemispherical chute, with one side of the panels in each gore being free, allowing a

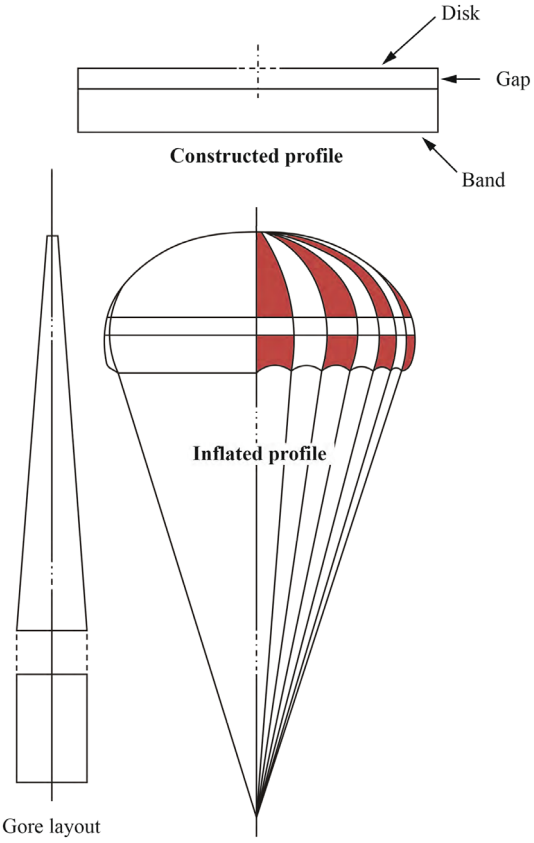


Fig. 9.22 Disk-gap band parachute

flow through the canopy. A schematic overview of the ring-sail parachute is provided in Fig. 9.23. A summary of its key characteristics is:

- High drag ($C_{D,0} \approx 0.8$) with good-to-moderate stability (α: $\pm 5°$ to $\pm 10°$)
- Design tailored for optimum performance by varying canopy shape and distribution of geometric porosity throughout canopy
- Currently limited to subsonic applications
- Time consuming fabrication
- Relatively light weight per unit drag area.

The ring-sail parachute has been used as space-vehicle recovery or landing system by, for instance, Mercury, Gemini, Apollo, and Beagle-2. The latest successful application was the one at the end of the Artemis-1 mission: the sea landing of the Orion capsule on December 11, 2022.

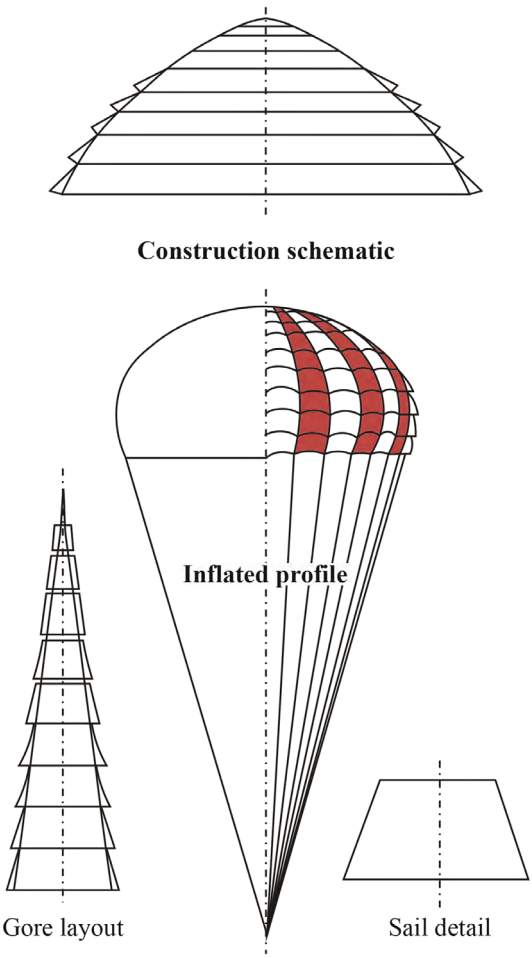


Fig. 9.23 Ring-sail parachute

9.2.4 Materials

Textile structures represent the primary components in parachute applications (parachute canopies, air bags, suspension members), and during the descent and landing sequence they are subjected to dynamic loads. Even though a detailed stress analysis is considered to be beyond the scope of this rather short discussion on parachute systems, we feel the need to give an overview of the most commonly used materials, and indicate how these could be used in the overall (conceptual) design process.

Material selection can be a tough subject during the design of a parachute. In fact, in general nylon and polyester fabric are suitable for typical parachute applications,

but when major issues arise, such as aerodynamic heating with deployment at high speed, friction-generated heat during deployment or storage entailing elevated temperatures, other considerations must be taken into account. Most applicable fibres are derived synthetically and special attention is given to fibres with a capacity to absorb energy of loading and which retain useful strength margins at elevated temperatures. Polymers are also useful as films, coatings, adhesives, rigid and semi-rigid foams, or special structural forms (Ewing et al. 1978).

Before we come to the presentation of different materials, first some definitions will be given.

A **fibre** is a slender, threadlike structure.
A **filament** is a man-made fibre having great length compared to natural fibres.
A **stample** is a measure of length and/or fineness of a fibre.
Yarn is a generic term for continuous strands of textile stample fibres or continuous filaments in a form suitable for knitting, weaving, or otherwise intertwining, to form a textile structure.
Denier is a term used to describe the weight and size of a fibre or yarn (weight in grams of 9,000 m of material).
Tex is the mass in grams per 1,000 m of yarn.
Tenacity is the strength of the fibre or yarn. Higher-tenacity yarn must be of smaller denier to maintain adequate flexibility for weaving (Wolf 2005).
Tensile strength, which is the absolute value of a fibre's strength, must be related to its cross-sectional area or to its linear density. Tensile strength is normally given in pounds per square inch (psi). However, the tensile trade uses the term "tenacity" to describe strength on a gram per denier (gpd) basis. Yarn weight is another important physical and economic textile factor. Since denier is based on weight per unit of length, tenacity is influenced by specific gravity of the fibre material, whereas strength per unit of area, or stress, is not. Specific gravity is defined as the ratio of the unit volume weight of a substance to that of water at 4 °C. To convert tenacity to stress, the following expression holds:

$$\text{Stress (psi)} = 12{,}800 \times \text{specific gravity} \times \text{tenacity (gpd)}$$

The **breaking tenacity** is the strength in grams per denier at the point of rupture. A fibre with high breaking tenacity is proper to use where weight efficiency is needed in fabrics, corage or threads.
The **breaking elongation** is a percent measure of the amount of stretch at the point of rupture. As a general rule, high strength fibre have low elongation.
The **modulus** is the ratio of change in stress to change in strain. High-modulus materials are stiff and show small elongation under load, whereas low-modulus materials are extensible and exhibit high elongation.

In the following, the types of materials that are used for parachute applications will be listed with their main characteristics.

Nylon

- Polyamide resin with high tenacity due to long, highly oriented molecules and high intermolecular forces
- A primary parachute material (Nylon 6-6)
- Tenacity is 2.5–9.5 gpd (6.6 for Nylon 6-6, which represents a tensile strength of about 662 MPa (96 ksi)
- Minimum elongation specified as 20%, actual elongation 30–40% at break
- Sensitive to ultraviolet light
- Nylon does not burn, but rather melts at about 250 °C (480 °F).

Dracon

- Polyester based man-made fibre
- Similar to nylon, but requires more treatment for stability
- Slightly lower elongation but better temperature resistance than nylon
- Slightly less weight and volume for comparable strength nylon.

Polytetrafluoroethylene, PTFE

- Fluorocarbon based man-made fibre
- Very low friction coefficient
- Very expensive
- Used for deployment bag liners
- Used as buffer material where cotton cannot be used
- Expanded PTFE (Goretex) available as a slit film fibre with a tenacity of 3 gpd.

Nomex

- Polyamide man-made fibre
- Moderate tenacity of 5.4 gpd (tensile strength of 620 MPa, or 90 ksi)
- Elongation at break about 16%
- Retains 50% of its strength at 260 °C (500 °F)
- Used for flight suits and high-temperature buffers
- Used as cloth in parachutes in high-temperature environments.

Kevlar

- Para-aramid fibre developed for use in radial tires
- “Intermediate-modulus para-aramid” version used for parachute materials
- High strength and modulus
- Each fibre is 0.00047 inches in diameter, approximatively 1.5 denier
- Tenacity of 20–22 gpd (tensile strength of about 2.5 GPa or 370 ksi)
- Elongation at break of 3–5%
- Kevlar does not melt or burn, and it is abrasion resistant, even if abrasive itself
- Sensitive to UV
- Cannot be cut or have the bitter ends seared.

Spectra

- High-strength polyethylene fibre
- Tenacity of 30–35 gpd (tensile strength of about −3.0 GPa or −440 ksi)
- Elongation at break of 2.7%
- Loss of strength at a certain temperature (10g/d at 100 °C or 212 °F)
- Low friction coefficient
- Used as a deployment-bag material for high-density packs.

Vectran-Liquid Crystal Polymer (LCP)

- Relatively new material entering the field of parachutes
- Tenacity of 23 gpd (tensile strength of about 2.0 GPa or 300 ksi)
- Elongation at break of 2–3%
- Friction properties similar to Spectra
- Immune to acids and UV
- Specific gravity slightly less than 1.0
- Experiences zero creep under constant load.

Zylon

- Tenacity of 43 gpd (tensile strength of about 5.8 GPa or 840 ksi)
- Elongation at break of 3.5–2.5%
- Decomposition temperature in air is 650 °C
- Much lower friction coefficient than Kevlar
- Significant strength reduction with prolonged exposure to light (65% reduction in 6 months) (Wolf 2005).

The material's choice will also have an effect on the mass and volume of the parachute system. Therefore, this section is concluded by discussing an easy way to assess the parachute mass. There are two main methods to compute the mass of the parachute system, the *preliminary design method* and the TWK Method. The former method uses data of past parachute designs to estimate the mass, whereas the latter is used when no detailed drawing is available, but the primary dimensions of the parachute are known. The mass of a parachute is then found to be (Knacke 1991):

$$M_p = S_0 w_c + d_0 n_g w_{rt} F_{rt} + n_s L_s w_s F_s \tag{9.1}$$

where

M_p = mass of the parachute (kg)
S_0 = surface area of the finished canopy (m^2)
w_c = specific canopy weight (kg/m^2)
d_0 = nominal canopy diameter (m)
n_g = number of gores in the canopy (-)
w_{rt} = specific weight of the radial tape (kg/m)
F_{rt} = strength of the radial tape (kg)

n_s = number of suspension lines (-)
L_s = length of suspension lines (m)
w_s = specific weight of suspension lines (kg/m)
F_s = strength of suspension lines (kg).

In Eq. (9.1), d_0, S_0, n_s, L_s, F_{rt} and F_s are known design data. The specific canopy weight, w_c, refers to the entire canopy, including vent and skirt tape, but is not related to the reinforcing radial tapes. The specific weight of lines and webbings decreases with increasing line/webbing strength. In addition, the mass calculation for suspension lines can be used for risers as well.

9.2.5 Operational Aspects

9.2.5.1 Inflation

The inflation of the parachute is in more than one way important. Of course, the most obvious one is that it has to fully open to become effective. But to go from a uninflated pack of small dimensions to a (fully) inflated canopy of several to many meters in diameter has a serious impact on the strength of the other elements of the parachute system as well. After all, the opening loads of the parachute are the largest ones occurring during the operational sequence of the parachute system, which means that the maximum parachute structural loads almost always occur during inflation. Therefore, structural analysis and performance calculations will frequently require accurate inflation-time predictions.

The inflation process has been schematised in Fig. 9.24, and, after ejection of the parachute from its storage canister, begins with the opening of the canopy mouth and initiates the inflation (number #1 in the figure). The air mass will now move along the canopy (#2) until it reaches the top of the canopy, the *crown*. There may be a venting hole in the crown, but it will be clear that any influx of air will be a lot larger than the possible outflow, such that the crown will expand (#4). After a while the structural tension in the canopy material, as well as its general inertial, will start resisting the expansion (#5). At the same time, the inlet at the bottom will now also expand, so that more air will flow inside the canopy, aiding the expansion (#6). The canopy will fill rapidly now, until it reaches its maximum expansion (#7), and the air pressure inside the canopy will get into equilibrium with the structural tension in the canopy "structure". When inflation is complete, the skirt is over-expanded, whereas the crown partially "collapses" by the momentum of surrounding air mass. The total time it takes, i.e., the inflation time, is defined as the time interval from the instant the canopy and the lines are stretched to the point when the canopy is first fully inflated.

During inflation, the parachute drag area increases from 0 to 100%. The drag-area versus time curve may be linear, convex, concave or even random, but it has proven to be constant and repeatable for known parachute types. In Fig. 9.25 some examples of drag-area curves are shown.

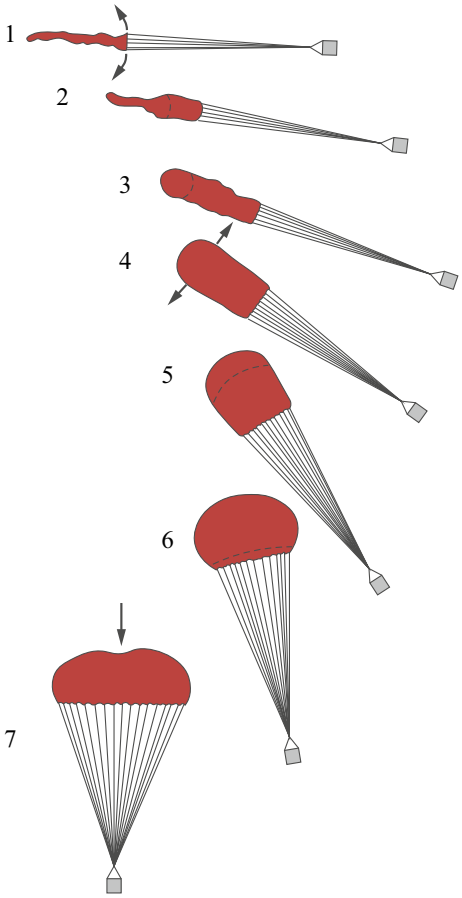


Fig. 9.24 The inflation process and opening of the parachute canopy

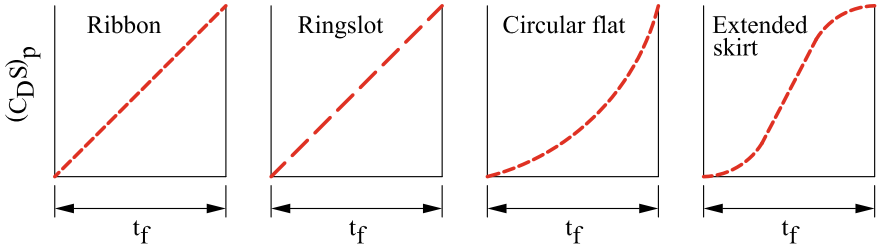


Fig. 9.25 Typical drag-area versus time increase for various parachute types

The parachute-deployment sequence is typically initiated by some trigger, e.g., reaching a certain atmospheric pressure (thus altitude) or at a certain maximum allowable dynamic pressure. Whenever this happens, the vehicle is never in vertical

flight, and may even be flying quite horizontally at only some tens of degree negative flight-path angle. At the beginning of inflation the parachute hardly generates a drag force. That changes rapidly when the crown starts expanding. The increased parachute drag will quickly force a more vertical flight direction. When the parachute is fully inflated, the drag force is so large that the system is usually on a vertical flight path, with the possible exception of Mars parachute descent.

Successful parachute inflation is a combination of many parameters, a few of them being: (1) fabric and geometric porosity (how much air will pass through the canopy having an averse effect on expansion), (2) Mach number (aerodynamic effects), (3) dynamic pressure (how fast the canopy will inflate), (4) forebody shape, the forebody-to-parachute area ratio, and the canopy trailing distance (which all determine the (turbulent) wake-effect on the parachute), and (5) the suspension-line length to reference-diameter ratio, which, if large, shortens the trailing distance, as well as increases the tension in the canopy. The amount of air moving toward the canopy vent should be small so that a high-mass shock, due to the air bubble reaching the vent of the parachute, can be avoided. Last but not least, the inflation of the canopy should be as axisymmetric as possible to prevent over-stressing individual canopy parts.

By definition, the canopy filling-time is the time from (suspension) line stretch to the moment that the canopy has fully opened for the first time (as it may contract and expand before it completely stabilises. Mueller and Scheubel, in Knacke (1987), reasoned that parachutes should open within a fixed distance, defined to be proportional to the nominal diameter, d_0. This assumption was confirmed by drop tests with ribbon parachutes, but also by Schilling (1957) and French (1968). The filling distance for a specific parachute canopy is defined as $s_f = k_f d_0$, with k_f being the fill factor.[5] This fill factor is not truly a constant, but may be influenced by canopy porosity, diameter, and opening Mach number. With s_f, the filling time can be written as (see also Fig. 9.26):

$$t_f = \frac{s_f}{V} = \frac{k_f d_0}{V} \tag{9.2}$$

where V is the velocity at line stretch. It is noted that some variations to the above equation exist, when filling times have been correlated with test data (Knacke 1987). He also provides expressions for the filling time of the various stages of reefed parachutes:

$$t_{f,1} = \frac{k_f d_0}{V_s}\sqrt{\frac{(C_D S)_r}{(C_D S)_0}} \qquad t_{f,2} = \frac{k_f d_0}{V_r}\sqrt{\frac{(C_D S)_0 - (C_D S)_r}{(C_D S)_0}} \tag{9.3}$$

[5] Ewing et al. (1978) lists the filling factors derived from measurement data for many parachute types, e.g., for the conical-ribbon parachute it is $k_f = 14$ ($M = 1.1$), for the ringslot $k_f = 14$, the ringsail $k_f = 7$ ($M = 0.2$, but $k_f = 1.4$ for $M = 0.3$), the ribless guide $k_f = 6$ ($M = 1.1$), and the disk-gap band $k_f = 6$ ($M < 2$; for $M = 3.2$ it jumps to $k_f = 51$). These numbers hold for unreefed openings, in case reefing is applied the numbers change.

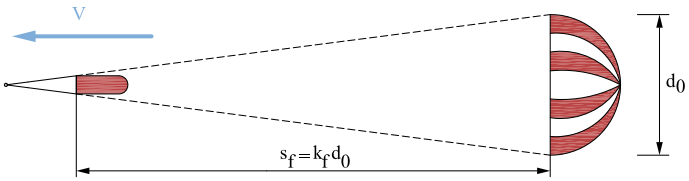


Fig. 9.26 Definition of filling distance

with $t_{f,1}$ and $t_{f,2}$ being the reefed and disreefed filling times in sec, V_s and V_r the velocities at line stretch and at disreef in m/s, and $(C_D S)_r$ and $(C_D S)_0$ being the reefed and full-open drag areas in m^2.

9.2.5.2 Opening Loads

In terms of parachute operations, those for missions on Mars are the most demanding. Due to the very thin atmosphere, the parachutes have to be opened supersonically, and, despite the large velocity, at low dynamic pressure due to the very low density. The envelope of opening conditions, i.e., the combination of Mach number and dynamic pressure, has been expanded over the years, with the availability of better manufacturing techniques and new materials. Inflation failure is a phenomenon that is not well understood. Inflation failure or *squidding* results in a partial inflated or collapsed parachute canopy. No sound model exists that can predict how and when this will occur. Facing an unknown Martian atmosphere in the early days of planetary exploration, successful inflation had to be demonstrated at real or near-flight conditions. To do so for a Mars mission is not trivial, because a Martian descent system has to operate at supersonic velocities, at low densities and in a low-gravity environment. This is further complicated by the shear dimensions of the parachute-capsule combination, which restricts wind-tunnel testing to low dynamic pressures at low velocities.

Therefore, in the late 1960s and early 1970s, many flight tests have been conducted, in parallel with (scaled-down) wind-tunnel tests, to support the development of the Viking-1 and -2 missions to Mars. The three main flight-test programs were named the Planetary Entry Parachute Program (PEPP), the Supersonic Planetary Entry Decelerator Program (SPED), and the Supersonic High Altitude Parachute Experiment (SHAPE), see, for instance, Murrow and McFall (1969), and Cruz et al. (2006) and the many references therein. Between the three programs, sixteen high-altitude supersonic flight tests were performed: five with ring-sail parachutes, three with cross parachutes, and eight with disk-gap-band parachutes. The decelerator vehicles were brought to high altitude by balloon, such that the low-density and dynamic-pressure conditions of Mars could be matched. After release, the test vehicles were rocket-accelerated to the right supersonic inflation conditions. In the end, the DGB parachutes were chosen for Viking-1 and -2, and have also been the parachute of

choice of most follow-on Mars missions, which would allow for extensive re-use of the Viking database to avoid the expensive flight testing.

In Fig. 9.27, the DGB deployment conditions of the test programmes, as well as the flown (NASA) Mars missions are plotted. Apart from the test programmes PEPP, SPED, SHAPE (Murrow and McFall 1969), and Viking-AV (Moog and Michel 1973), also the actual Viking opening conditions are shown (Moog et al. 1979). It is interesting to see that the Viking parachute was qualified up to Mach 2.1 at a dynamic pressure of 700 Pa, while the actual flights gave Mach 1.1 and 320 Pa. This was caused by the still largely unknown Martian atmosphere at that time and the relatively low ballistic coefficient of the descent module. Further, the conditions for Mars Pathfinder (MPF) (Spencer et al. 1999), the two Mars Exploration Rovers (MER) (Witkowski and Kandis 2005), Phoenix (Blanchard and Desai 2011), Mars Polar Lander (MPL),[6] Mars Science Laboratory (MSL) (Cruz et al. 2014), and InSight (Witkowski et al. 2019) are shown. It is clear that over the years the boundaries have been pushed, i.e., with stronger parachute designs, the parachutes could be opened at higher Mach numbers (and thus also higher dynamic pressure). The current front runner is MSL, which opened its DGB parachute with a nominal (reference) diameter of 21.35 m (33% larger than Viking's parachute) at Mach 2.1 to decelerate a payload close to 2,500 kg (compared to 930 kg for Viking). Viking, MER and MSL will be discussed in more detail in Sect. 9.7.

The opening of the parachute canopy will give the largest forces on both canopy, lines, attachment brackets, and the payload. During the design phase it is therefore necessary to accurately compute the forces that occur during parachute opening. In Fig. 9.28, a schematic overview is given of the time history of the opening force, including some key moments. In this figure it is assumed that deployment begins with mortar firing, which gives a reaction force on the structure, the *mortar recoil force*. The *snatch force* happens when the lines are stretched, and this induces a small shock load. When the inflation begins, the drag force slowly starts increasing until over expansion of the canopy, which results in the *peak opening force*. After finding its equilibrium, the drag force settles to its nominal value. For large parachutes, this process can last up to 7–10 s.

Several methods can be used to compute parachute opening forces. The most accurate one is to combine detailed models for dynamics, structural dynamics, and unsteady aerodynamics. Such a model-description is far beyond the goal of the current chapter, so we will restrict ourselves to a simple, analytical model. The most commonly used methods are the $W/(C_D S)$ Method and the Pflanz Method (Knacke 1987). Whereas the former is fast, but suitable only for preliminary design, the second is mathematically exact, but it has shown some limitations. Cruz et al.

[6] Even though MPL failed at some point during the entry, descent, and landing—it successfully separated from the orbiter, but it is believed to have crashed—the JPL special review board ruled out that the parachute system was the likely cause of failure (JPL Special Review Board 2005). It is thought that a premature shutdown of the descent engines was the most likely cause. It was estimated that parachute-deployment initiation occurred at a Mach number between 1.7 and 1.85, at a corresponding dynamic pressure between 440 and 564 Pa. Therefore, this range has been indicated in the figure as error bars.

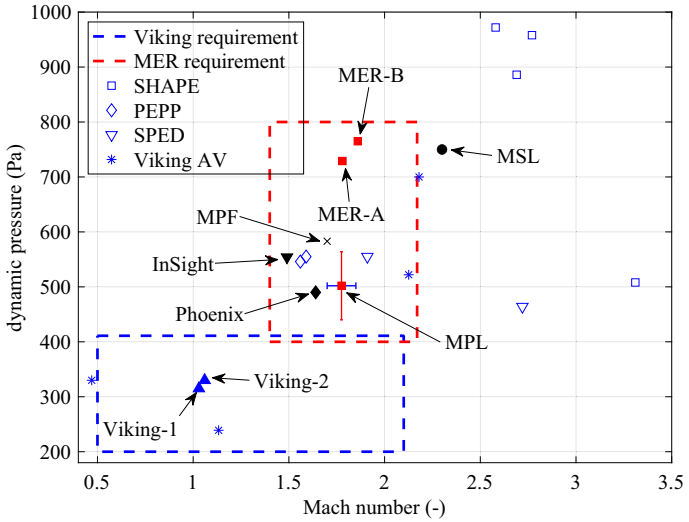


Fig. 9.27 Successful test and flight inflation of the disk-gap-band parachute. Data taken from several sources, indicated in the text

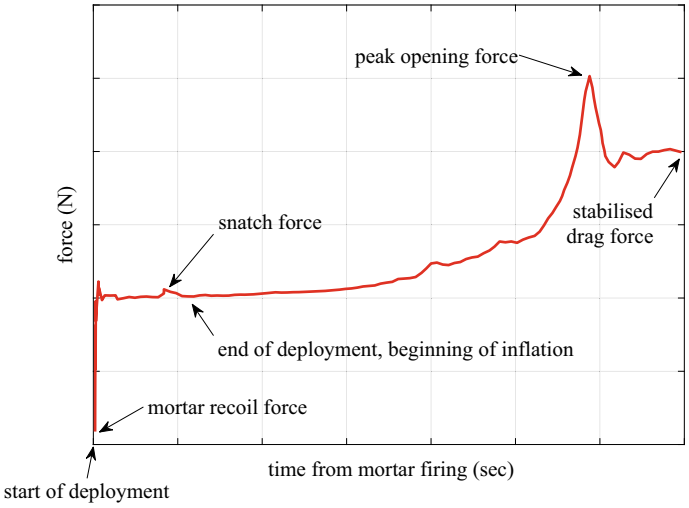


Fig. 9.28 Schematic overview of the forces occurring during the opening of the parachute

(2003) describe an analytical model, derived from an earlier model by McEwan (1970), to support the parachute drop and wind tunnel structural qualification tests for the Mars Exploration Rovers.

It is assumed that the parachute-payload system is in vertical motion, following mortar deployment of the parachute. The parachute is considered to be hemispherical. The force generated by the parachute (on the payload) from the start of inflation to full inflation is given by

$$F_p = \bar{q} C_D S + \frac{\mathrm{d}}{\mathrm{d}t}\left[\left(m_a + m_p\right) V\right] - m_p g = \bar{q} C_D S + \left(m_a + m_p\right) \frac{\mathrm{d}V}{\mathrm{d}t} + \frac{\mathrm{d}m_a}{\mathrm{d}t} V - m_p g \quad (9.4)$$

where $\bar{q}$ is the dynamic pressure, $C_D S$ is the parachute drag-area, m_a is the apparent mass (due to the enclosed air under the parachute), m_p is the structural mass of the parachute, V is the airspeed, and g is the acceleration of gravity. The apparent mass is assumed to be described by

$$m_a = k_0 \rho \mathcal{V}_0 S_r^{3/2} = k_0 \rho \pi d_0^3 / 12 S_r^{3/2} \quad (9.5)$$

with ρ being the atmospheric density and $\mathcal{V}_0 = \pi d_0^3/12$ the volume of a hemisphere. Furthermore, $S_r = S_r(t) = S_p(t)/S_{p,max}$ is the parachute projected area ratio (with $S_p(t)$ the projected parachute area at a given point in time during the inflation and $S_{p,max}$ the maximum projected parachute area). The constant k_0 is non-dimensional, which is commonly determined for the condition of peak-opening load by substituting the appropriate values of the other parameters.

Having an expression for m_a allows us to do the differentiation, required for Eq. (9.4):

$$\frac{\mathrm{d}m_a}{\mathrm{d}t} = k_0 \frac{\mathrm{d}\rho}{\mathrm{d}t} \mathcal{V}_0 S_r^{3/2} + \frac{3}{2} k_0 \rho \mathcal{V}_0 S_r^{1/2} \frac{\mathrm{d}S_r}{\mathrm{d}t} \quad (9.6)$$

The drag area $C_D S$ is approximated by $C_{D_0} S_0 S_r$, where C_{D_0} is the steady-state parachute-drag coefficient for a fully inflated parachute, and S_0 is the nominal parachute area. Using this approximation, substituting Eqs. (9.5) and (9.6) into Eq. (9.4), and neglecting smaller terms involving m_p and $\frac{\mathrm{d}\rho}{\mathrm{d}t}$, results in the basic equation for the calculation of opening loads:

$$F_p = \bar{q} C_{D_0} S_0 S_r + k_0 \rho \mathcal{V}_0 S_r^{3/2} \frac{\mathrm{d}V}{\mathrm{d}t} + \frac{3}{2} k_0 \rho \mathcal{V}_0 S_r^{1/2} \frac{\mathrm{d}S_r}{\mathrm{d}t} V \quad (9.7)$$

The reader is referred to the paper by Cruz et al. (2003) for an application of this model, and correlation with drop-test results

To limit the parachute opening forces to predetermined values, the increment of the drag area ($C_{D_0} S_0 S_r$ in Eq. (9.7)) has to be controlled in some way. To achieve this, one can think of two or more decelerators of different sizes (but all smaller than the case of using only one decelerator), which are deployed sequentially. An alternative is parachute reefing, which permits the step-wise opening of a parachute canopy (or restrains the parachute canopy from full inflation) at predetermined time intervals (reefing steps). Parachute reefing will be discussed next.

9.2.5.3 Reefing

In case a parachute is large, or even when a smaller parachute is opened at high dynamic pressure, the opening loads can be significant. An option to reduce the opening loads is to only partially inflate the parachute by temporarily restricting the canopy at the skirt: this process is known as reefing. One or more reefing lines are threaded through rings at the parachute skirt, and the free length of the line controls the degree of reefing and thus limits the canopy opening. With a smaller opening, also the drag area is reduced and so will be the opening force. Once the reefing line is cut with a so-called reefing cutter, it allows the parachute to continue inflation. When more than one line is included in the design, the reefing can be performed in multiple stages, thus showing a stepwise increase in drag area.

When parachutes are clustered, especially when they are large, uniform inflation is of importance to avoid widely varying parachute forces that may induce severe oscillations of the system, as well as asymmetric structural loading. Reefing all cluster parachutes close to stage #4 in Fig. 9.24 synchronises the inflation process. In this way, all parachutes can obtain an initial, uniform inflation and avoid a lead/lag inflation of any of the parachutes.

Furthermore, reefing can also be an effective method for drag-area control, i.e., to reduce it and so to temporarily increase the rate of descent. Obviously, to obtain a low-impact velocity the parachute has to be disreefed again, shortly before the impact with the ground. Moreover, a (slightly) smaller parachute diameter will typically also lead to an increased parachute stability.

When the parachute is trailing a large forebody that creates a significant wake, the turbulent flow entering the parachute may cause overinflation. In that case the installation of an overinflation control (OC) line might be necessary, which can decrease the parachute opening force. This line is installed around the skirt of the parachute canopy in the same way as a reefing line. It will ensure full canopy inflation, but will prevent overinflation.

Whatever the reason for reefing may be, including this in the parachute design adds complexity and possible failure modes. Of course, these aspects have to be analysed—and tested!—while designing the parachute system. For a first analysis of the dynamics involved, an approximation of the parachute drag area can be made to assess the effect on mechanical loads, descent rate, and descent time. In Fig. 9.29 the area ratio of the three clustered parachutes of the Space Shuttle's Solid Rocket Boosters (SRBs), obtained from test results, is shown. For one of the profiles, a simple

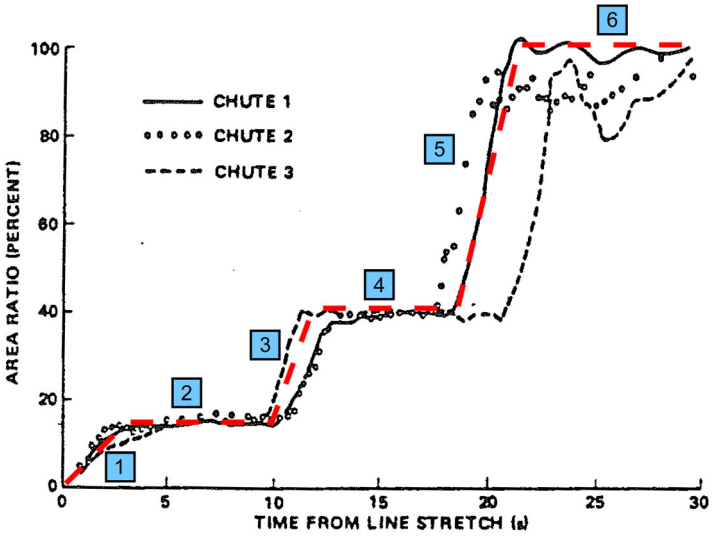


Fig. 9.29 Solid Rocket Booster reefing profiles, obtained from drop tests, and a simple approximation (red dashed line) for computer implementation. Image adapted from Moog et al. (1979)

approximation could be formed from constant and linear segments, as a function of time from line stretch:

(1) $x_f = 0.047t$, for $0 \leq t \leq 3$,
(2) $x_f = 0.14$, for $3 < t \leq 10$,
(3) $x_f = 0.14 + 0.13(t - 10)$, for $10 < t \leq 12$,
(4) $x_f = 0.40t$, for $12 < t \leq 19$,
(5) $x_f = 0.40 + 0.2(t - 19)$, for $19 < t \leq 22$, and
(6) $x_f = 1.0$, for $t > 22$.

During the simulations, offsets can be introduced to account for asymmetric inflation and/or reefing times. Note that the drag coefficient will behave in a non-linear manner that is not easily derived from test results. Of course, from measured accelerations the effect of drag could be derived, but the effect of inflated parachute area and drag coefficient may not be easily separated. It is better to simulate the effect of reefing on drag area $C_D S$ rather than the individual parameters.

Knacke (1987) distinguishes four reefing methods: skirt reefing, skirt reefing with control line, vent reefing, and slider reefing. Of these four methods, plain skirt reefing is the most commonly used and will be discussed below.

In skirt reefing (Fig. 9.30), reefing rings are attached to the canopy skirt at the connection point of each suspension line. This leads to the restriction of the canopy through a continuous line, called the *reefing line*, which is guided by the reefing rings and reefing-line cutters, each of them with a pyro-time train and a cutter knife.

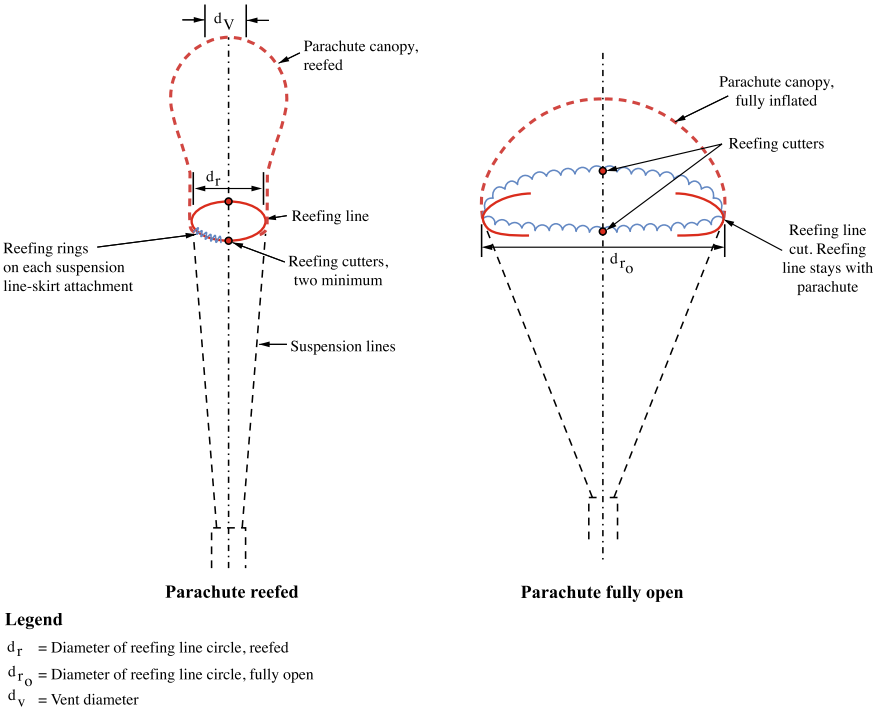


Fig. 9.30 Skirt reefing

After a pre-arranged time, the cutter fires and the knife severs the reefing line so that the parachute canopy either enter the next reefing stage or it can inflate to its full opening.

The length of the reefing line is determined by the reefing ratio, i.e., the predetermined reduction in parachute drag. This and other aspects that the designer has to consider during the design of the skirt-reefing system:

1. Define the reefing line length as the installed length,
2. Determine the forces in the reefing line and the resultant forces in the reefing system,
3. Ensure that the reefing system does not over-stress the parachute canopy, and
4. Ensure that the vent diameter of the canopy, d_v, is smaller than the diameter of the reefing-line circle, d_r.

The required reefing-line length, L_r, for a specific drag area, $(C_D S)_r$, is determined by the required drag-area ratio, ε,

$$\varepsilon = \frac{(C_D S)_r}{(C_D S)_0} \tag{9.8}$$

and the corresponding reefing-line ratio, τ:

$$\tau = \frac{d_r}{d_0} \tag{9.9}$$

Here, d_0 is the (known) nominal diameter of the parachute and d_r is the diameter of the reefing-line circle. L_r follows then with

$$L_r = d_0 \pi \tau \tag{9.10}$$

The forces in the reefing line and its components must be known before the design of the reefing system. For practical reasons, the designer can define the reefing-line force as a function of the known reefed parachute force, F_{rl}, for the particular reefing stage. This force is related to the radial force at the skirt of the parachute canopy. In turn, this radial force depends on the parachute force, F_r, and the angle between the canopy radials and the suspension lines, δ. There are few recommendations regarding the forces in reefing lines:

1. In a medium-to-high subsonic range, a reefing-line-to-maximum-reefed-parachute-force ratio of 0.03–0.04 can be used for the design of a first-stage drogue chute,
2. A reefing-line-to-maximum-reefed-parachute-force ratio of 0.025 minimum can be used for the design of large, main parachutes to be used for the final descent, and
3. A safety factor of 2.5 is recommended for the design of all components of the reefing system.

9.2.5.4 Clustering

Parachutes come in all kinds of sizes, and even a single parachute can accommodate very large payload masses. For instance, for the Soyuz entry capsule, a single ribbon parachute of 35.6 m diameter is used to carry a landing mass of about 3,000 kg. Also for the Booster Recovery System (BRS) of Ariane-5, only a single parachute was used as its main: a single ribbon parachute of 48-m diameter was used to carry the 31-m tall and 3-m wide empty boosters of 37,000 kg. To get an idea about the dimensions of such parachutes, in Fig. 9.31 the main parachute of the BRS is shown during one of the test campaigns. The size becomes even more staggering once one realises that a diameter of 48 m means that the surface area is 1,800 m^2. Compared to the field dimensions of a professional soccer field,[7] it means that four parachutes would be enough to completely cover the soccer field.

[7] The soccer association FIFA determined the range 100–110 m long by 64–73 m wide to be optimal, which means that the field area is between 6,400 and 8,030 m^2, or an average value of 7,215 m^2. The International Ice Hockey Federation (IIHF) specified the dimensions of an ice-hockey rink to be 60 by 30 m, which is the exact surface area of the BRS ribbon parachute.



Fig. 9.31 Testing the main parachute of Ariane-5's Booster Recovery System. Image: Airbus Defence and Space/ESA

However, it is not always a single parachute that carries the payload. An empty Solid Rocket Booster of 91,000 kg is carried by three conical ribbon parachutes of 35.1 m each, to evenly divide the weight. Also the Apollo Command Module descended under three ring-sail parachutes of 26.1 m each, but in this case the mass was "only" 5,900 kg (Apollo-11). In this particular case the parachute system was designed for two parachutes, and the third parachute was added for redundancy reasons, but at the same time for nominal configurations to reduce the impact load. In case of Apollo-15, the third parachute actually failed to open and the capsule sea-landed with only two functional parachutes (Fig. 9.32). The same redundancy principle is foreseen for the future Orion capsule, which will use three ring-sail parachutes of 35.4 m diameter, to land the capsule having a mass of 9,300 kg.

There are several advantages to the use of clustered parachutes:

1. is easier to fabricate, store, maintain, manipulate and ground-recover multiple smaller canopies compared to a single larger one,
2. the drag area can be adjusted by adding or deleting canopies,
3. the stability of the system is increased, even if the cluster can combine individually unstable parachute types, such as solid, flat and circular types,
4. the system redundancy can (easily) be increased by over-design, and
5. smaller canopies have a shorter inflation time and distance.

Of course, there are several disadvantages as well:

1. due to interference and the particular system geometry there will be a slight loss of $C_{D,0}$ ($\approx$5% for a three-canopy cluster),

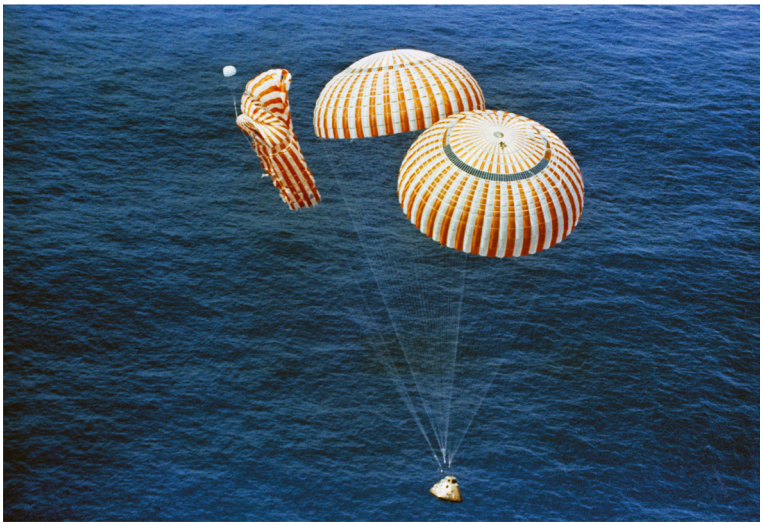


Fig. 9.32 Redundancy in Apollo 15's main parachutes allowed for a safe landing, despite a parachute-line failure. Image: NASA

2. there could be problems with asynchronous inflation, which might require an additional reefing stage,[8]
3. parachute clusters for low rates of descent have problems with inflating all parachutes,
4. variation in disreef times can occur, due to cutter-time inaccuracies, and
5. a clustered system is heavier than a single-canopy system.

The drag loss of clustered parachutes originates from two effects. First, the geometry of the cluster system forces parachutes to fly with a large angle of attack, as the individual parachutes are pushed away from the centre line. Second, there is the mutual interference, and canopy edges have been seen to be compressed a bit. A compensating effect is provided by long suspension lines or suspension-line-riser combinations, which cause a decrease of the individual angle of attack and lead to an increase of the cluster drag. In general, the drag loss in parachute clusters is proportional to the number of parachutes in the cluster. In Fig. 9.33 the drag loss as a function of number of parachutes is shown, with data taken from several sources listed by Knacke (1987). Plotted is the ratio of drag from the clustered parachutes to the undisturbed parachutes. For a cluster of three parachutes, the loss is between 5 and 15%, depending on type and operational conditions. For more parachutes, the loss can be severe, and values up to 30% were observed.

[8] Early tests with unreefed cluster configurations showed a very irregular parachute operation. Reefing the individual parachutes in the cluster assured a good inflation of the canopy into the reefed stage, a reasonably uniform, full inflation, as well as inflation forces.

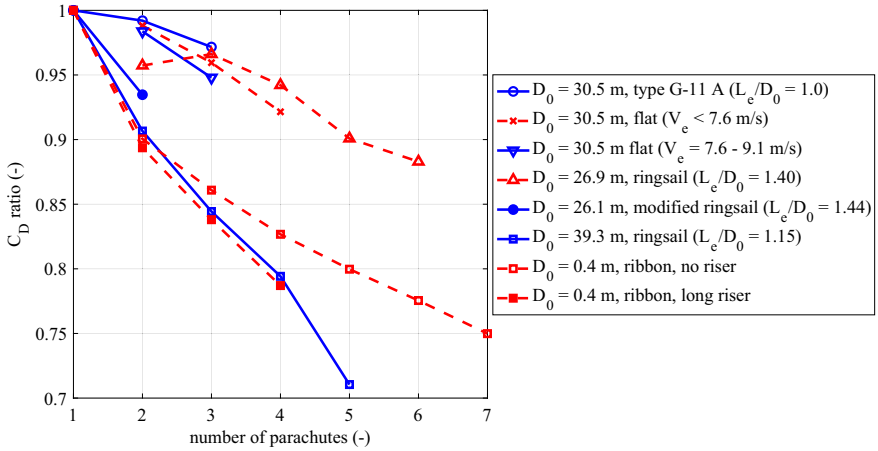


Fig. 9.33 Drag loss in parachute clusters (L_e = effective suspension-line length of a single parachute, V_e = design rate of descent). Data obtained from Knacke (1987)

During the development of a cluster system, several factors must be taken into account such as the configuration of the vehicle, the selection of the best-suited parachute type, the number of parachutes in the cluster, the individual parachute diameter, the selected cluster geometry, and the canopy loading (rate of descent) of the cluster. Canopy loading, $W/(C_D S)_p$, has a distinct effect on uniform, simultaneous canopy inflation and the corresponding opening forces. For detailed design procedures, the reader is referred to the design guides by Ewing et al. (1978) and Knacke (1987), whereas Wolf and Spahr (1975) present a cluster dynamics analysis.

9.2.5.5 Forebody Wake

When a parachute is used in combination with a relatively large forebody such as an Apollo capsule, as compared to a person, for instance, this forebody will produce a wake (highly turbulent air where the effective dynamic pressure is less than in the free stream). This wake influences canopy inflation and operational characteristics and, consequently, the motion and stability of the system as a whole. Depending on the distance between forebody and parachute, the drag loss can amount to 25%, see also Fig. 9.34. It is also clear from this figure that even when the distance seems large enough, there will still be a drag loss of a few percent.

Several tests were conducted for the DGB parachute to determine the drag and performance characteristic in the wake of a 0.10-scale Viking entry vehicle. The parachutes were also tested behind a small faired body to obtain minimum interference parachute performance characteristics. The results showed that in general the DGB parachute has good inflation and stability characteristics in the wake of the entry forebody at Mach from 0.2 to 0.8 and from 1.2 to 2.0. However, considerable loss was experienced in wake of the forebodies around Mach 1.0. Increasing the

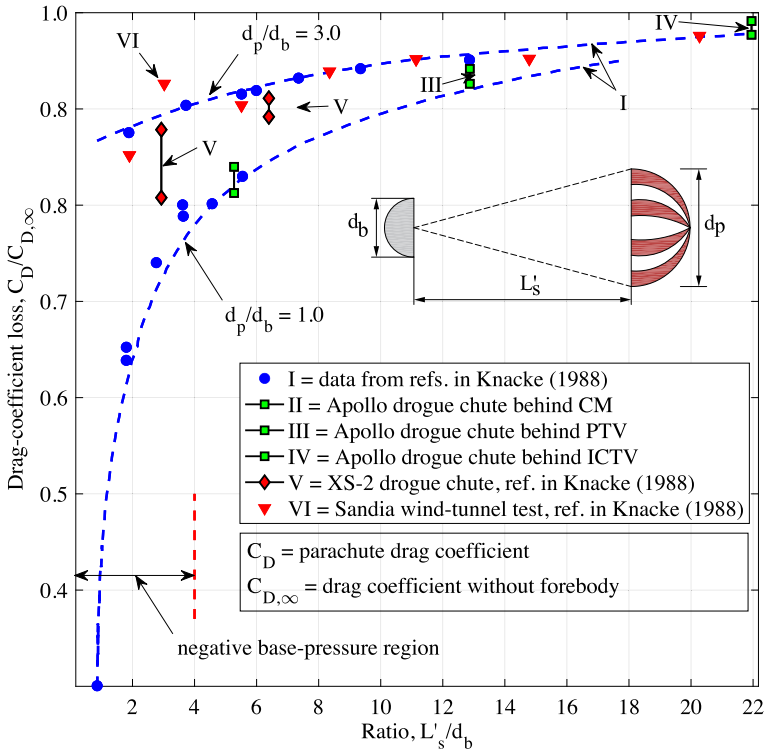


Fig. 9.34 Parachute drag loss due to the wake of the forebody (see for definitions Table 9.1). Data obtained from Knacke (1987)

Table 9.1 Configuration definitions for Fig. 9.34

Configuration			
Designation Vehicle	B/P CM Boiler-plate Command Module	PTV Parachute Test Vehicle	ICTV Instrumented Cylindrical Test Vehicle
Forebody diameter (m)	3.627	1.768	0.914
Drag area $C_D S$ (m^2)	10.3	11.7	12.3
$\frac{(C_D S)_p}{(C_D S)_\infty}$	0.82	0.92	1.0

parachute trailing distance and the suspension-line length increases the parachute drag coefficient for each Mach. Eventually, increasing the Mach number from 0.6 to 1.0 generated a 50% reduction in parachute drag coefficient, as the parachute became under-inflated in the wake of the Viking forebody (Relchenau 1972).

9.2.6 *Parachute Dynamics Simulation*

The complete simulation of parachute deployment, inflation with or without reefing, and descent in a possibly turbulent and windy environment, requires a suite of very advanced tools. To simulate the inflation process and address the forces acting in the lines and on the (flexible) canopy, a combination of computational fluid dynamics (CFD) and finite element analysis (FEM) would be required. Then, when the parachute-payload system is in "steady" descent, the oscillation of the payload under the parachute with flexible suspension lines under the influence of, for instance, wind, would require a multibody simulator with possibly flexible elements. After a brief overview of simulation tools at varying levels of complexity, in the remainder of this section we will provide a simple approach for a conceptual analysis only. The (complex) multibody stability analysis is beyond the scope of this book.

9.2.6.1 Simulation Tools

Over the years many simulation tools have been developed, at varying levels of complexity and for different purposes. To simply address the impact velocity, or to determine the required nominal drag area, an almost trivial one-degree-of-freedom model, simulating the vertical, terminal descent, would suffice. A simple two-dimensional model, consisting of planar motion and the in-plane rotation about the combined centre of mass would suffice to study the descent and *pendulum* motion of the system (see Fig. 9.35). The so-called *scissors* motion of the system cannot be studied with a single-body model; it would require two constrained bodies that can rotate with respect to each other. Hume (1973) developed such a model, where he neglected the aerodynamic forces acting on the re-entry vehicle, because in steady descent when the parachute(s) is (are) fully inflated, they are much smaller than the ones acting on the parachute(s). An extended seven-degrees-of-freedom model was used by Mooij (1992) to analyse the parachute descent of Huygens. This model represents a three-dimensional, rigid parachute-payload system that can address both pendulum and conical oscillations, and includes a swivel in the riser. Such a swivel allows for a spin rotation of the payload with respect to the parachute and allowed for the spin-vane analysis that will be discussed in more detail in Sect. 9.8.

In the past, studies of parachute stability have been performed as far back as the 1960s (Helmut and Lawrence 1965). This study looked at a simplified parachute point-mass system, and a closed-form solution was obtained by considering an aerodynamically neutral load and an aerodynamically stable parachute. In this model the motion of the parachute-load system has been restricted to one plane, which reduced the total number of degrees of freedom to three. A study on the dynamic stability for a three-dimensional parachute with a rigidly connected load included linearisation of the dimensionless equations of motion to estimate the stability characteristics of the parachute-load system (Wolf 1971). The total number of degrees of freedom of this model was reduced to five, since the roll about the parachute's

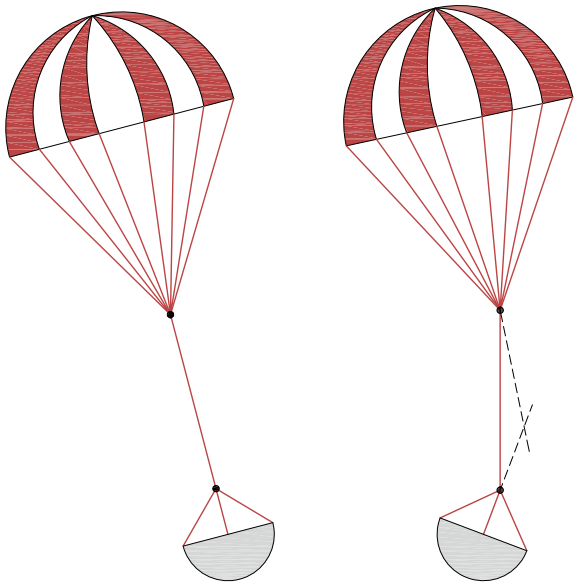


Fig. 9.35 Parachute-system pendulum and scissors motion

rotational axis was neglected. An interesting theoretical analysis of the longitudinal stability of the parachute system contained both the static and dynamic longitudinal analysis (Goodrick 1975). The simplification of the employed model was achieved by assuming a rigid connection between the parachute canopy and the load.

In all of the presented examples the dynamically complex motion of the parachute-load systems has been invariably simplified by the reduction of the total number of degrees of freedom from the maximum of nine to as little as five or even three. As long as the independent oscillatory behaviour of the load has little influence of the dynamic motion of the parachute-load system this assumption is fairly valid, especially if one is primarily interested in the performance characteristics, such as the deceleration efficiency of the parachute.

Finally, another important example of the stability analysis of the parachute system considered the six-degrees-of-freedom motion of both the parachute and the vehicle and, furthermore, included the elastic representation of the riser and suspension lines (Ibrahim and Engdahl 1974). This 15-degrees-of-freedom model has been applied to the study of stability for the Space Shuttle solid rocket booster (SRB) recovery system. The approach taken demonstrated a realistic mathematical description of the parachute descent, whereby the parachute and the load have been modelled as elements of the multibody system constrained by the suspension lines connecting the elements of the descent system. This approach clearly showed that, in general, a complete dynamical response and stability study of the parachute-descent system is closely related to the multibody modelling, since the separate motions of the parachute and the load are important to the overall dynamical behaviour of the

parachute-load systems. Similarly, Ke et al. (1991) have developed a novel algorithm for simulating the general parachute-payload system. Their multibody model allows for modelling parachute clusters, including the associated performance degradation, with elastic suspension lines.

Therefore, a consistent treatment of the parachute system as a multibody has been thought to provide an improvement and generalisation of the already existing models, especially in cases of the more complicated parachute systems consisting of three or even more elements. Examples of the parachute systems containing up to three elements (the parachute, the backshell and the airbags lowered on a bridle) can be found in the MPF and the MER missions. The parachute descent and landing phase of the Mars Pathfinder mission has been analysed with the Automated Dynamic Analysis of Mechanical Systems (ADAMS) multibody simulation program in which both the translational and rotational equations of motion were solved (Braun et al. 1997). Another example is the Mars Exploration Rover mission for which simulations based on the multibody parachute trajectory simulation tool developed in the *Program to Optimize Simulated Trajectories II* (POST II) were executed (Raiszadeh and Queen 2004).

The critical analysis of these Mars missions has revealed a general tendency toward increasingly complex landing systems consisting of multiple elements. Each of these elements (subsystems) has to perform optimally in a specified flight regime. This natural tendency was one of the reasons to look at the lander-system design from the point of view of multibody modelling. Another option provided by the multibody methodology is the possibility to simulate the dynamics of (pseudo-)flexible elements, which can be approximated by a large number of interconnecting subcomponents. For example, the flexible motion of the bridle connecting the backshell and the airbag can be analysed in this manner, if the subcomponents are taken sufficiently small.

There are in principle two fundamentally different ways of modelling multibody systems. The first approach considers each body individually (a rigid implementation would have six degrees of freedom per body), and requires a so-called confluence point to connect two bodies. The (mathematical) function of this confluence point is to enforce the kinematical constraints on the parachute-payload system. This type of modelling has the advantage that an existing 6-degrees-of-freedom parachute model can be applied without any major alterations within the framework of the parachute-load model (Mooij et al. 2003). Thus, existing software for the parachute models including those for added mass can be reused, as a direct consequence of the fact that the parachute and load are treated as independent bodies. The associated disadvantages are that the kinematical constraints are artificially enforced, the reaction forces should be explicitly calculated, and the complexity of an already complex system increases due to the additional degrees of freedom of the confluence point. Moreover, no general framework for the generation of multibody equations of motion can be set up, which means that each new system configuration should be considered separately.

In the second approach the constraint forces are eliminated, which makes this approach numerically efficient (minimum set of state variables) and more accurate

in the sense that the relevant constraints are directly incorporated into the equations of motion. Consequently, the reaction forces are eliminated and the kinematical description of the system is rigorous and correct for all types of motion. Additionally, this method can make use of the well-developed multibody formalisms for generation of the multibody equations of motion for an arbitrary number of bodies and joints, originally intended to describe (very) large interconnected mechanical systems, e.g., in the field of robotics. On the downside, though, practical implementation of the multibody formalisms from fields not related to the aerospace industry can be quite cumbersome.

For us to have a single modelling approach for our stability analysis has as an additional advantage that we can "scale down" the model, i.e., we go from three or two bodies (parachute, possibly a backshell, and a payload) to a single one (rigid-body description of the parachute-payload system), without having to develop new models for each configuration. Also, future extensions to systems with multiple parachutes would become relatively easy. A well-developed and documented method is the one by Wittenburg (2008), which, given the system description, will lead to a "self-formulation" of the equations of motion. This method was successfully applied to the stability analysis of the three-body Mars Pathfinder system, consisting of a disk-gap-band parachute, a backshell, and a landing-bag/rover combination (Mooij 2016).

In Fig. 9.36 both a three-body and two-body system are depicted, requiring to solve for 12 and 9 degrees of freedom, respectively. Note that each two-body connection is constraining three degrees of freedom, as the position of the connection for either body has to remain the same. For example, a three-body system has 18 degrees of freedom, of which the two connections will constrain three each. Of course, the number of constraints may change in case the connection type changes. A swivel, for instance, may allow for a single rotation only (along the spin axis), thus constraining five out of the six degrees of freedom.

9.2.6.2 Equilibrium Descent

With the discussion and provided references above, the reader has a starting point for developing and using a suitable model for stability analysis of a multibody parachute-payload system. In the context of the current discussion, we take a few steps back and present a very simple model to serve conceptual thinking about parachute systems. Let us consider the terminal descent of a parachute-payload system as shown in Fig. 9.37. The system is in vertical descent with velocity, V, and consists of a single parachute and a payload. Both elements experience a gravitational pull and aerodynamic drag, but for the sake of simplicity we consider the system to be a single body and add the forces together.

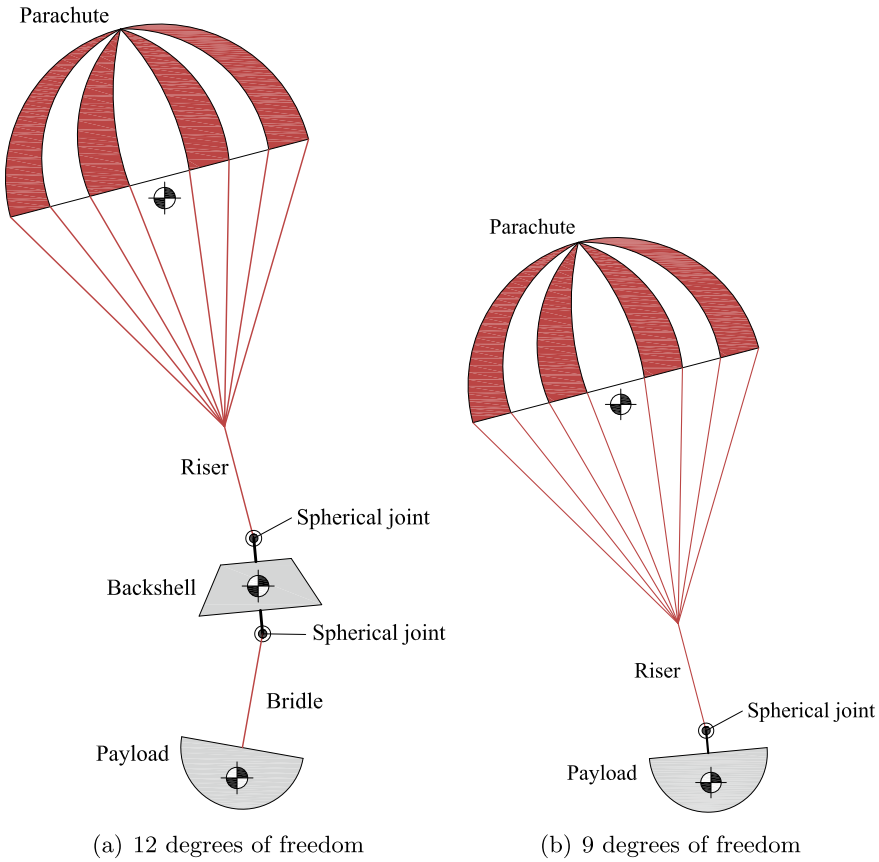


Fig. 9.36 Schematic representation of multibody parachute systems

Assuming vertical equilibrium, i.e., the system is in a steady descent, the dynamic equation of motion is[9]:

$$m_{system}\dot{V} = D_p + D_{ev} - m_{system}g_0 = 0 \tag{9.11}$$

with D_p and D_{ev} being the parachute and entry-vehicle drag, respectively:

$$D_p = C_{D,p}\frac{1}{2}\rho_0 V^2 S_0 \qquad D_{ev} = C_{D,ev}\frac{1}{2}\rho_0 V^2 S_{ev} \tag{9.12}$$

[9] In this model, we ignore the so-called *added mass*, i.e., the air mass that is trapped under the canopy. Later in this section we will discuss the added mass in more detail, and how it affects the translational and rotational motion.

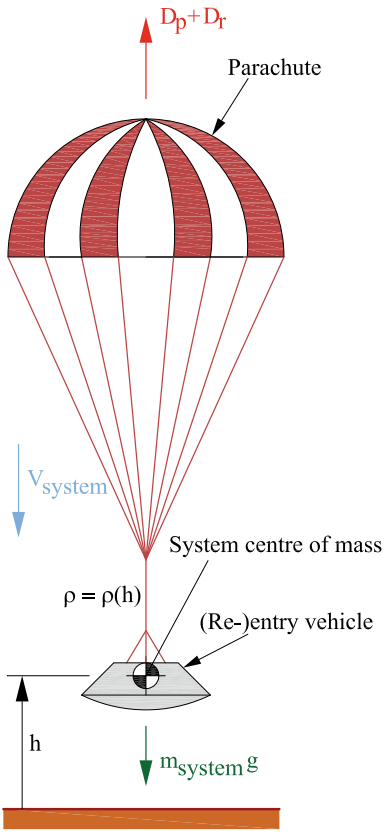


Fig. 9.37 Parachute model for terminal descent

and ρ_0 and g_0 being the (constant) density and gravitational acceleration close to the surface. With the parachute drag area, $C_{D,p}S_0$, being a lot larger than the vehicle drag area, $C_{D,ev}S_{ev}$, for a first assessment it is safe to ignore the latter, $C_{D,p}S_0 >> C_{D,ev}S_{ev}$, so:

$$C_{D,p}\frac{1}{2}\rho_0 V^2 S_0 = m_{system}g_0 \tag{9.13}$$

This equation allows us to determine the drag area of the parachute, when the final velocity, V_f, is known,[10] or, given a known parachute, what the final velocity will be. Let us consider the first case of determining the parachute drag area:

$$C_{D,p}S_0 = 2\frac{m_{system}g_0}{\rho_0 V_f^2} \tag{9.14}$$

[10] This velocity may be the landing velocity, or when retrorockets are used for the final deceleration, the velocity requirement for igniting the engines.

In case we have selected a parachute type, we typically know the nominal drag coefficient, as well as its geometry. As an example, assuming a hemispherical canopy, the surface area is given by $S_0 = \frac{1}{4}\pi d_0^2 = \pi r_0^2$. Equation (9.17) can thus be written as:

$$\frac{1}{4}\pi d_0^2 = \frac{2}{C_{D,p}}\left(\frac{m_{system}g_0}{\rho_0 V_f^2}\right) \quad \Rightarrow \quad d_0 = \sqrt{\frac{8}{\pi C_{D,p}}\left(\frac{m_{system}g_0}{\rho_0 V_f^2}\right)} \tag{9.15}$$

or

$$r_0 = \sqrt{\frac{2}{\pi C_{D,p}}\left(\frac{m_{system}g_0}{\rho_0 V_f^2}\right)} \tag{9.16}$$

In case one wants to refine the solution, the (known) drag area of the entry vehicle could be included:

$$C_{D,p}S_0 = 2\frac{m_{system}g}{\rho V_f^2} - C_{D,ev}S_{ev} \quad \Rightarrow \quad r_0 = \sqrt{\frac{1}{\pi C_{D,p}}\left(2\frac{m_{system}g_0}{\rho_0 V_f^2} - C_{D,ev}S_{ev}\right)} \tag{9.17}$$

Once the nominal diameter of the parachute is known, one could refine the model even more, e.g., by adjusting the system mass, $m_{system} = m_{ev} + m_p$, and iteratively re-calculate d_0.

9.2.6.3 Seven-Degrees-of-Freedom Model

The earlier discussed one-dimensional model will not provide sufficient insight in the dynamics of parachute-payload systems. In principle, for a performance analysis that allows for studying wind-gust responses, a single-body, six-degrees-of-freedom model would be sufficient, although it is noted that for a more detailed stability analysis one should most likely have to use a more-degree-of-freedom multibody system.

In the following the elements of a single-body model will be briefly introduced. A simulator based on these elements was developed for the analysis of the Huygens parachute-descent trajectory (Mooij 1992), of which a summary will be presented in Sect. 9.8.

The configuration of the parachute system is shown in Fig. 9.38. Both the parachute and the vehicle are considered to be rigid bodies, connected by a rigid bar (Fig. 9.38a). This rigid-bar connection prevents the two bodies from rotating with respect to each other (i.e., a scissor-type of motion). However, a (non-ideal) swivel may be implemented in the connection between parachute and vehicle to allow for a differential roll motion between the two bodies, and thus creating a seventh degree of freedom. Non-ideal here means, that the swivel causes a friction moment that counteracts the

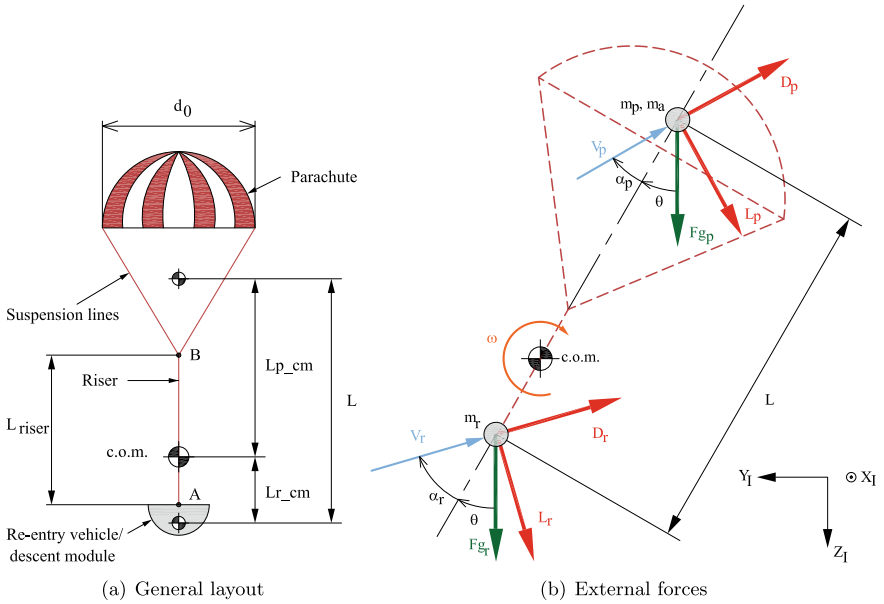


Fig. 9.38 Six-degrees-of-freedom parachute model

spinning motion. The air in and under the parachute is taken into account as added mass. The shape of the canopy should be derived from the applied parachute type and inflation state, but can ideally be approximated by a hemisphere. The model also allows for reefing (i.e., step-wise opening) of the parachute. Multiple parachute systems should be considered that can be deployed/release in sequence, with a minimum of three in case the system includes a pilot, drogue, and main parachute.

In terms of external forces, on both parachute and payload gravity acts as well as aerodynamic forces. Because the parachute/payload system can rotate around the common centre of mass (in Fig. 9.38b indicated by angular velocity ω), the angle of attack for parachute and payload, i.e., α_p and α_r, respectively, will in general be different. That means that the aerodynamic forces not only differ in magnitude (different properties, dynamic pressure, and reference area), but also in orientation. In Fig. 9.38b, L_p and D_p are the parachute lift and drag, and L_r and D_r the corresponding forces for the payload (with index r for *re-entry vehicle*).

One of the limitations of the applied rigid-body model is that only pendulum type of oscillations can be studied, i.e., the rotation around the common centre of mass of the parachute-payload system. The so-called scissor motion—where the parachute and payload rotate independently from, but also opposite with respect to each other—can therefore not be studied, and this is a motion that can occur induced by wind. In Fig. 9.38a the two points of rotation have been indicated by point A and B, i.e., the connection point of the harness (on the payload) to the riser and the suspension lines to the riser, respectively. The two rotations would be clockwise for the parachute and

counter-clockwise for the payload, or *vice versa*, instead of the common rotation ω indicated in Fig. 9.38b.

The drag force acting on the parachute is defined by

$$D_p = C_{D_p} \bar{q}_p S_0 \tag{9.18}$$

with C_{D_p} being the parachute drag coefficient, $\bar{q}_p$ the dynamic pressure at the location of the parachute,[11] and S_0 the nominal reference area of the parachute. The drag coefficient is primarily dependent on parachute shape, Mach number, Reynolds number, and relative canopy elasticity. However, there are also some other aspects to be considered.

In Fig. 9.34 the effect of the forebody's wake on the parachute drag was shown, with drag losses of as much as 25%. For a first analysis, the wake effect is typically taken into account as a reduction coefficient (=scale factor) of the parachute drag area. With a so-called wake factor, ϵ ($0 < \epsilon \leq 1$), the parachute drag is then:

$$D_p = \epsilon C_{D_p} \bar{q}_p S_0 = C_{D_p} \bar{q}_p S_0' \tag{9.19}$$

where $S_0' = \epsilon S_0$ is the effective parachute area. So, by introducing a smaller reference area S_0' (or diameter d_0') while leaving the other parameters unaffected, the wake effect can easily be taken into account.

The suspension-line length influences the canopy shape. When the lines are longer, the radial component of suspension line tension will be smaller, and therefore the projected parachute diameter will increase and thus the total drag area. Also canopy porosity will play an important role. An increase of porosity has a stabilising effect, which is good, but results in a decrease of the drag force at the same time, which is undesirable. The effects of these two phenomena on the performance of the parachute can be compensated for in the same way as for the wake effect. During a sensitivity study, the reduction factors can be varied to analyse their impact on (drag) performance.[12]

The last influence on parachute drag that is considered, is clustering, see also Fig. 9.33. Similarly to dealing with the wake effect, drag loss due to clustering can be taken into account in the form of a so-called cluster coefficient, $C_{cluster}$ ($0 < C_{cluster} \leq 1$). So:

$$D_p = C_{cluster} n_p C_{D_p} \bar{q}_p S_0' \tag{9.20}$$

with n_p being the number of parachutes in the cluster. It is noted that even though this simple approach would work for calculating the drag force, it will not be the right approach to calculate, for instance, the pitch (and yaw) moments. Since parachutes

[11] Due to the combined translation of and the rotation about the global centre of mass, the velocity at either the payload or the parachute will be different, and even though the density at either location might only be marginally different from each other, the resulting dynamic pressure will be different for varying locations of the parachute system.

[12] To study the effect of porosity on stability is a lot more complicated and would require more detailed CFD analysis of the local flow field and its effect on the moment coefficients.

in a cluster will no longer be aligned with the symmetry axis of the system, but have a certain angle, one should take this effect on the orientation of the individual forces into account to calculate the corresponding moment about the centre of mass.

9.2.6.4 Added Mass

When the parachute opens and starts inflating, it gathers air. This air has first to be accelerated to the speed of parachute/payload system. When this speed has been reached, and the system is decelerating, also the air under the parachute will decelerate again. The same applies to the air between the parachute and the re-entry vehicle. The extra mass has been given the designation apparent mass, virtual mass, or added mass (Ewing et al. 1978).

This acceleration and deceleration introduces an apparent force. The resistance of the air disappears under steady flow motion. It is an inertial effect which only takes place during unsteady motion. This may be a deceleration, but also an oscillation: when the parachute system is swinging, the added mass causes some extra damping. It can theoretically be shown that the added mass is also dependent on flow direction (Hume 1973). However, due to difficulties in measuring, the added mass is generally assumed to be independent from this. Since the added mass is dependent on altitude (in the form of atmospheric density), it increases when the altitude decreases.

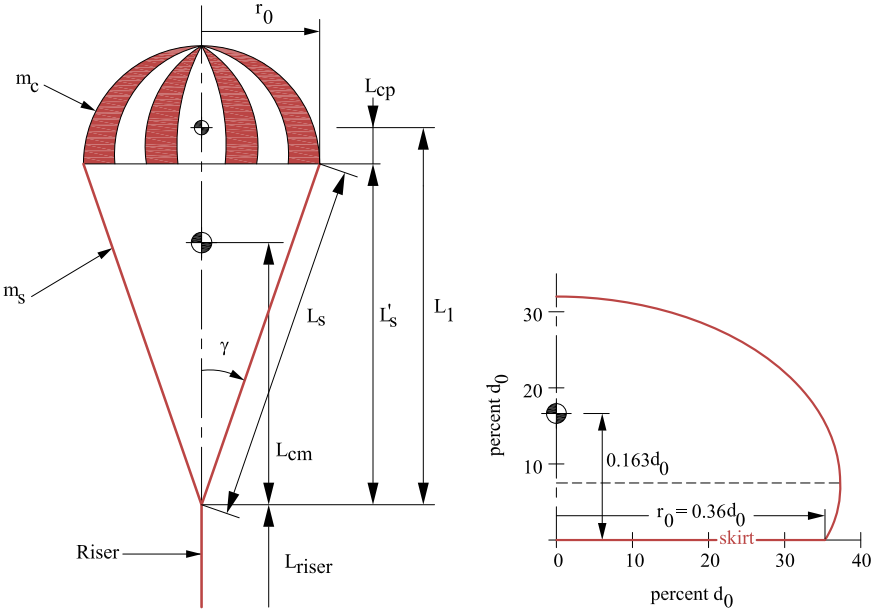
The apparent force due to the added mass, m_a, is an inertial force, and will not be part of the external-force vector. Newton's second law reads in this case:

$$m_{system}\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} + n_p m_a \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \sum_i \mathbf{F}_{\mathbf{ext,i}} \tag{9.21}$$

with $m_{system} = n_p(m_p + m_s) + m_r$ = mass of the system (kg), n_p = number of parachutes, m_p and m_s are the mass of a single parachute and corresponding suspension lines, respectively. It is stressed that the gravitational force does *not* include the added mass, gravity only acts on the "hardware" elements combined into m_{system}. Similarly, for the rotational motion the Euler equations read

$$\mathbf{I}\dot{\boldsymbol{\omega}} = \sum_i \mathbf{M}_{\mathbf{ext,i}} - \boldsymbol{\omega} \times \mathbf{I}\boldsymbol{\omega} \tag{9.22}$$

where the effect of added mass is included in the inertial tensor, **I**. As an example we will now derive the mass properties of a realistic conical-ribbon parachute system and show the simplifications for a hemispherical shape.



(a) Parachute and riser geometry

(b) Fully inflated conical ribbon canopy shape

Fig. 9.39 Parachute geometry

9.2.6.5 Global Mass Properties

The general geometrical layout of the parachute system is depicted in Fig. 9.38a. The system is assumed to be an axisymmetrical body (also called a rotational symmetric body), with the centre of mass (c.o.m.) of the parachute and the payload located on the axis of symmetry. As a result, also the parachute and the payload (from now on called the re-entry vehicle) must be axisymmetrical. The parachute has a fixed-shape canopy with rigid suspension lines. The parachute is thought to have a conical ribbon canopy. The geometry and mass properties of the parachute have been determined by Ibrahim and Engdahl (1974), who provide more details for the interested reader.

The geometry of the parachute and riser is shown in Fig. 9.39a. Typical dimensions of the fully inflated conical ribbon canopy shape are given in Fig. 9.39b. As can be observed, the canopy is over-expanded close to the skirt. In the figures, the following notations have been used:

m_c = mass of the canopy (kg)
m_s = mass of the suspension lines (kg)
r_0 = nominal radius of the inflated mouth of the canopy (m)

L_{cp} = distance between local centre of pressure (c.o.p) and canopy edge (m)
L_1 = distance between c.o.p. and convergence point of suspension lines (m)
L_s = length of suspension lines (m)
L_{riser} = riser-line length (m)
L_{cm} = distance between local c.o.m. and convergence point of suspension lines (m)
L'_s = projected distance of suspension lines (m)
γ = suspension line angle (rad).

Based on Fig. 9.39, the following relationships for a single parachute can be obtained.[13]

$$r_0 = 0.36 d_0 \quad (= 0.5 d_0 \text{ for a hemisphere}) \tag{9.23}$$

$$S_{p,ref} = \frac{\pi}{4}(0.72 d_0)^2 \quad (= \frac{\pi}{4} d_0^2 \text{ for a hemisphere}) \tag{9.24}$$

$$L_{cp} = 0.163 d_0 \quad (= 0.25 d_0 \text{ for a hemisphere}) \tag{9.25}$$

$$\gamma = \arcsin\left(\frac{r_0}{L_s}\right) \tag{9.26}$$

$$L'_s = L_s \cos\gamma = \sqrt{r_0^2 + L_s^2} \tag{9.27}$$

$$L_1 = L_{cp} + L'_s \tag{9.28}$$

d_0, L_s, m_c and m_s are supposed to be known geometry and mass data.

The c.o.m. location for the entire canopy, including the added mass, is given by:

$$L_{cm} = \frac{0.5 L'_s m_s + L_1 (m_c + m_a)}{m_s + m_c + m_a} \tag{9.29}$$

with

m_a = added mass (kg) = $\mathcal{V}_c \rho$
$\mathcal{V}_c$ = volume of canopy (m^3)
ρ = atmospheric density (kg/m^3).

[13] In the on-coming equations, the nominal parachute diameter is used as a reference value. However, when reefing is applied, the actual diameter is time dependent. In that case, the nominal diameter d_0 should be replaced by $d_0(t)$. Consequently, the radius, reference area and volume should be read as $r_0(t)$, $S_{p,ref}(t)$ and $V_0(t)$, respectively.

$\mathcal{V}_c$ is given by

$$\mathcal{V}_c = \frac{2}{3}\pi(0.325d_0)(0.36d_0)^2 \quad \left(= \frac{2}{3}\pi(0.5d_0)^3 \text{ for a hemisphere}\right) \tag{9.30}$$

To get an idea about the added mass, consider a single Apollo ring-sail parachute of 26.1 m. If we approximate the canopy by a hemisphere, the volume is $\frac{2}{3}\pi(0.5d_0)^3 \approx 4{,}655$ m^3. At the deployment altitude of about 3 km, the atmospheric density is $\rho \approx 0.9$ kg/m^3, such that $m_a \approx 4{,}190$ kg. This is quite close to the Apollo-11 mass of 5,900 kg, and a lot more if all three parachutes are considered. Mars Science Laboratory's main parachute had a nominal diameter of 21.35 m. Even though Mars' atmospheric density is a lot lower than Earth's, a surface density of $\rho \approx 0.020$ kg/m^3 will give an added mass of $m_a \approx 51$ kg for a canopy volume of $V_c = 2{,}548$ m^3. Compared to the payload mass this could be ignored, but when one considers the large distance to the global c.o.m. the contribution to the moment of inertia is very large and can simply not be ignored.

The moments of inertia of the parachute about the parachute c.o.m. are given by:

$$I_{xx_p} = \frac{m_s L_s^2}{12}\sin^2\gamma + \frac{2}{3}m_c(0.36d_0)^2 + 0.063\rho r_0^5 \tag{9.31}$$

$$\begin{aligned} I_{yy_p} = I_{zz_p} = \frac{m_s L_s^2}{12}\cos^2\gamma + m_c\left(L_{cm} - 0.5L_s'\right)^2 + \frac{m_c}{3}(0.325^2 + 0.36^2)d_0^2 + \\ + m_c(L_s' - L_{cm})^2 + 0.042\rho r_0^5 + m_a(L_1 - L_{cm}) \end{aligned} \tag{9.32}$$

The contribution of the added mass to the moments of inertia are represented by the last term on the right-hand side of Eq. (9.31) and the last two terms on the right-hand side of Eq. (9.32).

The inertia tensor of the re-entry vehicle (payload) with respect to the local c.o.m. of the re-entry vehicle is given by diag $\left(I_{xx_r}\ I_{yy_r}\ I_{zz_r}\right)$. To compute the total inertia tensor, including all n_p parachutes and the re-entry vehicle, we must know the location of the global c.o.m.. Therefore we need, apart from the locations of the local c.o.m. of both the parachute and the re-entry vehicle, also the location of the attachment point of the riser with the re-entry vehicle, x_{att}. From Fig. 9.38a, it follows that

$$L = L_{cm} + L_{riser} + (x_{att} - x_{cm}) \tag{9.33}$$

$$L_{p,cm} = \frac{m_r}{m_{tot}} \tag{9.34}$$

$$m_{tot} = n_p(m_c + m_s + m_a) + m_r = m_{system} + n_p m_a \tag{9.35}$$

$$L_{r,cm} = L - L_{p,cm} \tag{9.36}$$

The total moments of inertia can now be computed with[14]:

$$I_{xx_{tot}} = I_{xx_r} \tag{9.37}$$

when the parachute is connected to the riser by a swivel, or

$$I_{xx_{tot}} = I_{xx_r} + n_p I_{xx_p} \tag{9.38}$$

when no swivel is assumed, and

$$I_{yy_{tot}} = I_{yy_r} + n_p I_{yy_p} + I_{shift} \tag{9.39}$$

$$I_{zz_{tot}} = I_{zz_r} + n_p I_{zz_p} + I_{shift} \tag{9.40}$$

where, with the use of the parallel axis theorem (Steiner's rule)

$$I_{shift} = n_p(m_c + m_s + m_a)L_{p,cm}^2 + m_r L_{r,cm}^2. \tag{9.41}$$

9.2.6.6 Swivel Moment

When the payload is connected to the parachute by means of a riser including a swivel, a friction moment, $\mathbf{M_{sw}}$, is introduced, since the swivel is not an ideal bearing. Assuming that the riser is aligned with the (body-frame) X_B-axis, $\mathbf{M_{sw}}$ can be represented by

$$\mathbf{M_{sw}} = \begin{pmatrix} M_{sw_x} & 0 & 0 \end{pmatrix}^{\mathrm{T}} \qquad M_{sw_x} = -\text{sign}(p) C_f F_{axial} \tag{9.42}$$

where p is the spin rate (rad/s), C_f is the friction coefficient (m), and F_{axial} is the axial force in the riser (N). According to the above expression, the swivel moment is proportional with (but always opposite to) the spin rate, and is larger when the tension force is the riser is larger. It also implies that when $p = 0$, also $M_{sw_x} = 0$ (no generating moment). It is further assumed that forces in other directions than X do not influence the operation of the swivel.

For the axial (or tension) force, we can derive the following equation. Let us consider the situation of Fig. 9.40. On both the parachute and the re-entry vehicle an

[14] As an approximation, it is assumed that in case of multiple parachutes the geometry of each parachute is "overlapping", such that the mass properties for a single parachute are simply multiplied by the number of parachutes, n_p. This is obviously not correct, and for more accurate simulations the geometry and orientation of each individual parachute should be considered.

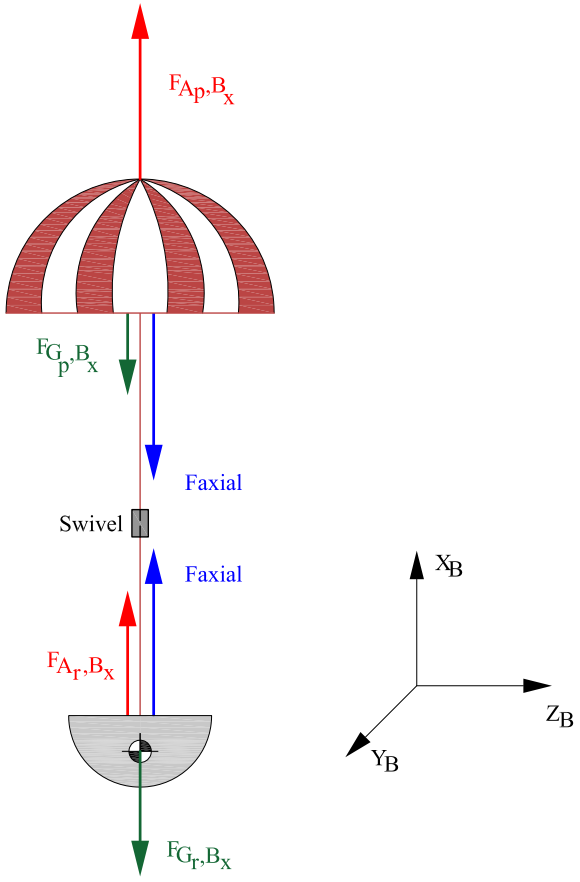


Fig. 9.40 The axial force in the riser due to aerodynamic and gravitational forces acting on parachute and re-entry vehicle

aerodynamic and gravitational force are acting. Looking only at forces in X-direction, we apply Newton's Second Law on the parachute, hereby obtaining

$$n_p(m_p + m_s + m_a)\left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_p = -F_{axial} + F_{A_p,B_x} - F_{G_p,B_x} \tag{9.43}$$

or

$$\left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_p = -\frac{F_{axial}}{n_p(m_p + m_s + m_a)} + \frac{F_{A_p,B_x} - F_{G_p,B_x}}{n_p(m_p + m_s + m_a)} \tag{9.44}$$

Similarly, while looking from the vehicle side

$$m_r\left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_r = F_{axial} + F_{A_r,B_x} - F_{G_r,B_x} \quad \Rightarrow \quad \left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_r = \frac{F_{axial}}{m_r} + \frac{F_{A_r,B_x} - F_{G_r,B_x}}{m_r} \tag{9.45}$$

In the above equations, $\frac{du}{dt}$ is the axial acceleration in the body frame, the indices A and G denote forces of aerodynamic and gravitational origin, and the subscripts p and r stand for *parachute* and *re-entry vehicle*, respectively.

Since the parachute and re-entry vehicle are assumed to be connected by a rigid bar, the acceleration for both elements must be identical:

$$\left.\frac{du}{dt}\right|_p = \left.\frac{du}{dt}\right|_r \tag{9.46}$$

Equating Eqs. (9.44) and (9.45) yields

$$\frac{F_{axial}}{m_r} + \frac{F_{axial}}{n_p(m_p + m_s + m_a)} = \frac{F_{A_p,B_x} - F_{G_p,B_x}}{n_p(m_p + m_s + m_a)} - \frac{F_{A_r,B_x} - F_{G_r,B_x}}{m_r} \tag{9.47}$$

or

$$F_{axial} = \frac{n_p(m_p + m_s + m_a)m_r}{n_p(m_p + m_s + m_a) + m_r}\left[\frac{F_{A_p,B_x} - F_{G_p,B_x}}{n_p(m_p + m_s + m_a)} - \frac{F_{A_r,B_x} - F_{G_r,B_x}}{m_r}\right] \tag{9.48}$$

Simplifying this equation leads to

$$F_{axial} = \frac{m_r(F_{A_p,B_x} - F_{G_p,B_x}) - n_p(m_p + m_s + m_a)(F_{A_r,B_x} - F_{G_r,B_x})}{n_p(m_p + m_s + m_a) + m_r}. \tag{9.49}$$

9.2.6.7 Reefing

To reduce the resulting opening loads during parachute deployment, parachute reefing can be applied. The inflation process can be divided into a number of stages, or reefing intervals. One of these intervals, the ith one, is shown in Fig. 9.41. In this figure, t is the independent variable, representing the time from parachute deployment. The values t_{i-1} and t_i are the boundaries of the ith reefing interval. For each of these boundaries, the percentage of deployment (x_{i-1} and x_i, respectively) is a known parameter. n_i is the exponent of a function, which describes the rate of deployment between t_{i-1} and t_i.

To compute the percentage of deployment x at time t, we can use the following formulae:

$$x = x_{i-1} + \Delta x \qquad i > 1 \tag{9.50}$$

with

$$\Delta x = \left(\frac{t - t_i}{t_i - t_{i-1}}\right)^{n_i}(x_i - x_{i-1}) \quad x_0 = 0 \quad t_0 = 0 \tag{9.51}$$

where t_0 represents a counter time that starts at parachute deployment/inflation.

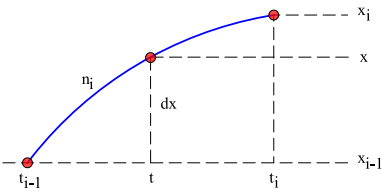


Fig. 9.41 Example of a reefing interval, with boundaries t_{i-1} and t_i. The parachute will deploy according to a function with exponent n_i, during which, for instance, the parachute area will increase from x_{i-1} to x_i

Table 9.2 Reefing-interval parameters for Fig. 9.29

i (–)	$t_i - t_{i-1}$ (s)	x_i (%)	n_i (–)
1	3	14	1
2	7	14	0
3	2	40	1
4	7	40	0
5	3	100	1

One could consider applying the above reefing concept to either the drag area, or to the drag coefficient and reference area separately. The latter method would provide more flexibility, but one should also have more information about the behaviour of these parameters. We will give an example of the latter method:

$$(C_{D_p})(t) = \frac{x_{C_D}}{100}(C_{D_p})_{fully\ inflated} \qquad (S_0)(t) = \frac{x_S}{100}(S_0)_{fully\ inflated} \tag{9.52}$$

Revisiting Fig. 9.29, the five reefing intervals, with the sixth one indicating the fully inflated condition, can be modelled with the parameters as listed in Table 9.2. The first column provides the number of the reefing interval; in the second column the length of the reefing interval is given, while the third one gives the percentage of deployment at the end of this interval. The last column, finally, shows us the exponent of the deployment function.

Note that reefing will not only affect the drag area, but also the added mass, as the volume under the canopy becomes also time dependent. For instance, for a hemispherical shape the parachute radius is obtained from the reference area with

$$S_0(t) = \frac{x_S(t)}{100}\pi r_0^2(t) \quad \Rightarrow \quad r_0(t) = \sqrt{\frac{x_S(t)}{100}}(r_0)_{fully\ inflated}. \tag{9.53}$$

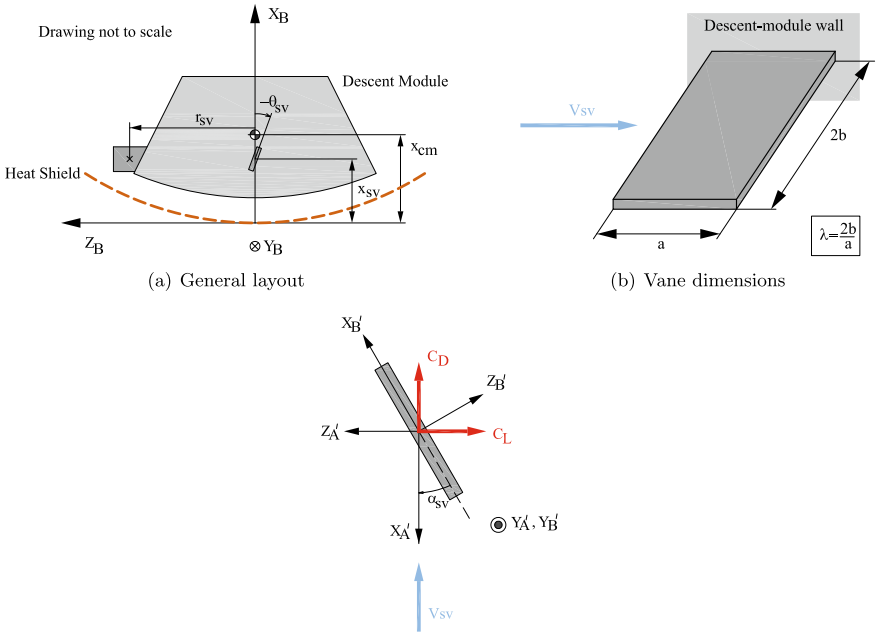


Fig. 9.42 Spin-vane model

9.2.6.8 Spin Vanes

When a planetary probe has to scan the horizon, which was the case for Pioneer Venus (Fig. 7.41a), the Galileo entry probe (Fig. 1.51b), and the Titan descent module Huygens (Fig. 7.41b), it may be equipped with so-called spin vanes to induce a relative spin rate of the vehicle with respect to the parachute. Spin vanes are small airfoils that are mounted on the vehicle wall, and act as a lifting surface while inclined to the velocity. Because there is an offset between the pressure point on the vane and the centre of mass, a roll moment will be induced. Of course, free rotation can only occur when a swivel is included in the riser connecting to the suspension lines to avoid entangling these lines.

To study the spin motion at a conceptual level, a simplified spin-vane model can be implemented. Each vane is considered to be a rectangular flat plate (Fig. 9.42). The lift and drag contribution for a single vane can be computed, based on flat-plate theory. These forces result in a roll moment that drives the vehicle. The contribution of all vanes is taken into account by multiplying the computed roll moment by the number of vanes. This implies, that the flow direction for each vane is assumed to be identical, which limits the application to symmetric flow conditions without transverse winds.

The spin-vane configuration is determined by the following parameters (Fig. 9.42):

n_{sv} = number of spin vanes
$\lambda = 2b/a$ = aspect ratio of spin vane
a, b = chord length and vane span (m)
θ_{sv} = (constant) offset angle (deg)
r_{sv} = mean local radius descent module (m)
x_{sv} = x-coordinate spin vane with respect to the nose of the heat shield (m)
η_{sv} = efficiency (scale) factor.

Limitations of the developed spin-vane model are the following:

1. Flow disturbances due to interactions between the descent-module wall and the vanes are not considered,
2. Partial visibility of a vane due to shielding of the descent module and a deviation in flow direction is not taken into account,
3. The flow direction for each of the vanes is not the same, unless we consider a vertical descent with no wind. For realistic computations, wind should be taken into account and therefore, the forces acting on each vane should be computed separately.

Due to the restrictions mentioned above, the model will only give an indication of the spin-rate profile. Detailed flow simulations should verify the assumptions and validate the results in the detailed design phase. An example spin-rate analysis is covered in Sect. 9.8.

9.3 Powered Descent

9.3.1 Introduction

Commonly, entry missions into the Earth's atmosphere are performed by unpowered vehicles, either capsules or winged vehicles. However, the Soyuz capsules ignite a landing engine a few seconds before touch down, to take out most of the velocity to lower the impact load on vehicle and crew. With the advent of reusable launcher stages, the use of propulsion systems for landing have become more prominent. In planetary missions, the use of propulsion systems has been more of a necessity, i.e., soft landing on the Moon is not possible without, and also for a soft landing on Mars engines working in tandem with parachute systems are required.

In the total chain of operations of (planetary) entry and descent missions, the propulsion system is used for (1) orbit insertion, although in many cases the engine is part of the orbiter rather than the lander, (2) deorbitation, after the descent module has separated from the orbiter, (3) attitude stability and control, by, for instance, the reaction-control thrusters, (4) deployment of aerodynamic decelerators by means of a mortars, tractor rockets, etc., (5) powered deceleration, descent and landing on an atmosphereless body, and (6) terminal braking to reach an acceptable landing velocity when parachutes alone cannot provide that. Not diminishing the importance of any of these operations, in the spirit of this chapter the focus in the current section will be on powered deceleration and descent, as well as terminal braking.

Historically speaking, the success rate of robotic planetary landers has been relatively low, which is partially due to the absence of sensors capable of identifying hazards, and a GNC system that could subsequently evade those hazards. To improve upon this, investigations for equipping future vehicles with more advanced sensors and navigation solutions are carried out by several agencies, for instance those part of NASA's *autonomous landing and hazard avoidance technology* (ALHAT) (Epp and Smith 2007b). To support these developments, also more advanced guidance algorithms are required, as these are particularly affected by new constraints on the trajectory, such as glide-slope constraints to keep the landing site in the *field of view* (FOV) of a navigation camera or a hazard-detection lidar.

A typical mission sequence to land on a(n atmosphere-less) body, let us assume this to be the Moon, may either begin in a low Lunar orbit or, as was the case for the Surveyor missions, be the crossing of a certain altitude when in a direct-impact trajectory. In the former case, the descent begins with the *descent orbit insertion* (DOI), which puts the spacecraft into a Hohmann-transfer orbit with a lowered periapsis above the landing site. After *powered descent initiation* (PDI), a few hundred km away from the landing site, the bulk of the orbital velocity is removed during this main braking phase. The so-called *high gate* (HG) is reached as soon as the landing site comes into the FOV of either the astronauts or the (hazard mapping) instruments. Once the landing site is actively targeted, the approach phase begins at the *approach gate* (AG), and it is during this phase that hazard detection and avoidance is executed, provided such a system is present, of course.[15] Finally, when visual contact with the landing site is lost, the *low gate* (LG) is reached. At some small distance above the landing site, all horizontal velocity should be nulled, and the approach phase ends at the terminal gate (TG). From there downwards, the vehicle descends at constant speed to the surface, up until touchdown.

The problem considered in this section is the trajectory segment from the approach to terminal gate, given initial and terminal conditions, and is thus a two-point boundary-value problem (TPBVP). This problem is solved by generating an achievable acceleration profile that connects these two points, which can be done, for

[15] In Sect. 9.5, advanced descent and landing systems, including hazard detection and avoidance, will be discussed in more detail.

instance, by minimising a pre-defined cost criterion (see also Sect. 7.5.4). The resulting acceleration profile is then executed by other elements of the GNC system (for instance, the combination of navigation and attitude control).

The mathematical formulation of the problem under consideration is as follows. Given a set of initial states at time $t = t_0$,

$$\mathbf{r}(t_0) = \mathbf{r_0} \quad \mathbf{v}(t_0) = \mathbf{v_0} \tag{9.54}$$

where $\mathbf{r}(t)$ is the position at time t and $\mathbf{v}(t)$ the velocity, find an acceleration profile and thus a trajectory

$$\mathbf{a}(t) = \frac{\mathbf{T}(t)}{m(t)}, \quad \forall \quad t \in [t_0, t_f] \tag{9.55}$$

where $\mathbf{a}(t)$ is the acceleration, $\mathbf{T}(t)$ is the thrust, $m(t)$ is the instantaneous spacecraft mass, and t_f is the final time, such that the spacecraft reaches a final state

$$\mathbf{r}(t_f) = \mathbf{r_f} \quad \mathbf{v}(t_f) = \mathbf{v_f} \tag{9.56}$$

with inequality and equality constraints $\mathbf{g}$ and $\mathbf{h}$

$$\mathbf{g} \leq \mathbf{0} \quad \mathbf{h} = \mathbf{0} \tag{9.57}$$

In case the above formulation would be given as an optimal-control problem, then, for instance, one could consider the objective to minimise the fuel mass (or equivalently the thrust force), i.e.,

$$\text{minimise } \mathcal{J} = \int_{t_0}^{t_f} \sqrt{\mathbf{T}(t)^{\mathrm{T}}\mathbf{T}(t)}\mathrm{d}t \tag{9.58}$$

where the equations of motion are given as path constraints.

Whether the applied guidance system is optimal or not, if we want to group the available algorithms together, typically three main categories can be identified: (1) open-loop guidance, (2) explicit algorithms based on analytical closed-form guidance expressions, and (3) algorithms that have numerical optimisation techniques as their basis. Gerth and Mooij (2014) provide an overview of many of the existing techniques. In this section, we will not try to copy this overview, but categorise the algorithms with some references, and focus on two of the simpler algorithms that will provide the reader with more physical insight.

Open-loop guidance methods are those that do not depend on state or output feedback, but can be treated as an initial-value problem: starting from an initial state, the provided acceleration profile will lead to a terminal state. The acceleration profile can follow from an (off-line) optimisation process, taking path and steering constraints into account, and during the mission a tracking system can follow the reference trajectory. Responding to perturbations and/or navigation errors, or retargeting in case of an unsafe landing situation, is obviously not possible. The second algorithm

that fits in this first category is the so-called gravity turn, which was the first active guidance law used for soft-landing of spacecraft by NASA in the Surveyor program (Cheng 1964). In its original form, the gravity turn targets a terminal altitude and velocity, and cannot precisely target a landing site. Since this guidance law is very illustrative for the level of discussion in this text, it will be discussed in Sect. 9.3.2.

The second category of explicit algorithms generate acceleration profiles based on the vehicle's state. They are computationally efficient, because they provide analytically-derived closed-form expressions. However, deriving the expressions can be mathematically cumbersome, and also the required assumptions and difficulties to include constraints, can render explicit methods sub-optimal in terms of propellant consumption and difficult to grasp. Yet, they are promising candidates for on-board precision guidance. Several techniques exist, the first one being *E-guidance* (Cherry 1964a), which forms the basis of several guidance algorithms used for planetary landings. It is based on an analytical solution of a two-point boundary-value problem, given the initial and terminal state. The *Apollo guidance* is an adapted version of *E*-guidance (Klumpp 1971) and has successfully landed the six Apollo missions on the Moon. An adapted version was also selected for Mars Science Laboratory (Wong et al. 2006). Whereas in the former two a linear acceleration profile is prescribed, the *modified Apollo techniques* are based on quadratic profiles, yielding the so-called quartic guidance algorithm, see, for instance, the paper by Lafontaine et al. (2004). D'Souza (1997), finally, derived an algorithm using optimal control theory, which is explicitly optimal. This *explicit mode with weighting on final time* is based on a flat, non-rotating and atmosphereless body, with a quadratic performance index on thrust acceleration and, as suggested by its name, a penalty term on final time that can be used to adjust the trajectory. Out of these four, because of its elegance and simplicity, the original *E*-guidance algorithm will be discussed in Sect. 9.3.3.

The third and last category is formed by numerical methods. Numerical methods often solve similar problem formulations as explicit methods, but do not rely so much on the same simplifications, and also constraints are more easily enforced. This kind of guidance algorithm is much more robust, and also more propellant-efficient. The drawback, though, will be a (significant) increase of computational cost. Fortunately, recent advances in optimisation theory and more powerful hardware will make on-board application of such algorithms feasible. Although computationally expensive, classical numerical optimisation techniques can be applied to solve the guidance problem. Steinfeldt et al. (2008) applied *gradient-based numerical optimisation* to the guidance problem, and using optimal-control theory minimised

$$\mathcal{J} = \int_{t_0}^{t_f} \sqrt{\mathbf{a}(t)^{\mathrm{T}}\mathbf{a}(t)}\mathrm{d}t \tag{9.59}$$

without a penalty of the terminal state. Even though the results showed promise, the high CPU load could possibly hinder on-board application.

So-called *neural guidance* is an innovative approach is based on training an artificial neural network (Gelly and Vernis 2009). A heuristic approach to train neurons with a genetic algorithm for 600 generations, using the complete non-linear equations of motion and incorporating boundary and path constraints. By penalising infeasible solutions, a "quasi-constrained" guidance algorithm is obtained, which can be used for systems that are not easily modelled analytically. Disadvantage of this approach is the CPU-heavy training of the algorithm, and also convergence is not guaranteed.

The most promising algorithm to mention is the one based on *convex optimisation* (Boyd and Vandenberghe 2004), of which a guidance implementation has been published some years back by Açikmeşe and Ploen (2007). If a problem turns out to be convex, or can be rewritten as a convex problem, the optimisation becomes easier and can be solved efficiently. Typically, the guidance problem is written down as a second-order cone problem, which makes it very computationally efficient for a numerical algorithm (Loboet al. 1998).

Over the years, many improvements and extensions have been developed, and describing the underlying mathematics is extensive, and not suitable for the current text. We will restrict ourselves to some key aspects:

1. Convex optimisation theory proves that a well-posed convex problem, such as the powered-descent problem, is *guaranteed* to converge,
2. Once the optimum is found, it can be proven that this is the *global* optimum,
3. There are a number of different, efficient solvers available for convex problems. For instance, Açikmeşe and Ploen (2007) use an interior-point method, which can be given pre-defined deterministic stopping criteria and a prescribed level of accuracy,
4. The performance index is defined to be $\mathcal{J} = \int ||\mathbf{T}||\mathrm{d}t$, and is thus propellant-optimal, contrary to, for instance, E-guidance, and,
5. Constraints and penalties can be imposed.

To conclude, De Ridder and Mooij (2017) analysed the powered descent of the ESA Lunar Lander, starting from PDI and the subsequent main braking phase (MBP), and ending at the TG.[16] The convex guidance system takes input from the navigation system, consisting of inertial sensors, a star tracker, and a simplified visual system for terrain-relative navigation, fused into state estimates by an extended Kalman filter. A basic quaternion-feedback proportional-derivative controller was implemented for attitude control.

To show some of the results, Fig. 9.43 indicates the targeting performance of 1,000 simulated flights, using noisy sensor measurements. As a result of the lower navigational performance during MBP, the figure shows a much larger standard deviation from the target at AG than for TG. This motivates re-optimisation of the

[16] The ESA Lunar Lander, with a dry weight of 900 kg, was designed to have throttleable engines, ranging from 416.5 N to 3,820 N with a specific impulse of 308 s. The PDI is defined at an altitude of 10 km, some 630 km away from the landing site. Approach gate is at 1.6 km altitude, and at 1.3 km downrange, with a targeted horizontal velocity of 67 m/s and −60 m/s vertical velocity. Final conditions are 15 m altitude, with a vertical velocity of −1.5 m/s. The entire simulated flight from PDI to TG lasts about 712 s.

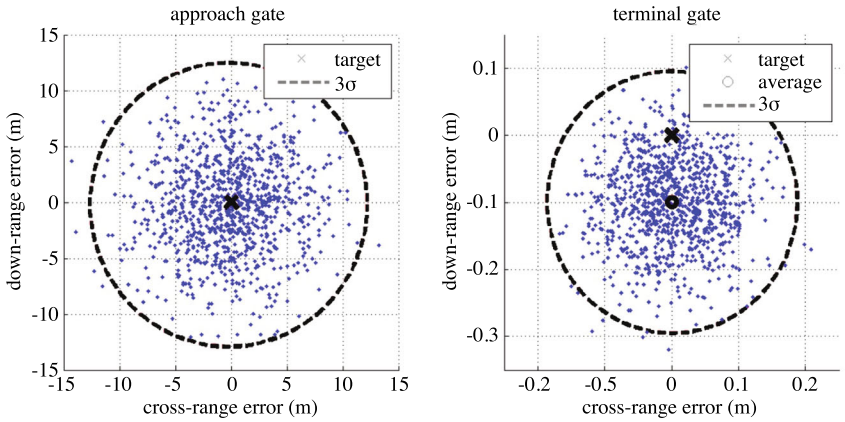


Fig. 9.43 Landing performance for 1,000 flights with navigation errors (De Ridder and Mooij 2017)

trajectory at AG and because more accurate navigation is available during AP, the spacecraft is able to reach TG within decimetres, even though it can be off by 10 m at AG. Without going into further detail, the integrated performance and robustness of the convex guidance system, including glide-slope and attitude constraints, was tested and it was found that convex guidance can practically be used for a pin-point Lunar landing. More details are available in the quoted reference.

To structure the remainder of this section, three more sub-sections are included. In the upcoming two sub-sections, the two already mentioned relatively simple guidance approaches for powered deceleration, descent and landing will be presented, i.e., Surveyor's gravity turn and Apollo's E-guidance. With a summary of terminal braking in Sect. 9.3.4, the discussion of powered descent will be concluded.

9.3.2 Gravity Turn

One of the main goals of the Surveyor program was to develop the technology—and to validate it!—for an automated soft-landing system for Lunar missions. The mission profile was on one hand rather straightforward, but, on the other hand, would also not leave much room for corrections in case something went wrong. Namely, after launch, the spacecraft was placed into a trajectory that would lead directly to a predetermined landing site within the ΔV constraints, while guaranteeing communication with the prime tracking station near Goldstone, California (Thurman 2004). During cruise, the spacecraft was three-axis stabilised using the Sun and the star Canopus as a reference, and a cold-gas reaction-control system. Only a single midcourse correction was planned to cope with launch errors. When reaching the vicinity of the Moon,

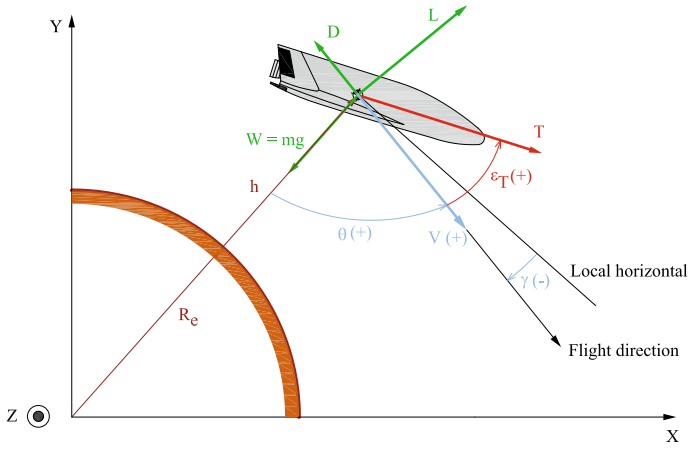


Fig. 9.44 Two-dimensional planar, thrusted motion

the spacecraft did not go into orbit, but immediately descended and had to reduce its velocity to zero to perform a soft landing.

The landing sequence was performed in two stages. This choice was based on the technological capabilities of the early 1960s. To land the spacecraft safely, a throttleable engine was required, which meant a liquid-propulsion system. However, at the time the available thrust level of thrusters with liquid fuel was only one-third of that achievable with solid fuels. Therefore, the major part of the deceleration was performed using the main retro engines with solid fuel, with the disadvantage that they could not be controlled and would induce some position and/or velocity error. At the end of main descent phase, a vernier engine with liquid fuel was used to perform the final descent. The design of the vernier engine obviously had to take the induced error from the previous phase into account.

To get some insight in the mechanics of the problem, we will derive a simplified version of the actual guidance system. The simplifications include a flat-Moon approach, a vehicle orientation close to vertical, and a constant spacecraft mass. It seems justified, though, as the final descent with the gravity turn does take place after the main descent, so at least the former two assumptions would not be too "far off". But for completeness, at the end of the section we will remove the introduced restrictions and show the result for a spherical Moon.

To link the starting equations with the planar equations derived in Chap. 5, we start with a full set of equations and simplify these step by step. Consider the two-dimensional situation depicted in Fig. 9.44. A spacecraft is flying over a spherical body that has an atmosphere. Apart from the external forces lift L, drag D, and weight $W = mg$, a thrust force, T, is exerted on the vehicle. With a thrust elevation ε_T, the equations of motion for unpowered flight, Eqs. (5.8)–(5.10), change into:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma + T\cos\varepsilon_T \tag{9.60}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L + T\sin\varepsilon_T - mg\cos\gamma\left(1 - \frac{V^2}{gR}\right) \tag{9.61}$$

$$\frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma \tag{9.62}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = V\cos\gamma \tag{9.63}$$

and have been extended with the equation of motion for the horizontal range, x. Finally, since the vehicle is now mass-varying, the standard equation for mass flow due to the use of a propulsion system needs to be added to complete the system of equations:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{T}{I_{sp}g_E} \tag{9.64}$$

Here, I_{sp} is the engine's specific impulse, scaled for the gravitational environment of Earth, and g_E is thus the nominal gravitational acceleration of Earth.

Atmosphereless flight obviously implies $D = L = 0$, and the assumption of a flat planet leads to $\frac{V^2}{gR} \to 0$, as the equatorial radius is effectively infinite. The final change that we make is that, because touchdown is vertically downwards, it makes more sense to use the flight-path angle's complementary angle, θ, i.e., the deviation from the local vertical. So, with $\theta = \pi/2 + \gamma$ (as θ is defined positive, whereas γ is negative), Eqs. (9.60)–(9.63) change into:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = mg\cos\theta + T\cos\varepsilon_T \tag{9.65}$$

$$mV\frac{\mathrm{d}\theta}{\mathrm{d}t} = mg\sin\theta + T\sin\varepsilon_T \tag{9.66}$$

$$\frac{\mathrm{d}h}{\mathrm{d}t} = -V\cos\theta \tag{9.67}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = V \sin\theta \tag{9.68}$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{T}{I_{sp} g_E} \tag{9.69}$$

where Eq. (9.69) has remained unchanged.

The so-called gravity-turn guidance was originally developed for the Surveyor missions, i.e., for an atmosphereless environment. Basic operating principle is that this guidance law aligns the thrust vector opposite to the velocity direction such that gravity "pulls" the flight-path into a vertical direction, to align the vehicle for its terminal, vertical descent. If the thrust level is chosen wisely, the gravity turn can in theory result in a vertical landing at zero speed. However, this would indeed only work in theory, as in practice the accumulated navigation and guidance errors are not part of a closed-loop feedback, which will prevent landing at or near the selected landing location.

To derive the guidance law, consider the situation depicted in Fig. 9.45. A powered vehicle is descending towards the (flat) surface of the target body with velocity V, which is assumed to be deviated from the local vertical by an angle θ. This deviation is assumed to be small. Further, the thrust acceleration $a_T = \frac{T}{m}$ is considered to be

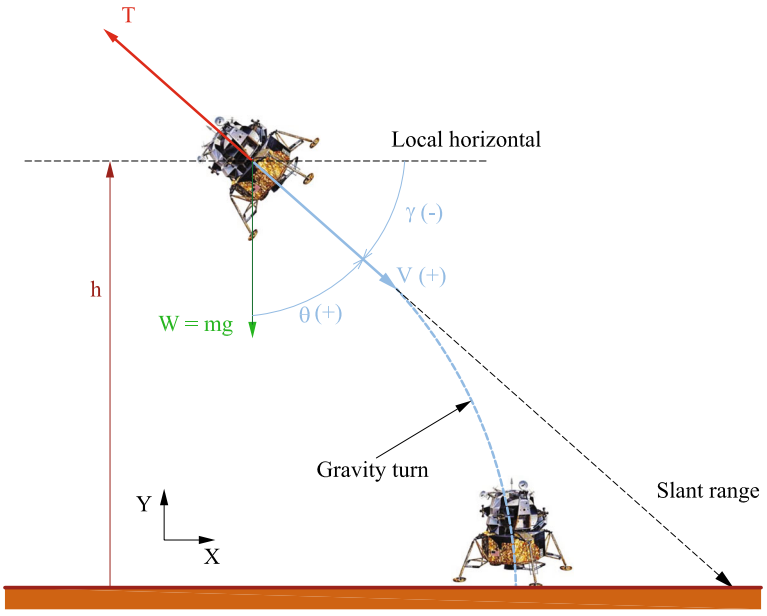


Fig. 9.45 Motion during gravity turn

constant, such that the effect of a variation of vehicle mass due to the fuel consumption is countered by a continuous throttling of the engines.[17]

Let us begin with setting up the two dynamic equations in terms of acceleration and angular velocity[18]:

$$\dot{V} = -a_T + g\cos\theta \tag{9.70}$$

$$\dot{\theta} = -\frac{g}{V}\sin\theta \tag{9.71}$$

Splitting the derivative, and substituting Eq. (9.71) yields

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{\mathrm{d}V}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\frac{\mathrm{d}V}{\mathrm{d}\theta}\frac{g}{V}\sin\theta \tag{9.72}$$

such that Eq. (9.70) can be written as:

$$\frac{\mathrm{d}V}{\mathrm{d}\theta}\frac{g}{V}\sin\theta = a_T - g\cos\theta \tag{9.73}$$

or

$$\frac{\mathrm{d}V}{V} = \left(\frac{a_T}{g\sin\theta} - \frac{1}{\tan\theta}\right)\mathrm{d}\theta \tag{9.74}$$

From the kinematic equation,

$$\dot{h} = -V\cos\theta \tag{9.75}$$

we obtain

$$\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\frac{\mathrm{d}h}{\mathrm{d}\theta}\frac{g}{V}\sin\theta = -V\cos\theta \tag{9.76}$$

or, with Eq. (9.71)

$$\frac{\mathrm{d}h}{\mathrm{d}\theta}\frac{g}{V}\sin\theta = V\cos\theta \quad \Rightarrow \quad \mathrm{d}\theta = \frac{g}{V^2}\tan\theta\mathrm{d}h \tag{9.77}$$

Substituting the above expression for $\mathrm{d}\theta$ into Eq. (9.74) results in

$$\frac{\mathrm{d}V}{V} = \left(\frac{a_T}{g\sin\theta} - \frac{1}{\tan\theta}\right)\frac{g}{V^2}\tan\theta\mathrm{d}h = \frac{1}{V^2}\left(\frac{a_T}{\cos\theta} - g\right)\mathrm{d}h \tag{9.78}$$

[17] In the Intermezzo at the end of this section, the complete model for a spherical target body with arbitrary thrust angle θ will be provided. In Example 9.2, the effect of mass variation when the engine cannot be throttled will be studied.

[18] For simplicity's sake and because the vehicle is treated as a mass point, ε_T is taken as 180°. However, for a finite-dimension body, the direction of the thrust force would come from the rotation of the vehicle to position its engine into flight direction. In that case, ε_T would be 0°, and the vehicle itself would have rotated 180°.

such that

$$V\,\mathrm{d}V = \left(\frac{a_T}{\cos\theta} - g\right)\mathrm{d}h \tag{9.79}$$

which, with the earlier stated assumption of small θ and thus $\cos\theta \approx 1$, is easy to solve. Assume that the terminal conditions are V_f and h_f, Eq. (9.79) is rewritten as:

$$\int_V^{V_f} V'\mathrm{d}V' = (a_T - g)\int_h^{h_f} \mathrm{d}h' \tag{9.80}$$

that solves into

$$\frac{V_f^2 - V^2}{2} = (a_T - g)(h_f - h) \tag{9.81}$$

With the goal to isolate a_T, the above equation is rewritten as

$$(V^2 - V_f^2) = 2g\left(\frac{a_T}{g} - 1\right)(h - h_f) \quad \Rightarrow \quad \frac{V^2 - V_f^2}{h - h_f} = 2g\left(\frac{a_T}{g} - 1\right) \tag{9.82}$$

and

$$\frac{V^2 - V_f^2}{h - h_f}\left(\frac{a_T}{g} + 1\right) = 2g\left(\frac{a_T}{g} - 1\right)\left(\frac{a_T}{g} + 1\right) = \left(\frac{a_T}{g}\right)^2 - 1 \tag{9.83}$$

which finally yields a quadratic solution for the normalised thrust deceleration, given the initial and final state:

$$\left(\frac{a_T}{g}\right)^2 - \left[\frac{V^2 - V_f^2}{2g(h - h_f)}\right]\left(\frac{a_T}{g}\right) - \left(1 + \frac{V^2 - V_f^2}{2g(h - h_f)}\right) = 0 \tag{9.84}$$

The applicable solution to the above equation is defined by the positive root, and is used as follows. During unpowered descent, the solution for $\frac{a_T}{g}$ is solved iteratively for each guidance step. As the altitude decreases, the *required* thrust deceleration a_T will increase. When the corresponding thrust reaches T_{max}, i.e., the design level of the thrust, the gravity turn is initiated where the maximum thrust deceleration is commanded. After completion of the gravity turn, $h = h_f$ and $V = V_f$ will have been reached. In an actual mission, the gravity-turn segment is generally followed by a constant-velocity, vertical-descent segment until $h = 0$, to account for errors and to avoid crashing into the surface. Note that when the gravity turn is initiated at a lower altitude, a larger thrust deceleration is required to reach the same final state.

On board, the propellant mass expended during the gravity turn may easily be approximated for each guidance interval, Δt. The ΔV acquired over each time step can be approximated with $a_T \Delta t$. The propellant mass used can then be calculated via the rocket equation:

$$\Delta m_{fuel} = m_i \left(1 - \frac{1}{e^{\frac{\Delta V}{I_{sp} g_E}}} \right) \tag{9.85}$$

where m_{fuel} is the propellant mass expended and m_i is the initial mass at the beginning of the time interval, both in kg.

Example 9.1 (*Gravity Turn*) The guidance system based on a flat-Moon approximation will be tested on a descending spacecraft, aimed at landing on the Moon with zero terminal velocity. This fictitious spacecraft has a mass of $m = 400$ kg, a maximum available thrust of $T_{max} = 1{,}500$ N, but to understand the effect of having a larger thrust value, also $T_{max} = 1{,}750$ and $T_{max} = 2{,}000$ N are considered. To calculate the required fuel mass for the braking manoeuvre, a specific impulse of $I_{sp} = 260$ s is assumed. Finally, the starting point for the simulation has been set to a point close to the end of the initial descent, i.e., $h_0 = 10$ km, $V_0 = 180$ m/s and $\gamma_0 = -75°$.

The simulation model is the one for planar motion over a spherical, non-rotating Moon, represented by Eqs. (9.60)–(9.64). It is stressed once more that the simulation is executed with constant thrust deceleration, a_T, with a magnitude determined at the beginning of the gravity turn. The decrease of mass is taken into account during the simulation, but it does not affect a_T, i.e., throttling of the engine is automatically included. In a real-life situation this may not be the case; this will be covered in a second example later on in this section.

Figure 9.46 shows the result of the three simulations with different T_{max}. It is clear that for larger T_{max}, the gravity turn begins later in the trajectory, because the time for deceleration can be shorter. However, even though the final velocity is close to zero in all three cases, the shorter the deceleration time the more the final descent deviates from vertical, indicated by the flight-path angle that does not quite reach $-90°$. The decrease of mass is of course proportional to the thrust force, which, given the constant thrust deceleration, will decrease as a function of current mass. The proportionality of this decrease is represented by the throttle setting of the propulsion system, i.e., the ratio of the required thrust to the maximum thrust to maintain a constant a_T. Not clearly visible in the graphs, the fuel-mass flow, i.e., the gradient of the mass-versus-time curves, is not constant, but gets smaller with time. Finally, a shorter burn with a larger thrust force requires less fuel mass, and could thus be used as design decision for the propulsion system.

All in all, the simplified gravity-turn guidance does what it should do, with only small errors in final altitude and flight-path angle. One should realise that the trajectory is simulated for a spherical Moon, whereas the guidance has been derived for a flat Moon. The assumed simplifications have undoubtedly contributed to these errors. Last but not least, while checking the workings of the algorithm it was shown that there is a certain thrust range for which the guidance works. A too low thrust force will not provide sufficient deceleration and with a too large thrust force the final flight-path angle deviates so much from vertical that it negatively affects the deceleration, leaving a significant velocity error at landing. ■

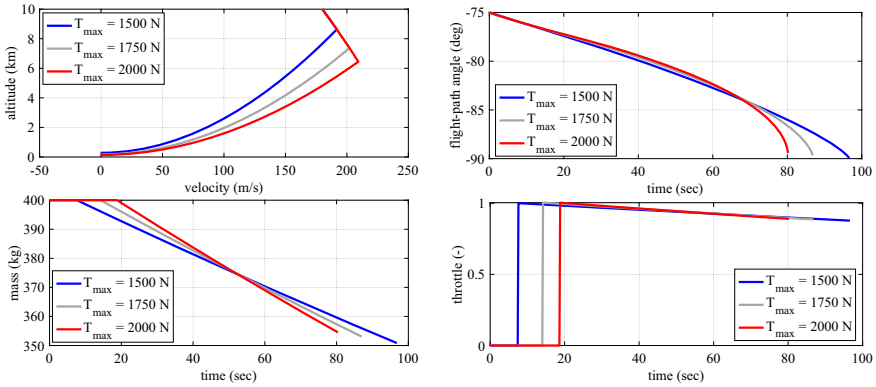


Fig. 9.46 The effect of multiple values of T_{max} on trajectory parameters for Lunar descent

*** Intermezzo 9.1 (Gravity turn for a spherical body)**

In this Intermezzo, the assumptions made earlier, i.e., a close-to-vertical orbit and a flat Moon, will be removed, such that the guidance can be used for non-vertical powered descent over a spherical body. The notion of a constant thrust deceleration opposite to the velocity remains a primary characteristic of the gravity-turn guidance, though. Once the conditions to execute the gravity-turn manoeuvre are met, the spacecraft will touch down with zero horizontal and vertical velocity. It can be applied to non-escape trajectories and is especially useful for landing on a body from a parking orbit or with a lower inclination. As mentioned earlier, for the Surveyor missions, the gravity-turn guidance was used for the maximum thrust Vernier phase.

The first ones to have expanded the guidance algorithm in this way were Citron et al. (1964), although they used cylindrical coordinates for their derivation. We stick to the spherical coordinates, and use Eqs. (9.60)–(9.62) as starting point, albeit without the aerodynamic forces. Also, to avoid the singularity in the equations for vertical flight ($\gamma = -90°$), we change to the deviation from vertical, θ, as was done before (Eqs. (9.65)–(9.68)). Introducing a central-field gravitational acceleration, $g = \frac{\mu}{R^2}$, and the thrust acceleration $a_T = \frac{T}{m}$, the equations of motion read:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -mg\sin\gamma + T\cos\varepsilon_T \quad \Rightarrow \quad \frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{\mu}{R^2}\cos\theta - a_T \tag{9.86}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = T\sin\varepsilon_T - mg\cos\gamma\left(1 - \frac{V^2}{gR}\right) \quad \Rightarrow \quad V\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\mu}{R^2}\sin\theta\left(1 - \frac{RV^2}{\mu}\right) \tag{9.87}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = V\sin\gamma \quad \Rightarrow \quad \frac{\mathrm{d}R}{\mathrm{d}t} = -V\cos\theta \tag{9.88}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = V\cos\gamma \quad \Rightarrow \quad \frac{\mathrm{d}x}{\mathrm{d}t} = V\sin\theta \tag{9.89}$$

which is the starting point for the derivation.

The descent trajectory resembles a parabolic arc, and is defined by the combination of altitude and downrange i.e., h and x. Along the path, these two parameters are combined into the path length, s. This is also the main parameter to determine the work done by the propulsion system, as it acts in opposite direction of the velocity. Therefore, it makes sense to change the independent variable from t to s, so with $V = \frac{\mathrm{d}s}{\mathrm{d}t}$, we find for Eq. (9.86):

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{\mathrm{d}V}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = V\frac{\mathrm{d}V}{\mathrm{d}s} = -\frac{\mu}{R^2}\cos\theta - a_T \quad \Rightarrow \tag{9.90}$$

$$V\frac{\mathrm{d}V}{\mathrm{d}s} = -\frac{\mu}{R^2}\cos\theta - a_T \tag{9.91}$$

Similarly, for Eq. (9.87) it follows that

$$V\frac{\mathrm{d}\theta}{\mathrm{d}t} = V^2\frac{\mathrm{d}\theta}{\mathrm{d}s} = \frac{\mu}{R^2}\sin\theta\left(1 - \frac{RV^2}{\mu}\right) \quad \Rightarrow \tag{9.92}$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}s} = \frac{\mu}{R^2}\sin\theta\left(\frac{1}{V^2} - \frac{R}{\mu}\right) \tag{9.93}$$

Finally, the two kinematic equations change to

$$\frac{\mathrm{d}R}{\mathrm{d}s} = \frac{\mathrm{d}h}{\mathrm{d}s} = -\cos\theta \tag{9.94}$$

$$\frac{\mathrm{d}x}{\mathrm{d}s} = \sin\theta \tag{9.95}$$

The rate of change of specific energy, $\mathcal{E}$, is equal to the work of the propulsion system done in flight direction, opposite to the direction of the velocity: $\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t} = -a_T V$. The change in specific energy with respect to s is:

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}s} = \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}s} = -a_T \tag{9.96}$$

After integration, this becomes:

$$\mathcal{E} = -a_T s + \mathcal{E}_0 = -a_T s + \frac{V_0^2}{2} + \frac{\mu}{R_0} \tag{9.97}$$

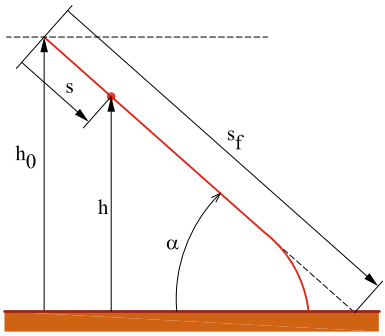


Fig. 9.47 Descent-trajectory curvature and approximation

with V_0 and R_0 being the initial velocity and radial distance, respectively. By definition, the specific energy is the sum of specific kinetic and potential energy (here, referenced to the centre of the Moon), i.e.,

$$\mathcal{E} = \mathcal{E}_k + \mathcal{E}_p = \frac{V^2}{2} + \frac{\mu}{R} \tag{9.98}$$

By equating Eqs. (9.97) and (9.98) and isolating the velocity, one gets:

$$V^2 = -2a_T s + V_0^2 + 2\mu\left(\frac{R_0 - R}{RR_0}\right) \tag{9.99}$$

For a spherical Moon, $R = R_e + h$, so we may write for the denominator

$$RR_0 = (R_e + h)(R_e + h_0) = R_e^2\left(1 + \frac{h}{R_e} + \frac{h_0}{R_e} + \frac{hh_0}{R_e^2}\right) \approx R_e^2$$

as terms with $\frac{h}{R_e}$ are much smaller than 1. With $g_M = \frac{\mu}{R_e^2}$ being the surface value of gravitational acceleration, Eq. (9.99) can thus be rewritten as:

$$V^2 = -2a_T s + V_0^2 + 2g_M(h_0 - h) \tag{9.100}$$

During descent from orbit, two opposing effects determine the trajectory shape in the altitude-velocity plane. The first one is the thrust deceleration, and since this deceleration acts in a direction opposite to the velocity, the resulting path would be a straight line if it had been the only effect. However, the second effect originates from gravity that pulls down the vehicle in vertical direction, and thus curving the trajectory (Fig. 9.47). The question is, how much is this pull, especially if we would relate it to the vehicle's initial velocity V_0?

With the descent vehicle starting from either from lunar orbit, or being on a direct entry path as was Surveyor, V_0 will at least have the magnitude of the local circular velocity. For Apollo, the powered descent was initiated at an altitude of 50,000 ft

(=15,240 m), and at that altitude $V_0 = 1{,}684.9$ m/s. The so-called free-fall velocity corresponding to that altitude is $V_f = 222.2$ m/s,[19] which gives a ratio $V_0/V_f = 7.58$. This means that the curvature of the trajectory will not be affected so much by gravity, and the path can be approximated by a straight line (in Eq. (9.100), the velocity appears quadratic, and in that case the ratio is 57.49 that gives an even smaller effect). Only towards the end of the descent will gravity become the dominant factor, and the trajectory will curve towards the local vertical.

By approximating the vehicle path by a straight line, the (local) altitude (as a function of s) in Eq. (9.100) can be easily derived. From Fig. 9.47, the following expressions hold:

$$\sin\alpha = \frac{h_0}{s_f} \quad \text{and also} \quad \sin\alpha = \frac{h}{s_f - s} \tag{9.101}$$

which yields an expression for $h(s)$:

$$h = h_0\left(1 - \frac{s}{s_f}\right) \tag{9.102}$$

Furthermore, the guidance system has to achieve specified terminal conditions of the landing, i.e., $V = 0$ and $h = 0$ (or, equivalently, some small values for either variable above the landing site, such that a purely vertical descent can conclude the landing procedure). Substituting the terminal conditions into Eq. (9.100) will directly lead to an expression for s_f:

$$s_f = \frac{V_0^2 + 2g_M h_0}{2a_T} \tag{9.103}$$

Substituting this result, together with Eq. (9.102), back into Eq. (9.100) yields a relation for $V(s)$:

$$V^2 = V_0^2\left(1 - \frac{s}{s_f}\right) \tag{9.104}$$

The next step in the derivation is to link the expressions for V and h to the equations of motion, Eqs. (9.91) and (9.94):

$$V\frac{\mathrm{d}V}{\mathrm{d}s} = -g_M\cos\theta - a_T \quad \text{and} \quad \frac{\mathrm{d}h}{\mathrm{d}s} = -\cos\theta \tag{9.105}$$

[19] The free-fall velocity is the final velocity an object would achieve when it is accelerated by gravity alone, starting from an altitude h_0 with zero velocity. Assuming constant gravity g, velocity will increase with $V = V_0 + gt_f = gt_f$, whereas altitude will decrease according to $h_f = h_0 + V_0 t_f - 0.5gt_f^2 = h_0 - 0.5gt_f^2$. From these two equations, the time-to-impact, t_f, can easily be found as $t_f = \sqrt{2(h_0 - h_f)/g}$. With $h_0 = 15{,}240$ m and $g = 1.62$ m/s^2, this gives $t_f = 137.2$ s, and thus $V_f = 222.2$ m/s. The local circular velocity is given by $V_c = \sqrt{gR}$, and with the equatorial radius of the Moon $R_M = 1737.1$ km, this gives $V_c = 1{,}684.9$ m/s.

The kinematic equation combined with Eq. (9.102) provides

$$\frac{\mathrm{d}h}{\mathrm{d}s} = -\frac{h_0}{s_f} \quad \Rightarrow \quad \cos\theta = \frac{h_0}{s_f} = \cos\theta_0 \tag{9.106}$$

such that with this expression for $\cos\theta$, the velocity derivative becomes:

$$V\frac{\mathrm{d}V}{\mathrm{d}s} = -g_M\frac{h_0}{s_f} - a_T \tag{9.107}$$

This equation is easily solved for V and s:

$$\int_{V_0}^{V} V'\mathrm{d}V' = -\left(g_M\frac{h_0}{s_f} - a_T\right)\int_0^{s_f}\mathrm{d}s' \quad \Rightarrow \tag{9.108}$$

$$\frac{V^2 - V_0^2}{2} = -\left(g_M\frac{h_0}{s_f} - a_T\right)s_f \tag{9.109}$$

To find an explicit expression for $V(s)$ is not trivial, but we can circumvent this by taking the second derivative of R with respect to s, i.e.,

$$\frac{\mathrm{d}^2R}{\mathrm{d}s^2} = \sin\theta\frac{\mathrm{d}\theta}{\mathrm{d}s} \tag{9.110}$$

and with Eq. (9.93) reformulated as

$$\frac{\mathrm{d}\theta}{\mathrm{d}s} = \frac{1}{RV^2}\sin\theta\left(\frac{\mu}{R} - V^2\right) \tag{9.111}$$

Eq. (9.110) is rewritten as

$$\frac{\mathrm{d}^2R}{\mathrm{d}s^2} = \frac{1}{RV^2}\sin^2\theta\left(\frac{\mu}{R} - V^2\right) \tag{9.112}$$

As simplification we use the fact that for a major portion of the flight $\frac{\mathrm{d}R}{\mathrm{d}s}$ is close to its initial value (as long as the velocity is larger than the free-fall velocity), and with

$$\left.\frac{\mathrm{d}R}{\mathrm{d}s}\right|_0 = \cos\theta_0 \tag{9.113}$$

it does mean that in Eq. (9.112) $\sin^2\theta \approx \sin^2\theta_0$. Typically, this initial angle is provided as flight-path angle, so with $\sin\theta = \cos\gamma$ we may write

$$\frac{R}{\cos^2\gamma_0}\frac{\mathrm{d}^2R}{\mathrm{d}s^2} = \left(\frac{\mu}{RV^2} - 1\right) \tag{9.114}$$

Last, but not least, since the radial distance will only change a few percent from the beginning of the descent to landing,[20] R can be approximated by R_M. Introducing $g = \frac{\mu}{R^2} \approx g_M$, Eq. (9.114) is rewritten as

$$\frac{R_M}{\cos^2\gamma_0}\frac{\mathrm{d}^2R}{\mathrm{d}s^2} = \left(R_M\frac{g_M}{V^2} - 1\right) \tag{9.115}$$

Finally, with Eq. (9.104) substituted for V^2, we have obtained an explicit relation for R as a function of s that we can now try to solve:

$$\frac{R_M}{\cos^2\gamma_0}\frac{\mathrm{d}^2R}{\mathrm{d}s^2} = R_M\frac{g_M}{V_0^2\left(1-\frac{s}{s_f}\right)} - 1 \tag{9.116}$$

Integrating the above equation, using the solution to the standard integral, $\int \frac{1}{1-ax}\mathrm{d}x = -\frac{1}{a}\ln(1-ax) + c_1$ yields

$$\frac{R_M}{\cos^2\gamma_0}\frac{\mathrm{d}R}{\mathrm{d}s} = -\frac{R_M}{g_M}\frac{s_f}{V_0^2}\ln\left(1-\frac{s}{s_f}\right) - s + c_1 \tag{9.117}$$

Using the initial condition for $s = 0$, $\frac{\mathrm{d}R}{\mathrm{d}s}\big|_0 = \sin\gamma_0$, the constant is found to be $c_1 = \frac{\sin\gamma_0}{g_M\cos^2\gamma_0}$.

Integrating once more, now using the solution to the standard integral

$$\int \ln(ax+b)\mathrm{d}x = \frac{(ax+b)\ln(ax+b) - (ax+b)}{a} + c_2$$

with $a = -\frac{1}{s_f}$ and $b = 1$, yielding

$$-\frac{\left(1-\frac{s}{s_f}\right)\ln\left(1-\frac{s}{s_f}\right) - \left(1-\frac{s}{s_f}\right)}{\frac{1}{s_f}} + c_2$$

and thus

$$\frac{R_M}{\cos^2\gamma_0}R = \frac{\sin\gamma_0}{g_M\cos^2\gamma_0}s - \frac{R_M}{g_M}\frac{s_f^2}{V_0^2}\left(1-\frac{s}{s_f}\right)\ln\left(1-\frac{s}{s_f}\right) - \frac{s^2}{2} + c_2 \tag{9.118}$$

[20] With the equatorial radius of the Moon being $R_M = 1737.1$ km, an initial altitude of 60–80 km represents a change of at most 4.6%.

With the initial condition $R = R_0$ for $s = 0$, the second constant of integration easily follows: $c_2 = \frac{R_0}{\cos^2 \gamma_0}$. Rearranging a few of the terms, the final solution is:

$$\frac{R - s \sin \gamma_0}{\cos^2 \gamma_0} = -\frac{\sin \gamma_0}{g_M \cos^2 \gamma_0} s - \frac{g_M s_f^2}{V_0^2} \left(1 - \frac{s}{s_f}\right) \ln \left(1 - \frac{s}{s_f}\right) - \frac{s^2}{2R_M} + \frac{R_0}{\cos^2 \gamma_0} \tag{9.119}$$

To obtain the thrust acceleration that fits the above path constraint, we substitute $R = R_0 - h_0$ and $s = s_f$ in the above equations, and substitute Eq. (9.103) for s_f, resulting in:

$$\frac{a_T^2}{g_M^2} + \sin \gamma_0 \left(\frac{V_0^2}{2h_0 g_M} + 1\right) \frac{a_T}{g_M} - \frac{\cos^2 \gamma_0}{4V_0^2 h_0 g_M} \left(V_0^2 + 2g_M h_0\right)^2 \left(1 - \frac{V_0^2}{2R_M g_M}\right) = 0 \tag{9.120}$$

For velocities below the escape velocity, this equation has only one positive root. This root is the solution and initiates the maximum vernier thrust phase. Citron et al. (1964) also provide an expression for the targeting final flight range, s_f, but derivation of said expression is considered beyond the scope of this introductory text, and the reader is referred to the quoted reference. This paper also compares this guidance law with numerical simulations, and it has been shown to be in exact agreement.

For implementation in a guidance system, the same approach as with the flat-Moon approximation can be followed. Replacing V_0 and γ_0 with the current values V and γ (which can be considered to be the "initial state" at the beginning of each guidance sample interval), the current level of a_T is evaluated with Eq. (9.120), and compared with the maximum available thrust deceleration, $a_{T,max} = \frac{T_{max}}{m}$. Once the current value reaches $a_{T,max}$, the propulsion system is switched on and will thrust at maximum capacity till the terminal conditions are reached. ■

Example 9.2 (*Surveyor Gravity Turn*) In this second example, the Surveyor gravity turn will be discussed. As mentioned, the gravity-turn guidance algorithm was originally developed for Surveyor (Citron et al. 1964), and it seems fit to simulate the last leg of its descent to the surface. The terminal descent phase of the Surveyor missions start after the main retro phase, during which the main velocity is taken out of the system. The initial conditions for the simulation are the same as the end conditions of this retro phase, and can be found in the published mission reports (JPL 1966). At the beginning of the so-called Vernier phase, the velocity is $V_0 = 119.36$ m/s (392 ft/s) at an altitude of $h_0 = 8{,}708.4$ m (28,571 ft), with a flight-path angle of $\gamma_0 = -83.98°$, which is thus $\theta_0 = 6.02°$ away from the local vertical. The minimum and maximum thrust are 480.4 and 1,187.7 N, respectively. The initial mass of the spacecraft is 296.4 kg plus 80.9 kg of fuel.

Initially, the spacecraft uses the minimum thrust of the Vernier thrusters, while the guidance system is establishing the moment when the switch to maximum thrust has to be made. When this happens, in the remainder of the flight the velocity is nulled for the target altitude of $h_f = 0$ m, while simultaneously reaching $\gamma_f = -90°$.

The resulting trajectory is shown in Fig. 9.48. At first glance, the resulting trajectory seems similar to the one discussed in Example 9.1. However, it is clear that at the end of the descent, when the velocity has reached 0 m/s, the altitude is non-zero, i.e., $h \approx 390$ m. The explanation for this residual altitude is simple, when we consider the fact that the Surveyor engine was not throttleable, so the gravity turn was executed with maximum thrust, from the moment guidance established the commanded thrust deceleration of $a_{T,max}$, based on the vehicle mass at that moment. Due to the ever decreasing mass, the actual thrust deceleration is always larger (as is shown in the bottom right plot), and thus the velocity will be zero before zero altitude. To solve this problem, either some form of predictor-corrector method could be used, taking the mass variation into account, or the engine should slowly be throttled down, as was done in Example 9.1. Note that the discrete jumps in the mass variation and the thrust-load plots are the cause of switching from T_{min} to T_{max}. It is pointed out that despite the fact that the system starts in a close-to-vertical trajectory, it takes about 35 s before the gravity-turn guidance is activated, enforcing a vertical descent some 60 s later. ■

9.3.3 E-Guidance

An open-loop guidance system, like the gravity-turn guidance discussed above will never be able to meet with all terminal conditions of velocity, altitude, and landing location. However, a crucial requirement to perform precision landing is that the

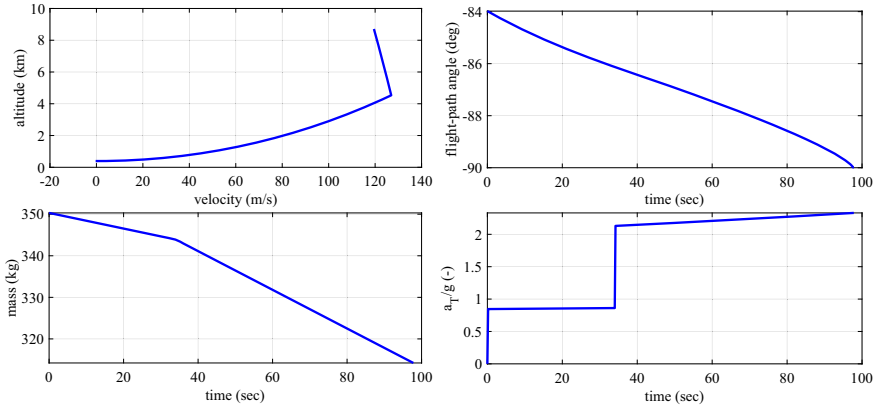


Fig. 9.48 Surveyor gravity-turn manoeuvre

guidance law is capable to land at a specified landing location, independent of the initial conditions. Such a system could then also be used for hazard avoidance, e.g., when the on-board cameras would identify a large boulder or crater at the designated landing site, such that an avoidance manoeuvre has to be executed that leads the vehicle to a new, safe landing site.

One of the guidance laws that can achieve this is E-guidance, sometimes also called Apollo-guidance or quadratic guidance. It is an explicit guidance method that solves the two-point boundary-value problem. It is a very simple method that can compute the trajectory in-flight. There are several ways to derive E-guidance, and two possible derivations are provided by Cherry (1964a) and Ohlmeyer (2003). These references follow a generalised algebraic approach, and one that proves optimality using optimal-control theory. A simpler method that gives insight from a slightly different angle has been developed by Sostaric and Rea (2005), and will be repeated here with slight modification.

The basic principle of E-guidance is to constrain the translational acceleration as a linear polynomial with time t. This acceleration profile, $\mathbf{a}$, would be the combination of thrust and gravity, so for an arbitrary gravity field, $\mathbf{a}$ is then defined as[21]:

$$\mathbf{a}(t) = \mathbf{a_c} + \mathbf{g} = \mathbf{c_1} + \mathbf{c_2}(t_f - t) = \mathbf{c_1} + \mathbf{c_2} t_{go} \tag{9.122}$$

with $\mathbf{a_c}$ being the commanded acceleration and $\mathbf{g}$ the local gravitational acceleration. The coefficients $\mathbf{c_1}$ and $\mathbf{c_2}$ and the time-to-go t_{go} are the quantities, which have to be determined such that the acceleration profile will lead the vehicle from its initial conditions to a final state. $\mathbf{c_1}$ and $\mathbf{c_2}$ are constants to be determined from the boundary conditions, and ideally remain constant throughout the descent. Also the time-to-go, t_{go}, follows from satisfying the the initial and final boundary conditions, and is used to shape the trajectory profile. The time-to-go is introduced for reasons of convenience, and will be discussed in more detail below.

Assuming now that $\mathbf{c_1}$ and $\mathbf{c_2}$ are known, the acceleration profile that will successfully realise a trajectory meeting the boundary conditions *until the specified time* t can be determined. Provided, of course, that t_{go} is known, which could be a user-specified variable or it could be estimated with several methods, known from literature. This will be discussed later in this section. It suffices to know for now that any future time t can be specified, and consequently *any* time-to-go, such that Eq. (9.122) will realise the trajectory to meet the boundary conditions starting from the current time t_0.

To obtain an analytic expression for the trajectory $\mathbf{r} = \mathbf{r}(t)$ and $\mathbf{v} = \mathbf{v}(t)$ that will be followed due to the acceleration profile, it is necessary to integrate Eq. (9.122) twice, and once we substitute the boundary conditions we can solve for $\mathbf{c_1}$ and $\mathbf{c_2}$.

---

[21] Because the flight-control system will only be dealing with the commanded thrust acceleration, a more logical formulation of Eq. (9.122) would therefore be

$$\mathbf{a_c} = \mathbf{a} - \mathbf{g} = \mathbf{c_1} + \mathbf{c_2} t_{go} - \mathbf{g} \tag{9.121}$$

Starting from this equation, however, will not alter the discussion in the text.

Thus, integrating once, from the initial time t_0 until an arbitrary future time t for which the required velocity shall be obtained, gives:

$$\begin{aligned}\mathbf{v} - \mathbf{v_0} &= \int_{t_0}^{t} \mathbf{a}(\tau)\,\mathrm{d}\tau = \int_{t_0}^{t} \{\mathbf{c_1} + \mathbf{c_2}(\tau - t_0)\}\,\mathrm{d}\tau \\ &= \mathbf{c_1} t_{go} + \frac{1}{2} t_{go}^2\end{aligned} \tag{9.123}$$

having assumed a constant t_{go} for now; integrating once more yields for the position

$$\begin{aligned}\mathbf{r} - \mathbf{r_0} - \mathbf{v_0} t_{go} &= \int_{t_0}^{t} \mathbf{v}(\tau)\,\mathrm{d}\tau = \int_{t_0}^{t} \left\{\mathbf{c_1}(\tau - t_0) + \mathbf{c_2}(\tau - t_0)^2 + \mathbf{v_0}\right\}\,\mathrm{d}\tau \\ &= \frac{1}{2}\mathbf{c_1} t_{go}^2 + \frac{1}{6}\mathbf{c_2} t_{go}^3\end{aligned} \tag{9.124}$$

The actual trajectory can thus be written as an explicit function of time:

$$\mathbf{v}(t) = \mathbf{c}_1 t_{go} + \frac{1}{2}\mathbf{c}_2 t_{go}^2 + \mathbf{v}_0 \tag{9.125}$$

$$\mathbf{r}(t) = \frac{1}{2}\mathbf{c}_1 t_{go}^2 + \frac{1}{6}\mathbf{c}_2 t_{go}^3 + \mathbf{v}_0 t_{go} + \mathbf{r}_0 \tag{9.126}$$

The latter equation is a third-order polynomial that connects the initial and final positions $\mathbf{r_0}$ and $\mathbf{r_f}$, and also fulfils the boundary conditions $\dot{\mathbf{r}}_\mathbf{0}$ and $\dot{\mathbf{r}}_\mathbf{f}$ for given coefficients.

To solve for $\mathbf{c_1}$ and $\mathbf{c_2}$, Eqs. (9.123) and (9.124) are written as a matrix-vector equation:

$$\begin{pmatrix} \mathbf{v} - \mathbf{v}_0 \\ \mathbf{r} - \mathbf{r}_0 - \mathbf{v}_0 t_{go} \end{pmatrix} = \begin{bmatrix} [1.3] t_{go} & \frac{1}{2} t_{go}^2 \\ \frac{1}{2} t_{go}^2 & \frac{1}{6} t_{go}^3 \end{bmatrix} \begin{pmatrix} \mathbf{c_1} \\ \mathbf{c_2} \end{pmatrix} \tag{9.127}$$

Until now we had assumed that any velocity $\mathbf{v}$ and position $\mathbf{r}$, to be reached within t_{go}, could be specified. At no loss of generality, this would thus also hold for the terminal conditions $\mathbf{v_f}$ and $\mathbf{r_f}$ that the guidance system shall achieve. Therefore, these conditions will be considered from now on, and thus, after inverting the coefficient matrix, the following closed-form solution is obtained:

$$\begin{pmatrix} \mathbf{c_1} \\ \mathbf{c_2} \end{pmatrix} = \begin{bmatrix} [1.3] - \frac{2}{t_{go}} & \frac{6}{t_{go}^2} \\ \frac{6}{t_{go}^2} & -\frac{12}{t_{go}^3} \end{bmatrix} \begin{pmatrix} \mathbf{v_f} - \mathbf{v_0} \\ \mathbf{r_f} - \mathbf{r_0} - \mathbf{v_0} t_{go} \end{pmatrix} \tag{9.128}$$

In the above formulation, the initial and final accelerations $\mathbf{a_0}$ and $\mathbf{a_f}$ are free. This means that a discontinuity may occur as soon as the E-guidance is activated, because suddenly a new thrust vector with corresponding magnitude and orientation will be

commanded. However, the consequences of this are neglected here, but in a real-life implementation this should be dealt with, of course.

In summary, the outcome of the above derivation is a closed-form solution to the guidance problem. Assuming perfect navigation and control, the coefficients $\mathbf{c_1}$ and $\mathbf{c_2}$ have to be computed only once. In practice, this is not the case and they need to be recomputed, reasons being that in reality the previous assumptions do not hold, navigation will not provide a perfect state knowledge, and the spacecraft cannot be treated as a mass point. The inertia of the system will prevent an infinite reorientation of the thrust force. But, typically the assumption of a constant time-to-go does not hold either and it may have to be continuously estimated, so even for the simplified model provided above, the coefficients will have to be recomputed. It may be clear from inspecting Eq. (9.128), though, that close to landing, when $t_{go} \to 0$, $\mathbf{c_1}$ and $\mathbf{c_2}$ will approach ∞. This would lead to an infinite thrust command and obviously has to be avoided. A practical solution is to stop recomputing $\mathbf{c_1}$ and $\mathbf{c_2}$ when $t_{go} \leq 1$ s. As t_{go} is an important parameter in E-guidance, we will discuss methods to obtain it in more detail.

Many methods exist for determining the time-to-go, some of which are covered by Steinfeldt et al. (2008), Wong et al. (2006), Cherry (1964b), Klumpp (1971), Lafontaine et al. (2004) and D'Souza (1997), who are discussing applications of E-Guidance and other explicit algorithms. A simple formulation is found by solving the rocket equation, which does approach the propellant optimum when compared to a numerical optimisation with t_{go} as independent variable. Cherry (1964b) describes this method for a *fixed*-thrust rocket, but it can also be adapted to a a throttleable propulsion system, as we will describe below.

Starting with a fixed-thrust rocket, the final vehicle mass m_f after a certain burn time t_b will be:

$$m_f = m_0 - \dot{m}_c t_b \tag{9.129}$$

where $\dot{m}_c$ is the constant mass-flow rate of the engine. In Eq. (9.121), the time-to-go was defined as $t_{go} = t_f - t$, and if it is assumed that time starts at $t = 0$ s, then the engine will burn for as long as the t_{go}, such that:

$$t_b = t_{go} \tag{9.130}$$

The rocket equation reads:

$$m_f = m_0 e^{-\Delta V / V_e} \tag{9.131}$$

with m_0 being the initial spacecraft mass, ΔV the change in velocity, and V_e the exhaust velocity. Combining Eqs. (9.129) and (9.131) yields:

$$m_0 - \dot{m}_c t_{go} = m_0 e^{-\Delta V / V_e} \tag{9.132}$$

With the introduction of the characteristic time, τ,

$$\tau = \frac{m_0}{\dot{m}_c} \tag{9.133}$$

Equation (9.132) changes to

$$\frac{m_0}{\dot{m}_c} - t_{go} = \frac{m_0}{\dot{m}_c} e^{-\Delta V/V_e} \quad \Rightarrow \quad t_{go} = \frac{m_0}{\dot{m}_c}\left(1 - e^{-\Delta V/V_e}\right) \tag{9.134}$$

or

$$t_{go} = \tau\left(1 - e^{-\Delta V/V_e}\right) \tag{9.135}$$

Note that the characteristic time $\tau = m/\dot{m}$ is commonly encountered, and can be interpreted as the time it would take to burn up the entire vehicle, were it made up only of propellant and the mass flow rate constant.

Equation (9.135) provides us with t_{go}, given the ΔV and assuming a rocket with a fixed thrust. In case this equation is applied to a variable-thrust rocket, however, the estimation is unstable, because of the dependency on the mass flow. With every iteration, t_{go} will become shorter, because the thrust, and hence mass-flow, will go up. To stabilise the process, τ is redefined as:

$$\tau = \tau_0 - t \tag{9.136}$$

where τ_0 is the initial value obtained at the first iteration after the initialisation, and t is the running variable, starting from $t = 0$ s at the first call of the guidance law. This redefinition is in principle a scaling with respect to the current vehicle mass, assuming a constant mass flow:

$$\tau = \tau_0 - t \quad \Leftrightarrow \quad \tau = \frac{m_0}{\dot{m}} - t \quad \Leftrightarrow \quad \tau = \frac{m_0 - \dot{m}t}{\dot{m}} = \frac{m}{\dot{m}} \tag{9.137}$$

ΔV is easily computed including gravity losses:

$$\Delta V = \sqrt{(\dot{x} - \dot{x}_f)^2 + (\dot{y} - \dot{y}_f)^2 + (\dot{z} - \dot{z}_f + g t_{go})^2} \tag{9.138}$$

whereas τ has to be initialised with τ_0 as the following value:

$$\tau_0 = \frac{m_0}{\dot{m}} \tag{9.139}$$

with $\dot{m}$ being the engine's mass-flow. There is one catch, though. Consider the commanded thrust at t_0, following from the commanded acceleration, a_c:

$$T(t_0) = m_0 a_c = \dot{m} V_e \quad \Leftrightarrow \quad \frac{m_0}{\dot{m}} = \frac{V_e}{a_c} = \tau_0 \tag{9.140}$$

This means that τ_0 is dependent on the commanded acceleration, which is a quantity that is to be computed by guidance and thus clearly not known at t_0. Therefore, t_{go} has to be initialised in a different way, after which the resulting thrust acceleration may be used to compute τ_0, and the procedure can continue as described above. The manner in which t_{go} is initialised is not that important, because the procedure will converge to a steady value very quickly. Test runs have shown that the following value results in a good first estimate that will not be too far from what the primary method will determine (Gerth and Mooij 2014):

$$t_{go} = 2\frac{x_f - x}{\dot{x}_f - \dot{x}}\bigg|_{\text{along track}} \tag{9.141}$$

where all quantities must be given in the along-track direction, such that the position vector of the spacecraft is in the x-y plane of the coordinate system (x along track, y cross track, z completes the triad). The switch from this initialisation method to the primary estimation technique occurs after a user-specified time.

It was shown by Gerth and Mooij (2014) that for the low-velocity Lunar descent the E-guidance had quite some difficulty to find the proper commands, i.e., the spacecraft was first commanded to fly upside down and to accelerate towards the landing site, only to pitch over again and brake towards the end. Even though the terminal conditions were met, this trajectory is clearly infeasible in a real-life scenario. The problem originated from the fact that the velocity is rather low, yielding a small ΔV and consequently a short time-to-go. To reach the landing site "in time", the vehicle must therefore accelerate towards it, so that it flies faster and arrives within the predicted t_{go}. The solution to this problem appeared to be a slight adjustment to the t_{go}-estimation, Eq. (9.135), by introducing a tuning factor Γ to the equation, an approach also proposed by D'Souza (1997), who adds a weighting factor Γ to the computation of t_{go} in his explicit law. Thus

$$t_{go} = \Gamma\tau\left(1 - e^{-\Delta V/V_e}\right) \tag{9.142}$$

A value of $\Gamma = 2.2$ works well for Lunar landing (Gerth and Mooij 2014).

To conclude the discussion on E-guidance, two examples showing the flight performance are provided. The first example shows the Surveyor descent for similar conditions as in Example 9.2. The second one shows the versatility of E-guidance for multiple starting conditions.

Example 9.3 (*Surveyor E-Guidance)* The gravity turn of Surveyor in Example 9.2 was started with initial conditions $V_0 = 119.36$ m/s at an altitude of $h_0 = 8{,}708.4$ m, with a flight-path angle of $\gamma_0 = -83.98°$. With a descent duration of about 98 s and a fuel spent of $m_f = 35.8$ kg, the down-range distance covered was 553.6 m. Without going too much into detail whether this gravity turn is representative or not, we will simulate the same descent with the E-guidance system. Different in the approach is, of course, that we assume the engine to be throttleable, and for the sake of simplicity the minimum thrust level is assumed to be zero. As the E-guidance system requires

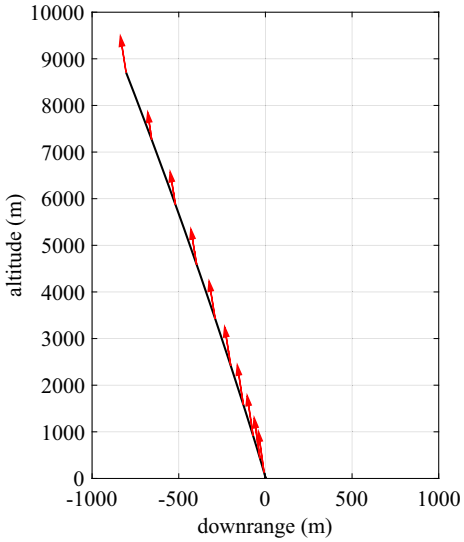


Fig. 9.49 Surveyor E-guidance trajectory: altitude versus downrange. Arrows indicate thrust direction

the initial down range, it is set to $x_0 = -800$ m, comparable to the distance covered in the gravity turn. Finally, the terminal-gate conditions are defined to be $h_f = 15$ m, $x_f = 0$, $v_{x,f} = 0$ m/s, and $v_{y,f} = -1.5$ m/s.

In Fig. 9.49, Surveyor's trajectory is shown, with arrows indicating the direction of the thrust. As the starting position was close to vertical, the trajectory itself is also close to vertical. All terminal conditions are met, which confirms the targeting capabilities of the guidance system. Figure 9.50 shows some characteristic parameters associated with this descent trajectory. These parameters can easily be compared with those in Fig. 9.48, with the difference that in the current simulation the thrust load is not kept constant, and has been determined by dividing the actual thrust by the instantaneous mass of the vehicle. The time-to-go shows a spike at the beginning, which comes from the initialisation of the algorithm. Of course, this results in a step command in the thrust (acceleration), to be avoided in a real-life implementation.

The decrease in mass is continuous, with a relatively constant mass flow after some 20 s, with a final mass of 310.1 kg. This means that $m_f = 40.2$ kg, which is about 15% more than was used during the gravity turn. This shows that, apart from the precision-landing capability, the gravity-turn guidance is not bad at all when it comes to fuel usage. Finally, the initial t_{go} estimation shows a value of about 100 s, whereas the final flight time is 120 s. This means that initially the estimate lacks accuracy, but this is updated during the descent and the estimate becomes more and more accurate. The time-to-go approaches zero close to landing. ■

Example 9.4 (*ESA Lunar Lander E-Guidance)* For this second example we look at the approach phase of the ESA Lunar Lander, which was studied by Gerth and

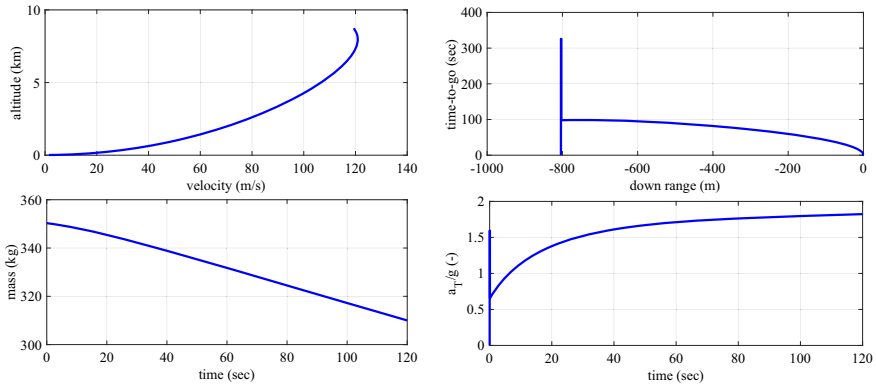


Fig. 9.50 Surveyor E-guidance trajectory characteristics

Mooij (2014). This vehicle has an initial mass of $m = 760$ kg, and a throttleable engine with a specific impulse of $I_{sp} = 325$ s. The nominal initial conditions are a downrange position $x_0 = -1{,}600$ m at an altitude $h_0 = 1{,}569$ m. The downrange speed is $v_{x,0} = 150$ m/s and the descent rate $v_{y,0} = -40$ m/s. Target conditions are the same as in the previous example, i.e., $h_f = 15$ m, $x_f = 0$, $v_{x,f} = 0$ m/s, and $v_{y,f} = -1.5$ m/s. Besides the nominal downrange position, also a short range will be considered, $x_0 = -600$ m, as well as to longer ranges: $x_0 = -2{,}600$ m and $x_0 = -3{,}600$ m. Finally, a very long range is studied: $x_0 = -10{,}000$ m. All other initial conditions remain the same.

Figure 9.51 shows all trajectories, the four low-range and medium-range ones on the left, and the long-distance one on the right. In all cases the trajectories end at the designated terminal conditions, with generally very small errors. For instance, for the nominal case, the achieved downrange position is $x_f = 0.03$ m (error 0.03 m), altitude, $h_f = 14.99$ m (error 0.01 m), downrange velocity $v_{x,f} = -0.01$ m/s (error -0.01 m/s), and descent rate $v_{y,f}$ 1.48 m/s (error 0.02 m/s). The used fuel mass $m_f = 53.68$ kg. A few observations. When the horizontal distance separating the spacecraft from the landing site is small, an "overshoot" trajectory is flown, and the spacecraft approaches the terminal gate from the right. For larger distances, the path followed resembles a parabolic arc, where the guidance aims at a vertical descent path close to the end, to avoid having any remaining (horizontal) velocity. In all cases, the trajectories end with very small terminal errors, see Table 9.3.

In this respect, the long-distance trajectory seems like it would have a larger error in horizontal velocity, but this is, in fact, not true. Even though moving more or less horizontally towards the terminal gate, the actual velocity is very small. As can be seen, the thrust direction is almost vertical, and is mainly used to counter the Moon's gravity (the vertical velocity is already almost equal to the terminal value of -1.5 m/s. The slight offset from vertical is required to finally negate the remaining horizontal velocity.

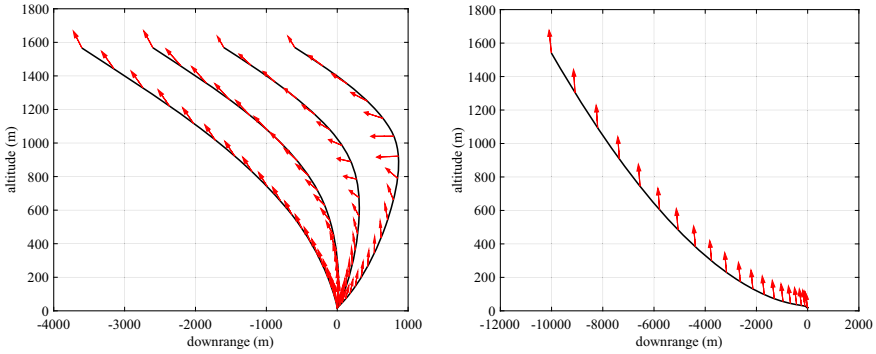


Fig. 9.51 Sensitivity of E-guidance performance for several initial downrange locations. The arrows indicate the thrust directions

Table 9.3 Summary of trajectory characteristics

x_0 (m)	m (kg)	x_f (error) (m)	h_f (error) (m)	$v_{x,f}$ (error) (m/s)	$v_{y,f}$ (error) (m/s)
−600	44.64	0.00 (0.00)	14.99 (0.01)	−0.01 (−0.01)	−1.48 (−0.02)
−1,600	53.68	0.00 (0.00)	14.99 (0.01)	−0.01 (−0.01)	−1.48 (−0.02)
−2,600	61.43	0.00 (0.00)	15.00 (0.00)	−0.01 (−0.01)	−1.49 (−0.01)
−3,600	68.12	0.00 (0.00)	15.00 (0.00)	−0.01 (−0.01)	−1.49 (−0.01)
−10,000	99.76	0.00 (0.00)	14.99 (0.01)	−0.02 (−0.02)	−1.50 (0.00)

However, there is one other aspect that needs to be discussed for these long-distance trajectories. Because the initial velocity is relatively low, compared to the distance away from the terminal gate, the vehicle has to accelerate first to have sufficient velocity to make it safely. However, the initial t_{go} calculation does not take this into account, and thus in the beginning a deceleration takes place (Fig. 9.52). Once the time-to-go estimate is more accurate a sudden increase in commanded acceleration occurs and thus an increase in velocity. A refinement of the guidance algorithm is required to avoid this unnecessary deceleration in the beginning of the final descent. The t_{go} remains nothing more as an instantaneous estimate that is a crucial input to the guidance algorithm. This parameter does not progress as quickly as "normal" time; the total time of flight until terminal gate is $t_f = 112.6$ s.

Last, but not least, the above simulations have been carried out with an unconstrained thrust force. Especially in the last simulation, the maximum commanded thrust exceeded by far the available thrust. Of course, for an on-board implementa-

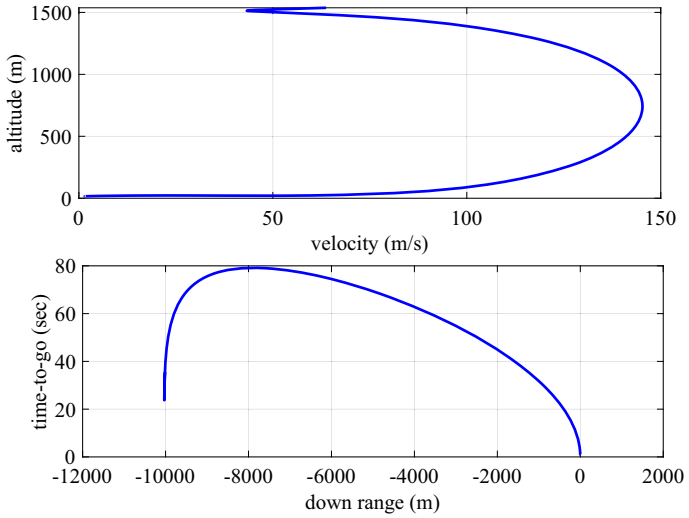


Fig. 9.52 Velocity and time-to-go for a long-range flight

tion the limits of the propulsion system need to be taken into consideration. Here, this is not pursued, as the example only meant to show the basic aspects of the algorithm. ■

9.3.4 Terminal Descent

The terminal descent is typically a vertical path to the surface, starting from a (very) low altitude. For Lunar landings such a descent could be incorporated to account for guidance and navigation errors. This last leg could be done in free fall, if the terminal velocity at the end of the guided phase would be below the accepted threshold, and calculation of the landing velocity is easily done using an energy balance (see Example 9.5).

Example 9.5 (*Apollo Lunar Module*) The Lunar Module (LM) used in the Apollo Moon landings was designed for a maximum (vertical) touchdown velocity of $V_{touchdown} = 5$ m/s. Suppose that the descent guidance is targeted towards a certain altitude with a final velocity of $V_f = 2$ m/s. At that moment the engine is shut off, and the LM will descent in a free fall to the surface. Given that the acceleration due to gravity at the Moon's surface is $g_M = 1.62$ m/s^2, what is the altitude h_f that the engine has to be shut off?

To solve this problem, we may use conservation of energy, since no external forces are at work and the mass of the vehicle is constant. Let state 1 be the state at the end of the guided phase, and state 2 the one at the surface; the surface is also considered

to be the reference plane. Total energy of the system, i.e., the sum of kinetic and potential energy (with respect to a flat Moon) is given by

$$\mathcal{E}_t = \frac{1}{2} m V_1^2 + m g_M h_f = \frac{1}{2} m V_2^2 + 0 \tag{9.143}$$

with $V_2 = 5$ m/s and $V_1 = V_f = 2$ m/s. Mass cancels out, so

$$h = \frac{1}{2 g_M} (V_2^2 - V_1^2) \tag{9.144}$$

which gives us $h_f = 6.48$ m.

Suppose now that the guidance system is not optimal and the switch-off moment is delayed and the thrust level is increased to compensate for this error. In that case, the vehicle not only decelerates, but starts accelerating again and has a velocity of $V_f = 2$ m/s *upwards* when it passes h_f for the second time. At that moment the engine is switched off. How does this effect the subsequent vertical descent?

The moment the engine is switched off, the vehicle moves upwards and converts kinetic into potential energy until the vehicle reaches a zero-velocity state. At that moment, it will start descending and when it reaches h_f again it will have the same velocity, but now directed downwards. Therefore, the altitude is not affected to enforce the 5 m/s touchdown velocity. This, of course, is easily confirmed by inspecting Eq. (9.144): the starting velocity V_1 appears in quadratic form, so the sign does not matter.

Example 9.6 (*Asteroid Approach*) During a future asteroid mission, a spacecraft is on an approach trajectory to asteroid's surface. Just prior to touchdown, the spacecraft is moving downward at a constant velocity relative to the surface, and has a downward thrust of 0.03 N. At that moment the guidance system commands a decrease in downward thrust, with a level of 0.016 N, and the onboard IMU determines that the acceleration of the spacecraft relative to the surface becomes $6 \cdot 10^{-6}$ m/s^2 downward. What is the gravitational acceleration of the asteroid near its surface?

In the initial phase, when the spacecraft has a constant downward velocity, the thrust balances the gravitational force:

$$0.016 - m g_0 = 0 \quad \Rightarrow \quad m g_0 = 0.016\text{N}$$

where m is the mass of the lander. After the change in thrust, this becomes

$$0.008 - m g_0 = m(-6 \cdot 10^{-6})\text{m/s}^2 \quad \Rightarrow \quad 0.008 - 0.016 = m(-6 \cdot 10^{-6})$$

This gives $m = 1{,}333$ kg and thus $g_0 = 1.2 \cdot 10^{-5}$ m/s^2. ■

In case of Mars landings, retro rockets are crucial to lower the velocity to an acceptable level. The Viking landers, for instance, used three terminal descent engines with a maximum thrust of about 2,700 N each to reduce the velocity from about 66 m/s

to about 2.44 m/s (Ingoldby 1978). The powered descent was initiated at an altitude of 1.49 km and lasted about 40 s, where the engines were throttled by the guidance system to meet with both velocity and altitude constraints. Also, the pitch and yaw orientation was controlled such that the flight path would be vertical close to the surface.

The model that we can use to analyse the (vertical) powered descent is a simplification of the model that we used earlier, i.e., Eqs. (9.65)–(9.69). This model assumes a flat, non-rotating planet, and in addition we impose the conditions of vertical flight ($\gamma = -90°$ or, equivalently, $\theta = 0°$) and the thrust acting vertically upwards. To be able to use this model also in an atmospheric environment, we should in principle also add a drag force, and, if we would allow for a (horizontal) wind, also a lift force. However, in the simplified analysis that we present here, we assume a steady atmosphere, and also that the drag force is (very) small compared to the thrust, and therefore $D \approx 0$ and $L = 0$. So, the model to be used is (see also Fig. 9.54):

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = D - mg + T \approx T - mg \tag{9.145}$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = 0 \tag{9.146}$$

$$\frac{\mathrm{d}h}{\mathrm{d}t} = -V \tag{9.147}$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{T}{I_{sp}g_E} \tag{9.148}$$

Note that in general the thrust consists of two components, i.e., the impulse thrust, equal to $\dot{m}c$, and (in the atmosphere) a pressure term, $A_e(p_e - p_0)$. In these expressions, $\dot{m}$ is the change of mass per unit time, and is equal to the mass expelled by the engine, c is this mass's velocity, the so-called exhaust velocity, A_e is the exhaust area, p_e is the exhaust pressure, and p_0 is the ambient pressure. Combining the two terms, we can write for the thrust

$$T = \dot{m}c + A_e(p_e - p_0) = \dot{m}c_{eff} \tag{9.149}$$

where the effective exhaust velocity, c_{eff}, has been introduced. Inspecting Eq. (9.148) and equating the relevant terms we find that (Fig. 9.53)

$$c_{eff} = I_{sp}g_E \tag{9.150}$$

In case a constant thrust deceleration $a_T = \frac{T}{m}$ is considered, Eq. (9.145) changes to $\frac{\mathrm{d}V}{\mathrm{d}t} = a_T - g$ and the solution is simply found:

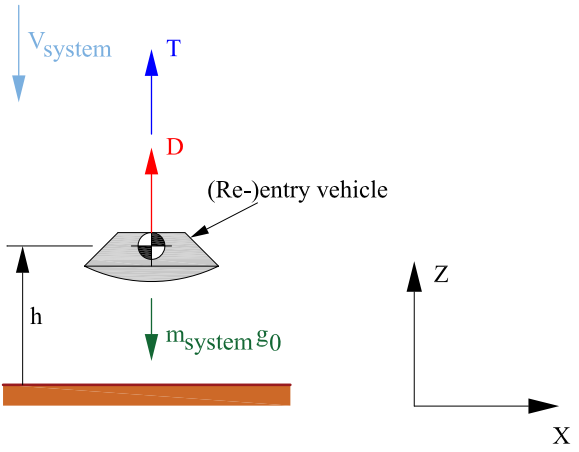


Fig. 9.53 Terminal-descent layout

$$V_f - V_0 = \int_0^{t_b} (a_T - g)\mathrm{d}t' = (a_T - g_0)t_b \tag{9.151}$$

In this equation, the final time t_f is equal to the burn time t_b, and close to the surface $g = g_0 \approx$ constant. The term $g_0 t_b$ is called the gravity loss. So, given the initial and final velocity, the burn time can easily be determined for a given thrust load:

$$t_b = \frac{V_f - V_0}{a_T - g_0} \tag{9.152}$$

The change in altitude follows from Eq. (9.147):

$$\mathrm{d}h = -V\,\mathrm{d}t = -(V_0 + a_T t - g_0 t)\mathrm{d}t \quad \Rightarrow \quad h = -V_0 t - \frac{1}{2}(a_T - g_0)t^2 \tag{9.153}$$

and thus for the total change in altitude for the complete burn: $h_b = -V_0 t_b - \frac{1}{2}(a_T - g_0)t_b^2$. The fuel mass can be estimated with Eq. (9.148):

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{T}{I_{sp} g_E} \quad \Rightarrow \quad \frac{\mathrm{d}m}{m} = -\frac{T}{m}\frac{1}{I_{sp} g_E}\mathrm{d}t = -\frac{a_T}{I_{sp} g_E}\mathrm{d}t \tag{9.154}$$

such that

$$\int_{m_0}^{m_f} \frac{\mathrm{d}m'}{m'} = -\frac{a_T}{I_{sp} g_E}\int_0^{t_b} \mathrm{d}t' \tag{9.155}$$

and

$$\ln\frac{m_f}{m_0} = -\frac{a_T}{I_{sp} g_E} t_b \quad \Rightarrow \quad \frac{m_f}{m_0} = e^{-\frac{a_T}{I_{sp} g_E} t_b} \quad \Rightarrow \tag{9.156}$$

$$m_f = m_0 - m_{fuel} = m_0 e^{-\frac{a_T}{I_{sp} g_E} t_b} \quad \Rightarrow \quad m_{fuel} = m_0 \left(1 - e^{-\frac{a_T}{I_{sp} g_E} t_b}\right) \tag{9.157}$$

Alternatively, for a system with constant thrust, the characteristic parameters can be found as follows. For a constant thrust the mass flow is constant, as given by Eq. (9.148), thus the current mass as a function of time can be determined through

$$m = m_0 + \dot{m} t \tag{9.158}$$

so the burn time is directly linked to the amount of fuel spent:

$$t_b = \frac{m_f - m_0}{\dot{m}} = \frac{I_{sp} g_E}{T} (m_0 - m_f) = \frac{I_{sp} g_E}{T} m_{fuel} \tag{9.159}$$

In the above equation, the used fuel is $m_{fuel} = m_0 - m_f$ (with $m_f = m_0 + \dot{m} t_b$ and $\dot{m} < 0$). Note that at this moment, both m_f and t_b are unknowns, unless all available fuel will be used. Equation (9.145) can be rewritten as

$$m \frac{\mathrm{d}V}{\mathrm{d}m} \dot{m} = (T - mg) \quad \Rightarrow \quad \mathrm{d}V = \frac{(T - mg)}{\dot{m}} \frac{\mathrm{d}m}{m} \tag{9.160}$$

Integrating yields

$$\int_{V_0}^{V} \mathrm{d}V' = \frac{1}{\dot{m}} \int_{m_0}^{m} \left(\frac{T}{m'} - g\right) \mathrm{d}m' \quad \Rightarrow \tag{9.161}$$

$$V - V_0 = \frac{T}{\dot{m}} \ln \frac{m}{m_0} - g \frac{m - m_0}{\dot{m}} = \frac{T}{\dot{m}} \ln \frac{m}{m_0} - gt \tag{9.162}$$

where we have substituted Eq. (9.158). For a required final velocity, V_f, this means that

$$V_f = V_0 + \frac{T}{\dot{m}} \ln \frac{m_f}{m_0} - g t_b \tag{9.163}$$

For the corresponding altitude, i.e., the moment when the engines have to be switched on

$$\mathrm{d}h = -V \mathrm{d}t = \left(-V_0 - \frac{T}{\dot{m}} \ln \frac{m}{m_0} + gt\right) \mathrm{d}t \tag{9.164}$$

so

$$\int_{h_b}^{0} \mathrm{d}h = -V_0 \int_{0}^{t_b} \mathrm{d}t' + \frac{T}{\dot{m}^2} \int_{m_0}^{m_f} \ln \frac{m'}{m_0} \mathrm{d}m' + \int_{0}^{t_b} g t' \mathrm{d}t' \tag{9.165}$$

The second integral on the right-hand side is rewritten as $\int_{m_0}^{m_f} (\ln m' - \ln m_0) \mathrm{d}m'$, which integrates into[22]:

$$\int_{m_0}^{m_f} (\ln m' - \ln m_0) \mathrm{d}m' = (m_0 - m_f) + m_f \ln \frac{m_f}{m_0}$$

Thus, the complete integration results in

$$h_b = -V_0 t_b + \frac{T}{\dot{m}^2} \left[(m_0 - m_f) + m_f \ln \frac{m_f}{m_0} \right] + \frac{1}{2} g t_b^2 \tag{9.166}$$

Finally, as we mentioned earlier, both t_b and m_f are not known beforehand, so without some form of guidance system like E-guidance we cannot readily solve this. The simplest approach for the current discussion is to do this iteratively.

The procedure begins with initialising m_f with the dry mass of the vehicle, thus assuming that all fuel is used, and V_f with V_0. Then, calculate t_b with Eq. (9.159) and update m_f with Eq. (9.163), given the current V_f. For the new final mass we then update t_b and V_f, and we continue this until convergence of V_f. For an accuracy of 0.001 m/s, we found that 10 iteration steps were required. Once t_b and m_f have been established, h_b can be calculated with Eq. (9.163).

Example 9.7 (*Terminal Descent on Mars*) Let us consider a Viking-like lander, with a total available thrust of $T = 8{,}100$ N and $I_{sp} = 260$ s. The mass at terminal-descent initiation is $m_0 = 660$ kg (dry mass $m_{dry} = 576$ kg and $m_{fuel} = 84$ kg). In this example we will look at vertical descent only, and compare the analytical solutions with those obtained with a numerical simulation. Two cases are considered, i.e., the one with a constant thrust load and one with constant thrust.

For the constant thrust load, the initial thrust deceleration is $a_{T,0} = \frac{T}{m_0} = 8{,}100/660 = 12.27$ m/s^2. This value is close to that of the Viking landers, which were designed to decelerate with a maximum of one Earth g. This means that when compensating for Mars' gravity the total thrust deceleration would be $a_{T,Viking} = 9.81 + 3.71 = 13.52$ m/s^2. Using this thrust load $a_{T,0}$, the following analysis can be done.

Given an initial and final velocity of $V_0 = -66$ m/s and $V_f = -2.44$ m/s (note that the minus sign is introduced to be in line with the frame definition shown in Fig. 9.54), the burn time, fuel-mass consumption and covered altitude range follow from Eqs. (9.152), (9.153) and (9.157), respectively: $t_b = 7.42$ s, $h_b = 254.0$ m, and $m_{fuel} = 20.12$ kg.[23]

[22] Here, we use the solution to the standard integral $\int \ln x \mathrm{d}x = x \ln x - x$.

[23] The Viking landers started the terminal descent at about 1,500 m altitude, and used about 65 kg fuel during an engine burn of about 40 s. However, the engines were throttled, such that on average the deceleration was smaller, and also the horizontal motion had to be controlled. No specific information could be found on the actual trajectory flown. Note that enforcing a burn time of 40 s

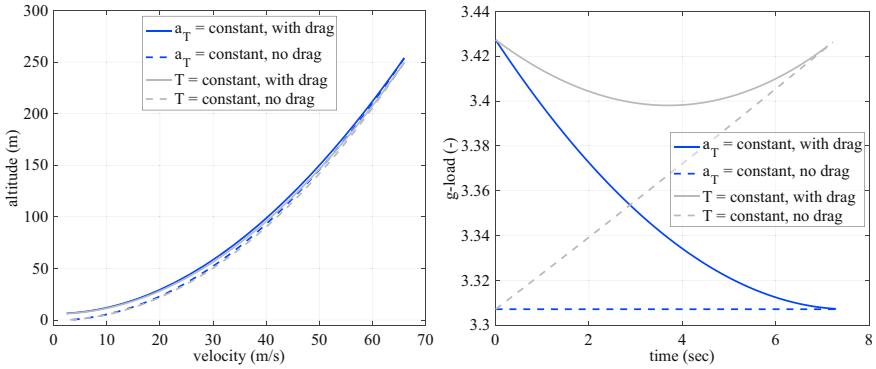


Fig. 9.54 Terminal-descent results (numerical simulations)

In case the lander descends with constant thrust, using the iterative procedure discussed above the corresponding parameters are $t_b = 7.24$ s, $m_f = 637.0$ kg, so $m_{fuel} = 23.0$ kg, and $h_b = 249.6$ m. As can be seen, slightly more fuel is used, because the engines are not throttled back.

To compare the analytical results with those obtained with a numerical simulation we will take V_0, m_0, and h_b as initial condition, and look at the effect of including a drag force as well ($C_D = 1.2$, $S_{ref} = 5.76$ m^2). The model used is that of a spherical, non-rotating Mars, with a central gravity field (but with the same surface value of $g_0 = 3.711$ m/s^2). Stop criterion for all simulations is either a zero altitude, or reaching the required final velocity. The result of the four simulations (two for constant thrust load and two for constant thrust) is shown in Fig. 9.54.

At first sight, the trajectories do not seem to be so much different, which in a way is true. Comparing either the two solid lines or the two dashed lines in the $h - V$-diagram, shows that either a constant thrust-load or a constant-thrust trajectory will give closely-related results. This is also confirmed by the characteristic values, listed in Table 9.4. The inclusion of drag gives a slightly larger difference. The *g-load*, referenced to a Martian g_0, is a bit more interesting. The two curves with drag included begin at a higher value, as at $t = 0$, when the velocity is 66 m/s, the drag is at its maximum ($D = 294$ N, which is about 3.6% of the thrust force). However, in absolute sense the difference with the case when drag is ignored is quite small, so it seems that it would be justified to ignore drag in the analysis of terminal descent. In case the thrust is kept constant, the decrease in mass will give a linear increase in *g-load* when drag is ignored, but otherwise the decreasing drag force dominates this increase and only in the second half of the short flight the *g-load* increases again. The effect of drag is most clear in the case of constant thrust load, and shows as a quadratic effect. With decreasing velocity the gradient of the *g-load* is slowly approaching zero.

---

gives $a_T = 5.3$ m/s^2, $h_b = 1{,}368.8$ m, and $m_{fuel} = 52.7$ kg. It may be clear that extending the burn duration will also increase the gravity loss, and thus the required fuel mass.

Table 9.4 Summary of terminal-descent characteristics

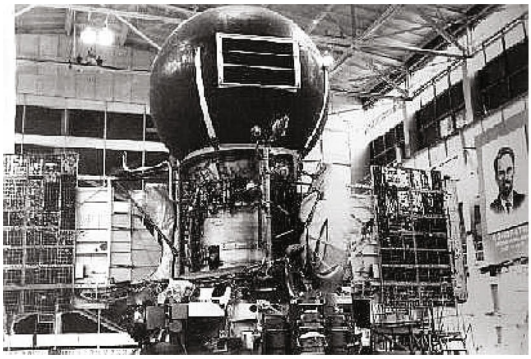
Run	t_b (s)	h_f (m)	V_f (m/s)	m_{fuel} (kg)
a_T = c, analytical	7.42	0.00	2.44	23.17
T = c, analytical	7.24	0.00	2.44	23.00
a_T = c, numerical, no drag	7.41	0.00	2.55	23.13
a_T = c, numerical, with drag	7.29	6.46	2.44	22.76
T = c, numerical, no drag	7.23	0.00	2.56	22.95
T = c, numerical, with drag	7.12	6.20	2.43	22.60

Looking in some more detail at the characteristic values (Table 9.4), the two numerical cases when drag is included reach the terminal velocity above the surface, due to the integrated decelerating effect of D. This is something to take into consideration. But, with an actual guidance system this should not be a problem, as the difference is small.

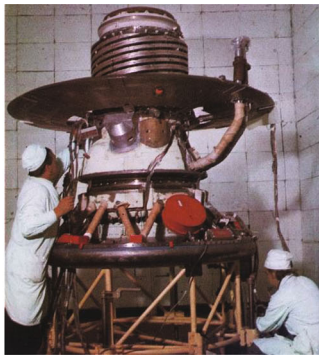
A final remark should be made about the duration of the burn. Suppose we would simulate the earlier mentioned Viking-case with $a_T = 5.3$ m/s^2 (which represents a net deceleration of 1.59 m/s^2) and $t_b = 40$ s, starting from $h_0 = 1{,}368.8$ m. The numerical simulation (without drag) yields $t_b = 39.97$ s, $h_f = 1.29$ m, $V_f = 2.44$ m/s and $m_{fuel} = 52.62$ kg, which seems close enough. However, including drag shows a larger difference: $t_b = 36.73$ s, $h_f = 157.30$ m, $V_f = 2.44$ m/s and $m_{fuel} = 48.52$ kg. Especially the large difference in final altitude is worrying. The much longer flight through the atmosphere gives a significant integrated contribution of the drag force, meaning that the terminal velocity is reached sooner. For comparison, a(n analytical) thrust load of $a_T = 6.89$ m/s^2 with a burn time of $t_b = 20$ s, requires a starting altitude of $h_0 = 684.2$ m. The corresponding results from the numerical simulation are $t_b = 19.99$ s, $h_f = 0.18$ m, $V_f = 2.44$ m/s, and $m_{fuel} = 34.71$ kg for the no-drag case, and $t_b = 19.10$ s, $h_f = 43.81$ m, $V_f = 2.44$ m/s, and $m_{fuel} = 33.19$ kg for the simulation with drag. Even though the "error" in final altitude has decreased, it may be clear that the difference is still too large. Therefore, the analytical model should only be used for a very first estimate, or for situations where the burn time of the engines is relatively short. ■

9.4 Other Decelerators

Even though the main decelerator systems are formed by parachutes and propulsion systems, over the years several other decelerators have been proposed and operated. In this section we will briefly introduce a few of them, most notably the drag skirt, the ballute, and the inflatable aerodynamic decelerator (IAD). The ballute is, in fact,



(a) Venera-9 Orbiter



(b) Venera-11 Lander

Fig. 9.55 The Russian Venera orbiter, with on top the spherical entry capsule, and lander with drag skirt. Images: NPO Lavochkin/NASA

also an IAD, but since it was the first one in a long line of developments it will serve as an introduction to the broader family of IADs.

9.4.1 Drag Skirt

The Russian Venera program was a very successful planetary exploration program of Venus, and began in 1961 with the first flyby of Venera-1, followed in 1965 with the flyby of Venera-2. From Venera-3 onwards, a small entry probe descended in the atmosphere, extending its descent time by parachute. However, from Venera-9 the design of the entry probes changed, and a much larger lander was designed. Because of increased understanding of the (thick) atmosphere and to make sure that the lander would reach the surface as soon as possible, the probes would only descend by parachute to an altitude of about 50 km. The Venusian atmosphere is so thick that a simple flat plate mounted on top of the lander was sufficient to guarantee a stable descent to the surface. This disk-like aerobrake is also called a *drag skirt*, see Fig. 9.55. The sharp edge of the drag plate would lead to a stable flow separation, and thus minimising the attitude excursions during the descent. This new disk brake was developed and tested in 1973–74 in wind tunnels, and was also dropped from helicopters and aircraft from altitudes up to 14,000 m (Harvey 2007a).

9.4.2 Ballute

In the late 1950s and early 1960s the use of parachute for supersonic applications was not a reliable solution, due to the instability of these parachutes at supersonic

speeds. This stimulated the search for alternative means, and NASA and the Air Force began researching and developing trailing-edge decelerators for launch vehicle and entry vehicle recovery. These decelerators were the first generation of the so-called inflatable aerodynamic decelerators, and the key characteristic is that it has an *enclosed* volume, inflated by either a gas generator, by air forced into the structure by ram air inlets, or a combination of both. Note that by this definition, a parachute does *not* qualify to be such an inflatable decelerator, as it is open on one side.

The initial concepts focussed on simple geometries, like a cone or a sphere, and one aimed at characterising their aerodynamic properties in extensive wind-tunnel campaigns (McShera 1963). Later geometries evolved to more structurally optimal shapes, the prime example being the so-called *isotensoid*, see also Fig. 9.56. An eye-catching detail is the burble fence that is an inflated toroidal structure on the ballute, intended to ensure flow separation. This stabilises the ballute as it decelerates through different flow regimes (from supersonic to subsonic). The isotensoid was initially designed by the Goodyear Aircraft Corporation (Nebiker 1961), and from theory it would create a stress state that is equal in both radial and circumferential directions. However, due to the induced drag and the burble fence, some stress concentrations will appear, but overall the resulting geometry will be relatively low stress. This particular trailing isotensoid concept was name *ballute* by Goodyear, its name is formed by a combination of "balloon" and "parachute", and can be considered to be a parachute braking device that is optimised for use at high altitudes and high supersonic velocities.

Goodyear continued working on ballutes, amongst others through the *aerodynamic deployable decelerator performance evaluation program* (ADDPEP), which matured the concept significantly. The program included significant analysis, materials research, supersonic wind-tunnel testing, and multiple sounding rocket flights of 1.5 m diameter test articles.

After the initial development, there were quite some applications of the ballute, either as a supersonic decelerator or as a stabilisation device. Examples are the meteorological sounding rocket decelerator device, the Gemini ejection seat stabilisation, and it was proposed as a pilot for the Viking missions by Martin Marietta (Cruz et al. 2006). The application of Gemini ejection seat stabilisation is shown in Fig. 9.57. The ballute was deployed from the ejection seat back pocket and inflated through ram air inlets, and could stabilise the astronaut (in their seat) and provide initial deceleration at high velocity and low density. Once the astronaut had reached lower speeds the ballute would be jettisoned and a conventional recovery parachute was to be deployed by a drogue gun and pilot chute, and used for landing. NASA went as far as drop testing the system with dummies, and initial tests revealed several problems with the ram-air inflation at subsonic speeds, whereas at supersonic speeds the ballute would often fall. After a redesign of the system to ensure more positive inflation at very low dynamic pressures, a series of very successful dummy and test-pilot drops from between about 4 and 12 km was performed (Ray and Burns 1967; Grimwood et al. 1969).

Besides their military application, i.e., the ballute has been used on free-fall bombs dropped from an aircraft, helping to both decelerate and stabilise the descent, it has

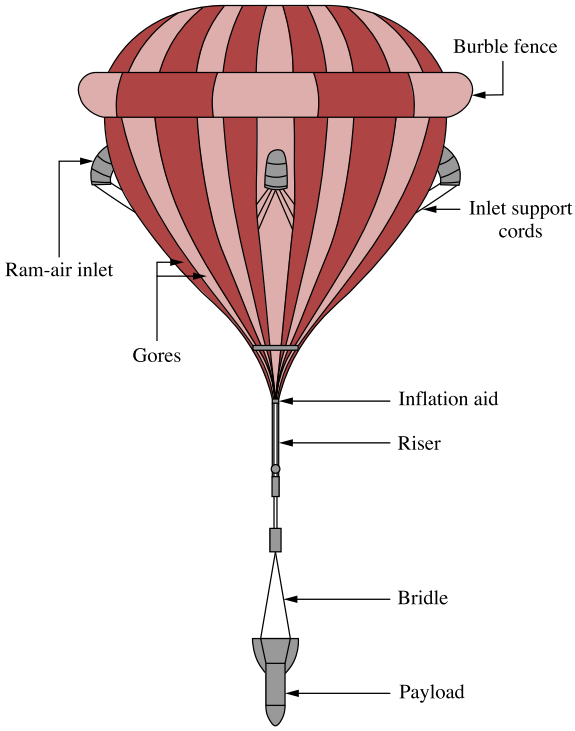
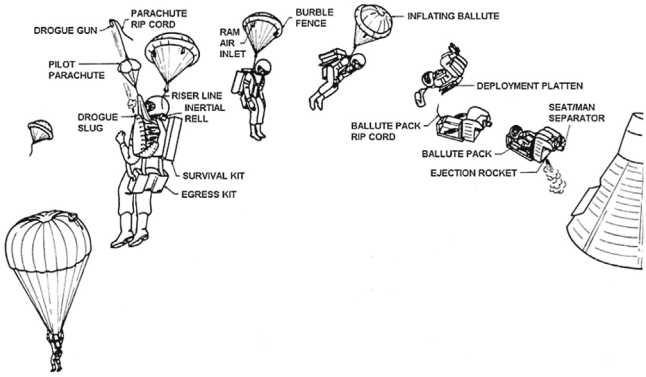


Fig. 9.56 The trailing-edge isotensoid ballute

also seen many aerospace applications, albeit with a break of several decades (Fallon II 1995). Due to the advances in design and construction technology, related to burble-fence and air-inlets design, materials and general construction, the superior aerodynamic performance in the supersonic regime has made the ballute a serious competitor of supersonic parachutes, even though traditionally the disadvantages, being bulky, heavy and costly, have been clear.

The most recent successful use of a ballute was on NASA's low-density supersonic-decelerator (LDSD) project as a pilot device to deploy the 30.5 m diameter supersonic parachute during the supersonic flight dynamics test (SFDT) (Clark et al. 2015). Even though not the primary objective of the tests, ballute deployment and operation was successful. The ballute, a trailing isotensoid with a diameter of 4.4 m and consisting of 16 gores, was mortar deployed at a Mach number of 2.6 and inflated at line stretch in 0.5 s (Fig. 9.58). During the first test, the ballute performed above expectations in both measured drag and observed stability.



(a) Schematic



(b) Dummy test

Fig. 9.57 The Gemini escape system, using a ballute. Images: Goodyear/NASA

Following the experience gained in the LDSD project, Brandeau et al. (2016) discuss the application of ballutes for supersonic deceleration at Mars. With the increase of payload masses to be delivered to the Martian surface, they give several reasons why the use of ballutes is favourable. Beginning with supersonic deceleration prior to parachute deployment and use as a pilot device for deploying main parachutes, ballutes will increase stability as a drogue device and it will give mass savings compared to a single mortar-fired main parachute. And finally, for those vehicles with a high ballistic coefficient that cannot deploy a supersonic parachute, using a ballute as aerodynamic decelerator may reduce the mass of the terminal-descent system.

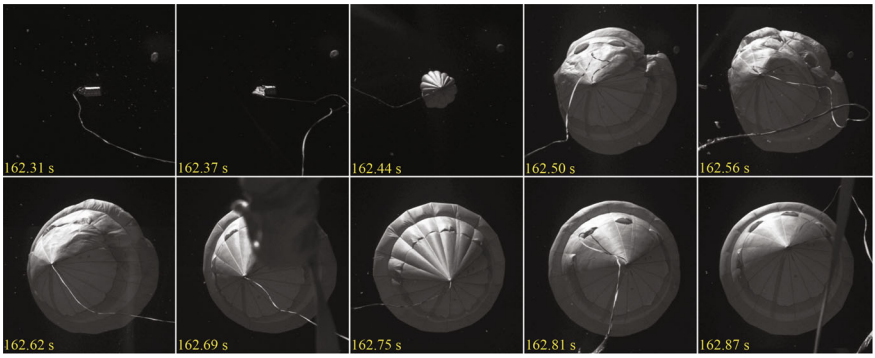


Fig. 9.58 Ballute-inflation sequence (LDSD project), from just before line stretch to full inflation. The reconstructed mission time (in seconds) is indicated in the images. Image: NASA, taken from Clark et al. (2015)

9.4.3 Inflatable Aerodynamic Decelerator

Whereas the ballute discussed in the previous section is a trailing-edge inflatable decelerator, the larger family of inflatable aerodynamic decelerators consists of many different shapes and attachments, of which the ballute is only one. Common to all IADs is that they are inflatable devices designed to greatly increase drag of, for instance, an entry vehicle. As we discussed before, its shape is maintained by a mostly closed gas-pressurised, three-dimensional body, which is inflated by an internal gas-generating source, ram-air, or a combination of both.

Development and testing of IADs began in the late fifties to early sixties, but it was not until the mid-seventies that IADs reached their peak technology-readiness. This was related to the Viking, Pioneer Venus, and Galileo missions, which would need deployable decelerators during atmospheric descent. In parallel, the disk-gap band parachutes were under development, but these were having inflation and oscillation issues at Mach numbers larger than two. Wind-tunnel tests had proven the supersonic capabilities of IADs and confirmed an operating range beyond the DGB performance thresholds. However, when it became clear that Viking did not require a parachute deployment above the Mach-2 threshold, and that the Pioneer Venus and Galileo missions were possible with Earth-based parachute technology, further development to mature the IAD technology was halted and thus many IAD design issues remained unsolved (Cruz et al. 2006; Smith et al. 2010).

To give an indication of how well the IAD development was progressing, Fig. 9.59 shows the deployment range of IADs obtained from successful flight tests. It shows that most tests happened at much larger dynamic pressures than the DGB boundaries set for Mars Science Laboratory. Also the Mach number at deployment during most notably the airforce-sponsored ADDPEP, increased significantly, from Mach 2 to a very impressive Mach 10. This program was executed using trailing isotensoids,

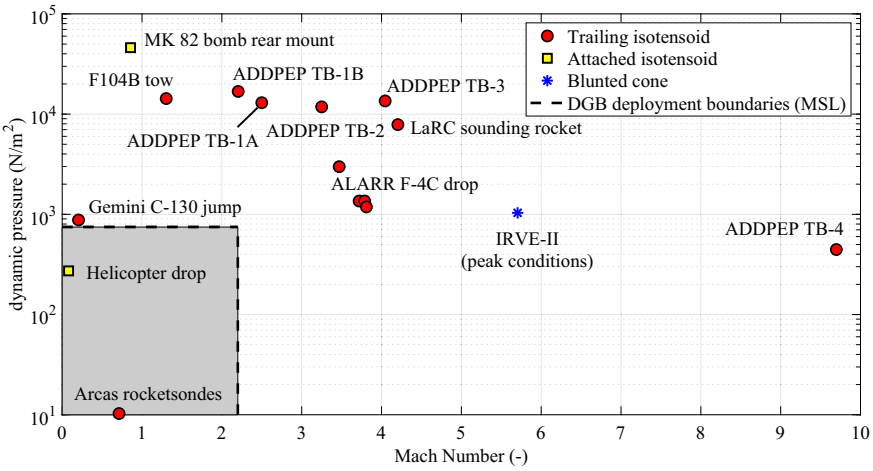


Fig. 9.59 Deployment range for inflatable aerodynamic decelerators. Data taken from Smith et al. (2010)

for most of the tests. The few tests done with attached isotensoids were much more limited in operational range.

Nonetheless, dozens of IAD geometries have been tested in wind tunnels and even more conceived in system studies. An overview of the most common types is shown in Fig. 9.60. All trailing IADs are detached from their payloads and connected through a riser and possibly bridles, as has also been depicted in Fig. 9.56. Trailing IADs must operate in the wake of the the leading body, which is a disturbed flow at a lower Mach number than the free-stream Mach number. Consequently, the IAD bow shock is weaker and thus also the induced drag is lower. This could lead to a larger and heavier design. The wake effect may also cause instabilities and a large trailing distance would be required for efficient and stable operation. On the plus side, the design and performance (analysis) of a trailing IAD can be decoupled from that of the entry vehicle. Also, when the IAD is no longer needed it can easily be separated. The attached IADs, on the other hand, show a more favourable ratio between drag area and mass. Countering this, though, is the required integrated design approach with the vehicle itself.

In terms of operational applicability, the IADs can be seen as an alternative to (supersonic) parachutes or as (hypersonic) aerodynamic decelerator. Thus, the two types are the SIAD (supersonic IAD) and HIAD (hypersonic IAD). The SIADs are inflated at supersonic speeds ($M \leq 5$) at moderate to high dynamic pressures, and can use an internal gas source, ram-air, or a combination of both for inflation. Due to the (relatively) low velocities, they must be capable of sustaining moderate levels of aerodynamic heating. The category of HIADs include those that are inflated at hypersonic speeds ($M > 5$) or even outside the atmosphere (exo-atmospherically).

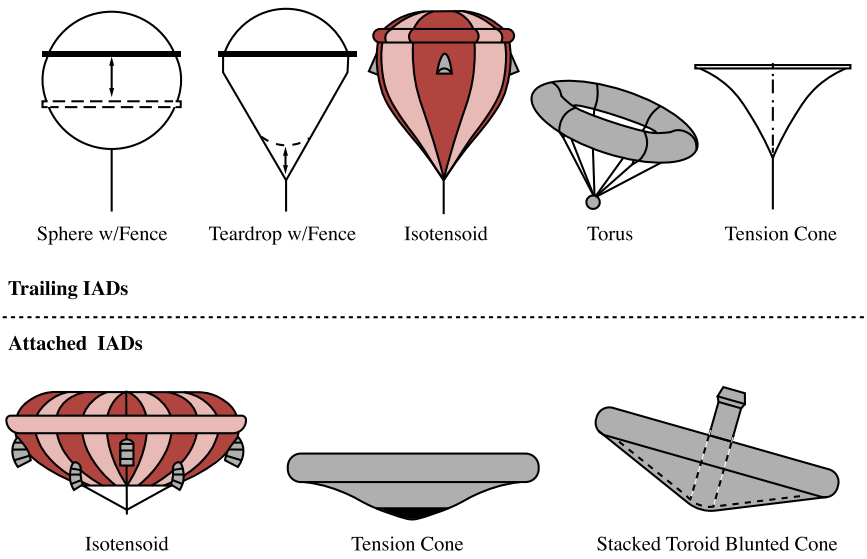


Fig. 9.60 Examples of feasible trailing and attached inflatable aerodynamic decelerators

Thus, they can be used for aerocapture or entry to landing, but must be able to survive high heating.

The trade-off between an SIAD and parachutes should include the following considerations. Advantages of SIADs are deployment Mach numbers up to $M \approx 5$ and having a large drag coefficient at supersonic Mach numbers. Disadvantages, on the other hand, are the relatively large mass per unit frontal area as compared to parachutes and the lower effectiveness, i.e., the ratio between drag area and mass, at subsonic speeds. Working in favour of parachutes is that they are flight qualified and a proven technology for applications up to $M \approx 2.2$, and that their drag-area-to-mass ratio at subsonic speeds is large. The downside is that they show a rapid reduction in C_D for $M > 1.5$ as well as becoming much more unstable. Finally, at the low atmospheric densities encountered in planetary missions, the upper deployment Mach number is about 3.

When HIADs are compared with the rigid aeroshells with a thermal protection system, the latter are based on proven technology and the analysis tools have reasonably matured. Disadvantages are that the maximum diameter and area are constrained by the launcher fairing and the heat-shield's mass is large for a given ballistic parameter. For the HIAD, the advantages and disadvantages are more or less reversed. The required technology is largely unproven, which means a low technology readiness level, and from the structural-design side, aeroelasticity is a major issue. The launcher fairing, on the other hand, hardly puts a constraint on the diameter and area, and because of this (and the relatively low mass), very low values for ballistic parameter can be achieved. Last but not least, for some missions the HIAD may be the only option to boost the performance in terms of increase in useful orbit (aerocapture) or

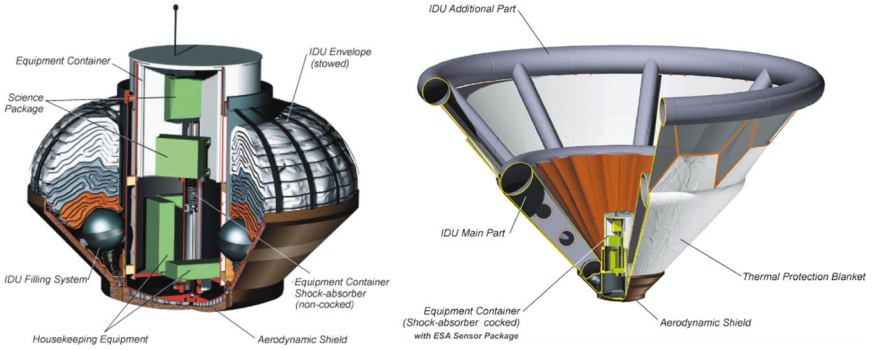
the landed mass. More details on the use of IADs and/or parachutes for planetary missions can be found in, for instance, the papers by Fallon II (1995), Cruz et al. (2006), Rohrschneider and Braun (2007), and Smith et al. (2010).

Interest in the IAD has renewed in recent years from numerous atmospheric entry trajectory analyses at Earth, Mars, Titan, and Neptune calling for deployable decelerator operations outside of the DGB performance envelope. But it was not limited to just studies. Since 2000, several flight-test programs have been initiated and carried out. We will briefly introduce the most important ones.

On 9 February 2000, the inflatable re-entry and descent technology (IRDT) demonstrator was launched on the first qualification flight of the Soyuz Fregat launcher (Marraffa et al. 2000). Officially an ESA project, the IRDT project was executed in close collaboration with Astrium-I (formerly DASA), NPO Lavochkin and the European Commission, and aimed at developing new re-entry technology. After entry into the Earth's atmosphere, the demonstrator used an inflatable aeroshell to land without the aid of a parachute or a conventional heat shield. The 122 kg demonstrator saw a very short development time and launch of under one year, at a very low budget of less than 2 M€, because of Lavochkin's expertise in inflatable heat shields, developed for the Mars-96 project.[24] After entry, the flexible entry shield was deployed, increasing the diameter from 0.8 to 2.3 m (see also Fig. 9.61 for an overview of both the stowed and the deployed configuration). This entry shield was covered by an insulating layer (multi-layer insulation, MLI) protected by a silica-based fabric, impregnated with an ablative material. Unfortunately, the second envelope that was to increase the diameter to 3.8 m and effectively act as the parachute system, failed to deploy, and the impact velocity was about 60 m/s, rather than the designed 13–15 m/s. Despite the incurred damage, the flight data from the on-board computers and the sensor package could be retrieved, and also the post-flight hardware inspection yielded valuable results.

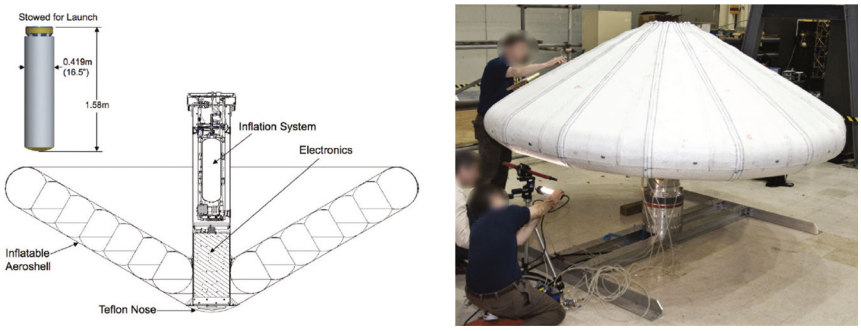
In 2004 NASA began the *inflatable reentry vehicle experiment* (IRVE), a mission to demonstrate that an inflatable structure can be used as a heat shield through the hypersonic flight regime, all the way down (Hughes et al. 2005). The experiments concerned a stacked toroid blunted cone HIAD geometry (Fig. 9.62). The first flight, launched in 2007, suffered a flight mishap while separating from the sounding rocket. The next IRVE-2, with a deployed diameter of 3 m and fitted into a 38 cm payload container, was launched on 17 August 2009 with a Terrier-Black Brant 9 sounding rocket from NASA's Wallops Flight Facility on Wallops Island, Virginia. Outside the atmosphere, the aeroshell was successfully inflated (with nitrogen) and entered the atmosphere unguided. O'Keefe and Bose (2010) discuss the post-flight analysis and confirmed that the IRVE-II flight dynamics were in line with expectations. Overall, the test flight confirmed the feasibility of using HIADs in terms of inflating

[24] Mars-96 was a failed, Russian Mars mission, launched in 1996. Failure of the second fourth-stage burn prevented the probe assembly from leaving Earth orbit, which caused it to re-enter the Earth's atmosphere, breaking up over a 200-mile long portion of the Pacific Ocean, Chile, and Bolivia. Besides an orbiter, it also consisted of two surface stations and two penetrators. Whereas the surface stations would land by parachute, part of deceleration of the penetrators would have been by inflation of a braking device.



(a) Stowed (b) Deployed

Fig. 9.61 ESA's inflatable re-entry and descent technology (IRDT) demonstrator. Images: ESA

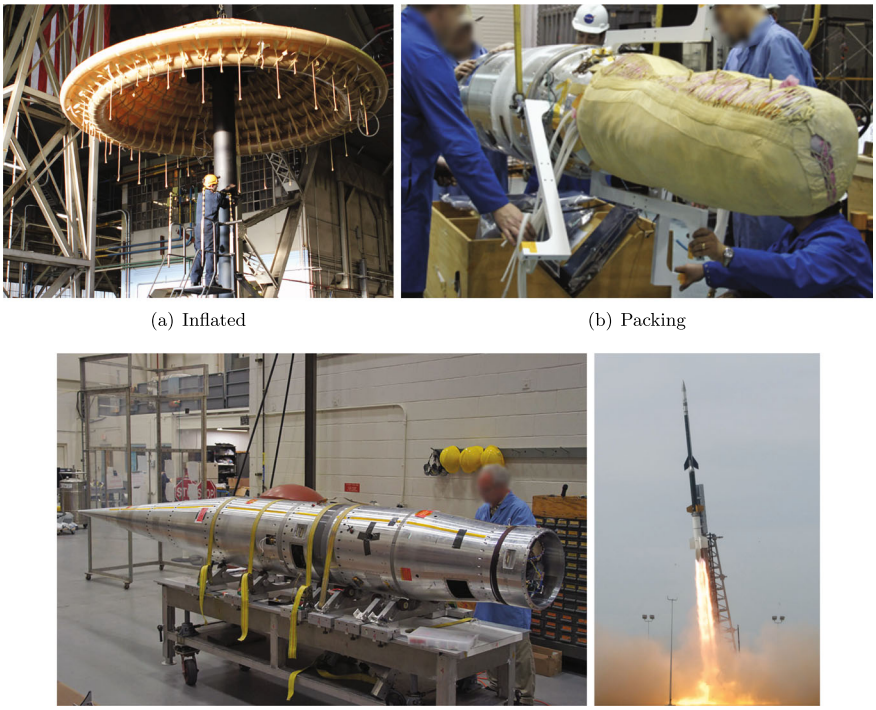


(a) Schematic (b) Flight model

Fig. 9.62 Inflatable reentry vehicle experiment. Images: NASA

the aeroshell exo-atmospherically, maintaining pressure during the peak dynamic pressure, and demonstrating that a re-entry heating of 20 kW/m^2.

The next vehicle to be tested was IRVE-3 (Fig. 9.63), which was launched July 23, 2012 on a Black Brant XI suborbital sounding rocket, yet again from NASA Wallops Flight Facility. With a mass of 281 kg it was heavier than its predecessor, but it had the same geometrical shape of 3 m diameter and a 60° sphere-cone aeroshell. The nose cone with its payload was separated at 176 km altitude, 102 s after launch. After a coasting phase of 262 s it reached its apogee of 469 km. Inflation of the aeroshell started another 72 s later. After an attitude reorientation and actuation of the centre-of-mass offset system (such that the L/D could be controlled), the vehicle reached the atmospheric interface, defined to be at 25 N/m^2 (664 s after launch, altitude 85 km), and the re-entry experiment could begin. The maximum heat flux experienced was 144 kW/m^2, resulting in a maximum temperature of about 1000 °C, whereas the re-entry deceleration reached 20.2 g, both of which the TPS and inflatable structure handled as planned. Forty three seconds later, when the vehicle had



(a) Inflated (b) Packing

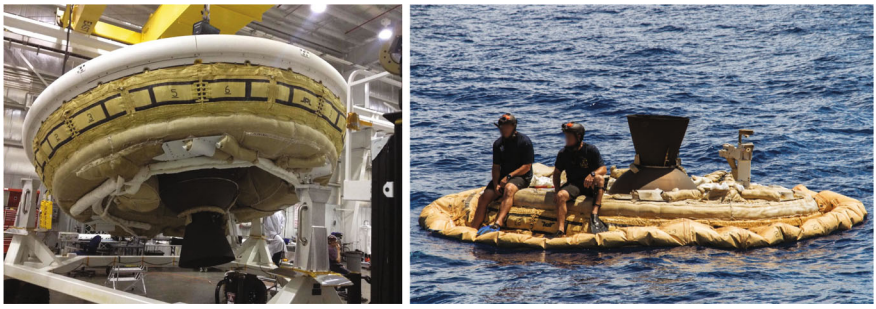
(c) Experiment section (d) Launch

Fig. 9.63 IRVE-3 test on July 22, 2012. Images: NASA

descended to 28 km (and reached a Mach number of 0.7) the re-entry experiment had officially ended. Splashdown in the ocean occurred at 447 km downrange. Recovery of the vehicle failed, because unfortunately the re-entry vehicle had sunk before the recovery boat was able to reach it. Despite that, the flight experiment was a success (Dillman et al. 2013).

NASA was planning to follow this up with three more vehicles, all with expanding flight envelopes and to test increasingly-larger inflatable heat shields with other types of heat-resistant fabrics (Litton et al. 2011). Unfortunately, the successor missions were all cancelled/delayed. IRVE-4, an almost copy of the IRVE-3 vehicle, would focus on characterising HIAD response to control inputs through more elaborate centre-of-mass control, in preparation of the High Energy Atmospheric Re-entry Test (HEART). HEART was planned to enter the atmosphere at orbital velocities, demonstrating the whole system for a relevant-scale HIAD in relevant environments. With a scale of 8–9 m diameter, HEART would be ballistic (unguided) and provide proof-of-concept that the HIAD can perform on this scale. HEART-2, finally, would then have the ability to control the HIAD to land at a targeted location.

The second last project to discuss is the low-density supersonic decelerator (LDSD), a re-entry vehicle designed to test various techniques for atmospheric entry



(a) Vehicle (b) Recovery

Fig. 9.64 The low-density supersonic decelerator (LDSD). Images: NASA

on Mars (Clark et al. 2015; Brandeau et al. 2016). The goal of the project was to develop a reentry system capable of landing 2-3-ton payloads on Mars, as opposed to the 1-ton limit of the currently used systems (such as MSL and Mars2020). The disc-shaped LDSD uses an inflatable structure, i.e., a supersonic inflatable aerodynamic decelerator (SIAD), which is essentially a donut-shaped balloon (Fig. 9.64a), to create atmospheric drag to decelerate the vehicle before deploying a large supersonic parachute. Launch mass of the system was 3120 kg, and the total project cost was US$ 230 million.

The first test flight took place on June 28, 2014, see Fig. 9.65 for a visualisation of the mission. A high-altitude helium balloon took the vehicle to around 37 km, after which the balloon separated and dropped back to be recovered. Four small, solid-fuel rocket motors spun up the vehicle to provide stability. Next, the vehicle's Star 48B solid-fuel motor powered it to Mach 4, at which moment the test period began. The SIAD de-spun and was then inflated. This inflation occurred nominally in 0.4–0.6 s. At Mach 2.7 (dynamic pressure 500 N/m^2), the descent system was activated by deploying a ballute in a blunt-body wake, acting as a parachute-deployment pilot device, see also Fig. 9.58. Creating an 8000-N drag force, it slowed down the system to Mach 2.35 (around 107 s after SIAD deployment), and the supersonic disk-sail (SSDS) parachute was deployed. This parachute had a diameter of 33.5 m (compared to the MSL disk-gap band parachute with a diameter of about 21 m, this is very large for a supersonic parachute!). Unfortunately, deployment did not go as planned, as it began tearing apart after deployment. As a result, the vehicle impacted the Pacific Ocean, but fortunately all hardware and data were recovered (Fig. 9.64b). A second test followed on June 8, 2015, but also during this test, the parachute tore apart. A third test was planned for 2016, but in the end this one was cancelled.

Despite the failing test of the disk-sail parachutes the test flights were considered to be very successful, amongst others, because the SIAD was the largest blunt-body aeroshell ever that was flown supersonically. The entry and descent had caused no observed aerothermal damage or degradation, and these flight tests have raised the technology-readiness level to TRL-6. Concerning the use of the ballute, it was

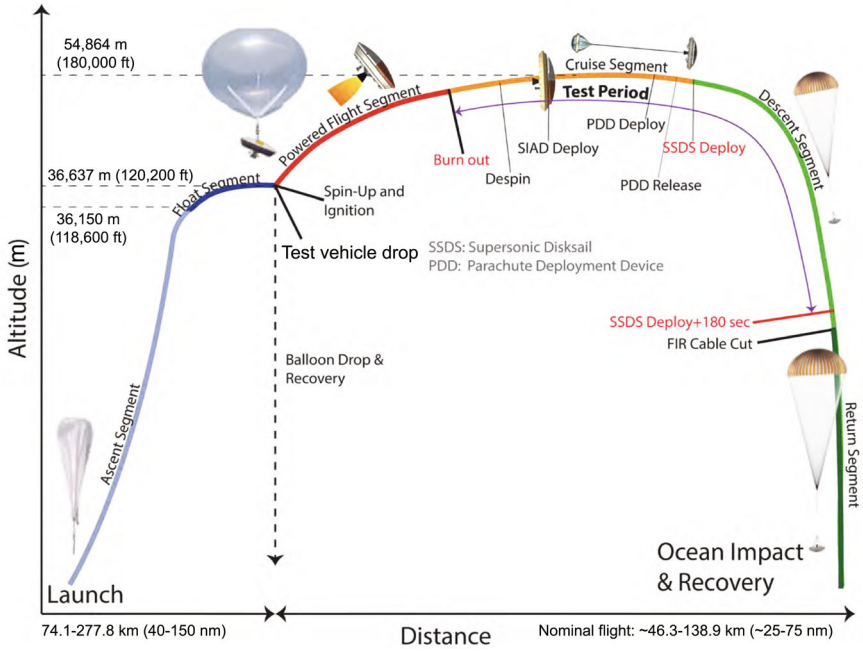


Fig. 9.65 The low-density supersonic decelerator (LDSD) mission profile. Adapted from original image by NASA

the largest one ever successfully flown at supersonic conditions, and was at the same time also the first-ever supersonic pilot deployment. The parachute itself, even though it was torn, was the largest supersonic parachute ever deployed, and all sensor (including visual) data collected gave a wealth of information for further supersonic parachute development.

Finally, the LDSD project was followed up by the Low-Earth Orbit Flight Test of an Inflatable Decelerator (LOFTID),[25] a project aimed at testing inflatable aeroshells to support heavy-payload delivery (experiments, equipment, and people) into those atmospheres of Venus, Mars, and Titan, as well as return to Earth. The launch on November 10, 2022, successfully showed that after inflation the 6-m-diameter aeroshell could enter at about Mach 25, survive the 1,600 °C peak temperature, and slow down to subsonic speeds. The next step in the program is to design, build, and test a 12-m diameter inflatable aeroshell.

To finalise this section, we would like to mention the paper by Rohrschneider and Braun (2007), who discuss the application of IADs to aerocapture. They provide an overview of the development and the tests involved, and indicate the potential mass savings that can be achieved using the HIADs. They highlight the challenges, and

[25] https://www.nasa.gov/mission/low-earth-orbit-flight-test-of-an-inflatable-decelerator-loftid/, date accessed: December 27, 2023.

discuss what would be required in the coming years. A similar overview is found in an earlier paper; Hall (2000) goes into quite some detail of the technological challenges and other issues to be addressed to bring IAD development to the next level of maturity:

1. *Optimal IAD shape.* The shape of the IAD (sphere, disk, toroid, or any other) will have an impact on system mass and aerodynamic characteristics, thus at the very least the interaction with the mission shall be analysed. Even though many shapes have been studied in the past, they should be revisited once the mission has been selected.
2. *Survivability of the IAD material.* The induced drag and heating will to a large extent determine the thickness and selection of the membrane material. Expanded material testing is needed, and aero-thermodynamic prediction tools should be tailored to the specifics of IAD designs. The larger diameters will affect the shock standoff distance, and thus also the heating environment around the stagnation area. Radiative heating, in this sense, should possibly be considered. As an example of aero-thermodynamics analysis, the reader is referred to the paper by Mazaheri (2013) on the analysis of HEART.
3. *Flow stability.* Large-scale flow instability, amongst others vortex shedding, can cause motion and orientation changes of the IAD/payload system. Shock-shock interaction may cause smaller-scale instabilities, but these effects can possibly be avoided by carefully selecting the geometry. Aeroelastic effects should be included, as they effect the geometry, so a rigid shape can no longer be assumed.
4. *Aerocapture system mass.* Will IADs have a sufficiently low mass to be competitive with propulsive orbit insertion and rigid aeroshell aerocapture? Conceptual studies indicate that significant mass savings can be achieved, but a more detailed multi-disciplinary design approach is required to analyse this concept—taking shape, size, material selection and level of gas pressurisation for each of the candidate missions into account.
5. *Trajectory robustness.* The IAD entry corridor width shall be studied in detail, and how the system can compensate for atmospheric and vehicle uncertainties. The IAD design should have a sufficient design margin to accommodate it.
6. *Structural integrity.* Because thin film material is used to minimise the system mass, structural integrity must be ensured, even though the dynamic pressure and heating are lower than for rigid aeroshells. Detailed finite-element analysis may be required to address the reduced design margins.
7. *Trailing ballute tether design.* Can the (small-diameter) tether(s) be designed to take the heating and stress? Additional points of attention in this case will also be, for instance, shock impingement from the spacecraft and attachment to the ballute.
8. *Parent spacecraft protection.* For attached IADs, the heat throughput to the parent spacecraft shall be analysed, such that it can be determined what auxiliary thermal and aerodynamic protection the spacecraft needs. One should not forget that the spacecraft stagnation radius may be smaller than the IAD nose

radius, which leads to different thermal loading. Of course, any induced mass penalty shall be considered in the overall vehicle design.

9. *Deployment and inflation.* It must be studied whether this particular technology is already mature enough, or can be transferred from other inflatable efforts, such as the inflatable antenna experiment.
10. *Experimental verification.* Because the complete aerothermodynamic environment for planetary missions cannot be reproduced in ground-based facilities, the proper combination of (flight) testing and analysis shall be found, such that proof of concept can be confirmed. In case of sub-scale tests, can the appropriate scaling parameters be matched for testing?

The short description in this section does no justice to the importance of IADs, both in terms of design and operation. The provided references—and the many references contained therein—provide an excellent starting point for further study, though.

9.5 Advanced Descent and Landing Systems

Since the start of the space age, various spacecraft were designed and send out to land on Solar-System bodies other than Earth. The first attempts suffered quite some failures, but most missions developed and executed in the past couple of decades completed their mission successfully. And it was not just limited to the "big players", the United States and Russia, but also China, Europe, and Japan have achieved landing missions to the moons, Mars, comets and asteroids. One possible conclusion from this could be that more than a few players master the task of planetary descent and landing.

However, during the last years, many researchers continued to investigate how landing on other bodies can be improved. And rightfully so, one may wonder why there is a need for improving landing capabilities, even though most landings of the last decades have been successful. An obvious answer could be that even if one has shown the ability to land on Mars, one might have no idea how to land on, for example, Europa, one of Jupiter's moons, for the simple reason that each body in the Solar System is fundamentally different. Moreover, mass is a very important factor when designing space missions. Depending on the scientific yield, more instruments may be required, thus adding to the dry (and wet!) mass of the lander. The very large mass of the *Curiosity* rover (almost 900 kg), payload of the Mars Science Laboratory (MSL), led to the development of the sky crane as the best option to land the rover. If the rover had been lighter the same airbag system as used for the two Mars Exploration Rovers (MER), could have been used (each of the MER rovers weighing less than 200 kg).[26]

It may be clear that the more advanced the lander systems get, the less likely it is that the mission design can solely depend on heritage systems. Some abilities we

[26] Both MSL and MER missions will be discussed in Sect. 9.7.

currently do not have, because they were not needed yet, but which might be required for future landers are (1) to land in a hazardous environment, (2) to land at higher altitudes on Mars, and (3) to perform a precision landing (landing at exactly the specified location). These are just a few of the challenges researchers are currently trying to tackle. Undoubtedly there are many more, which may not even have been identified yet. Some of the above mentioned abilities will be discussed in the course of this section.

We will begin the discussion in Sect. 9.5.1 by looking at landing-site selection, mainly at the size of landing regions and the constraints on the terrain. From this we will be able to derive needs for new technologies to advance current landing capabilities in Sect. 9.5.2. In this context we will focus on very precise landing and landing in unsafe environments. It is noted that a major part of this section is based on the papers by Woicke and Mooij (2014, 2016, 2017). For a detailed description of a complete hazard-relative navigation system, including hardware verification and validation, the reader is referred to Woicke (2018).

9.5.1 Landing Site Selection

To better understand why there is a need to develop more advanced landing systems, let us first have a look at how landing sites, or rather *landing regions*, are selected. For all past landers the landing region was selected prior to entry, descent, and landing (EDL), mostly even in a very early design phase. Landing regions are basically decided on science requirements and are thus determined by scientists, who will support a certain mission. Then, engineers become involved to make it work. For instance, when scientists want to investigate ice on the Moon, this will lead to the requirement that landing has to be performed somewhere where ice is known to be present, or likely to be found. This could mean that such a location is in a shadowy region, which poses some constraints on the landing location. Other missions would be downright impossible (for now), from an engineering point of view. This could be landing at high elevation on Mars, for example, on top of *Olympus Mons*. Due to the very low density at those heights, parachutes cannot sufficiently decelerate descent modules for retro rockets to take over. In other words, the lander would simply shatter during touchdown.

Usually, when identifying the mission need, space agencies will ask scientists to identify regions of interest and why this would be so. Of course, it only makes sense that they identify potential landing sites that are within the engineering limits of current technology. To assist the scientists, engineers will create maps that indicate what is possible and what is not. For example, Fig. 9.66 shows an elevation map of Mars, which can be used to constrain the possible landing sites with respect to the maximum allowable elevation. An MSL-sized lander cannot land at elevations above 2 km.

Still, it is possible that scientists will ask for a landing site that cannot be achieved with current technology, but would be within reach if certain new technology could

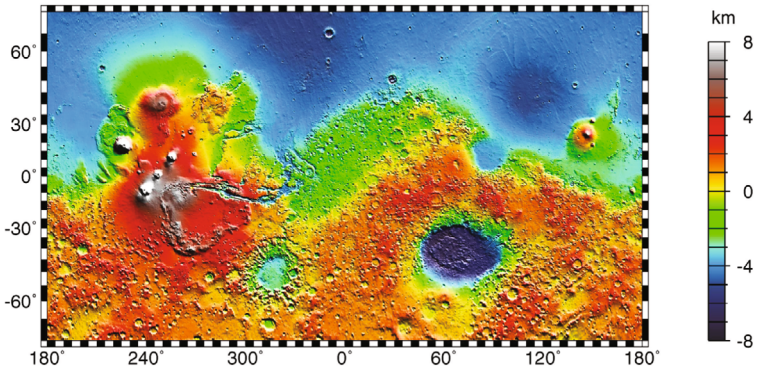


Fig. 9.66 Mars elevation map. Image: NASA

be developed. Such a request stimulates technology-development programmes that will lead to an advancement of the current state-of-the art. But let us keep this for later and discuss how landing sites are selected to be reachable with the current state-of-the-art landing technology.

At present, landers cannot land precisely at a pre-defined location. Because of inaccuracies in modelling the atmospheric entry, as well as limitations in navigation filters to accurately predict the vehicle's location, such pinpoint landings are not possible. Therefore, the landing region will be described by an ellipse, which is defined such that with a probability of 3σ (or 99.7%) the vehicle will end up inside it. The reason that the landing region is an ellipse and not a circle is easily understood, when one considers the flight dynamics involved. The vehicle's projected motion on the surface can be described by two components, i.e., down-range that is in the direction of flight, and cross-range, which is perpendicular to this direction. Because velocities are substantially larger in down-range direction, small errors will thus have a larger impact, and thus is the landing region elongated in that direction.

To get some feeling about the dimensions of landing ellipses, we will compare the landing ellipse of three different classes of Martian landers: Viking, the Mars Exploration Rovers, and Mars Science Laboratory.[27] The Viking landers were the first soft-landing Mars missions, and both spacecraft landed in 1976. Their landing ellipse was a very large 280 × 100 km. *Mars Pathfinder* followed in 1997, and with increasing technology and knowledge about the atmosphere, the ellipse had already reduced to 200 × 70 km. In 2003 and 2004 the Mars Exploration Rovers landed on Mars. Using (improved) technology developed originally for *Pathfinder*, the landing ellipse was reduced significantly, i.e., to *only* 150 × 20 km. The navigation system MER carried was called DIMES (Cheng and Matthies 2005), which was able to estimate the velocity of the lander with respect to the surface, based on three consecutive images. This system was a necessity for safely operating its air-

[27] We use the term *classes* here, because both Viking landers and MER, were not just a single lander, but two identical ones, putting the two of them in the same class.

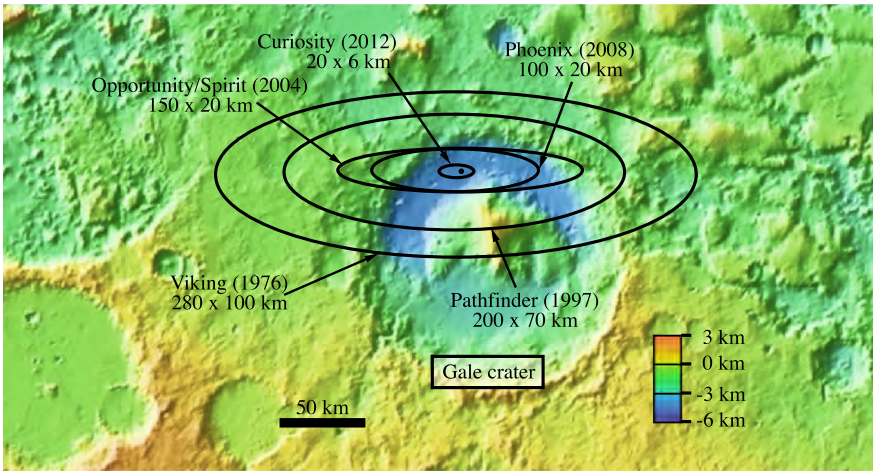


Fig. 9.67 Decrease of the landing-ellipse size over the past decades. The dot in the centre ellipse is the landing location of Curiosity. Data taken from original image of NASA/JPL/GSFC/Arizona State University

bag system, because a too large vertical velocity during touch down could possibly rupture the air-bag. Thus, with more accurate knowledge about the vertical velocity, it could better counteract any errors using its thrusters, and the velocity would be lowered in case it was too large. This reduced the uncertainty in landing location, but only from the moment the system was switched on. The Mars Science Laboratory was able to significantly reduce its landing ellipse with respect to its predecessor by performing a guided entry, which removed one source of inaccuracies, as it could respond to atmospheric uncertainties. Its resulting landing ellipse was *only* 20 × 6 km. Thus for an MSL-like landing, scientists will have to select a landing region that will fit their requirements, meets the engineering constraints and is at least 20 × 6 km large. For MSL, this meant it could land *Curiosity* in the *Gale crater*. Selecting a suitable landing region is clearly more challenging the larger the landing ellipse gets.

Even though landing-ellipse sizes have significantly decreased over the past years, we are still far away from the capability of performing pinpoint landings, which will certainly be required in the future. For example, if landing next to a surface asset is necessary, it will be very beneficial if the spacecraft lands *very close to* the asset, while at the same time avoiding to land *on* the asset. Therefore, there is a need to develop systems that can decrease the size even further to enable pin-point landings.

At this point it is worth mentioning the JAXA Hayabusa-2 mission. This mission, launched on 3 December 2014, is an asteroid mission that rendezvoused with near-Earth asteroid 162173 Ryugu on 27 June 2018. Amongst others, it took (sub-surface) samples that it returned to Earth on 5 December 2020. The mission was extended through at least 2031, when it will rendezvous with a small, rapidly-rotating asteroid (1998 KY26). It is the sampling method of the asteroid that we want to mention

specifically.[28] The first step in the process was to separate an impactor from the spacecraft. When the spacecraft had orbited to a safe distance, the impactor was fired into the asteroid surface by detonating a 4.5 kg shaped charge of plasticised high melting explosive (HMX) for acceleration. The impactor accelerated from a distance of 500 m and excavated a crater of about 10 m in diameter, exposing pristine material. The second step was the deployment of a reflective target marker in the area near the crater to assist with navigation and descent. The third and final step was the touchdown of and sampling by the main spacecraft. It is obvious from this description that this was indeed a pin-point "landing", because markers could aid the descent. The deployment of visual markers or radar beacons is, of course, not always (or mostly not) possible, so we will focus on different means to achieve pin-point landing.

So far, we have mentioned engineering constraints at several occasions. But, only one example of such a constraint was given: the maximum elevation of the landing site. However, there is another very important constraint that landing regions have to meet: they have to be inherently safe. If a region is inherently safe, it describes a region that is free of all possible surface features or conditions that could pose a hazard to the lander during descent and touch-down. Of course, these conditions cannot be looked at in isolation, but have to studied in an integrated approach with vehicle geometry, landing-system configuration, and lander centre-of-mass location.

What kind of features or surface conditions are actually posing a hazard to a vehicle depends on the design, but also on the target body. Let us discuss the most common hazards one by one.

Slope

The slope between two points describes how steep the terrain is in between these two points. The slope is typically not constant over an entire landing region but can vary. Usually, the slope is computed by fitting a mean-plane trough multiple neighbouring points, as this will exclude smaller outliers, such as rocks, from the slope evaluation. Figure 9.68 shows an image of the Lunar surface. Multiple different slopes are present in this scene, two of these are indicated. If the local slope is too large it can pose a threat to the lander, as the lander can either slip away or, even worse, tip over. If and how extreme slipping occurs depends on the composition of the ground, as well as the mass and velocity of the lander. The stability with respect to tipping over is based on the location of the lander's centre of mass, in combination with the landing-leg geometry. Landing legs, shock-attenuation systems and a (simplified) stability analysis will be discussed in more detail in the next section on landing systems, more specifically in Sect. 9.6.2.3.

[28] https://phys.org/news/2019-07-japan-hayabusa2-probe-touchdown-asteroid.html, date accessed August 8, 2021.

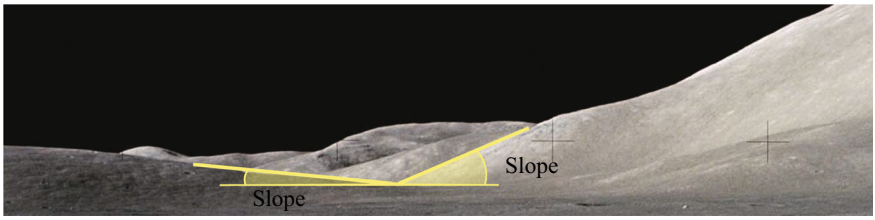
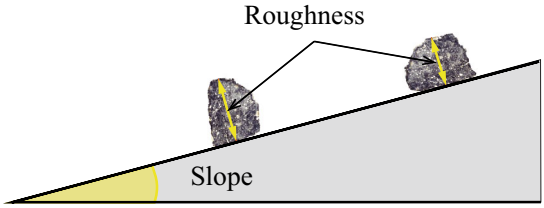


Fig. 9.68 Two out of several slopes indicated on the Lunar surface

Fig. 9.69 Roughness as opposed to slope



Roughness

Roughness describes the features that deviate from the mean plane. Simply speaking, roughness is created by small features lying on the surface, such as rocks and boulders. The difference between slope and roughness is depicted in Fig. 9.69. Larger rocks are often referred to as boulders. Therefore, roughness is equivalent to rocks and boulders lying on the surface. How much roughness a lander can handle, depends on its design. The ground clearance, a measure that indicates how much clearance there is in between the vehicle's underside (or protruding nozzles!) and the surface, usually determines the maximum allowable size for roughness features.

Craters

Craters do not necessarily pose a hazard. Whether or not this is the case, strongly depends on the mission itself, but also the age and size of the crater. On one hand, for a rover mission it might be very unfavourable to land inside a (smaller) crater, as the rover would be confined to the inside of the crater. On the other hand, younger craters might be very interesting geologically and could be a very interesting landing site. However, given the landing accuracy of the system, one may want to include extra margins if the nominal landing site is close to a crater rim. As a reminder, *Curiosity* did land in a crater, although truth be told, the *Gale crater* has a diameter of 154 km, and the total distance that the rover has covered to date, during its 9-year (and on-going) mission, is just over 25 km.

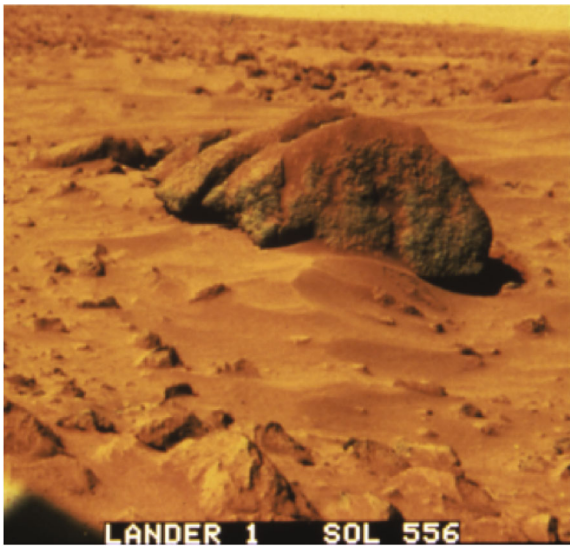


Fig. 9.70 A 2-m boulder named "Big Joe", right next to the Viking-1 lander. Image: NASA

Illumination

As for craters, the hazardousness of certain illumination conditions depends on the mission itself. If a mission receives all its energy from solar power, it is of utmost importance that the landing region is illuminated. The same goes for landing systems that rely on visual imagery for certain descent systems, such as terrain-relative navigation (to be discussed later). Last but not least, public outreach is also important: if the landing site is in total darkness, no descent images can be broadcasted.

Figures 9.70 and 9.71 provide two examples of surface hazards. Figure 9.70 shows a boulder, which is located right next to the landing site of the Viking-1 lander. This boulder is 2 m high and would have certainly caused a landing failure if the landing would have been performed on top of it. Viking-1 landed in 1976, back then orbiter maps were not of sufficient quality to detect this size of surface features. It is important to remember that engineers and scientists thought that no such features would be present in the entire landing region, based on statistical analysis. Analysing the image closely, it is apparent that this was a wrong assumption, as Viking clearly landed in a boulder field. A second example, Fig. 9.71, is the asteroid Itokawa, the destination of the Japanese Hayabusa lander back in 2005. The longest axis of this asteroid is approximately 500 m long. In the image we can see that the asteroid's surface is full of boulders and rocks. Clearly, this can cause a potential danger to a lander mission.

Summarising, one can say that what amount of slope or roughness, which size of craters, and how much darkness is allowable in a landing region to be inherently safe, depends on the mission. For the Mars Science Laboratory, slopes larger than 15° and roughness larger than 0.5 m were specified as hazardous. Therefore, any region with

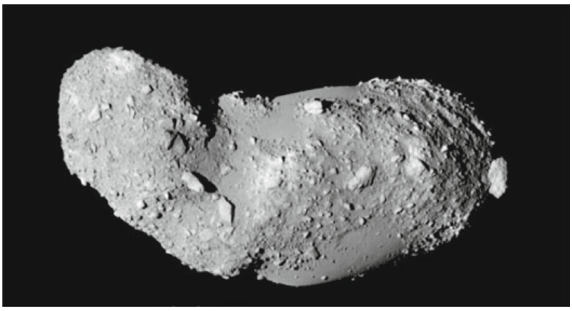


Fig. 9.71 Astroid Itokawa. Image: JAXA

at least the size of the landing ellipse, which would only contain slopes smaller than 15° and rocks or boulders smaller than 0.5 m, would classify as inherently safe. However, one should keep in mind that surface features can tell a lot about a planet's past. Therefore, hazardous regions can be very interesting from a scientific point of view.

So to land safely, an inherently safe landing region is necessary. But how can engineers and scientists determine whether a landing region classifies? As we discussed earlier, landing regions are usually chosen prior to launch or while the vehicle is still in orbit around the body. Thus, orbital imagery is used to assess the safety of possible landing regions. If rocks larger than 0.5 m pose hazards to the lander, this means that the images used need to have sufficient resolution to resolve these small features. For Mars, the Mars Reconnaissance Orbiter (MRO) actually mapped the surface at sufficient resolution. For the Moon, maps of sufficient resolution are only available for certain regions, taken by the Lunar Reconnaissance Orbiter (LRO). Most importantly, both MRO and LRO used very large and advanced instruments to generate these maps! It is very unlikely that a lander mission would bring along an equally good instrument. More importantly, one should also not forget about opaque atmospheres: for example, Venus or Titan cannot be mapped from orbit. In addition, it is possible that the terrain changes in between analysis of the images and landing, for example, due to a sand storm or seismic activity.

Figure 9.72 shows two good examples for how the surface can change over time (these pictures were taken by the High Resolution Imaging Science Experiment (HiRISE) camera on NASA's MRO). In the top figure, a boulder, approximately 3.5 × 6 m tall, is indicated by the red circle. This boulder moved recently, prior to taking the picture. The track it left while moving, resembling a dotted line from the upper left to middle right of this image, is indicated by the red arrow. Based on the looks of the tracks, scientist concluded that the movement happened recently. The second example in the figure at the bottom is a boulder that rolled down a slope, leaving a trail of about 500 m long. The trail has an odd repeating pattern, suggesting the boulder could not roll straight due to its shape. An interesting feature is coming from the moment the boulder encountered a small crater. It flew up in the air, and bounced off at the crater rim at the other side. After a few more bounces it continued its rolling path until it came to a halt.

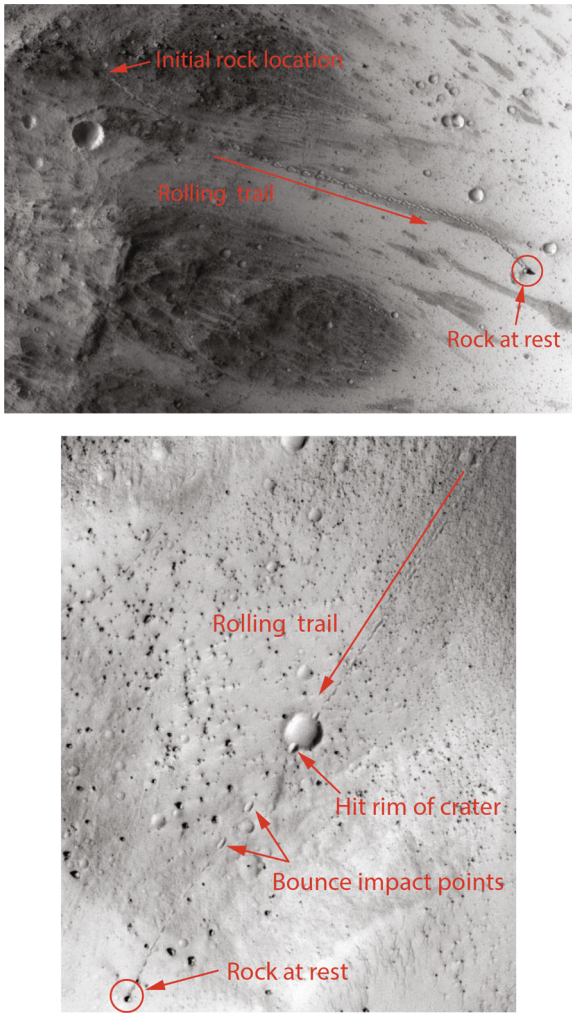


Fig. 9.72 Surface changes on Mars, as observed by MRO. Images: NASA/JPL-Caltech/Univ. of Arizona

Concluding, for Mars and certain regions of the Moon it is possible to select inherently safe landing sites prior to the mission, for all other bodies, this is (currently) not possible. Moreover, as we have seen there is no 100% guarantee that the orbital images and the surface conditions during landing will fully match. In addition, the constraint that the entire landing ellipse has to be inherently safe will greatly limit the number of possible landing regions. Even more importantly, landing in unsafe regions might even be very desirable, as crater rims, boulder fields and young craters are very interesting locations from a science point-of-view. Therefore, there is a need to find a system that can decrease the risk of a landing failure due to surface hazards.

9.5.2 Integrated Navigation Approach

Even though many of the past missions to other Solar-System bodies were successful, advancing current systems is still very important to push the current boundaries. In the previous section it was established that to increase the number of possible landing sites, and by that to increase the potential scientific return, two main points should be addressed:

1. The need to develop systems that can further decrease the landing-ellipse size, and
2. The need to develop a system that can decrease the risk posed by undetected surface hazards.

We will discuss some of the current developments to address that address these (two) needs. Obviously, there is more to it than just these two points, and more areas in which advancements are needed could be identified, but in the spirit of this chapter we will restrict ourselves to just these two.

Let us look at the first need, i.e., the one to decrease the size of the landing ellipse. Earlier, it was established that the large landing ellipses were caused by errors in the entry flight-path modelling, due to insufficient knowledge about the atmosphere, not having a closed-loop landing guidance, and finally navigation errors, due to limited sensors (or none at all) and the appropriate filters. In the following the focus will be on decreasing the landing-ellipse size by decreasing the navigation error. As landing in exactly one spot is almost impossible, we will focus on precision landing, i.e., a landing within a few meters of the selected landing site.

To achieve the second need, it is mandatory to solve the precision-landing problem first. If, on one hand, a vehicle would be able to detect all hazards within its landing region, but would be unable to land accurately enough to avoid these, the system would be useless. On the other hand, if a spacecraft *can* perform a precision landing, it can also be designed such that it can detect and avoid hazards.

9.5.2.1 Precision Landing

Today, the main sensor for obtaining navigation measurements during descent and landing is the inertial measurement unit (IMU). The IMU can detect changes in angular rate and (non-gravitational) acceleration, from which the current position and orientation can be determined. Combining these measurements with information about the past state will yield the current, estimated, state. However, both the measurement and the past state are not error free. Therefore, usually different kinds of filters, such as the extended Kalman filter, are used to decrease the errors, provided sufficient measurements are available. But still a substantial error may remain. For orbiting spacecraft, often additional sensors are used, for example, star sensors, However, these cannot be used during descent, as vibrations will be too large and will render the measurements useless.

Furthermore, there are two different kinds of precision. The first one is with respect to the entire body, and describes how well one can land with respect to the orbiter map. This is an inertial accuracy. For example, if one selects that the lander is supposed to land in Delft (52° 1' north and 4° 2' east), the inertial accuracy describes how far the lander deviated from this location. Next to this, there is also the accuracy with respect to a selected landing site inside its field of view. This is usually not an inertial measurement. Therefore, it does not depict the absolute navigation error, but only the relative one. As an example, imagine one wants to pour some freshly brewed tea into a cup located on the table. First, we will move towards the table without looking at the cup yet. This is what we would call *absolute* navigation. Here, we will use the information we have on where the cup is located on the table, based on where we initially put it. As soon as we are close enough we will focus on the cup only and will navigate the tea pot relative to the cup. This is "terrain" *relative* navigation. We will explain this in more detail later. To put the concepts of navigation into a "space perspective", for current Lunar-landing mission studies, inertial accuracies of 200 m and relative accuracies of 10–20 m are desired.

The landing ellipse always describes the inertial accuracy of the system. If we can increase the inertial accuracy, the landing ellipse will thus decrease in size. So how can the inertial error be reduced? By introducing terrain absolute navigation into our landing system.[29]

9.5.2.2 Terrain Absolute Navigation

To decrease the inertial error a system called *terrain absolute navigation* (TAN) can be used. TAN tries to locate the position of the spacecraft with respect to a map of the surface. Usually, this is done based on a crater map. A crater map (or list) is a map that contains a large number of the craters on a planet's or other body's surface (or at least in the vicinity of the landing site). As these maps are created based on orbital images, the positions of the individual craters can be linked to an inertial position with respect to the planet's or moon's inertial reference frame. The crater map can be created prior to launch of the mission, and may be updated in case an orbiter is part of the mission. Figure 9.73 shows a crater map for a small part of the Lunar surface.[30] Besides the craters themselves, this map contains other features as well, e.g., cracks (Rima Reumur, Rima Oppolzer, and the multiple Rimae Triesnecker), planes (Sinus

[29] Note that there is more to precision landing than simply decreasing the navigation error. It is also necessary to have a guidance system that can influence the landing location and react if the current path starts deviating too much from the planned part (see also Sect. 9.3). Moreover, a propulsion system that cannot just brake the vehicle, but also counteract undesired changes in the state, has to be available. This would (preferably) require a throttleable engine of which also the orientation could be changed. However, these problems are already solved to some extent. Therefore, we will focus on the navigation problem. Still, in the course of this section we will mention one guidance law that can be used for precision landing.

[30] https://www.lpi.usra.edu/resources/lunar_orbiter/bin/info.shtml?264, date accessed August 10, 2021.

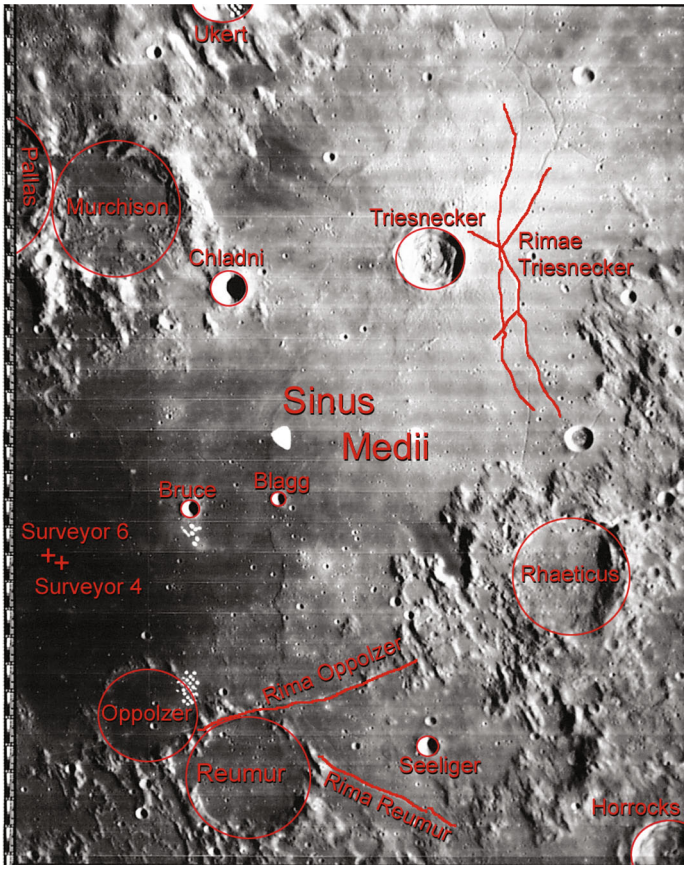


Fig. 9.73 Example of a crater map, including some other features (ridges and landers). Photo centre-coordinates: 12.96° north, 3.63° east, taken from an altitude of 2,699.13 km, with a Sun angle of 68.4°. Image: Lunar and Planetary Institute (LPI)

Medii), and in this particular case the crash location of Surveyor-4 and landing site of Surveyor-6. The visible craters could be stored in a data table, combining location and size information, see Table 9.5.

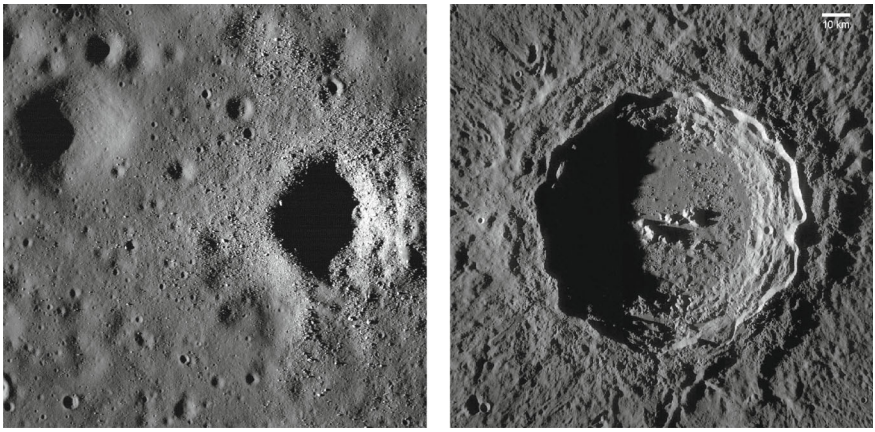
To obtain information on the lander's position with respect to the primary body's inertial frame, it is necessary to investigate which of the craters are currently in the field of view of the spacecraft. Thus, a camera is required to perform TAN. At each instance of the state evaluation, one camera image will be used. By means of a crater-detection algorithm, the craters in the image will be searched for, in a way similar to the workings of a star tracker's software scanning a star database. Based on the dimension and the relative location of the craters the crater catalogue will provide the relevant craters. If a match is found, the spacecraft's location can be derived from the location of the craters in its field of view. At least four craters have to be

Table 9.5 List of craters visible in Fig. 9.73. Data: Lunar and Planetary Institute (LPI), see Footnote 30

Feature name	Feature latitude/longitude	Feature size (km)
Blagg	1.3° N/1.5 ° E	5
Bruce	1.1° N/0.4° E	6
Chladni	4.0° N/1.1° E	13
Horrocks	4.0° S/5.9° E	30
Murchison	5.1° N/0.1° W	57
Oppolzer	1.5° S/0.5° W	40
Pallas	5.5° N/1.6° W	46
Reumur	2.4° S/0.7° E	52
Rhaeticus	0.0° N/4.9° E	45
Rima Oppolzer	1.7° S/1.0° E	94
Rima Reumur	3.0° S/3.0° E	30
Rimae Triesnecker	4.3° N/4.6° E	215
Seeliger	2.2° S/3.0° E	8
Sinus Medii	2.4° N/1.7° E	335
Surveyor 4	0.45° N/1.39° W	–
Surveyor 6	0.49° N/1.40° W	–
Triesnecker	4.2° N/3.6° E	26
Ukert	7.8° N/1.4° E	23

detected unambiguously, to obtain a good estimate of the vehicle's position. The computed state will be passed on to the navigation filter and can be used to update the information obtained from the IMU and the previous state (refer to Sects. 7.3.1 and 7.3.2 for the operational principle of the Kalman filter).

How does such a crater-detection algorithm work? Usually, it first tries to identify the location of the craters and will then fit ellipses, such that these represent the outline of the crater. Craters can be identified, using different methods. For instance, one can use the knowledge that craters usually draw shadows. Usually, about half of the inside of a crater appears very dark in images, while the other half is very bright. By looking at Fig. 9.74, it is clear that this phenomenon is very obvious. Based on these image regions of extreme contrast, craters can be identified. However, it is also possible to identify the rim of the crater using a texture-detection algorithm. An example of this is shown in Fig. 9.75, for which a texture-detection algorithm was run on an image of a single crater; the algorithm managed to detect the rim very well. Often, combinations of such algorithms are used. Unfortunately, it is beyond the scope of this text to discuss the working principle in more detail. An example can be found in Mammarella et al. (2011), though. For further study it is also advised to have a look at the paper by Maass et al. (2020), who discuss a crater navigation system for autonomous precision landing on the Moon.



(a) Oceanus Procellarum (b) Copernicus

Fig. 9.74 Craters on the Moon, with shadows cast inside. Images: NASA/GSFC/Arizona State University

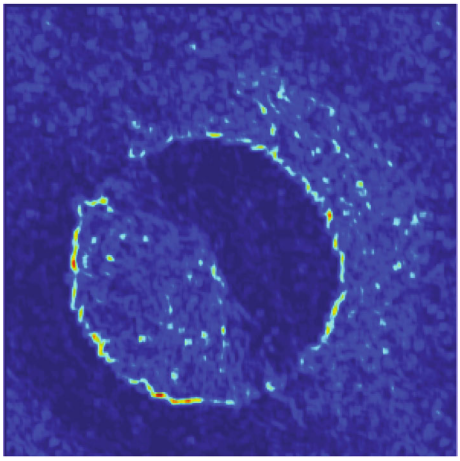


Fig. 9.75 Texture map identifying the rim of a crater (Woicke and Mooij 2016)

The accuracy of the estimate is mainly determined by the resolution of the crater map, the accuracy of the positions in the maps, and the errors in the crater detection. One should also keep in mind that these algorithms have to run in real-time on-board the spacecraft, so computation power will be limited.

Terrain absolute navigation is performed at higher altitudes. If altitudes are getting too low, crater maps are not available any more, due to the limited resolution. Therefore, the inertial state cannot be estimated. For a Lunar-landing scenario TAN could be performed for altitudes of approximately 100 km down to 10 km.

9.5.2.3 Terrain-Relative Navigation

If the altitudes are getting too low to use TAN, the error in the inertial state will not be further improved and thus a different method has to be found to improve the IMU measurements. Such a method does exist, i.e., *terrain relative navigation* (TRN), with which it is possible to compute the change in position and orientation of the spacecraft's state over a measurement interval. Thus, the outcome of TRN should be combined with the IMU measurements to decrease the localisation error.

TRN requires at least two consecutive images of the surface underneath the lander, and preferably more. These images do not necessarily need to be visual images, but can also be radar or lidar images. From these images "features" are extracted, which are then tracked over multiple images. Using the information of this tracking, in combination with the time between two consecutive measurements, the motion of the vehicle can be derived. Clearly, multiple features should be tracked at a time. For each feature, the corresponding movement is computed. Since incorrect associations of features are possible, the magnitudes and directions of the movements are compared to remove outliers. For two reasons it is important that the imaging interval is small enough: (1) at least parts of the features should be present in consecutive images, and not have disappeared from the field of view, and (2) the estimation process is linear, and a smaller time interval means more accurate estimates with which to update the guidance and control algorithms. The estimation improves if features can be tracked over multiple images. The new state of the spacecraft can then be computed using the old state and the outcome of the combination of IMU and TRN.

This working principle is also depicted in Fig. 9.76 for a very basic case of pure lateral motion. First, three consecutive images are shown. It can be seen how new features enter the image from the right and "old" features leave the image at the left-image boundary. The direction of flight is from left to right. In the second set of images, four features are selected and extracted from all three images. In real life, more features would be extracted, for reasons of redundancy and to remove outliers. Quite a few of those will then not be found back in the next image. For the sake of the current example, such features are not included in the figure. Finally, it is shown how the moved distance can be computed from the extracted and tracked features. A nice overview on different TRN methods and sensor options is given by Johnson and Montgomery (2008).

Until now, the word "features" has been mentioned several times, but what exactly are those? Truth be told, they can be almost anything, but the most obvious choice would be, of course, landmarks that can also be identified by visual inspection, such as craters and rocks. However, more abstract features exist, for example, edges and corners. Usually, information on the relative position of features is used to get unambiguous matches. Similar to TAN, also TRN has to run on-board and must therefore be fast and not too demanding with respect to computation power.

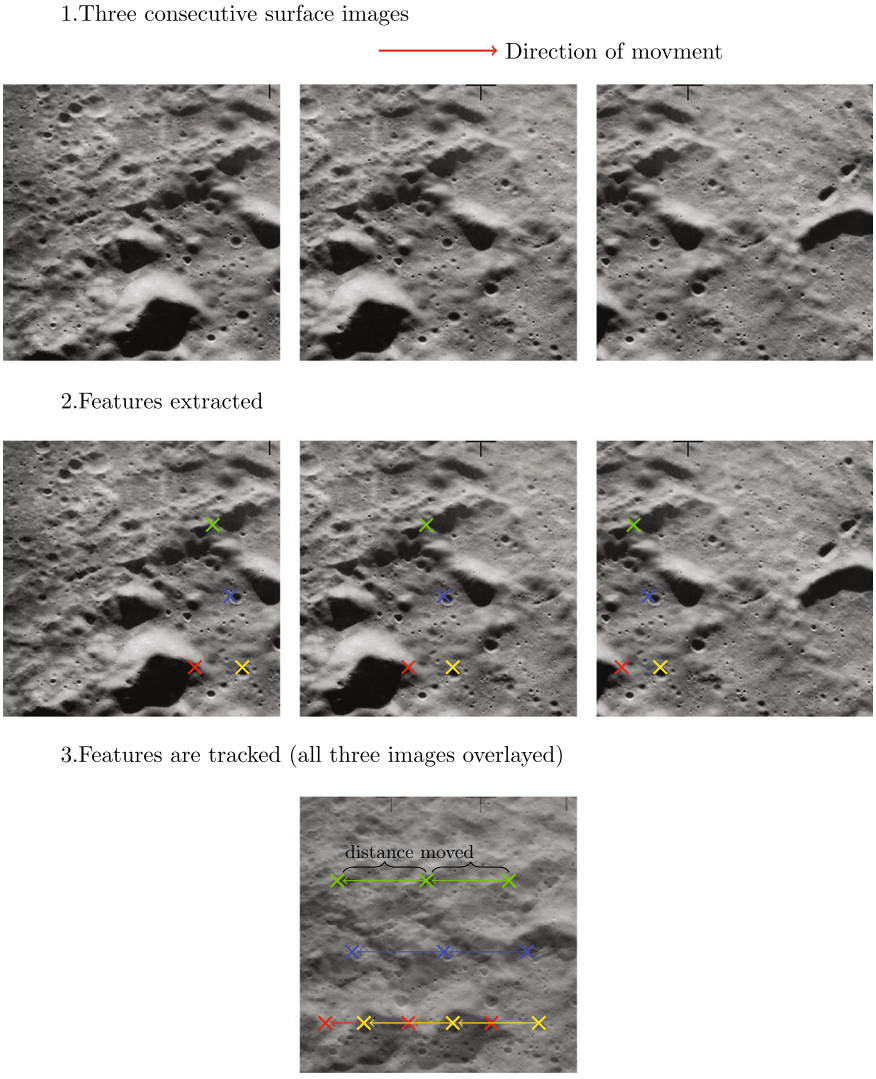


Fig. 9.76 Terrain relative navigation. Images: courtesy dr. Svenja Woicke

9.5.2.4 Hazard-Detection and Avoidance

Successfully combining TRN and TAN makes it feasible for a landing system to perform a precision landing. But this capability opens up more possibilities: in case a spacecraft is able to accurately steer towards a certain landing location, it is also capable to avoid certain regions. This means that a so-called hazard-detection and avoidance system (HDA) can be integrated on the spacecraft. Such a system enables

the lander to "see" hazards and protects the vehicle from landing on top of, or right next to, such dangerous locations.

Hazard-detection and avoidance can be split into two main parts, hazard mapping and piloting. Guidance is sometimes included as a task of the HDA system, although it is common to see the HDA system itself as being independent from the guidance system. It goes without saying that there is a direct interface between the two, as the guidance system should be designed such that it can react to the outputs of the HDA system. But, despite this interface the guidance system will be able to work independently from this system.

During the hazard-mapping phase, maps of the surface of the landing region are created. After this, these maps are passed on to the piloting function, which assesses all possible landing sites based on the level of the hazards as indicated by the hazard maps, combined with additional maps. These maps, for example, contain information about the distance to the closest hazard, the reachability in terms of the guidance algorithm, and the fuel cost. The first step is usually to determine whether the nominal landing site is safe. If so, no diversion has to be executed, even if other sites would be safer. The reason for this is that each diversion will always add risk to the mission. If, on the other hand, the nominal landing site is deemed to be unsafe, a better and safe landing site will be selected by the piloting function. The piloting function has to be set up such that this selection is unambiguous. This new landing site will then be send to the mission manager, which will tell the guidance function to adapt the trajectory accordingly.

We have established that to assess the safety of the terrain, hazard mapping is necessary. So let us have a look into how hazard mapping works. Note, that some people also refer to hazard mapping as hazard detection. The main task of this part of the algorithm is to determine the hazardousness of the landing region. In Sect. 9.5.1 we already discussed which surface features and conditions can pose hazards. To recap, the most important hazards identified are: (1) too large slope, (2) too large roughness, (3) craters, and (4) insufficient illumination. See also Figs. 9.68 and 9.69 for the concept of the different surface hazards.

However, not all of these apply to all mission scenarios and spacecraft designs. Therefore, there are many different system options for hazard mapping. Generally speaking, one can say the newer systems tend to evaluated all of the above mentioned hazards (Epp and Smith 2007b; Epp et al. 2014), while older systems often only detected roughness or craters, or even only assessed the safety based on some texture measures (Pien 1991; Cuseo and Dallas 1989). This was mainly driven by the limitations in computing power of the on-board computers. The newer algorithms do rely on the same techniques (Howard et al. 2011; Yan et al. 2013; Cohanim et al. 2012), albeit that with the increase in performance of flight computers the capabilities of HDA systems have been increased.

To evaluate the surface with respect to the presence of hazards, a sensor observing the surface is needed. The task of this sensor is to create a map of the surface, which can be used as an input to compute slope, roughness and illumination. Here, two different kinds of sensors can be selected: active and passive sensors. Active sensors emit and receive a signal, while passive sensors only receive. In the case of hazard

mapping, active sensors would directly generate a three-dimensional map, a so-called digital elevation model (DEM) of the surface, while passive sensors can only generate a projection of the surface (a picture). The primary option for an active sensor is lidar, with two different variations: either a scanning or an imaging lidar. In general, an imaging lidar is the faster system, whereas a scanning lidar might be able to create more pixels even at larger distance. A passive sensor would be an imaging camera.

What will drive the decision to choose either a passive or an active sensor? As was already mentioned above, active sensors measure the three-dimensional shape, while passive sensors need some algorithm to reconstruct the surface map from the input image(s). From this point of view it seems that active sensors are superior to passive sensors; however, passive sensors are way superior when it comes to system budgets. As an example, the DragonEye 3D Flash Lidar Space Camera™ of Advanced Scientific Concepts, Inc., has dimensions of 11.2 cm × 13.2 cm × 11.9 cm, a mass of 3 kg, and a power consumption of 35 W, while the EACM space imaging system of Malin Space Science Systems only has dimensions of 7.8 cm × 5.8 cm × 4.4 cm, a mass of 256 g, and a power of only 2.5 W. Since for space mission each gram and Watt can have a substantial influence, it has to be decided for each mission individually whether an active or passive sensor should be chosen. If cameras are chosen, multiple different methods are available to compute the surface map from the camera image(s). However, it is beyond the scope of this text to discuss each of them individually. The interested reader is referred to Câmara et al. (2005), Woicke and Mooij (2014), and Strandmoe et al. (1999), amongst others. It is important to note that with lidar illumination conditions cannot be evaluated. Therefore, an active system would always require an additional camera to assess these. Camera images can also be used for hazard detection, without retrieving the three-dimensional scene. This process is called texture detection, and will be discussed later.

For both active and passive sensors, the computation of slope and roughness from three-dimensional maps can be done in the same manner. This, first of all, involves a plane fit. Here, different kind of fits can be used. An extensive discussion would be beyond the scope of this introductory text. However, one important fact to remember is, that there is something called an *intelligent mean plane*, which can exclude outliers. The difference between this plane and a normal mean pane is shown in Fig. 9.77.

This is important for both the presence of errors in the 3D maps and for scenes of very high roughness. Knowing the slope, the roughness of the surface can then be computed as the deviation from the mean plane. This means that individual (sets) of pixels are compared to the local mean-plane. If the Z-coordinate has a considerable off-set from the local mean-plane, a roughness feature has likely been found.

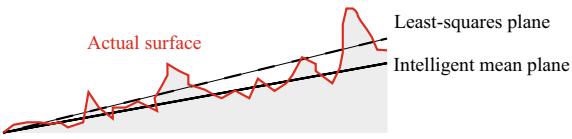
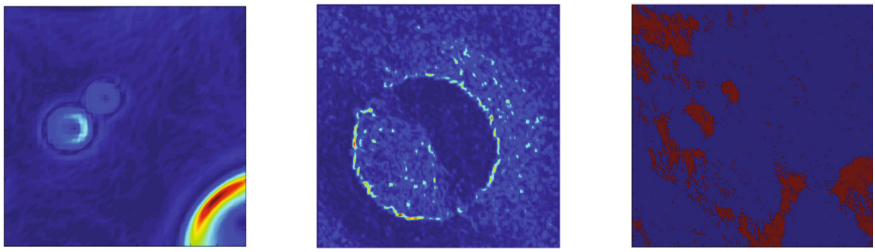


Fig. 9.77 Plane fitting and robust/intelligent plane-fitting



a) Slope b) Roughness from texture c) Illumination

Fig. 9.78 Example maps. Images courtesy of dr. Svenja Woicke

Notwithstanding, if passive sensors are used it is also possible to infer surface roughness from texture levels in the image. This process is called texture detection, or roughness-from-texture. For very noisy DEMs this method might be preferable over roughness estimations based on the DEM. As for slope estimations, there are various methods to perform texture detection, each with different strengths and weaknesses with respect to certain texture variations. The appropriate method for a certain landing scenario has to be selected on a case-by-case basis and, for example, depends on the reflectivity and uniformity of the surface. A more detailed discussion on slope computation and roughness estimation methods can be found in Woicke (2014).

Figure 9.78 shows an example slope map (a), an example of a roughness map, obtained using texture detection (b), and a map indicating which locations are insufficiently illuminated (c). In (a) we can distinguish three craters, two smaller ones in the middle of the map and part of a larger one at the bottom-right corner. The colour shading in the bottom-corner one indicates that there are larger slopes in that crater. Figure (b) shows one big crater, which is detected by the algorithm as a region of higher texture at its rim. Lastly, (c) shows regions that are too dark for landing in red. The blue regions are sufficiently illuminated and safe for landing. These maps would be passed on to a piloting function, which would then be able to select a new landing site based on this information, if required.

Hazard detection and avoidance would be performed once or twice during a descent. Mostly, the first run is executed at higher altitudes, roughly at 1–2 km, where the slopes can already be assessed, but roughness features cannot be identified yet. The second run is then executed at lower altitudes where the resolution is sufficient to resolve roughness features. This is done at altitudes of about 200 m. One hazard evaluation can take up to 2 s. Therefore, continuous evaluations are not possible. Furthermore, continuous evaluation could also lead to problems with the guidance system, if the landing location would change frequently. But it would also not be required to keep updating the hazard maps, because as soon as the resolution is sufficient to resolve the maximum allowable roughness size, further mappings should not lead to better (different) decisions.

Mentioning *resolution*, how can it be computed? Generally speaking, computing the resolution is the same procedure for both active and passive sensors. The reso-

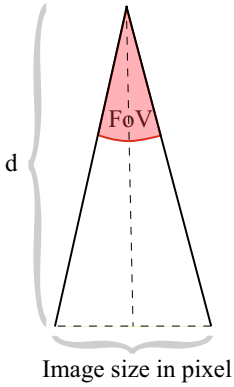


Fig. 9.79 Computing the resolution

lution depends on the field-of-view (FOV) of the sensors, the distance between the sensor, and the surface (d) and the size of the image in pixels (px). This geometrical problem is presented in Fig. 9.79.

Based on this figure, it can be derived that the resolution of the image in one dimension can be computed from:

$$\text{res} = \frac{\tan\left(0.5 \cdot \text{FOV}\right) \cdot d}{0.5 \cdot \text{px}} \tag{9.167}$$

The resolution can be different in x- and y-direction depending on the sensor design. Using this formula it is possible to compute the achievable resolution in both x- and y- direction. However, for detecting roughness, the resolution in z-direction, sometimes also called depth resolution, is a very important factor. Unfortunately, this resolution cannot be derived from geometry, but is rather a sensor property. Therefore, it is very important to select a sensor that can reconstruct the depth well enough to resolve the desired maximal allowable roughness.

9.5.2.5 Complete System

In the previous sections three advanced system have been discussed that can be included in a landing system to improve the landing accuracy and enable landings in terrains that are not entirely free of hazards or even unknown prior to landing. These three systems are: (1) terrain absolute navigation for increasing the inertial navigation error during the earlier approach phase at high altitudes, (2) terrain relative navigation, to counteract the IMU error at lower altitudes, where terrain absolute navigation is no longer possible, and (3) hazard detection and avoidance, which can enable landing in unsafe regions. If you recall the flow chart of the GNC system from Chap. 7 (Fig. 7.1 for a complete overview and Fig. 7.2 for a simplified representation), one may remember that this system consists of one block each for guidance, navigation, and control. The new, more advanced system will include additional blocks for tasks

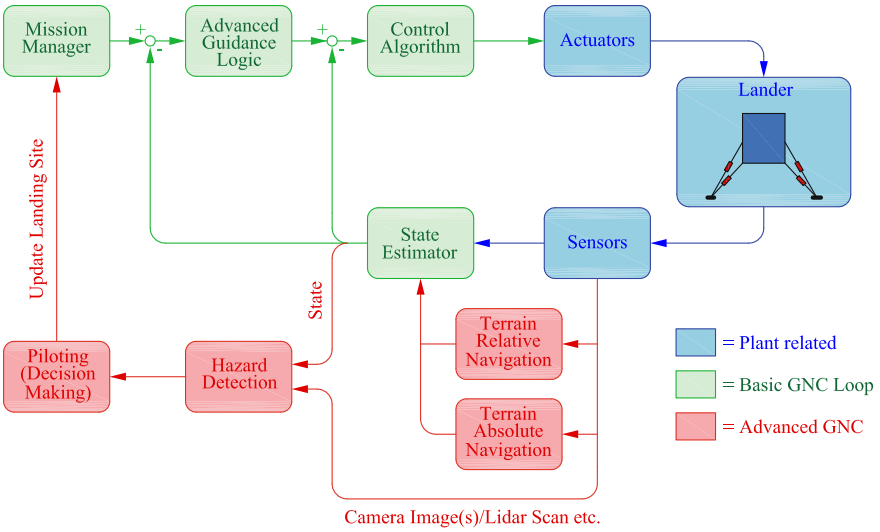


Fig. 9.80 Flowchart of a GNC system including HDA, TAN, and TRN

of TAN, TRN, and HDA. Also it shows that each of these systems needs inputs from the spacecraft's sensors. The extended flow chart is shown in Fig. 9.80.

As one might already have read in between the lines, to date no mission performed the full set of TAN, TRN, and HDA, although this could very well be partially true. In 2013, the Chinese Chang'e-3 mission landed on the Moon. It was claimed that this mission performed HDA during the final stages of its descent. Li et al. (2016) describe the mission and system design of Chang'e 3, as well as the guidance algorithms of powered descent and soft landing, which are also assessed in detail. The hover and hazard-avoidance phase began 100 m above the nominal landing site. After detecting a hazard the lander had the freedom to move horizontally. It would then continue in a slow descent, and from 4 m altitude the free-descent phase started, ending with the landing. The elevation map obtained by the 3-D laser imaging sensor showed several potential hazards: some 1–2 m craters in the northwest, north, and southeast of the landing site, and some 0.2–0.5 m holes and stones in the southwest, south, and east of the landing area. Chang'e-3 autonomously chose the relatively flat middle zone as the safe landing site, which was 4.5 m away from the nadir point (=nominal landing site).

The German Aerospace Center (DLR) developed their own crater navigation system (CNav) for precision landing on the Moon (Maass et al. 2020). Using actual navigation camera images acquired during the Chinese Chang'e-3 landing mission they tested their algorithm down to an altitude of 350 m. Without having access to the camera-calibration data, nor any IMU data, the algorithm successfully estimated the entire Chang'e-3 kinematic vehicle state during the powered descent. Having reached TRL 6 of this TAN algorithm, the next step would be to flight-qualify the system.



(a) Landing in boulder field (b) Vehicle after landing

Fig. 9.81 *Morpheus* test vehicle. Images: NASA

For the ESA Lunar Lander the full set of TAN, TRN, and HDA was supposed to be flown as a technology demonstration, but unfortunately, due to funding problems, this development was stopped at the end of phase B2, in 2012. Some details were published, and can be found, for example, in the papers by De Rosa et al. (2011), Carpenter et al. (2012) and Parreira et al. (2013).

As was the case with ESA, NASA is also working on a technology demonstration of TAN, TRN, and HDA. However, they perform their development outside of the scope of a real mission. Their system is called ALHAT which stands for *autonomous landing and hazard avoidance technology* (Epp and Smith 2007b). In spring 2014, the ALHAT hardware and software suite was flown on the *Morpheus* test vehicle, using a lidar as the active sensor (Epp et al. 2014). Carson III et al. (2015) discuss some of the results of the program, focussing on interfacing and verifying ALHAT safe precision landing systems with the *Morpheus* vehicle. The very successful free-flight test campaigns provided valuable performance data on the ALHAT sensors and raised the overall technology readiness level of the ALHAT system to 6. Figure 9.81 shows the *Morpheus* test vehicle, both during one of the flight tests, i.e., it landing in a boulder field (and autonomously identifying and avoiding the boulders), and after landing.

9.6 Landing Systems

Landing on the surface of a celestial body can either be on a hard surface, the “land”, or a liquid surface, the “sea”. Obviously, the terms land and sea may have a different interpretation for different bodies, but for the sake of the discussion we will be using these terms. A land landing is most commonly done, giving the examples of Lunar, Mars, Venus, and asteroid landings, but it is also the preferred landing environment for all Russian capsules on Earth. And, obviously, an aircraft-like vehicle, such as the Space Shuttle, requires a runway for its final mission phase.

Russia (and the former Soviet Union) is know for landing its capsules on land. Starting with the flight of Vostok-1 that would bring Yuri Gagarin into space, all Soyuz capsules that came after were designed for a land landing, usually somewhere

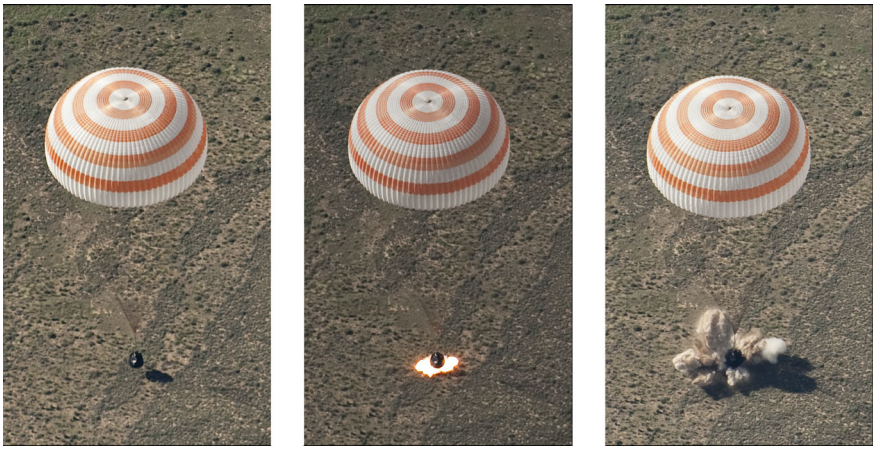


Fig. 9.82 Soyuz TMA-17 firing its retro-rockets just prior to landing. Image: NASA (through WikiCommons)

in the deserts of Kazakhstan in central Asia. Major part of the re-entry capsule's velocity is taken out by its large main parachute, yet its landing velocity would still be around 8 m/s, which is too uncomfortable for the astronauts to endure. Therefore, a sequence of events is initiated at an altitude just under 6 km. The first of these events is the jettisoning of the heat shield that has protected the capsule against the brutal entry environment, which, in fact, will also reduce the fall rate under the parachute due to the lower capsule mass. Once removed, the capsule's six solid-propellant landing engines are exposed, and will operate just before it touches the ground (at an altitude between 0.8 and 1.1 m). The landing engines reduce the capsule speed to a value in the range between 0 and 3 m/s. In Fig. 9.82, this sequence has been depicted, whereas Fig. 9.83 shows the discarded heat shield and exposed landing engines, integrated in the capsule's bottom structure.

Even though landing engines are often used to reduce the impact velocity, they will not be further discussed here. In Sect. 9.3, the powered descent of planetary probes and Lunar landers has been discussed, although a hardware-related discussion was considered to beyond the scope of that section. General details on space-related propulsion systems can be found in the book by Turner (2006), whereas more specific details on engines used on planetary landers are treated by Ball et al. (2007).

Given the option for either landing technique, there are obvious advantages and disadvantages to be identified for the selected one. The advantages of water landing are (1) good impact attenuation, (2) a large landing area is available, which eliminates the need for a precision landing, (3) if proper reorientation for water impact is achieved, failure tolerance is no problem, and (4) precise timing might not be needed. There are, of course, also notable disadvantages, which are related to (1) a dependency on the sea state, (2) the integrity of the hull after landing to provide additional flotation, (3) breaching of the TPS for the flotation devices to be deployed,



(a) Jettisoned heat shield. Image: ESA

(b) Landing engines Soyuz-TMA-13 (Adapted from original image: Maxim Marmur, Space Adventures)

Fig. 9.83 Soyuz capsule after landing

(4) complex sealing, (5) search and rescue facilities (ships, helicopters), as well as dedicated equipment, (6) possible reorientation to reduce water impact (e.g., flare manoeuvre), and (7) design and integration of flotation devices.

Independent from the type of landing, they have in common that after touching the landing surface, the vehicle's terminal velocity will be reduced to zero, mostly in an extremely short time. This can cause a large impact load on the system, which has to be as low as possible to avoid damage to the vehicle and/or its payload. After landing, when it concerns a return mission to Earth, the vehicle will have to be recovered, either for post-flight analysis or for payload extraction.

This section will focus on the different options of landing systems, typically related to the landing of capsules and planetary landers; the rolling landing of spaceplanes is considered to be beyond the scope of this introductory text. We will start off with sea landing (Sect. 9.6.1), and discuss a simple analytical model to determine the impact loads during vertical landing. In Sect. 9.6.2 an overview of several shock-attenuation systems for land landing will be given: starting off with the aeroshell penetrator, it is followed by crushable materials, landing legs, and landing bags. A good overview of the latter three is given by NASA (1970), which gives structures-related space-vehicle design criteria from the era of US manned space capsules (Mercury, Gemini, and Apollo). More details on this will be given later on.

9.6.1 Sea Landing

Besides the sea state, also several mission and vehicle characteristics are important to consider when determining the impact loads on the system. From the mission point of view, the flight-path direction and velocity will directly determine the magnitude and

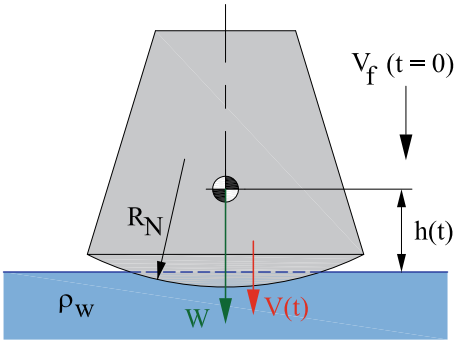


Fig. 9.84 Impact configuration of a capsule on a liquid surface

direction of the occurring deceleration of the centre of mass. The vehicle's orientation will affect the location on the vehicle where the loads will act on the structure, and is important input to determine the load paths. The vehicle's mass and geometry are obvious aspects from a vehicle point of view.

One can imagine that a full analysis requires a very advanced hydrodynamic analysis, which is far beyond the scope of the current discussion. For a basic understanding the reader would be best served with simplified analysis models that could also be used for a first estimate in conceptual design studies. We will present such models here, derived from impact theory using the principle of conservation of momentum. Linking these simple models with more advanced finite element analysis is done by Wang and Lyle (2007), and confirms that the analytical models serve well as first-order approximations.

It was Theodore Von Kármán, in the 1920s, who laid the foundation for classical impact theory for the vertical impact of a wedge at zero trim, to determine the loads on seaplanes when touching down on the water surface (Von Kármán 1929). Many scientists and engineers followed suit in extending the theory. Milwitzky (1952) generalised the impact theory for V-bottom-shaped seaplanes during a step-landing impact, and he compared the theoretical results of the motions, hydrodynamic loads, and pitching moments with experimental data. Derived from the non-linear models that can only be solved numerically, Milwitzky derived a simplified expression, which was also used by McGehee et al. (1959) to determine the water-landing characteristics of conical-shaped entry capsules having a segment of a sphere as the bottom. Part of the extensive experimental programs of the 1950s and 1960s, multiple drop tests provided sufficient experimental data to support the validity of the theoretical approach.

The model derived from the generalised setup of Milwitzky (1952) assumes a vertical velocity with respect to the liquid's surface (so $\gamma = -90°$), and such an orientation that the vehicle's X-axis is parallel to the velocity (and the bottom of the vehicle is parallel to the liquid's surface). The sea is assumed to be at rest, i.e., no waves or otherwise asymmetric loading on the vehicle. This baseline configuration is shown in Fig. 9.84.

Given the impact velocity V_f, the instantaneous capsule's velocity, $V(t)$, while traversing through the liquid with density ρ_w is given by

$$V(t) = \frac{WV_f}{W + 0.75\frac{2\pi}{3}\rho_w g_0 R_N^3 \sin^3\left[\arccos\left(1 - \frac{h}{R_N}\right)\right]} \tag{9.168}$$

where $W = mg_0$ is the weight of the capsule (with $g_0 = 9.80665$ m/s^2 is the (constant) gravitational acceleration), R_N is the equivalent radius of the bottom part of the capsule (typically the nose radius), and h is the current penetration depth in the liquid ($h = 0$ m at $t = 0$ s):

$$h = \int_0^t V(t')\mathrm{d}t' \tag{9.169}$$

The deceleration can subsequently be derived from numerically differentiating the velocity profile. In case the velocity would be determined at intervals Δt, then the deceleration is

$$a(t) = -\frac{V(t + \Delta t) - V(t)}{\Delta t} \tag{9.170}$$

As mentioned, the above model compared well with the test data (for vertical fall direction only), provided by McGehee et al. (1959), despite the fact that the model was originally derived for seaplanes with a V-shaped bottom. Hirano and Miura (1970) derived a model to study the water impact accelerations of axially symmetric bodies, i.e., spheres and cones. With test results they checked the validity of the model, and the results were quite satisfactory. The model for a rigid capsule with a spherical bottom will be presented here. Starting point is yet again based on the Von Kármán approach, with the following equation expressing conservation of momentum:

$$mV_f = (m + m_w)V(t) \tag{9.171}$$

where m_w is the "virtual" mass of the displaced water.

Assume the situation that the (spherical) bottom part of the capsule is partially immersed in water. The virtual mass is in that case given by

$$m_w = \frac{4}{3}\rho_w h^{1.5}(2R_N - h)^{1.5} \tag{9.172}$$

where, as before, ρ_w is the density of the liquid (=water), h is the penetration depth into the water, and R_N is the radius of the spherical bottom. The instantaneous velocity, $V(t)$, of the centre of mass of the rigid body is simply

$$V(t) = \frac{\mathrm{d}h}{\mathrm{d}t} = \frac{V_f}{1 + \frac{m_w g_0}{W}} = \frac{V_f}{1 + \frac{m_w}{m}} \tag{9.173}$$

After substituting Eq. (9.172) into Eq. (9.173), $V(t)$ can be written as

$$V(t) = \frac{V_f}{1 + \frac{8\rho_w g_0 R_N^3}{3W}\left(\frac{h}{R_N}\right)^{1.5}} \tag{9.174}$$

which, after differentiating, yields an expression for the overal acceleration, $a(t)$ (Hirano and Miura 1970):

$$a(t) = \frac{\mathrm{d}^2 h}{\mathrm{d}t^2} = -\left\{ \frac{3\sqrt{2}\left(\frac{3W}{4\pi\rho_w g_0 R_N^3}\right)^2 \frac{V_f^2}{g_0 R_N}\sqrt{\frac{h}{R_N}}}{\pi\left[\frac{3W}{4\pi\rho_w g_0 R_N^3} + \frac{2\sqrt{2}}{\pi}\left(\frac{h}{R_N}\right)^{1.5}\right]^3} \right\} g_0 \tag{9.175}$$

Assuming that $\frac{b}{R_N} << 1$, the maximum acceleration can be found as

$$a_{max} = -\frac{256}{243}\left(\frac{4\rho_w g_0 R_N^3}{3W}\right)^{\frac{2}{3}} \frac{V_f^2}{R_N} \tag{9.176}$$

with the corresponding impact time being

$$t_{max} = \frac{21}{160}\left(\frac{3W}{4\rho_w g_0 R_N^3}\right)^{\frac{2}{3}} \frac{R_N}{V_f} \tag{9.177}$$

and penetration depth

$$h_{max} = \frac{1}{8}\left(\frac{3W}{4\rho_w g_0 R_N^3}\right)^{\frac{2}{3}} R_N \tag{9.178}$$

One possible extension to this model is to include the rise of water due to the splash up, which was not considered by Von Kármán. Wagner (1932) found that splash up has a significant effect on the impact force. Let this impact force be defined by

$$F_{impact} = C_s \frac{1}{2}\rho_w V_f^2 \pi R_N^2 \tag{9.179}$$

where C_s is the so-called non-dimensional *slamming* coefficient,[31] and in line with, for instance the definition of the aerodynamic forces, πR_N^2 is defined to be the reference area. Following the approach of Wagner, Milo (1991) analytically derived an expression for C_s, as a function of h/R_N:

[31] During higher sea states, a ship typically experiences large pitch and heave motions. The front part of the ship can then rise above the water surface. When it re-enters the water, large impulse pressures are transmitted to the hull structure, as a result of the relative motions of the sea and ship. The hull literally slams into the water surface.

$$C_s\left(\frac{h}{R_N}\right) = 5.50\left(\frac{h}{R_N}\right)^{0.5} - 4.19\left(\frac{h}{R_N}\right) - 4.26\left(\frac{h}{R_N}\right)^{1.5} \tag{9.180}$$

This expression was derived under the assumption that the impact would take place vertically, and that during the impact the velocity would remain constant. This means, that effectively the equation only holds for the initial-stage slamming. With Eq. (9.180), the maximum deceleration can be estimated as:

$$a_{max} = -\frac{g_0}{2W} C_s\left(\frac{h}{R_N}\right) \rho_w \pi R_N^2 V_f^2 \tag{9.181}$$

In general, Eq. (9.181) will lead to larger decelerations than Eq. (9.176).

As mentioned, the expression for C_s has been derived for constant velocity during impact. This means that the deceleration profile cannot be determined, beyond the maximum value given by Eq. (9.181). However, with a small adjustment a reasonable deceleration profile can be obtained as follows. Using the definition of F_{impact}, Eq. (9.179), at each time step let us recompute the velocity according to Eq. (9.174), and with an updated value for h, Eq. (9.169), C_s is re-evaluated with Eq. (9.180). The deceleration can then simply be determined by

$$a(t) = \frac{F_{impact}}{m} \tag{9.182}$$

One other aspect that appears to be important to take into consideration when going into more advanced analysis models, is the flexibility of the structure. Studies from the Apollo era indicated that the interaction between the water and a flexible heat-shield structure would generate much larger impact forces than those of a rigid capsule (Benson 1966; Stubbs and Hathaway 1969). Consequently, the deceleration of the centre of mass would be proportionally larger. However, this would only occur up to impact angles of about 30°; impacting at larger angles did not show a visible increase. This was also confirmed by Wang and Lyle (2007) with their simulated finite-element method.

Example 9.8 (*Deceleration Loads During Sea Landing*) To get a feeling for some numbers, consider the following example. An Apollo-like capsule with mass $m =$ 4,976 kg and a heat-shield radius of $R_N = 4.7$ m impacts the ocean with a final velocity of $V_f = 8.3$ m/s. Let us assume a (salt-) water density of $\rho_w = 1{,}025.6$ kg/m^3. Subsequently, the deceleration profile is calculated with the "seaplane" model, Eqs. (9.168) and (9.170), the model of Hirano and Miura (1970), Eq. (9.175), and the one adapted from Wagner (1932) that includes slamming, but varies the velocity during the impact, i.e., Eq. (9.182). Moreover, the points of maximum acceleration will be evaluated, using Eqs. (9.176) and (9.181).

The results are shown in Fig. 9.85. It shows that the seaplane model with V-shaped bottom and the model with spherical bottom are, in fact, quite close. The model including the slamming exhibits a significantly larger deceleration, which confirms the importance of slamming. In any case, the duration of the impact until

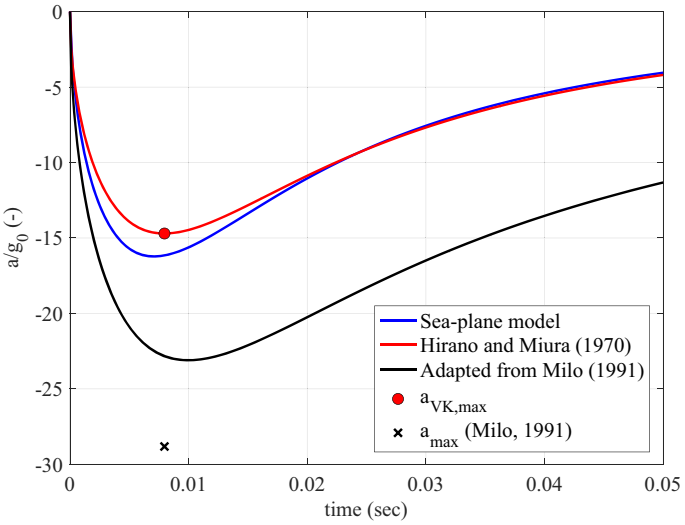


Fig. 9.85 Deceleration profile for the sea impact of an Apollo-like capsule: model comparison

the deceleration reduces to acceptable values is very short ($t \approx 0.05$–0.1 s). The two points for a_{max} can be considered as two limit cases, in between which the real maximum deceleration will occur. However, keeping the velocity constant at the (largest) velocity value seems to be given an overestimate, when compared with the model that adjusts the velocity. This curve seems to be confirmed by Wang and Lyle (2007), who did a finite-element calculation of the maximum deceleration. Their conclusion was indeed that the real value would be in between the two limit cases, but their results show that in fact it is reasonably close to the peak value with the adjusted velocity. Our recommendation would therefore be to use this model for conceptual design studies.

To study the sensitivity to some key parameters, we will indeed use this model and vary the impact velocity, $V_f \in [6, 8, 10]$ m/s, the bottom radius, $R_N \in [1.25, 2.5, 5]$ m, and the capsule mass, $m \in [1{,}250, 2{,}200, 5{,}000]$ kg. Each parameter will be varied separately, whereas the other two will be kept at their centre value.

The results of this (limited) analysis are plotted in Fig. 9.86 (it is noted that the centre lines, colour coded "red", are the same in each plot). From these results, some obvious conclusions can be drawn. It is clear that lowering the impact velocity reduces the kinetic energy that needs to be absorbed, which lowers the deceleration peak. Between the lower and higher V_f (6 m/s vs. 10 m/s), there is a difference in *g-load* of about 17 g. Also the gradient is significantly smaller, and the peak occurs slightly later (although with this time scale it would not be noticeable).

A similar conclusion can be drawn when varying the bottom radius: a larger radius will significantly increase the deceleration. For the current range of R_N, the difference in *g-load* is even larger than the one for V_f. Enforcing a smaller bottom radius could be enforced by an asymmetric impact attitude, which is easily achieved

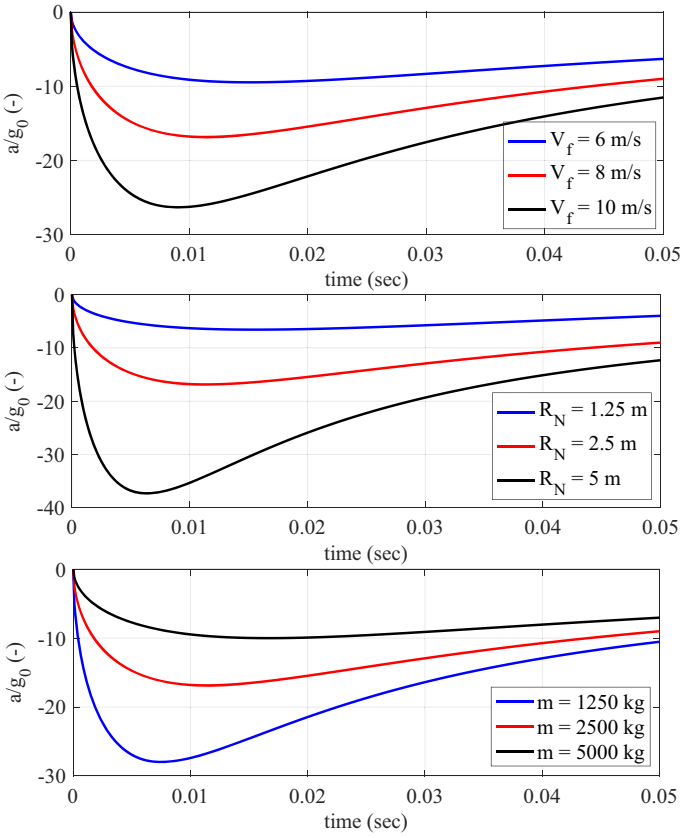


Fig. 9.86 Deceleration profiles for the sea impact of an Apollo-like capsule: variation of V_f, R_N, and m

by the connection to the parachute. In Fig. 9.87, the asymmetric attitude of a model of the Orion capsule during a drop test can be seen. The (much) smaller corner radius will significantly lower the impact loads. It should be noted, though, that when the capsule is oscillating under the parachute, or just be tilted by a horizontal wind, this might negate the effect.

Last, but not least, the variation of mass, ranging from 1,250 to 5,000 kg, has an inverse effect on the deceleration. The heavier vehicle will penetrate deeper into the water, and will slow down less rapidly compared to the lighter vehicles. Thus, the corresponding *g-load* will also be lower.

To conclude, it is obvious that the impact deceleration can become quite large, and in particular for manned missions special care should be taken to lower this as much as possible. Different options can be considered, for instance, to include shock absorbers in the support structure of the astronauts' seats, or to use some form of crushable structure in between the heat shield and the upper part of the capsule. These



Fig. 9.87 Orion parachute drop test on May 1, 2013 (Image: NASA)

particular design aspects will not be treated here, though, but some shock-attenuation techniques will be covered in the next section. ■

9.6.2 Shock-Attenuation Techniques

Lander shock-alleviation techniques are used to absorb some of the energy while touching down or impacting. An active way of doing this is to use landing engines or rockets to give a short boost prior to landing, such as Soyuz capsules do. When the landing rockets are used for a longer time, and/or at a higher altitude than just a few meters, these rockets are also called retro rockets, and they are typically used for Mars missions. The application of these have been discussed before, in Sect. 9.3.

Passive means can take a variety of forms, but in essence they can be grouped into hard and semi-hard landing assists. Main type of the former group is formed by so-called penetrators that will be discussed first (Sect. 9.6.2.1). Common options for the semi-hard landers are crushable materials, landing legs, and landing bags. Each of these concepts will be briefly discussed in the sub-sections following, i.e., Sects. 9.6.2.2–9.6.2.4.

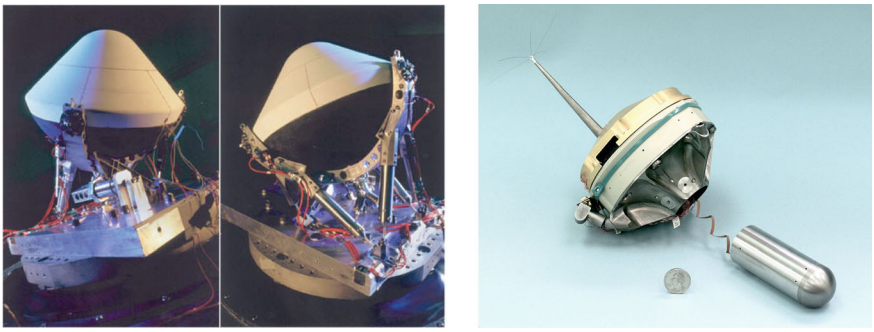
9.6.2.1 Aeroshell Penetrator

Aeroshell penetrators are not actual deceleration mechanisms, at least not meant for aerodynamic braking. Typically, they are light-weight "pointy" vehicles that will marginally decelerate during their atmospheric flight, and their kinetic energy will only be absorbed during impact with the surface. The more sensitive equipment can be mounted on a platform and the loads on this platform can be limited through, for instance, the crushing of an aluminium honeycomb skirt (see later in this section for more discussion on crushable structures). The design of the vehicle should aim for very low mass, as to reduce the impact load, and the vehicle should preferably be passively stable, such that it is guaranteed that the front part of the vehicle will impact the surface. Of course, the use of penetrators is not confined to operation in an atmospheric environment, e.g., the Japanese Lunar-A penetrators were meant to deploy seismometers and make a heat-flow measurement on the Moon.

Over the years, many penetrator concepts have been proposed, and in several cases they had gone into hardware testing before the projects were cancelled or (indefinitely) delayed. Lorenz (2011) provides an extensive overview of the history and future prospects of penetrators, covering (early) Mars proposals, and missions to Mercury, the Moon, Titan, asteroids, comets, and the Galilean satellites. We will not go into great detail and discuss all of them, but restrict to two missions that have actually been launched: Mars-96 and Deep Space 2.

The Russian Mars-96 mission carried two 1.7 m long penetrators to be released from orbit around Mars (Surkov and Kremnev 1998). Even though Mars-96 did not manage to leave Earth orbit, it is worthwhile to briefly mention this mission, as the penetrators were built and tested. The descent and landing of each penetrator would be identical, beginning with it spinning up for stability, followed by separation from the orbiter. At a safe distance from the orbiter a solid-propellant engine would ensure deorbitation and targeting towards the atmosphere, which it would reach after 20–22 h. Before entry, a 3.6 m inflatable braking device would be deployed that would enforce an impact velocity of about 80 m/s. At impact, the 14-cm diameter forebody would separate from the aftbody and delve 5–6 m into the surface, while the after-body remained on the surface connected to the forebody by wires. Deceleration loads on the aftbody would be limited to 500 g by a shock attenuation system. Each penetrator was powered by a radio-isotope thermoelectric generator (RTG) and a battery, and its expected lifetime was one year.

The Deep Space 2 probes were two miniature space probes the size of a basket ball that were flown as payload aboard the Mars Polar Lander in January 1999 (Fig. 9.88). Very light-weight—each probe weighed 2.4 kg—they were the first spacecraft to use only a rigid aeroshell for stability and (marginal) deceleration. The probes were designed such that they could passively reorient themselves from almost any entry attitude, including entering completely backward or with a small initial tumble rate. The aeroshell remained attached all the way to the surface and would shatter at impact. Then, the penetrator would separate into two parts: an aftbody that remained at the surface due to its larger diameter, and a forebody that disappeared under the



(a) Probe with aeroshell (b) Probe with penetrator

Fig. 9.88 Deep Space 2. Images: NASA

surface due to its narrow, bullet-like shape and the high impact velocity of about 180–200 m/s (see also Fig. 9.89).

It was predicted that the deceleration acting on the forebody could be as high as 30,000 g (with g being normalised to Earth conditions) and that it would penetrate between 0.2 and 0.6 m, depending primarily on the ice content of the soil (Smrekar et al. 1999). Because the aftbody would come to a standstill more abruptly, the deceleration would be up to 60,000 g, and coming to rest in a crater with a diameter of roughly 0.5 m, and it could be 0.1 m deep. A 1-m tether for relaying power and data connected the two sections of the probe.

Unfortunately, both probes failed to establish contact after landing, and the mission was declared a failure on March 13, 2000. Possible causes for the failure may have been (JPL Special Review Board 2005):

1. Due to the high impact load, the probe radio equipment and/or batteries may not have survived,
2. The probes may have bounced on impact or may have landed on their sides (causing a problem with the antenna performance or radio link geometry),
3. The probes may have simply hit ground that was too rocky (hard) for survival, or
4. The batteries on the probes might not have retained sufficient power.

although the actual cause could not be established.

9.6.2.2 Crushable Material

The total energy change involved with a landing on a hard surface is obviously linked to the velocity the vehicle has just prior to impact and its mass. The load that this energy change induces is also depending on the planetary surface, in the sense how much energy is absorbed by the surface. The softer or more elastic the surface material is, the more energy will be absorbed. Ever since the design of the first

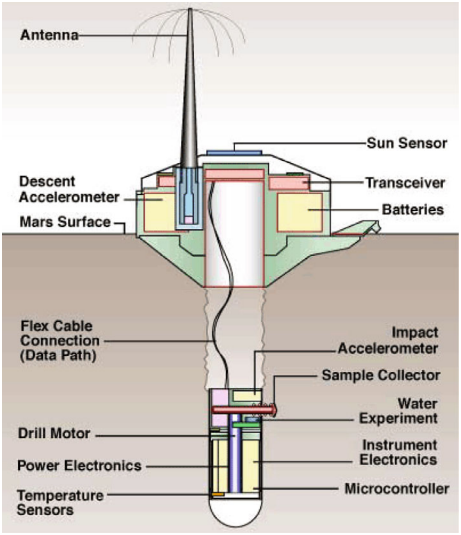


Fig. 9.89 Layout of the major Deep Space 2 components. Image: NASA

planetary landers, many concepts of reliable, lightweight, energy-absorbing elements have been designed and tested. This not only involved more advanced mechanisms, such as landing legs, but also a simple layer of crushable material that absorbs the landing energy by the plastic-deformation process. Or, to be more precise, assuming that there is no change in potential energy, the conversion is that of kinetic energy directly to thermal energy.

To get an impression on how the material would deform, Fig. 9.90 shows the result of an impact test with a loose-pack tubular cell honeycomb crushable structure, one of several honeycomb structures that were tested in the 1960s (McFarland 1964). Several conclusions were derived concerning wall thickness, packing density, and the relation with internal loads and energy absorption, which are beyond the scope of the (very) brief introduction provided here. An example application is shown in Fig. 9.91, though. This figure shows the use of honeycomb crushable structures just under the bottom plate of the Mercury space capsule. At sea impact, this would reduce the impact forces felt by the astronaut, who was seated directly above it, see, for instance, the report by NASA (1970).

A nice overview of the different techniques is also provided by Doengi et al. (1998), who did a theoretical and experimental evaluation of several shock alleviation techniques in the framework of ESA's Technology Research Programme. A more recent test campaign is given by Schröder et al. (2020), who used DLR's landing and mobility test facility (LAMA) to test a crushable shell for small-body landers.

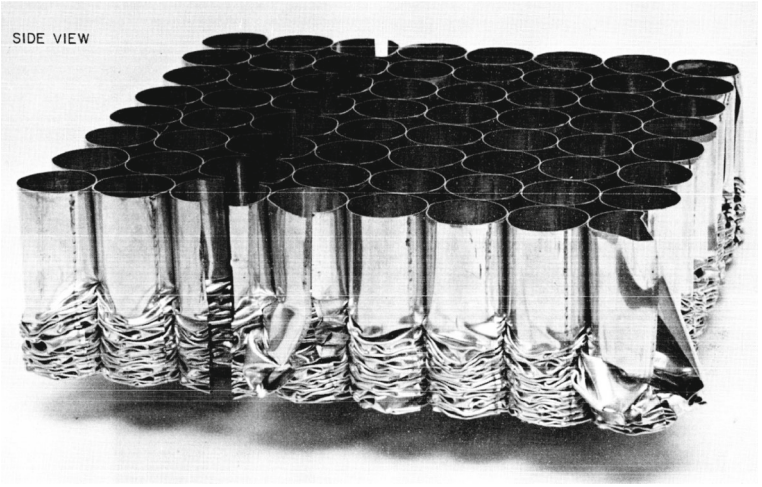


Fig. 9.90 Testing of a crushable structure, partially collapsed: low-unit-weight, loose-pack tubular cell honeycomb (3003 H-19 aluminium), with a wall thickness of 0.229 mm (0.009 inch). Image: JPL, taken from McFarland (1964)

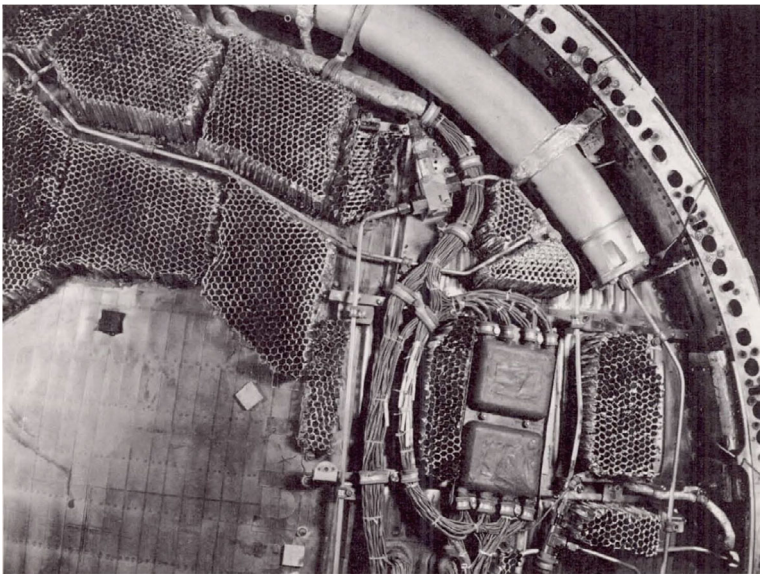


Fig. 9.91 Honeycomb structure just below the bottom plate of the Mercury capsule, to absorb part of the impact forces. Image: NASA, taken from Grimwood (1963)

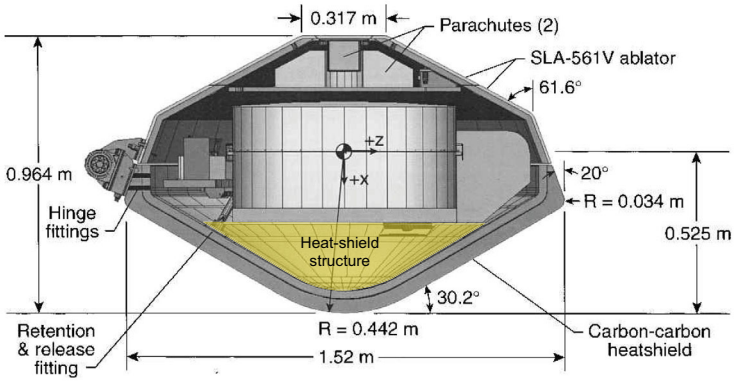


Fig. 9.92 Genesis sample return capsule, with the underlying heat-shield structure highlighted (Original image: NASA)

A good example of such a "crushable" material is found in the sample-return capsule of the Genesis mission (see also Fig. 9.119 for the launch configuration). After completion of Genesis' mission, i.e., the collection of solar-wind particles for two years, the collected samples were stored in a sample container inside the sample return capsule (SRC), see Fig. 9.92. Such a design is often used for sample return capsules, e.g., Stardust and Osiris-REX, although, the heat-shield structure could also be replaced by a structural foam. The nominal landing operation is typically by parachute, but even then the structure will absorb some of the landing energy. In case the parachute system fails, as was the case with Genesis, a properly designed heat-shield structure can provide a lot of extra protection.

Of course, Genesis was not the first to benefit from energy-absorbing structures. Apart from energy absorbers in landing legs, which will be discussed shortly, in the structural design of the Russian Venera landers of the 1960s and 1970s a crushable structure was included: the base of the lander was a shock-absorbing, crushable ring (see Fig. 9.93). Schroeder et al. (2018) revisited the structural design of Venera, to understand how the the kinetic energy on landing was dissipated, such that the gained insight can be valuable for designing a new Venera-class lander for future missions. They developed a low-fidelity analytical model and a high-fidelity FEM model, allowing variation of geometrical parameters, material selection, and leg configuration. Once promising designs were identified with the analytical model, verification and detailed analysis was done with the FEM model.

A (very) simple analysis model that can give a first-order idea of the decelerations involved, is derived as follows. Let us assume that all energy is absorbed by the crushable structure, e.g., the surface is infinitely hard and will not deform. Further, the vehicle with mass m has an impact velocity, V_f. Thus, the kinetic energy at impact is equal to $\mathcal{E}_k = \frac{1}{2} m V_f^2$. The impact will crush the structure to a stroke length, s, and the lander will not bounce up. Impact direction is perpendicular to the surface.

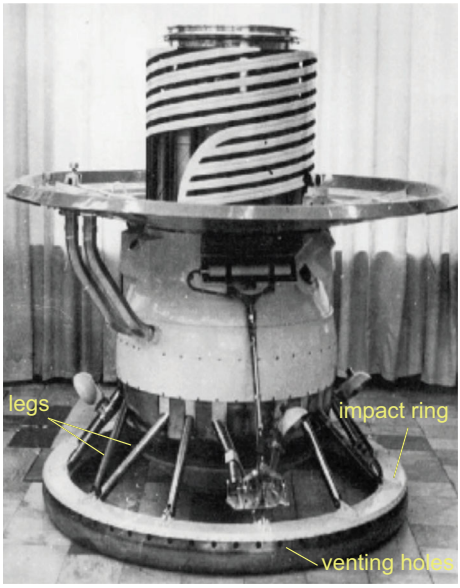


Fig. 9.93 Venera-9 landing structure, with a crushable impact ring and connecting energy-absorbing legs (Original image: NASA)

The work exerted on the crushable structure will in that case be equal to

$$W = \int_0^s F(y)\mathrm{d}y \tag{9.183}$$

where the impact force may vary with the amount of compression. However, if we assume that this force is constant, considering the small time scale, the above expression changes to $W = Fs$. If the vehicle has come to a stand-still, this means that the work should be equal to the kinetic energy, thus

$$\mathcal{E}_k = \frac{1}{2} m V_f^2 = F_{impact} s \tag{9.184}$$

and with $F_{impact} = m a_{impact}$ this means

$$\frac{1}{2} m V_f^2 = m a_{impact} s \quad \Rightarrow \quad a_{impact} = \frac{1}{2} \frac{V_f^2}{s} \tag{9.185}$$

such that the impact acceleration can be determined, given the impact velocity and a maximum (available) stroke length.

Duration of the impact follows from

$$t_{impact} = \frac{a_{impact}}{V_f} \tag{9.186}$$

However, suppose that the impact force is not constant, but would follow, let us say, a sinusoidal pattern (half a wave), then the maximum deceleration can be obtained as follows. We know from integrating a sine function that half a wave has a surface area of 2, and we can easily show that the average height of half this wave is the surface area divided by the interval width, which is equal to π.[32] Thus, it would suffice to multiply the (average) acceleration (corresponding to the height of the sine wave), Eq. (9.185), by $\frac{\pi}{2}$ to get the maximum acceleration.

Example 9.9 (*Deceleration Loads with Crushable Materials*) In this example we will have a look at the effect of impact velocity and available stroke length on the duration of impact and corresponding deceleration. Note that the vehicle mass does not explicitly appear in the simple model derived above. However, one can easily understand that a heavier (lighter) vehicle will lead to a larger (smaller) impact velocity, if all other descent-system parameters remain the same. Any other mass effect will not be considered here.

The velocity range that is considered is $V_f \in [10, 50]$, whereas the stroke lengths are $s = 0.1$ m, 0.2 m, and 0.3 m. Calculating the occurring (constant) deceleration and duration, using Eqs. (9.185) and (9.186), as well as the maximum deceleration from a sinusoidal deceleration. The result is shown in Fig. 9.94. It is clear that the deceleration is large and the duration very short, just what one would expect. A longer stroke length will lower the loads, and the quadratic increase due to a larger impact velocity is also clear. It was shown that while analysing the load profile, the choice of deceleration profile is important: in case of a sinusoidal profile the load will be about 50% larger compared to a constant one. ■

On a final note: in Sect. 9.6.2.1 the penetrator was discussed, with very large impact loads. With the above simplified method we can estimate these loads as follows. Given $V_f = 200$ m/s (the maximum value for Deep Space (2), it was predicted that the penetration depth in the soil would be between 0.2 and 0.6 m. Consider this to be the stroke length, and the soil the "crushable material", Eq. (9.185) gives a maximum deceleration of about 10,200 g, but with a sinusoidal deceleration profile this becomes already 16,000 g. One can imagine that when the surface is "rockier", the deceleration profile may be a lot more abrupt, and the estimated 30,000 g does not seem to be unrealistic. What is certain, though, is that using this simplified approach may give an underestimation of the actual load on the system, and that it is important to use more high-fidelity models once the design goes beyond the conceptual level. Also, any crushable structure supporting the equipment platform should then be included in the analysis.

Such an analysis involves many steps, beginning with the entry vehicle's descent profile, velocity, orientation, and the properties of the landing site (e.g., soil properties

[32] By definition, the average value of a function over an interval $[a, b]$ is

$$\bar{f} = \frac{1}{b-a} \int_a^b f(x) \mathrm{d}x. \tag{9.187}$$

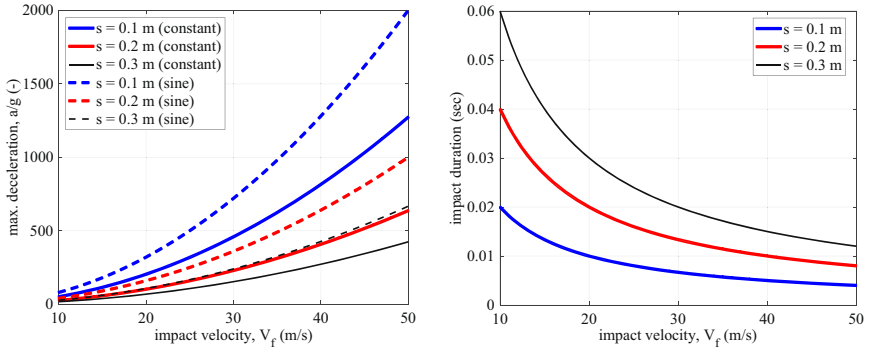


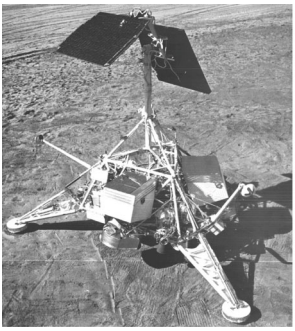
Fig. 9.94 Maximum deceleration and impact duration as a function of impact velocity and maximum stroke length

or ocean conditions). The simulation predicts how the vehicle's structure deforms upon impact, the distribution of impact forces, and the resulting energy absorption. The crushable structure that will absorb the energy should be modelled with the appropriate material properties and deformation characteristics. Analysis must show how it compresses and deforms during impact, including the effects of buckling and and potential failure. Sub-scale and possibly full-scale tests shall validate the analysis results and confirm the functionality of the design.

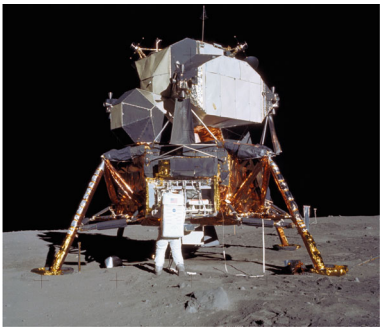
9.6.2.3 Landing Legs

Landing legs have been used on many missions, whether it was to the Moon, Mars, or even more exotic bodies, such as the *Rosetta* mission to comet *Churyumov-Gerasimenko* with a dedicated lander called *Philae*. In Fig. 9.95, four landers are shown, each with a different landing-leg configuration. What is obvious is that the Apollo Lunar Module has four legs, whereas Surveyor, Viking, and Philae each have three. Main consideration for the number of legs is how much impulse is to be transferred during the landing, depending on the lander's mass and the impact velocity. Four legs are typically used when the lander is heavy and/or the touch-down velocity is rather high. Having only three legs has the additional advantage, besides being a lighter structure, that when the landing area is more irregular a tripod has more chance to be stable than a four-legged configuration, something that we may have experienced while "wobbling" on a chair.

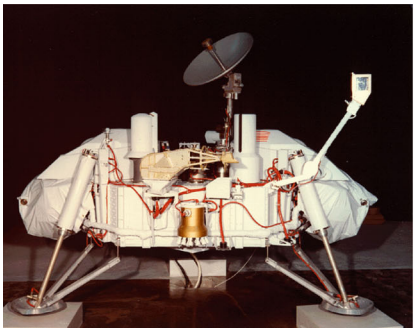
In the images of the Lunar Module and the Viking lander, the cylindrical shock absorbers are clear to be seen. But also the different ways of connecting the struts to the main body of the lander can be discerned. For instance, the primary strut with the absorber connects high up the Viking lander, but the three-legged configurations of Philae joins in a single point that is connected to the bottom plate of the lander's bus.



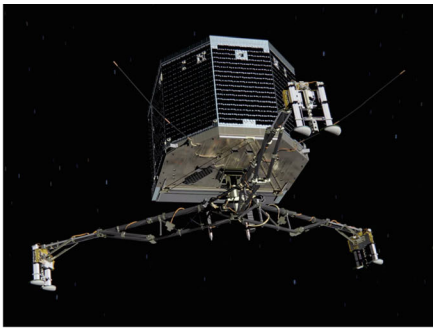
(a) Surveyor



(b) Apollo's Lunar Module



(c) Viking



(d) Philae. Image: ESA.

Fig. 9.95 Severeal landers with different landing-leg configurations. Images: NASA, except where indicated

We will not go into detail about the different design options and will also not discuss load cases in detail. Instead, a global overview of the main issues is provided, i.e., the different configurations from a geometrical point of view, how an initial sizing analysis of the absorbers can be done, and the stability of landers while touching down.

Figure 9.95 has displayed different landing-gear configurations. For (very) low gravity environments, in combination with a low lander-mass, the tripod configuration with a single-point fixture of Philae could be selected, with a wider footprint of legs. Of course, ground clearance taking rocks and/or boulders into account, should always be taken into account. For larger gravity and/or vehicle mass, connection to the side of the lander is usually chosen, either with three or four legs. For this type, the cantilever and the inverted tripod landing-gear designs remain the most common design for planetary landers, and a simplified two-dimensional representations of these has been illustrated in Fig. 9.96.

For the cantilever design, the secondary struts are connected to the primary strut, more or less half way, and to the main body structure. The upper section of the primary strut connects higher up to the lander's core structure, which could be as

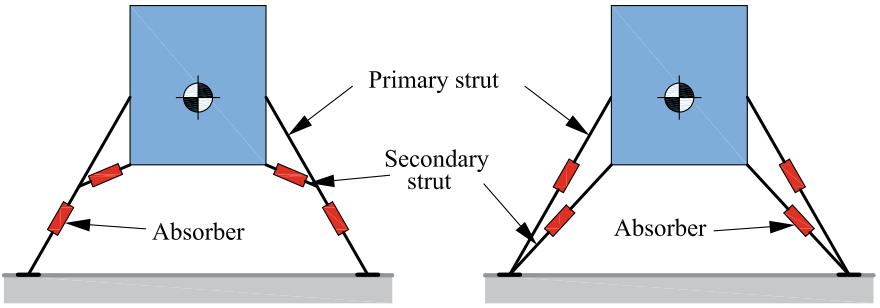


Fig. 9.96 Landing-gear design: cantilever (left) and inverted tripod (right)

high as the top of the structure. Both the lower section of the primary strut and the secondary struts have energy-absorber elements incorporated within the internal cylinder of the struts. The joints connecting the body structure and both primary and secondary struts are pivoted on ball joints for landing flexibility, or to allow for folding of the complete landing gear. A bending moment is the main load acting on the primary strut, because of the secondary-strut connection and the compressive loading from the energy absorber in the primary strut. The secondary strut is mainly loaded in axial direction.

The inverted tripod design has both the primary and secondary struts connected to the footpad. Also in this design, both struts have energy absorbers, although sometimes they are missing in the secondary strut. Similar to the cantilever design, the joints connecting the body structure and both strut are pivoted on ball joints. In terms of loading, both primary and secondary strut are subjected to axial loads.

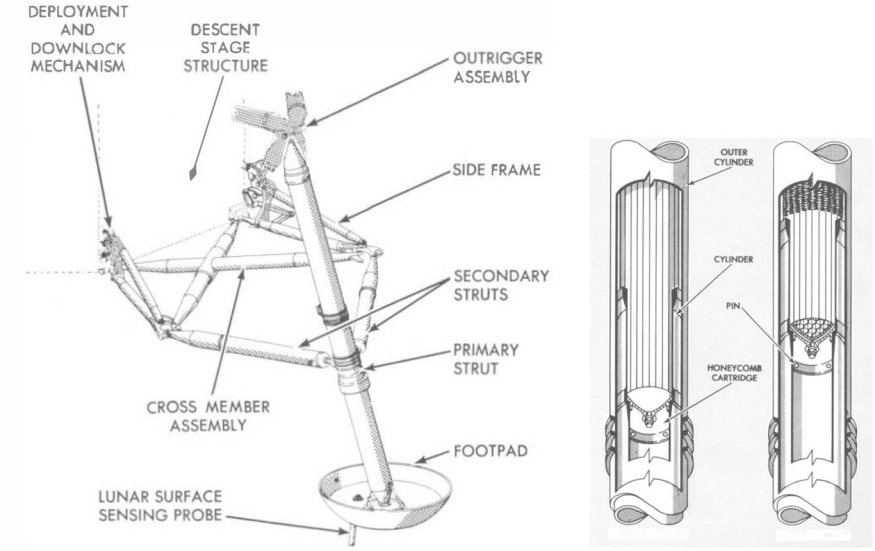
The energy absorbers in the struts can take different forms. Energy absorption can take place by crushing a (honeycomb) structure inside the absorbers, which obviously restricts the usage to a single landing only. Alternative, reusable solutions for use in the harsh space environment are considered as well, such as metal-bellows shock absorbers,[33] which have two sealed gas chambers and a piston connected to the metal bellow, electromagnetic absorbers, using a passive electromagnetic system as damping and a spring for resetting, and a variation that was implemented in the Philae lander to absorb kinetic energy by driving a damping motor (Witte et al. 2014).

An example of the more conventional crushable-structure absorber is shown in Fig. 9.97. The landing-velocity envelope for this Lunar Module's landing gear is a vertical velocity of 3 m/s (10 ft/s), with no horizontal velocity, or a 2.1 m/s (7 ft/s) vertical velocity with a horizontal velocity of maximum 1.2 m/s (4 ft/s), with a vehicle attitude within 6° of the local horizontal when the angular rate is 2°/s or less.

A simple evaluation of the absorber's stroke length can be done as follows. Given a vehicle with a mass m and impact velocity V_f will have an impact impulse p of

$$p_{impact} = mV_f \tag{9.188}$$

[33] https://www.metalbellows.com/assets/OTS_Bellows-1.pdf, accessed January 2021.



(a) Configuration (b) Energy absorber

Fig. 9.97 Apollo's Lunar Module landing-gear configurations and energy absorber. Images: NASA, taken from Bryan and Strasburger (1965)

The crushable legs will have a minimum force, a so-called threshold force of $F_{threshold}$, such that to equate p, the impact duration and acceleration will follow from

$$t_{impact} = \frac{p_{impact}}{F_{threshold}} \quad \text{and} \quad a_{impact} = \frac{F_{threshold}}{m} \tag{9.189}$$

Assuming that the work done by crushing the material is equal to reducing the kinetic energy to zero, it follows that

$$\mathcal{E}_k = F_{threshold} s_{crush} \tag{9.190}$$

Thus, as an example, when a lander with mass $m = 1{,}500$ kg touches down on a single leg, with velocity $V_f = 0.8$ m/s, the impact impulse to completely stop the motion is $p_{impact} = 1{,}200$ Ns. If a crushable leg with $F_{threshold} = 2{,}400$ N is used, the impact will occur in $t_{impact} = 0.5$ s and the deceleration the lander will experience is $a_{impact} = 1.71$ m/s^2. Finally, with a kinetic energy of $\mathcal{E}_k = 480$ J, the required crush length needed results in $s_{crush} = 0.2$ m. Obviously, when the lander would be landing on four legs (perfect alignment, only vertical velocity, and no rotation of the lander), the impact impulse will be distributed over the four legs, i.e., $p_{impact} = 300$ Ns per leg. With the same limit force of the crushable leg, the impact time is of course only 25%, i.e., $t_{impact} = 0.125$ s. The deceleration remains the same, but the

required crush length is also just 25% of the earlier found value: $s_{crush} = 0.0125$ m. Note that the legs are assumed to be a "vertical equivalent", which means that no inclination of the legs with respect to the local vertical is considered.

In the above discussion we have assumed that the lander either lands vertically on a flat surface, or possibly on one or two legs, if the lander is somewhat tilted. However, in reality the landing surface will never be perfectly flat, and in a worst-case scenario the lander could even touch down on a slope. Therefore, the inclination of the sloping surface is an important input to the (geometrical) design of the lander. To consider a three-dimensional configuration is not a trivial issue, and will be too complicated to discuss for the limited space that we have available. However, as with so many design issues, in the 1960s many of these issues were addressed in an approximate way for the design of the Apollo missions, and in this particular case, the landing on the Moon of Apollo's Lunar Module and its predecessor, Surveyor. Despite the perhaps "dated feel" of the (mostly analytical) methods, they are still valuable (conceptual) design methods.

Walton et al. (1964) performed a study on the touchdown stability for Lunar landing vehicles, to establish practical procedures for predicting tumbling motion of a vehicle that lands on a non-level surface. Even though landers typically have three or four landing legs, two-dimensional inelastic impact theory can be used for a first assessment, and the authors present some experimental results that confirm that this simple theory is indeed accurate. Therefore, the two-dimensional analysis will provide a good approximation to the real, highly complex three-dimensional case and can be used to estimate the performance and restrictions of a landing gear. Following this performance analysis it is possible to size the landing gear and determine its limits in terms maximal allowable horizontal and vertical speed, as well as the maximum allowable slope of the landing site.

The following assumptions underly the presented model:

1. The lander is a rigid body,
2. After touching the surface, one foot is fixed to the ground, forming a hinge around which the lander starts rotating, until the second impact takes place and the second foot forms a hinge, and
3. The landing is considered to be a two-dimensional case, such that all forces occurring after impact as well as the resulting motion are symmetrical about a vertical plane through the centre of mass of the vehicle.

To determine whether the lander (with mass m and moment of inertia I) will turn over after impact of the first leg, a critical angular velocity has to be defined, see the layout and definitions in Fig. 9.98. In this figure, $\dot{\gamma}$ is defined as the angular speed of the axis between centre of mass and the point of impact, A; the critical rate $\dot{\gamma}_{cr}$ would then just be sufficient to lift the c.o.m. to the highest point possible, while maintaining ground contact. This means that the centre of mass is directly over the axis of rotation, and if it would move further potential energy is gained. Thus, if the kinetic energy associated with $\dot{\gamma}$ exceeds this change in potential energy, the vehicle will overturn.

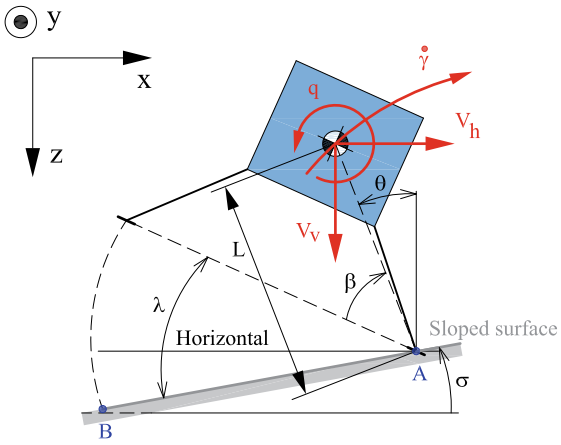


Fig. 9.98 Two-dimensional landing dynamics of a vehicle after first impact of the landing gear

From the condition that the kinetic energy associated with $\dot{\gamma}_{cr}$ is equal to the critical change of potential energy, the critical rate can be derived as follows. Let the change in potential energy be given as

$$\Delta\mathcal{E}_p = mg_0L(1-\cos\theta) \tag{9.191}$$

and, after touch down, the kinetic energy due to the rotation about A is

$$\mathcal{E}_{k,rot} = \frac{1}{2}I_A\dot{\gamma}^2 \tag{9.192}$$

with $I_A = I + mL^2$ the moment of inertia about the contact point A, I being the moment of inertia about the vehicle's centre of mass, m is the vehicle mass, and L the distance between the c.o.m. and A. If $\mathcal{E}_{k,rot} < \Delta\mathcal{E}_p$, then the rotation will stop just short of the critical point. Thus, if we equate both energies an expression for $\dot{\gamma}_{cr}$ can be derived.

$$mg_0L(1-\cos\theta) = \frac{1}{2}I_A\dot{\gamma}_{cr}^2 \quad \Rightarrow \tag{9.193}$$

$$mL^2\frac{g_0}{L}(1-\cos\theta) = \frac{1}{2}(I+mL^2)\dot{\gamma}_{cr}^2 \quad \Rightarrow \tag{9.194}$$

$$\bar{m}\frac{g_0}{L}(1-\cos\theta) = \frac{1}{2}(1+\bar{m})\dot{\gamma}_{cr}^2 \quad \Rightarrow \tag{9.195}$$

$$\dot{\gamma}_{cr} = \sqrt{\frac{2g_0}{L}(1-\cos\theta)\frac{\bar{m}}{1+\bar{m}}} \tag{9.196}$$

where $\bar{m}$ is a dimensionless inertia parameter, defined by $\bar{m} = \frac{mL^2}{I}$. In the above equations, θ is the angle between the line connecting the c.o.m. and A, and the local vertical. The surface gravitational acceleration is indicated by g_0.

The angular velocity of the lander about A, directly after impact, depends on the horizontal and vertical velocity of the lander, V_h and V_v, as well as its angular pitch rate, q, just prior to impact, and can be derived by representing the forces in A as impulses.[34] By decomposing the velocity components in the centre of mass into components along the line connecting the c.o.m. and A, and the line perpendicular to that, we find for the angular-momentum equilibrium:

$$I_A\dot{\gamma} = mLV_h\cos\theta - mLV_v\sin\theta - Iq \quad \Rightarrow \tag{9.197}$$

$$(I + mL^2)\dot{\gamma} = mL^2\left(\frac{V_h}{L}\cos\theta - \frac{V_v}{L}\sin\theta\right) - Iq \tag{9.198}$$

Dividing both sides by I, and substituting the definition of $\bar{m}$ yields

$$(1+\bar{m})\dot{\gamma} = \bar{m}\left(\frac{V_h}{L}\cos\theta - \frac{V_v}{L}\sin\theta\right) - q \quad \Rightarrow \tag{9.199}$$

$$\dot{\gamma} = \frac{\bar{m}\left(\frac{V_h}{L}\cos\theta - \frac{V_v}{L}\sin\theta\right) - q}{1+\bar{m}} \tag{9.200}$$

If the lander does not tip over, the lander will fall back and land on the opposite foot (or pair of feet), in Fig. 9.98 by the point B. In case the touchdown takes place on an adversely inclined surface, the lander will gain additional angular speed due to the height difference. The new angular velocity is thus not $\dot{\gamma}$, but

$$\dot{\rho} = -\sqrt{\dot{\gamma}^2 + \frac{\frac{2g_0}{L}\bar{m}\left[\cos\theta - \cos(\theta+\lambda)\right]}{1+\bar{m}}} \tag{9.201}$$

The critical angular velocity for the second impact can be determined in a similar manner as for the first one. If both $\dot{\gamma}$ and $\dot{\rho}$ are below $\dot{\gamma}_{cr}$, and the total kinetic energy within the maximal allowable energy limit set by the dampers, the lander will not tip over.

Using the model described above, one can show that is sufficient to consider only four critical modes that lead to a lander tumbling over. These four cases have been identified by Walton et al. (1964), and visualised in Fig. 9.99. It implies that within the limitations of the two-dimensional impact theory, that if the vehicle is stable for these four modes, it will be stable for any approach. The nomenclature of the critical modes is derived from the number of feet which impact and the order in which the

[34] In case the legs have energy absorbers, the impulse at landing will be reduced and thus the lander would have a larger range of V_h and V_v to remain stable after touch down. In the current model, this effect is not included, meaning that the provided stability criteria represent a worst-case scenario.

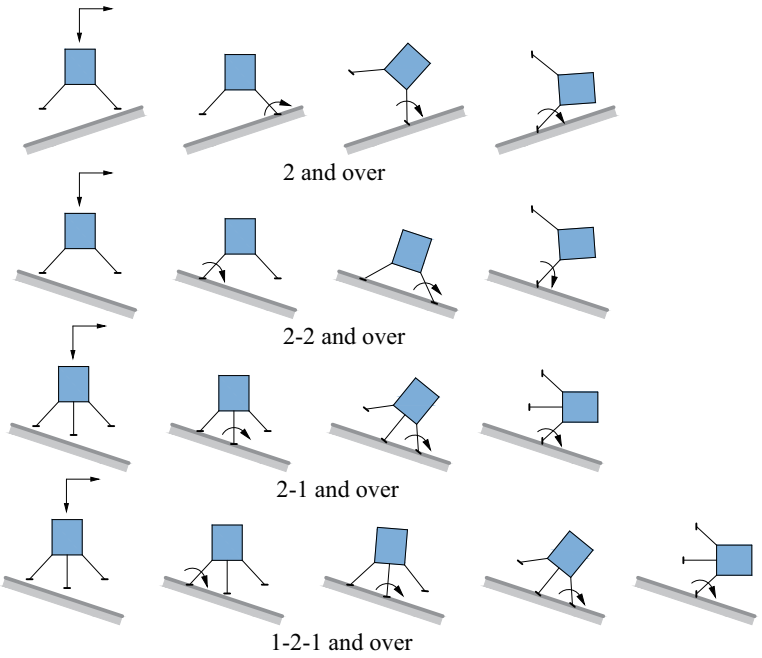


Fig. 9.99 Lander critical tumbling modes: for a landing gear with four legs, but also holds for a three-legged configuration. Image inspired by Walton et al. (1964)

impacts occur. They are explained as follows, assuming a vertical and horizontal velocity component as indicated in the figure:

1. Two and over. Two feet touch down simultaneously, the vehicle starts rotating about an axis connecting the two feet, and the centre of mass will then be carried over this axis of rotation.
2. Two-two and over. Two feet touch down simultaneously, the vehicle will start rotating until the other two feet hit the surface, and the vehicle's centre of mass is carried over this second axis of rotation.
3. Two-one and over. Two feet touch down simultaneously, and the vehicle will rotate until a single other foot impacts. A rotation about an axis through the single pinned foot is initiated, which will carry the centre of mass over this axis.
4. One-two-one and over. A single foot touches down first, the vehicle then rotates about an axis through this foot, which successively leads to a two-one and over impact involving the remaining feet.

Example 9.10 (*Landing-Dynamics Stability*) For a future landing mission to Saturn's moon Enceladus to investigate its geysers, a lander with mass $m = 300$ kg has been designed. The lander, with a central bus structure of 1.6 m wide and 0.6 m high, has a three-legged landing gear, with each leg being 1.6 m long. In the nominal upright position, the legs make an angle of 45° with the local vertical. The centre

of mass is located 0.332 m above the bottom plate of the bus, and the corresponding moment of inertia is $I = 150$ kg m^2.

From the given geometry and mass properties, the following (required) parameters (see also Fig. 9.98) can be derived: $L = 2.0965$ m, $\beta = 45°$, $I_A = 1{,}434.9$ kg m^2, and $\bar{m} = 8.5658$. Using the two-dimensional model described above, we will look at the tumbling rates of the lander for several slope and touch-down angles, i.e., σ and λ, respectively. To do so, $\dot{\gamma}$ and $\dot{\rho}$ (Eqs. (9.200) and (9.201)) are evaluated for a range of touch-down velocities, and compare those with $\dot{\gamma}_{cr}$, Eq. (9.196). In case both rates are smaller than the critical rate, the lander is assumed to be stable, and if either or both is/are larger, the lander is labelled "unstable". The considered velocity range at impact is $V_h \in [-0.6, 0.6]$ m/s and $V_v \in [0, 1]$ m/s, whereas at any case the lander's pitch rate, q, is zero. It is noted though, that for a realistic range of q, the stability properties were only marginally affected. Last, but not least, Enceladus' surface gravity is $g_0 = 0.113$ m/s^2.

The results of the evaluation are shown in Fig. 9.100. The four cases considered are (1) $\sigma = 0°$ and $\lambda = 0°$, (2) $\sigma = 0°$ and $\lambda = 5°$, (3) $\sigma = 5°$ and $\lambda = 5°$, and (4) $\sigma = -5°$ and $\lambda = 5°$, yielding values for θ of 45°, 40°, 45°, and 35°, respectively (note: $\theta = \pi/2 - \beta + \sigma - \lambda$). In the figure, the stability bands are clear. For opposing slope and touch-down orientation the stability band is the smallest, partially driven by the fact that the lander will gain momentum after touching down in point A and rotating back to point B for the second impact. In that case, $\dot{\rho}$ is more likely to exceed the maximum allowable tumbling rate. ■

To conclude the discussion on landers with a landing gear, Witte (2015) provides an in-depth analysis of landing safety, using a combination of high-fidelity numerical multi-body modelling and an experimental data of a legged-landing platform test campaign, used for validation purposes. Outcome of the study was a mathematical

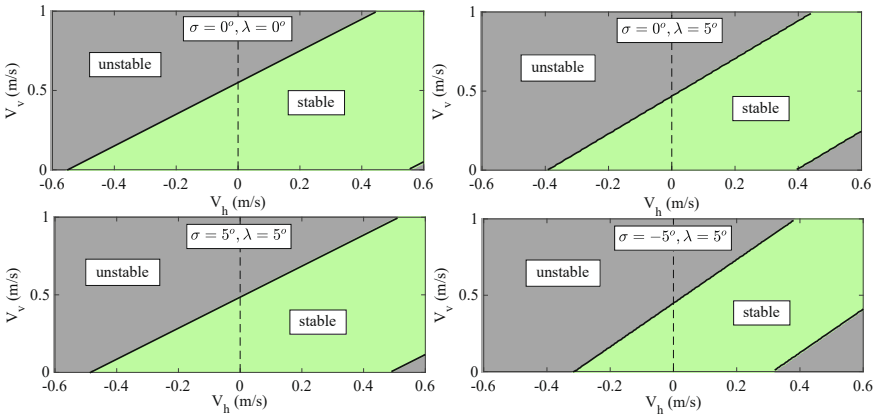


Fig. 9.100 The touchdown stability range for a planetary lander on Enceladus, for different surface inclinations and touch-down orientation

method to determine terrain-related failure probabilities from touchdown dynamics data. Such a study is the logical next step for landing-safety assessment.

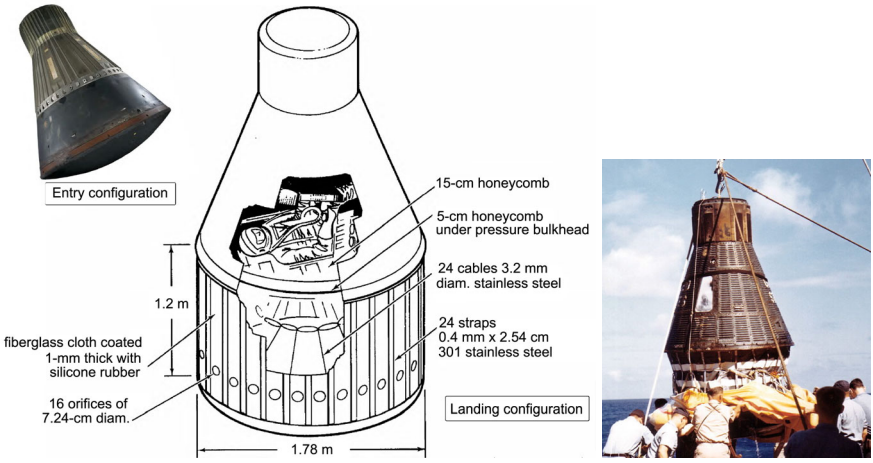
9.6.2.4 Landing Bags

Impact-attenuating airbags, also known as landing bags, are a good alternative for certain weight categories of planetary and robotic payload landings. Depending on surface roughness, a tough fabric airbag can tolerate small rocks and inclines. Two types exist: airbag systems that fully enclose the payload, such that after landing the system can bounce and tumble, as well as a system designed for a "no-bounce" landing, in which case the payload is only protected on one side and after landing it will "hit and stick" on the surface and not tumble.

The packed volume of the landing bags is small and they can be tailor-made for complex payload geometries. Apart from requiring an inflation system, they provide weight savings in the overall landing system, and compared to propulsive landing systems they are relatively low cost. Limitations in their use are a maximum payload weight and the surface state, in terms of roughness and slopes.

The use of landing bags dates back to the beginning of the space era. They were prominent in the early days of Soviet space exploration, with the Luna-9 lunar lander using airbags to achieve the first ever soft landing of a human-made vehicle on an extraterrestrial planetary body in 1966 (see Fig. 9.12). To achieve a safe landing, at an altitude of 75 km above the landing site (and 48 s before impact), a protective airbag was inflated to enclose the capsule (Harvey 2007b). A descent engine slowed down the vehicle from 2.6 km/s down to 6.1 m/s. Five meters above the surface the engine was shut down, and the landing capsule was ejected. When touching down on the surface, the nitrogen gas inside the airbag compressed and expanded, causing the entire system to experience several bounces before finally coming to rest. After severing the landing system, the airbags deflated and flew off to the side, leaving the instruments ready to work. The achievement of successfully soft landing on the Moon is even more impressive when one realises that the design of the landing system was based on photographs from Earth-based observations, and the final design could be lowered onto any soil from hard rock to loose dust. The design was put through extensive testing on Earth by dropping a breadboard model onto different compact and loose materials (Cherkasov et al. 1986). The same landing-system design and operational procedures were repeated with Luna-13, launched in December of the same year. Three days after launch, the vehicle also achieved a successful soft-landing on the lunar surface.

Also in the beginning of the US manned space program, airbags were used. In the Mercury program in the 1960s, a passive system was used to lower the loads during water landing. In between the heat shield and the capsule a skirt was placed that would inflate when the heat shield was separated from the capsule, see also McCullough et al. (1964) and NASA (1970). Figure 9.101 depicts Mercury's shock attenuation system, not just the skirt, but also the location of the crushable honeycomb structure shown in Fig. 9.91. After freeing the skirt, air was drawn in through holes in the



(a) Overview (Image adapted from NASA (1970))

(b) Mercury-8 recovery (Image: NASA)

Fig. 9.101 Mercury's shock attenuation system

skirt. Upon landing, the air was forced back out the holes, dissipating a portion of the kinetic energy of the vehicle, and thus lowering the impact loads. The figure on the right shows the Mercury-9 capsule after recovery: the deployed skirt with the white bands is visible, as well as the flotation devices at the bottom.

For the Mars Pathfinder and the two Mars Exploration Rover missions, NASA decided to use a large cluster of all-encapsulating inflatable airbags (Fig. 9.102). Motivation for this was the relatively unknown surface conditions and the possibility of a large horizontal velocity component. After release of the landing system, it rolled and bounced for quite some time, before coming to rest in exactly the right orientation. After deflation of the airbags, the platform with the rover on top was freed, and after releasing a ramp the rovers could drive down and begin their mission. The reader is referred to Sect. 9.7.5, where the Mars Exploration Rover missions are discussed in more detail.

One of the most recent applications of landing bags is the one for the Orion Crew Module. The design discussed by Willey et al. (2007) and Johnson and Hixson (2008) includes the use of multiple discrete airbags for reasons of redundancy. Towards the end of the parachute descent, at about 1,500 m (5,000 ft) altitude and a descent rate of 7.6 m/s (25 ft/s), the heat shield is jettisoned. At around 300 m altitude (1,000 ft), the bags are inflated using an active cold gas system, and landing occurs about 40 s later. The airbags use actively controlled vent orifices to dissipate energy during landing, which will keep the orientation of the capsule fixed: it thus results in a "stuck" landing, compared to the bouncing landing used for the Mars landers, discussed above. In Fig. 9.103, the landing-bag configuration is shown while being subjected to a drop

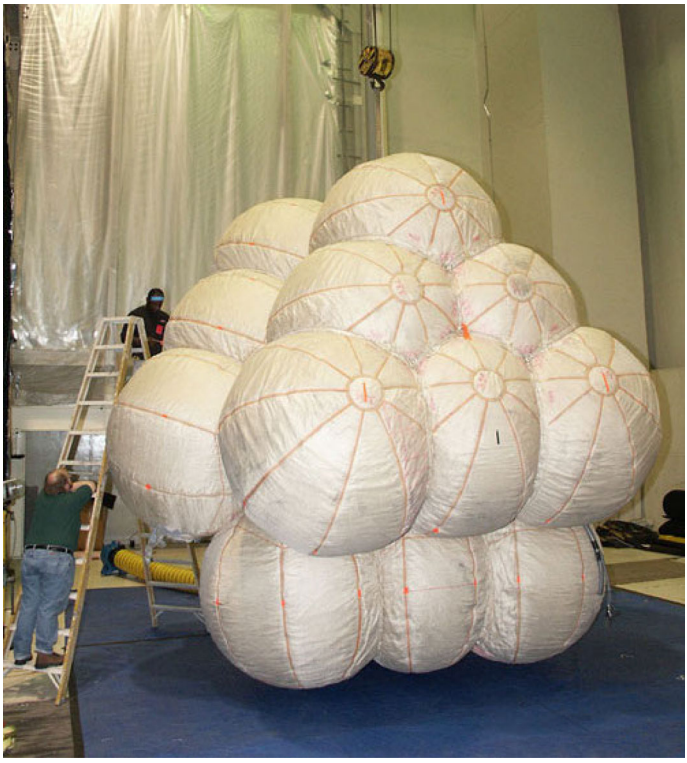


Fig. 9.102 Mars Exploration Rover airbags during a test. Image: NASA

test at Nasa Langley Research Center. In their paper, Willey et al. (2007) support the test results with finite-element calculations, showing a good correlation.

The operating principle of airbags is that kinetic energy of the lander is converted to potential energy of the enclosed gas through compression. In case of a closed system, after rebound this process is reversed, and energy can only be dissipated through various sources of loss, such as heat dissipation and internal vents. This process can be sped up by venting the gas after it has first been compressed, such that the energy is dissipated to the atmosphere or simply released from the system.

Providing a detailed model including thermodynamics, fluid mechanics, and structural dynamics is well beyond the scope of this section. However, it does serve the purpose of conceptual understanding that a simplified and idealised approach is introduced, so we will discuss the three main dynamic phases from an energy point of view: (1) free fall until just before touch down, (2) compression of the airbag after contact with the planetary surface, and (3) the venting of the airbag from the moment that the internal pressure has reached its maximum.

Let us consider the (closed) system as shown in Fig. 9.104a. The airbag, attached to the payload with mass m, is considered to be a single system. Close to the surface,



Fig. 9.103 Drop test at NASA Langley Research Center of the Orion Crew Module. Images: NASA

this system is moving with downwards velocity, V. Any venting mechanism in the airbag is closed. This means that the total energy of the system, $\mathcal{E}_t$, consisting of the system's kinetic and potential energy, $\mathcal{E}_k$ and $\mathcal{E}_p$, respectively, and the internal energy of the gas inside the airbag, $\mathcal{U}_g$, can be written as:

$$\mathcal{E}_t = \underbrace{\mathcal{E}_{k,m} + \mathcal{E}_{p,m}}_{\text{payload}} + \underbrace{\mathcal{E}_{k,g} + \mathcal{E}_{p,g} + \mathcal{U}_g}_{\text{gas}} \tag{9.202}$$

Any other energy contribution, e.g., internal energy of the payload, is neglected.

For the sake of the analysis, we assume that the gas in the landing bag will behave as an ideal gas, so the ideal gas law will hold. Earlier, in Chap. 3, Eq. (3.1), this law was introduced, and is here slightly reformulated by replacing the density of the gas, ρ_g with $\frac{m_g}{v_g}$, i.e., the ratio of the gas mass with its volume, v_g.

$$p_g v_g = m_g R_{air} T_g \tag{9.203}$$

Let us assume that an Earth landing is considered, and that the operating medium in the airbag is air, with a temperature of $T_{g,0} = 20\,^{\circ}\text{C}$. The internal pressure is taken as $p_{g,0} = 130$ kPa.[35] The (specific) internal energy of air at $T_g = 293$ K is $u_{air} = 210.49$ kJ/kg, the enthalpy is $h_{air} = 295.17$ kJ/kg, whereas the specific gas constant is $R_{air} = 0.287$ kJ/kg/K (Çengal et al. 2016).

[35] The internal pressure of 130 kPa is taken from a paper by Cooper et al. (2005), who did testing on the airbag landing-attenuator system of the Orion capsule.

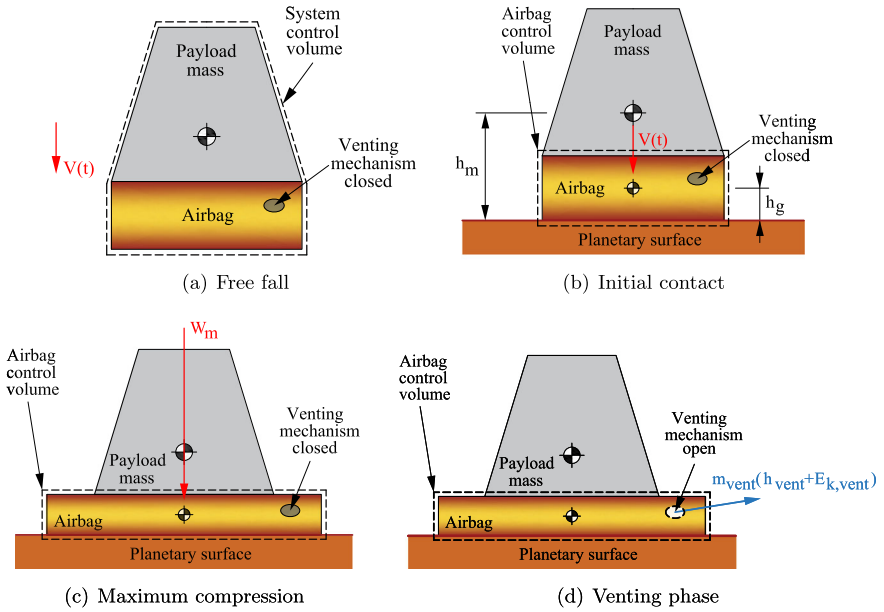


Fig. 9.104 Different phases for an airbag landing (drawings not to scale)

Substituting the values of $T_{g,0}$, $p_{g,0}$, and R_{air} into Eq. (9.203) yields a specific gas mass of $m_g = 1.54$ kg per cubic meter of volume, v_g. Let us assume $v_g = 1.25$ m^3, so $m_g = 1.93$ kg.[36]

To get a feeling for the different levels of energy, suppose that we have a payload mass of $m = 150$ kg, and an impact velocity of $V = 8$ m/s (Fig. 9.104a). Consider the situation in Fig. 9.104b, where a payload with a landing bag attached to the bottom has just made initial contact with the planetary surface. With the mass of the payload and the system velocity given, we find $\mathcal{E}_{k,m} = 4{,}800$ J. Concerning the potential energy of the system, we need to consider the distance between the centre of mass of the payload and the planetary surface. Without going into detail of the dimensions of this payload, we suppose that this distance is $h_m = 1.5$ m. Then, with $g = g_0 = 9.81$ m/s^2 and $m = 150$ kg, we find $\mathcal{E}_{p,m} = 2{,}207$ J. Moving on to the gas, the kinetic energy is obviously a lot smaller, due to the much lower mass: $\mathcal{E}_{k,g} = 62$ J. The potential energy is likewise small, i.e., with the dimensions of the cylindrical landing bag given and its (local) centre of mass being equal to the radius of the cylinder ($h_g = 0.445$ m), it follows that $\mathcal{E}_{p,g} \approx 9$ J. What remains is the internal energy of the gas, which is equal to $\mathcal{U}_g = m_g u_g = 406.76$ kJ. Looking at individual terms, it is obvious that the contribution of $\mathcal{E}_{k,g}$ and $\mathcal{E}_{p,g}$ can easily be ignored. The point-mass energy is 7.07 kJ, which contributes less than 2% to the total. Thus, the internal energy of the gas represents by far most of the energy of the system.

[36] For the sake of argument, we will assume that the landing bag has a cylindrical shape in unloaded condition. With a volume of 1.25 m^3, and assuming that the landing bag has a width of 2 m, the diameter will be 0.89 m.

Next step in the analysis, as shown in Fig. 9.104c, is from the moment of initial contact with the planetary surface onwards, when the payload mass acts as a compressor and does work, $\mathcal{W}_m$, on the airbag. We will analyse what happens until the moment of maximum compression. As the airbag is being compressed we need to do a thermodynamics analysis of the gas, using the First Law of Thermodynamics. Since energy is conserved (any venting hole is closed, so the gas remains within the control volume), we can write:

$$\mathcal{E}_{in} - \mathcal{E}_{out} = \Delta\mathcal{E}_g \tag{9.204}$$

but for the closed system $\mathcal{E}_{out} = 0$. Furthermore, we assume that the process is adiabatic and no work is done by the gas. Thus, energy is conserved and the work done on the gas will lead to a change in internal energy:

$$\mathcal{W}_m = \mathcal{U}_2 - \mathcal{U}_1 \tag{9.205}$$

with index 2 indicating the moment of maximum compression. Obviously, the work done cannot exceed the initial energy content of the payload, i.e., the sum of its kinetic and potential energy. At maximum compression the kinetic energy is zero, as the payload is at its lowest point. The potential energy will not be zero, but depends, of course, on how much the airbag has been compressed. What we have analysed earlier is that the potential energy cannot be ignored with respect to the kinetic energy, but can we do that now, when we have to consider the *change* in potential energy? Even though it would be smaller than the kinetic energy, we could do a simple analysis. From the earlier results we see that $\mathcal{E}_{p,m}$ (2.2 kJ) is about 46% of $\mathcal{E}_{p,m}$ (4.8 kJ). Suppose that the landing bag is compressed by 25%, this means that the contribution of $\Delta\mathcal{E}_{p,m}$ to $\mathcal{W}$ will be 11.5%, which is commonly a contribution too large to ignore. Therefore, we consider that $\mathcal{W}_m = \mathcal{E}_{k,m} + \Delta\mathcal{E}_{p,m}$, and thus

$$\mathcal{U}_2 = \mathcal{U}_1 + \mathcal{E}_{k,m} + \Delta\mathcal{E}_{p,m} \tag{9.206}$$

Note that the increase in internal energy is, of course, directly related to the change in geometry of the airbag, i.e., a decrease in v_g and thus an increase in p_g (as we considered an adiabatic process, so T_g = constant):

$$\mathcal{W}_m = \int_1^2 p_g \mathrm{d}v_g \tag{9.207}$$

This will require a more advanced analysis than what we do here, including the efficiency of the energy-transfer process.[37] We will continue with the energy approach, Eq. (9.206), though. Using the values found earlier ($\mathcal{E}_{k,m} = 4.8$ kJ, $\Delta\mathcal{E}_{p,m} = 0.2$ kJ

[37] Following the simplified analysis, a 25% reduction of v_g will lead to a pressure increase to $p_g =$ 173.3 kPa.

and $\mathcal{U}_1 = \mathcal{U}_g = 406.76$ kJ), we find that $\mathcal{U}_2 = 411.56$ kJ, which is the energy content of the gas at the moment at which the airbag has reached the state of maximum compression.

At that moment, a pressure trigger could open a vent, to release energy and to reduce the bouncing back of the airbag. The system is no longer a closed system, as the gas can escape to the environment. This final phase has been depicted in Fig. 9.104d. The energy of the system can yet again be determined by using the First Law of Thermodynamics. If we assume that at this moment no work is done anymore, this means that $\mathcal{E}_{in} = 0$. With gas escaping from the vent with velocity V_{vent}, which may be large, it will have a specific kinetic energy $e_{vent} = 0.5V_{vent}^2$ per kg of escaping gas. The internal energy of this amount of gas is $m_{vent}h_{vent}$, with h_{vent} being its enthalpy per unit mass. Thus, the energy balance for the moment the system comes to rest is

$$-m_{vent}\,(h_{vent} + e_{vent}) = \mathcal{U}_3 - \mathcal{U}_2 \tag{9.208}$$

with the subscript *3* indicating the state of the system having reached rest.

Rearranging Eq. (9.208) gives

$$\mathcal{U}_3 = \mathcal{U}_2 - m_{vent}\,(h_{vent} + e_{vent}) \tag{9.209}$$

thus, in words, the final energy of the system is the energy of the system before venting minus the energy "leaving" the system, i.e., the sum of internal and kinetic energy.

Suppose now that the attenuation of the system is optimal, and the energy left in the system is zero, i.e., $\mathcal{U}_3 = 0$.[38] This will allow us to calculate the maximum amount of vented gas required to reach a zero-energy state:

$$m_{vent} = \frac{\mathcal{U}_2}{h_{vent} + e_{vent}} \tag{9.210}$$

The specific enthalpy for the vented air is, as before, $h_{vent} = 295.17$ kJ/kg (adiabatic process, so temperature is constant). The specific kinetic energy is smaller than this and ignoring it would give a conservative estimate of m_{vent}. However, one should realise that the velocity of the vented gas can be high and even approaching sonic speed: with $V_{vent} = 340$ m/s, this means that $e_{vent} = 57.8$ kJ/kg. So, for a more detailed analysis one should not ignore this term. With $\mathcal{U}_2 = 411.56$ kJ we find $m_{vent} = 1.39$ kg while ignoring e_{vent}, and $m_{vent} = 1.17$ kg, when not. These values represent about 72% (60%) of the gas stored in the airbag prior to impact, which shows that complete energy attenuation using venting is possible.

Of course, the above analysis is only a simplified one, merely touching upon some of the aspects involved. For a more accurate analysis, non-linear analysis taking

[38] This is, of course, yet another assumption that will have to be verified during a full analysis of the system. In the spirit of the simplified analysis we do assume so.

into account aspects, such as landing orientation, landing-bag geometry, material deformation, non-ideal venting, spillage, and thermodynamics effects. The earlier cited reference by Willey et al. (2007), as well as the one by Do and De Weck (2012), provide a good starting point to do such an analysis. This would be required to face the challenges inherent in airbag system design, configuration trades, and material selection.

9.7 Missions

In this section a limited number of (personally selected) planetary missions will pass review, to address typical mission-operations aspects and hardware elements. By no means pretending to give a complete overview, subsequently we will address the following missions: Surveyor and Apollo (Sect. 9.7.1), Viking (Sect. 9.7.2), Galileo (Sect. 9.7.3), Stardust (Sect. 9.7.4), the Mars Exploration Rovers, MER-A and MER-B (Sect. 9.7.5), and finally Mars Science Laboratory (Sect. 9.7.6).

9.7.1 Surveyor and Apollo Lunar Lander

In support of the American Moon Program, the Surveyor was a series of seven unmanned missions that were designed to reach the surface of the Moon by a soft landing (Fig. 1.47). On 2 June, 1966, Surveyor-1 landed on the Moon, and even though it was about four months after the Russian Luna 2, it was considered a major stepping stone for the Apollo program. The Surveyors were launched with the Atlas Centaur, and with a launch mass varying between 995 and 1039 kg, the landing mass was between 294 and 306 kg. Out of the seven spacecraft launched, five made successful landings (see Table 1.3). The Surveyor program had the goal to select suitable landing sites for the Apollo program, with the exception of the one for Surveyor-7: because sufficient data was already gathered for Apollo, its landing site was chosen on a scientific basis.

Surveyor-1 made use of a direct transfer between the Earth and the Moon, whereas Surveyor-3, -5, -6 and -7 began their journeys in an Earth parking orbit and landed on the Moon via a lunar orbit. Surveyor's propulsion system consisted of a solid retro-rocket, used during the main descent towards the lunar surface, and vernier engines that were used for midcourse velocity-vector corrections, attitude control during main retro-rocket engine burning, and velocity vector and attitude control during terminal descent. The propulsion subsystem was controlled by the flight-control system through pre-programmed manoeuvres, commands from Earth, and manoeuvres initiated by flight-control sensor signals. A typical example of a pre-programmed manoeuvre is the earlier discussed gravity turn (Sect. 9.3.2).

Figure 9.105 shows a typical descent and landing scenario, which began with a *mark* signal from the altitude-marking radar, when the spacecraft was at a 111 km (60

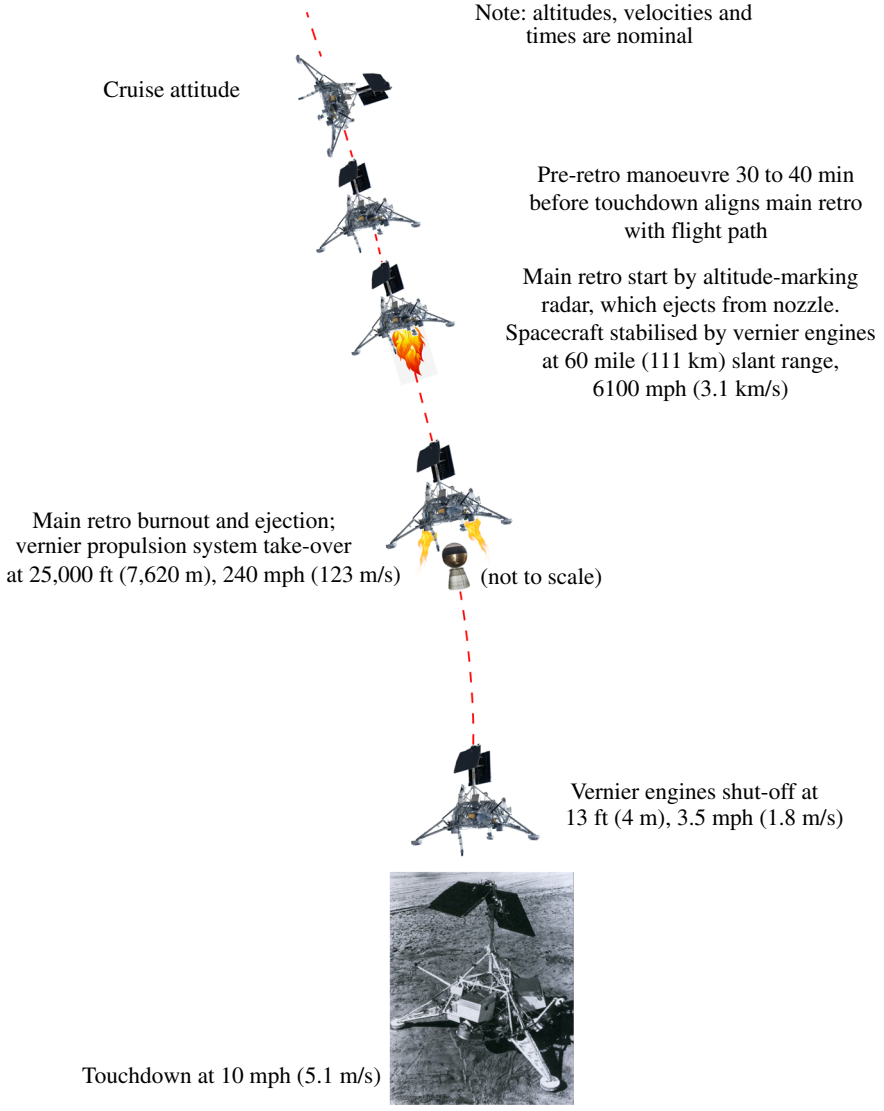


Fig. 9.105 Surveyor descent scenario (Source material: NASA)

miles) slant range from the lunar surface, with a velocity of about 3.1 km/s.[39] After a preset, timed delay of 5.9 s, the three liquid-propellant vernier engines ignited, shortly followed (after 1.1 s) by ignition of the main retro-rocket, using a carboxyl-terminated polyhydrocarbon composite-type solid propellant and providing between

[39] Even though the current description is done for Surveyor-6 (Surveyor Project Staff 1968), it is representative for the other Surveyors, albeit that sometimes slightly different values would apply.

35 and 45 kN of thrust. Burnout of this engine takes place some 40 s, when the velocity has reduced to 123 m/s at an altitude of about 7,600 m. Twelve seconds later, the main engine was jettisoned to reduce the decelerating mass.

During the powered descent, the vernier engines served as attitude-control thrusters, maintaining an inertially fixed attitude. After burnout of the main engine, however, these engines, using a combination nitrogen tetroxide (N_2O_4) as oxidiser, and mono-methyl hydrazine mono-hydrate (72% MMH-28%H_2O) as fuel, stored in three tanks each, took over the deceleration. At that time, the radar altimeter and doppler velocity sensor (RADVS) had been activated to provide continuous range and velocity information. The RADVS, of which the beam orientation is shown in Fig. 9.106, could compute the slant range from 40,000 ft (12,192 m) to 13 ft (=4 m), and generated the 1000-ft and 13-ft *mark* signals during the descent. The Doppler velocity sensor, working on the principle that a reflected signal has a Doppler frequency shift proportional to the approaching velocity, could compute the spacecraft's velocity along its X-, Y- and Z-axes. Using the measurements, the spacecraft's longitudinal axis could be pointed along the flight path, while maintaining a total thrust level of about 0.9 Lunar g, by continuously differentially throttling the vernier engines.

Mapping the current velocity and slant range to predefined descent contours, the RADVS controlled the spacecraft to closely follow the velocity/slant-range contour down to the 13-ft altitude mark. Reaching that point, the engines were shut off, and the Surveyor free-fell the last few meters to the lunar surface. This resulted in a landing speed of about 3 m/s. The free-fall to the surface was in an attempt to avoid surface contamination by rocket blast. Total duration of the descent, starting at the 111 km slant-range mark till initial touchdown was 3 min 10 s.

The Apollo project aimed at bringing man to the Moon "before the end of the decade", as president Kennedy so aptly stated back in 1961. A number of predecessor projects would secure the technology development, i.e., the previously discussed Surveyor, Mercury and Gemini, to gain experience with manned spaceflight, rendezvous and docking and advanced entry technology (see also Sect. 1.3.2), but also the earlier Apollo flights served as technology-development missions. The program became a huge success, when Neil Armstrong became the first human who walked on the surface of the Moon with Apollo 11. More than 500 million people watched the landing live on television on that day, 20 July, 1969. Apollo-11 was followed by another five successful landing missions, and in total 12 men walked on the Moon. The astronauts descended to the Lunar surface in the so-called Lunar Module, which made a powered descent and landing. Being the only planetary lander to have transported men down to the surface of another celestial body, it does not require much imagination to realise Apollo has been one of the most influential programmes for planetary landers to date. Therefore, this part of the Apollo mission will be the focus of the current discussion.

The spacecraft involved in the Apollo landing missions are very similar, differing in some sub-systems only and having slightly different mass properties. They all consisted of two parts: the Command Service Module (CSM) and the Lunar Module (LM). The CSM itself consisted of two parts, i.e., the Command Module (CM), a

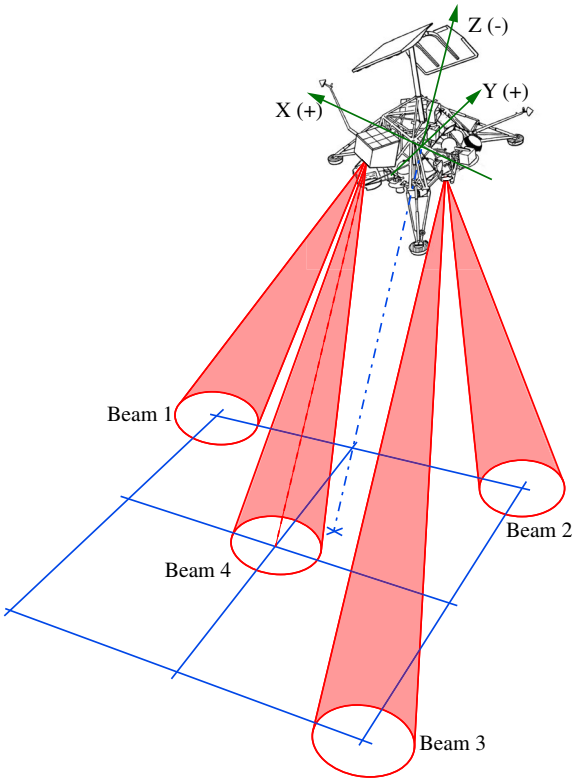


Fig. 9.106 Surveyor radar altimeter and doppler velocity sensor beam orientation. Surveyor image: NASA

compartment and a control centre for the astronauts and the module used for re-entry at the end of the mission, and the Service Module (SM), which was responsible for the primary propulsion and had the ability to manoeuvre the spacecraft. The CSM would remain in low Lunar orbit for the full duration of the Lunar surface mission. The LM consisted of a descent and an ascent stage, each having their own propulsion system. The ascent stage had the reaction control system (RCS) mounted to it, which was used during both phases of the powered flight. Engine and RCS specifications have been listed in Table 9.6. The main descent engine had a nominal thrust of 43,903 N, equal to 92.5% of the maximum thrust of 45,040 N. However, the engine was only throttleable between 4,671 and 30,248 N (10–65%), a higher setting would lead to engine failure. The engine was restartable, and could be gimballed with $\pm 6°$ motion capability (Grumman Aerospace Corporation 1971).

The descent of the Apollo missions all started in a low Lunar orbit, with minor variations in the eccentricity and semi-major axis in between missions. For the sake of discussion, we will focus on the Apollo-11 mission, covered in detail by Bennett (1970). The coupled CSM/LM combination was delivered into a 62 $\times$ 58 nautical mile (114.8 by 107.4 km) Lunar orbit, see also Fig. 9.107. After de-coupling from the CSM, a *descent orbit insertion* (DOI) with a ΔV of 75 ft/s (22.86 m/s) brought

Table 9.6 Apollo LM propulsion system (Grumman Aerospace Corporation 1971)

Descent stage engine	1x UDMH-NTO ($C_2H_8N_2/N_2O_4$) T = 45 kN, I_{sp} = 311 s
Ascent stage engine	1x UDMH-NTO 15.6 kN, I_{sp} = 311 s
RCS	16x UDMH-NTO thrusters 441 N per thruster, I_{sp} = 311 s

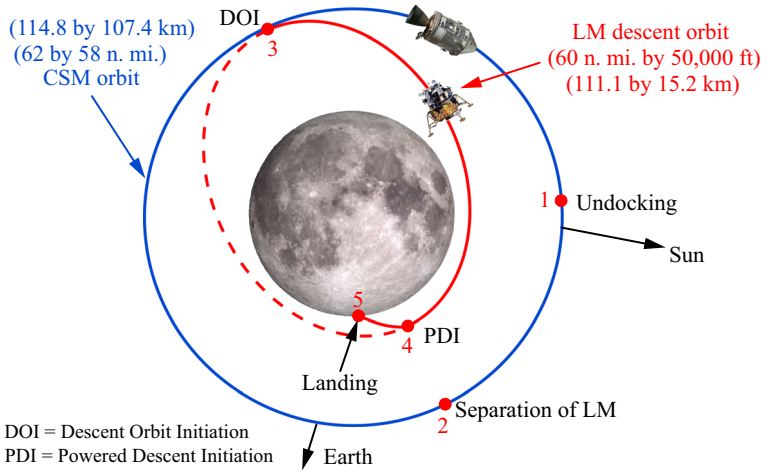


Fig. 9.107 Deorbitation of the Apollo's Lunar module from low Lunar orbit. Source images: NASA

the LM[40] to a lower altitude were it could begin the powered descent. This Hohmann transfer was an elliptical orbit with its apolune at 60 nautical mile (111.1 km) and perilune at 50,000 ft (15.2 km). Reaching the perilune was the moment that the *powered descent initiation* (PDI) took place for the final, powered descent to the Lunar surface. This altitude was selected as being the best possible compromise between efficiency (and therefore not higher) and safety (and therefore not lower).

From PDI downwards, the flight was divided into three phases, each satisfying a key operational requirement: the braking phase, the approach phase, and landing phase (Fig. 9.108). Total duration of the powered descent was about 12 min, covering a distance just over 480 km. Table 9.7 gives an overview of the pre-mission altitude-velocity relations for certain key points in the descent trajectory of Apollo-11.

The objective of the first phase, the braking phase, was to reduce orbital velocity, and was designed for efficient propellant usage. However, in case the throttle com-

[40] In Fig. 1.34c the LM is shown, consisting of the ascent and descent stage. With a total mass of 16,471 kg (including fuel), this mass was divided over the ascent and descent stages as 4,805 kg and 11,666 kg, respectively. The dimensions of the LM were a width at the tanks of 4.22 m, a height of 7.04 m and a bottom diagonal of 9.45 m.

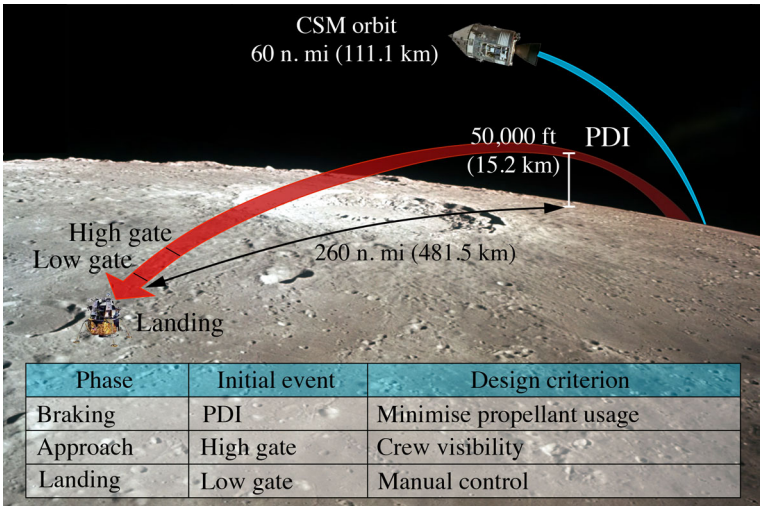


Fig. 9.108 Powered descent of Apollo's Lunar module down to landing. Source images: NASA

Table 9.7 Apollo-11 pre-mission powered-descent event summary. Data converted from Bennett (1972)

Event	TFI[a] (min:sec)	V_I (m/s)	$\dot{h}$ (m/s)	h (m)	ΔV (m/s)
Ullage	−0:07				
Powered descent initiation	0:00	1694.7	−1.2	14,878.5	0.0
Throttle to maximum thrust	0:26	1685.2	−0.9	14,851.4	9.4
Rotate to windows-up position	2:56	1219.2	−15.2	13,696.9	479.1
LR altitude update	4:18	934.2	−27.1	11,948.5	773.0
Throttle recovery	6:24	443.8	−32.3	7,510.0	1292.0
LR velocity update	6:42	400.8	−38.7	6,901.9	1340.8
High gate	8:26	154.2	−44.2	2,290.6	1638.3
Low gate	10:06	16.8(20.7)[b]	−4.9	156.1	1882.4
Touchdown (probe contact)	11:54	−4.6(0)[b]	−0.9	3.7	2065.0

[a]Time from ignition (TFI) of the descent propulsion system
[b]Horizontal velocity relative to the lunar surface

mands would exceed the 10–60% range, maximum thrust would be commanded. The LM was guided to high-gate conditions for initiation of the approach phase, at about 8.3 km downrange from the landing site. Note that the term *high gate* is derived from aircraft-pilot terminology and refers to the begin of the approach to an airport. During the approach phase the pilot used visual monitoring to approach the landing site (as we will show later). The landing phase began at so-called *low-gate* conditions (150 m altitude), which was also taken from aircraft-pilot terminology.

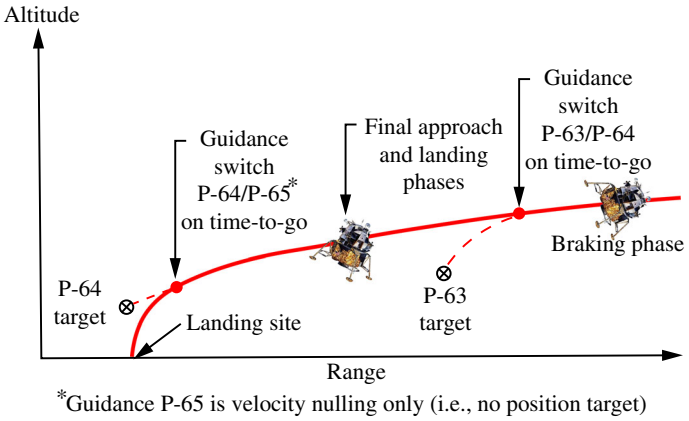


Fig. 9.109 Target sequence for Lunar Module automatic descent guidance. Lunar Module image: NASA

During the landing phase the pilot looked out the window at the landing site and took over from the automatic control for the final touchdown.

Several sequential programs were used during Apollo's powered descent for guidance and control operations (Fig. 9.109). The programs were indicated by the labels P63 to P67, and would allow for several landing options. These multiple options were included to account for unknown surface conditions, and could be enforced by the crew; for instance, P63 would lead to a landing early in the approach phase. The guidance system used explicit laws (very much alike the E-guidance discussed in Sect. 9.3.3) and variable thrust, enabled during the last two minutes of main braking to correct for errors. At 10 s from the target, nulling of the horizontal velocity was activated, until, finally, a vertical descent law was called and executed until touchdown.

The different guidance programs will now be briefly described, and can be reviewed in more detail in the papers by Bennett (1970, 1972) and the references therein. The first program, P63 *braking-phase guidance*, used an ignition algorithm to determine the time to ignite the descent propulsion system, based on a stored surface range to the landing site. After ignition, the basic quadratic guidance law is used to guide the lander to the desired start conditions for the approach phase. To ensure that the targets will be met before the time-to-go reaches zero, the braking-phase target is a point below the surface and further downrange than the start of the approach phase. When the time-to-go reaches a preselected value, the guidance program switches automatically from P63 to P64.

The conditions at reaching *high gate* have been visualised in Fig. 9.110. Besides an IMU, including accelerometers and gyros to sense changes in velocity and attitude, and aided by a simple periscope to measure star direction for IMU alignment, the navigation system used a landing radar (LR), similar to the RADVS of the Surveyor. The LR was mounted at the bottom rear of the LM and consisted of four radar

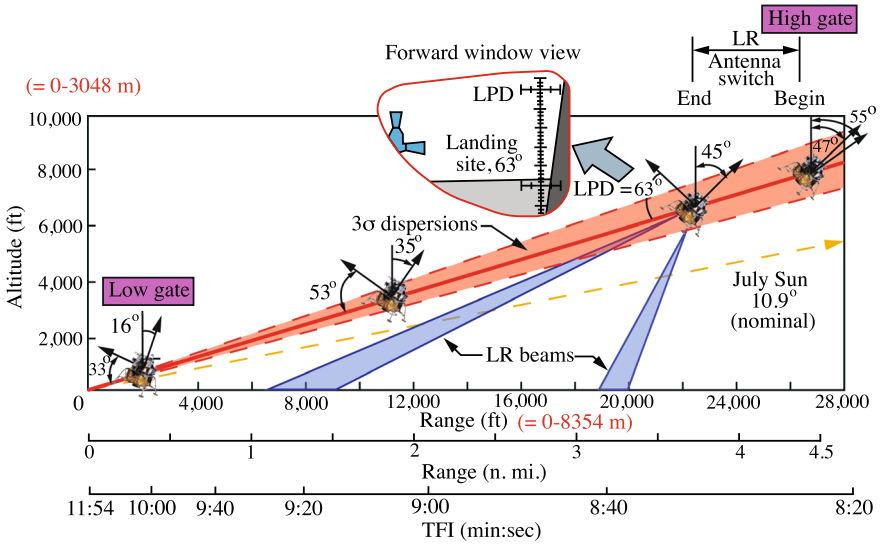


Fig. 9.110 Surface scanning and guidance for the Lunar module: approach phase (high gate to low gate). LPD stands for landing-point designator. Data taken from Bennett (1970), Lunar Module image: NASA

beams, one to provide ranging measurements and three to provide velocity measurements. Shortly after reaching *high gate*, an antenna switch was made to account for the change in attitude. Crucial input to the LM commander was the landing-point designator (LPD), a grid on his forward window. He used this device to aim for the landing site.[41] During the descent, the attitude of the LM was slowly (but surely) changed into a more vertical one.

The P64 *approach-phase guidance* also uses the quadratic guidance law, albeit with different settings. The desired targets were selected to provide trajectory shaping throughout the approach phase and to get suitable conditions for the beginning of the vertical descent. In addition, the trajectory angle was kept at such a value that the crew could use the window as a guide to approach the landing site. In case the crew decided to change the landing site they could do so with a manual override, and change the trajectory. The target point for the guidance law was yet again a point further than the landing site, but was meant to steer the lander to the desired point near or directly above the landing site. As before, when the time-to-go reached a pre-selected value the guidance logic would switch automatically from P64 to P65.

[41] During the approach and landing phases, the computer calculated the view angle to the landing site (relative to the forward body axis, Z_B), and displayed it on the display and keyboard. An angle of zero corresponded to a view along the Z_B-axis. The commander could then monitor the landing site along this view angle. If, for some reason, it was desired to change the landing site, incremental changes could be made in-plane or in cross-range direction, by moving the hand controller in the appropriate direction and thus providing input to the computer. The cross-range position was changed in 2°-increments, and the in-plane position in 0.5°-increments.

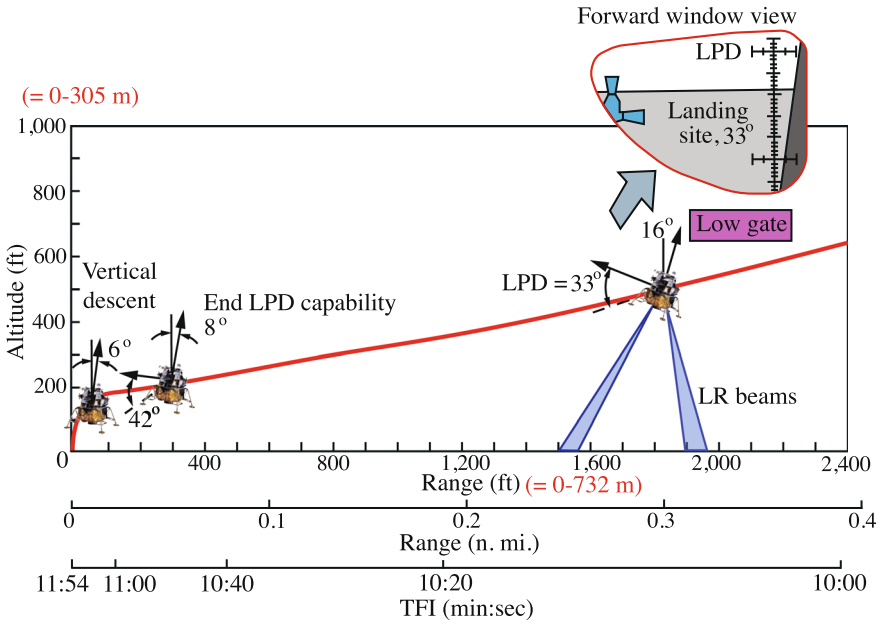


Fig. 9.111 Surface scanning and guidance for the Lunar module: landing phase (low gate to landing). LPD stands for landing-point designator. Data taken from Bennett (1970), Lunar Module image: NASA

In Fig. 9.111, the flight phase from high to low gate is shown. The view angle towards the landing site is initially 63°, and has reduced to 33° when low gate was reached. Indicated in the figure is also the Sun direction: all the time the Sun was "at the back", giving the most optimal lighting conditions.

The P65 *velocity nulling guidance* reduced the horizontal velocity to (a value close to) zero, and prepared the conditions for an automatic vertical descent to the surface. No position control was used during this guidance mode, and both vertical and horizontal velocity components were controlled separately. The vertical guidance algorithm controlled the altitude rate by adjusting the throttle, while the horizontal guidance algorithm nulled the horizontal velocity by making small changes to the thrust angle (with a limit of 20°). A constant vertical thrust component was assumed. Figure 9.111 shows the final phase from *low gate* to landing. Visibility to the landing site is at 33° at the beginning. Once the LM has turned to an angle of 8° deviating from the local vertical, the vehicle was in a too much upright attitude for the landing-point designator to be effective anymore, and was no longer used. At that point, the LM was about 100 m from the landing site, at an altitude of about 60 m.

The P66 *rate of descent* and P67 *manual guidance* were optional modes that could be used by the crew at any time during the automatic modes P63, 64, or 65. During P66, the crew would control the LM attitude, while the computer kept the throttle of the descent propulsions system at a desired value. This mode would mostly be used

Table 9.8 Touchdown-limitation velocities of Apollo's Lunar Module (Rogers 1972), data converted to metric

Touchdown direction	Touchdown magnitude
Vertical velocity	3.0 m/s
Horizontal velocity	For $V_v \leq 2.1$ m/s, $V_h = 1.2$ m/s;
	For $2.1 \leq V_v \leq 3.0$ m/s, $V_h = 4/3 - 4/3 \cdot V_v$ m/s

during late P64 operation (near *low gate*), prior to P65 switching for manual control of the final touchdown position. During P67 the crew had complete manual control of throttle and altitude, but the on-board computer retained control over navigation and display operations. This mode could be considered a backup mode in case P66 did not function well.

The LM managed to achieve soft landing in all six missions. The crew looked outside the module during the landing to detect potential hazards, like craters and boulders. To let the crew watch out the window the approach angle of the vehicle relative to the target was low (nominally 16° for Apollo-11-14 and 25° for Apollo 15-17). Also, the sunlight elevation angle was low (5–13° for Apollo-11-14 and 7–23° for Apollo 15-17) to create shadows that provided the crew with sufficient contrast in the terrain. The Sun had to be behind the crew and the LPD view angle needed to be larger than the Sun angle, such that the crew could see the shadows. The hazard tolerance for the LM was defined to be 0.5 m and the effective slope tolerance was 12°. The touchdown limitations can be found in Table 9.8.

9.7.2 Viking

The objective of the Viking mission was to advance the knowledge of planet Mars by means of observations from orbit as well as from measurements of atmosphere and on the surface of Mars. During the entry and descent flight the measurements of the atmospheric composition and structure were carried out. The data from temperature and pressure sensors mounted on the lander was combined with accelerometers measurements to obtain the temperature, pressure and density profiles of the Martian atmosphere. Once landed, the study of biology, geochemistry commenced. The physical properties of the Martian soil and meteorological conditions were also studied during the Viking mission. It was the most expensive and ambitious mission ever sent to Mars, with a total cost of roughly US$ 1 billion ($5 billion in 2019 dollars). It was highly successful and formed most of the database of information about Mars until the late 1990s and early 2000s.

The US Viking program consisted of two space probes, Viking-1 and Viking-2. Each spacecraft was composed of two main elements, an orbiter, designed to photograph the surface of Mars from orbit, and a lander that would study the planet

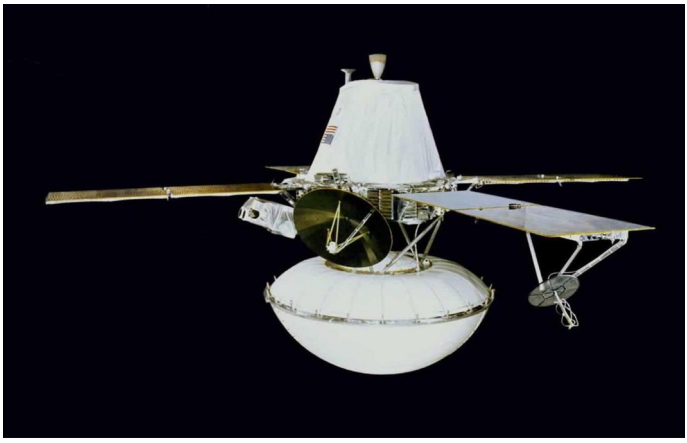


Fig. 9.112 The Viking orbiter with lander in its aeroshell. Image: NASA

from the surface (see Fig. 9.112 for the orbiter/lander-aeroshell combination and Fig. 9.95 for the deployed lander). The orbiters were also the communication platform to relays the landers' science data after touch down. Viking-1 was launched on 20 August, 1975 (reaching Mars orbit about 10 months later), whereas Viking-2 was launched on 9 September, 1975, with Mars-orbit insertion about 11 months later. Both spacecraft were launched with a Titan III-E rocket with a Centaur upper stage.

Landing-site selection was made during Mars orbiting. Before separating from the orbiter, final updates to the lander's computer and a detailed subsystem checkout were performed, and the lander batteries were completely charged. After separation, the lander performed the necessary de-orbit, coast, and entry operations; Lander 1 touched down safely on 20 July, 1976 at *Chryse Planitia* (22.46° north, 48.01° west), whereas Lander 2 performed a safe landing on 3 September, 1976 at *Utopia Planitia* (47.96° north, 225.77° west).

The Viking landing subsystems consisted of several main elements (Holmberg et al. 1980). To save on mass, deorbitation was done with the 12 aft-pointing RCS engines, clustered in four groups of three. These engines were nominally low-thrust, monopropellant hydrazine (N_2H_4) engines of 36 N at a tank pressure of 24 bar. During entry the lander was protected by an aeroshell and thermal protection system. The aeroshell consisted of a 70° half-angle blunted nose cone with diameter size of 3.5 m. To induce lift on the otherwise axisymmetric vehicle, the centre of mass of the aeroshell was shifted by 4.65 cm along the negative Z_B-axis. Note that both RCS thrusters and their fuel tanks were mounted at the edge of the (backside of the) aeroshell, such that at parachute deployment their mass could be discarded. Thermal protection was provided by a heat shield, which was made of phenolic honeycomb, filled with ablative material. The back cover of the aeroshell was constructed of the laminated glass fabric and phenolic resin. One of the main elements of the EDLS was the parachute system, which came up for discussion many times in Sect. 9.2. The

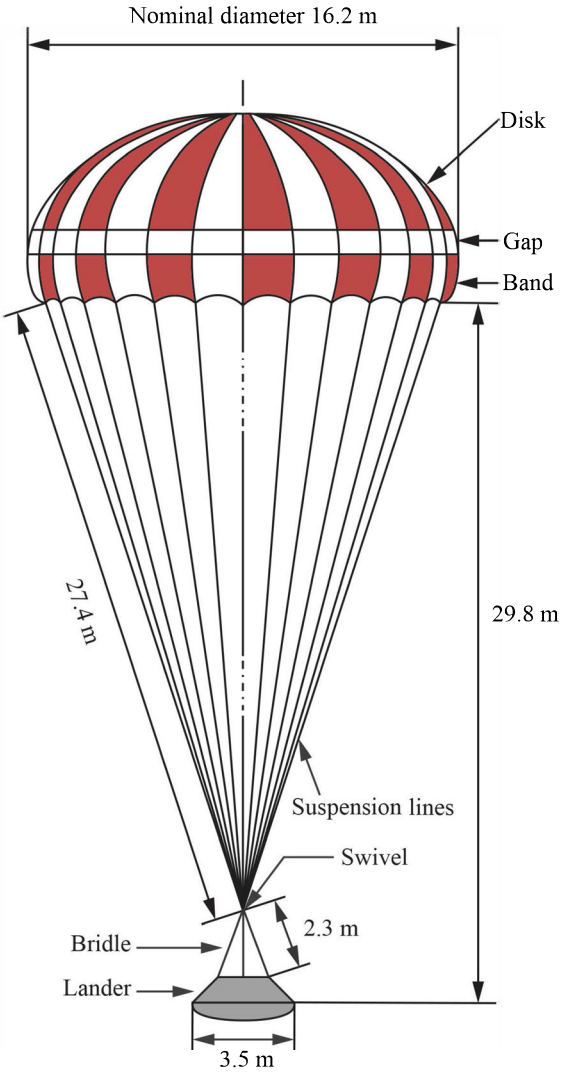


Fig. 9.113 Viking-lander disk-gap-band parachute. Data taken from Lundstrom et al. (1974)

parachute was a disk-gap band with a nominal diameter of 16.1 m (see Fig. 9.113), constructed mainly from Dacron, with a total mass of 44 kg. It was deployed through the base cover by a pyrotechnic mortar device.

The *Terminal Descent Propulsion (TDP) system* consisted of three engines, four roll thrusters and two fuel tanks inside the lander body (see Sect. 7.4.3, and in particular Fig. 7.36, for a schematic layout of the the Viking TDP system). The TDP system was a blow-down pressure-fed hydrazine-monopropellant system. The fuel tanks were made from titanium containing propellant and nitrogen pressurant. Each of these spontaneous catalytic TDP engines delivered 2670 N of thrust. Thrust control

was accomplished by a throttle valve which regulated the amount of hydrazine fuel. Finally, the *Lander* had a body made of aluminium and titanium, and a triangular shape, with the landing legs located at every corner. The lander's legs had the following functions: to provide upright and stable position on the surface, to minimise the shock impact at final touch down, to activate immediate descent engine shut down at first contact with ground, and to support the lander for the full mission duration, satisfying stiffness requirements imposed by cameras and high-gain antennas.

The mission of the lander would officially start after separation from the orbiter, pyrotechnically initiated and with springs imparting a small relative velocity (Cooley and Lewis 1977). Before the actual deorbitation, the lander's accelerometers were calibrated, and its attitude-control system was activated, such that the descent module could be oriented for the manoeuvre. As previously mentioned, the 12 RCS thrusters were used as deorbitation system. During a burn of close to 30 min they provided a ΔV of 180 m/s, using about 73 kg of fuel. The atmospheric entry of the Viking lander took place approximately 3 hr after the separation from the orbiter. At that moment, the entry velocity was about 4.6 km/s, with the entry interface defined at 243.8 km (chosen such that about 7–10 min would remain between entry and touchdown). The inertial flight-path angle at entry was approximately $-17°$. Figure 9.114 shows the complete entry and descent profile, which will be discussed below.

The guidance and control system had been activated before, and had reoriented the module with the RCS to properly position the heat shield (Ingoldby 1978). The attitude controller would keep the pitch attitude fixed to maintain the nominal angle of attack of $-11.1°$, which, as we mentioned above, is non-zero due to the offset of the centre of mass. Initially, velocity navigation was started from bootstrap[42] estimates of inertial velocity and position. Eventually the position and velocity data were corrected, once additional sensors were switched on: the radar altimeter at the entry interface and the terminal descent and landing radar (TDLR), also known as landing velocity radar, during the parachute phase at approximately 4.57 km. Once the atmospheric density had increased sufficiently, i.e., when a deceleration of 0.05 g was sensed at around 91 km altitude, the attitude-hold mode of control in pitch and yaw was released and control was shifted to the aeroshell dynamic-stability control, with additional rate damping.

The deceleration in the hypersonic phase of the descent was provided by the aeroshell. Maximum dynamic pressure was encountered at about 206 s after entry, giving a maximum deceleration of 7.28 g (7.14 g for Viking-2). Just prior to that the maximum stagnation heat flux was encountered, i.e., $q_c = 237.4$ kW/m^2 for Viking-1 and 243.5 kW/m^2 for Viking-2. The deployment altitude of the disk-gap band parachute was at an altitude of approximately 5.9 km at Mach 1.1 (see also Fig. 9.27 for the deployment requirements of the Viking parachutes: Mach numbers less than or equal to 2.1, at dynamic pressures from 239 to 413 Pa); Table 9.9 shows the flight data of parachute deployment. Seven seconds later the aeroshell was jettisoned, and

[42] The process of bootstrapping means simply integrating the acceleration data, without any filter correction from external measurements.

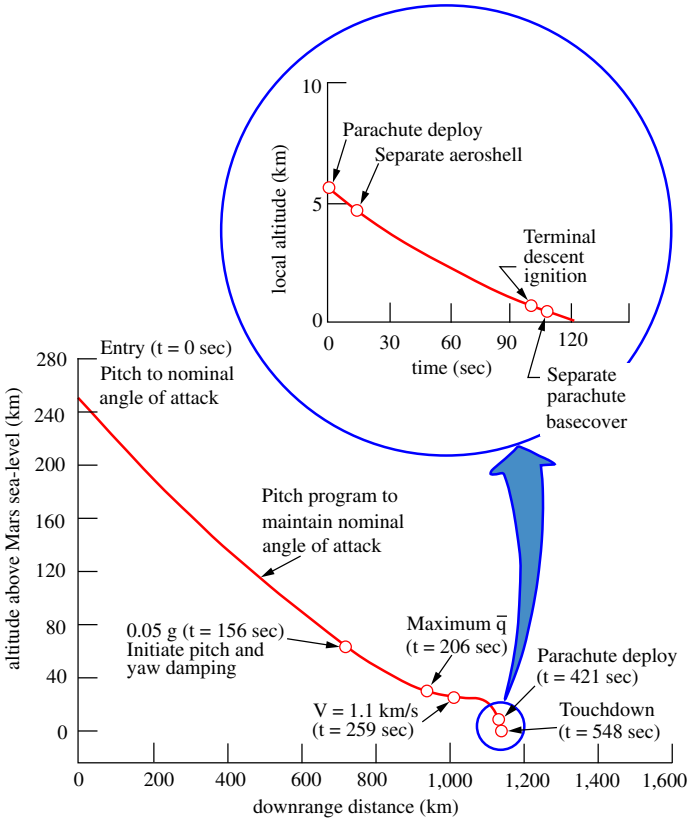


Fig. 9.114 Lander entry sequence. Data taken from Cooley and Lewis (1977)

Table 9.9 Observed conditions for Viking parachute deployment. Data taken and converted from Cooley and Lewis (1977)

Item	Viking-1	Viking-2	Design limit
Radar altitude (m)	5,874.4	5,859.5	5,791.2 ±167.6
Relative velocity (m/s)	232.6	237.1	n.a.[a]
Flight-path angle (deg)	−53.4	−50.8	n.a.
Mach number (–)	1.1	1.1	<2.1
Dynamic pressure (Pa)	320.8	378.3	<413

[a] n.a. = not applicable

8 s later the landing legs were deployed. The deceleration under parachute continued until an altitude of about 1,490 m was reached, with a velocity of about 66 m/s. The last portion of the descent was carried out by the lander released from the backshell, some 45 s later. The drag provided by the parachute also contributed to the separation between the parachute/back- shell combination and the lander.

It was during the parachute phase (subsequent to aeroshell separation) that the terminal descent and landing radar was switched on, to provide measurements through touchdown. The TDLR was a continuous-wave, four-beam Doppler radar, very similar to the system installed on Surveyor and Apollo. It measured the lander velocity relative to the surface. The antenna-phased arrays were an integral part of the component structure, and was mounted on the bottom of the lander for proper field of view. The four beams each made an angle of 21.5° with respect to the lander's vertical axis. Details on the navigation equations and the actual guidance algorithms have been discussed at length by Ingoldby (1978) and Holmberg et al. (1980).

The terminal engines of Viking-1 were ignited at 1,459.0 m (1,438 m for Viking-2, with desired values of $1{,}462.4 \pm 91.4$ m), and 2 s later, the parachute and base cover were separated. The landers then began their terminal descent manoeuvres, such that the flight path became vertical. The terminal-descent propulsion system operated for about 40 s (using about 61 kg of fuel), with the three terminal descent engines being differentially throttled to control both the lander pitch and yaw attitudes, and the lander's velocity as a function of altitude. A constant-velocity descent began at an altitude of just 19.2 m (for either lander). A pyrotechnically operated valve shut down the engines, when the landing legs sensed touchdown. The axial velocity of the landers was 2.4 (2.5) m/s.

More information about the actual performance of the GNC system, and the analysis of the entry and descent trajectory with flight data can be found in the papers and reports by Ingoldby (1978), Ingoldby et al. (1976), and Euler et al. (1978).

9.7.3 Galileo

Galileo—named after the first modern astronomer, the Italian Galileo Galilei—was one of the four flagship missions of NASA, sent out to explore Jupiter and its moons.[43] Launched on 18 October, 1989, aboard Space Shuttle Atlantis, Galileo arrived at Jupiter on 7 December, 1995, via gravity-assist flybys of Venus and Earth, becoming the first spacecraft to orbit this gas giant. En route to Jupiter the spacecraft conducted the first asteroid flyby near 951 Gaspra (29 October, 1991) and discovered the first asteroid moon, Dactyl, around the asteroid 243 Ida (28 August, 1993). It furthermore launched the first probe into Jupiter's atmosphere, which we will discuss in more detail below. On 21 September, 2003, after 14 years in space and 8 years in the Jovian system, Galileo's mission was terminated by sending the orbiter into Jupiter's atmosphere at a speed of over 48 km/s. This was done to eliminate any chance of a future impact with Jupiter's moon Europa that could contaminate the icy moon with terrestrial bacteria. The mission's total cost was estimated at approximately US$ 1.4 billion.

The scientific yield of the Galileo mission was tremendous. Galileo changed the way we look at our solar system. As said, the spacecraft was the first to fly past an

[43] http://solarsystem.nasa.gov/galileo/.

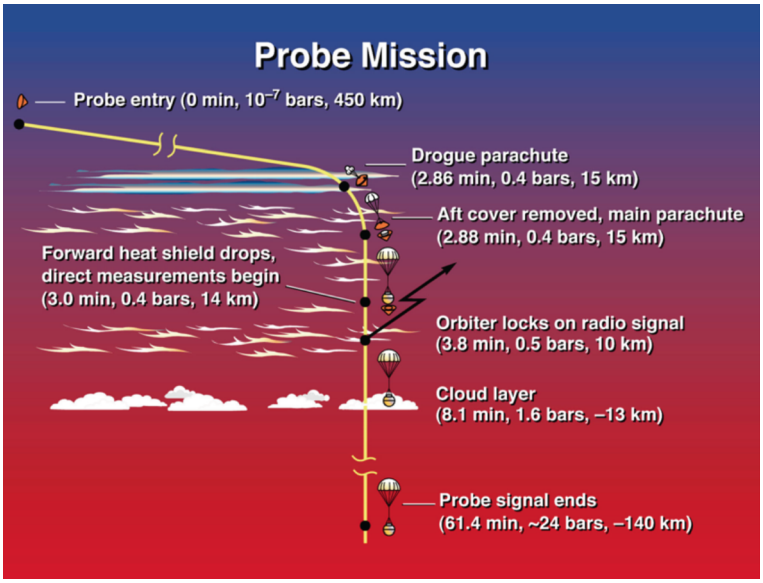


Fig. 9.115 Mission profile of the Galileo entry probe. Image: NASA

asteroid and the first to discover a moon of an asteroid. It provided the only direct observations of a comet (Shoemaker-Levi 9) colliding with a planet (over a 5.6 day period in July 1994). Galileo was the first to measure Jupiter's atmosphere with a descent probe and the first to conduct long-term observations of the Jovian system from orbit. It found evidence of subsurface saltwater on Europa, Ganymede and Callisto and revealed the intensity of volcanic activity on Io.

Upon arrival at Jupiter in December 1995, the Galileo spacecraft delivered a probe that descended into the giant planet's atmosphere. In Fig. 9.115 the nominal mission profile of the probe is shown. The official entry interface was defined to be at an altitude of 450 km above the 1-bar pressure level in the atmosphere. At that moment the atmospheric pressure was just 10^{-7} bar. During the following entry, the pressure steadily increased, as well as the density. The increasing density, in combination with the large velocity, caused a severe thermal environment. The probe, however, was protected by a heat shield (see Figs. 1.51 and 9.116). After passing the peak heat flux, at an altitude of 15 km (pressure level: 0.4 bar) and 2.86 min into the mission, the drogue chute was released. The drogue chute stabilised the probe, provided initial deceleration and pulled away the aft cover ($t = 2.88$ min, still close to $h = 15$ km). A little later the main parachute was deployed, and the forward heat shield got jettisoned ($t = 3$ min, $h = 14$ km), see also Fig. 9.116.

As the instruments were now no longer covered, at this moment the direct measurements could begin. Of course, the measurement data had to be transmitted back to the Galileo orbiter more than 209,215 km overhead for storage and transmission to Earth. Radio contact between orbiter and probe was established after 3.8 min, at

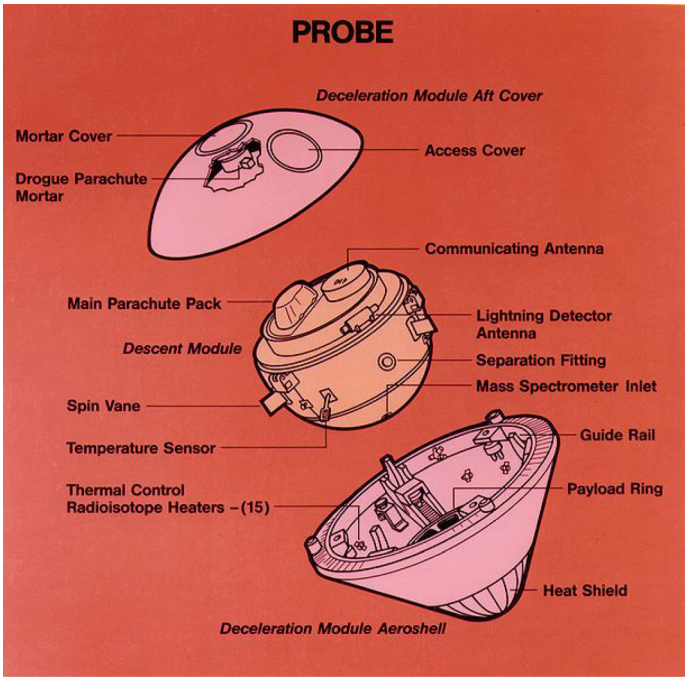


Fig. 9.116 Major elements and subsystems of the Galileo atmospheric probe. Image: NASA

an altitude of 10 km (pressure level: 0.5 bar). For close to an hour the instruments of the probe did their job, all the time transmitting data back to the orbiter. The Galileo probe (Fig. 9.116) was equipped with three spin vanes to guide its spin rate during parachute descent (see also the discussion in Sect. 7.4.4). The probe's spin design was mainly to ensure a spin of less than 50 RPM, to avoid excessive Doppler shift on the radio link (Lorenz 2006). At an altitude of -13 km (pressure level: 1.6 bar), 8.1 min after entry, the probe past the cloud layer and continued descending down. Finally, the probe signal ended after 61.4 min after entry at a pressure level of about 24 bar ($h = -140$ km). It had descended deeper into the atmosphere than expected, but was finally overcome by Jupiter's high temperatures and pressures.

The parachute subsystem consisted of a pilot parachute packed inside a mortar assembly and a main parachute mounted on the probe descent module (Rodier et al. 1981). The mortar assembly was an integral part of the probe aft-cover structure and the phenolic nylon heat shield. The pilot parachute was dimensioned to ensure extraction of the aft cover through the worst-case reverse flow region. The main parachute was sized to meet two criteria: (1) a descent time no greater than 45 min to the 15-bar pressure altitude in the worst-case model Jovian atmosphere, and (2) the ballistic coefficient of the forward aeroshell without the heat shield attached must be greater than 1.05 times that of the descent module and parachute combination. The final design of both pilot and main parachute is shown in Fig. 9.117.

Design type	20° conical ribbon
Constructed diameter (D_0)	1.14 m
Projected diameter (D_0)	0.74 m
Drag area	0.49 m^2
Horizontal ribbons	8
Porosity (λ_G)	16.4%
Gores	8
Material (canopy and lines)	Dacron
Material (riser and bridle)	Dacron
Weight fabrics	0.413 kg

Design type	20° conical ribbon
Constructed diameter (D_0)	3.80 m
Projected diameter (D_0)	2.47 m
Drag area	5.46 m^2
Horizontal ribbons	25
Porosity (λ_G)	24.8%
Gores	16
Material (canopy and lines)	Dacron
Material (riser and bridle)	Kevlar
Weight fabrics	3.674 kg

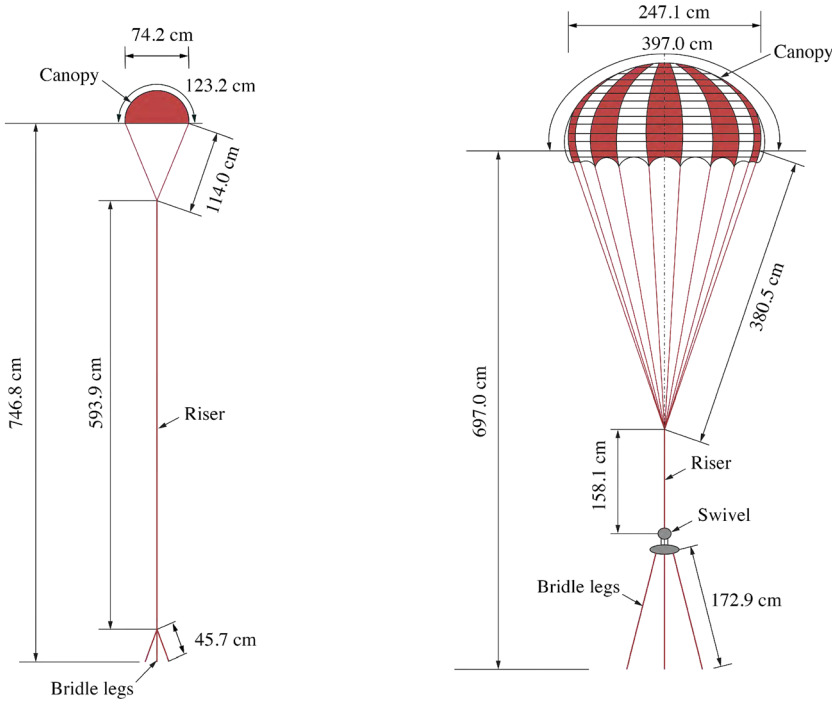


Fig. 9.117 Configuration of the Galileo probe pilot (left) and main (right) parachute. Data taken and converted from Rodier et al. (1981)

The pilot chute of the Galileo entry probe deployed about one minute later than anticipated, resulting in a small loss of upper atmospheric measurements. After a detailed review of the flight and support data sent back to Earth the conclusion was drawn that faulty wiring in the parachute control system was the most likely cause. The fact that the parachute opened at all was attributed to luck, although some people claim that a second design mistake counteracted the first one. It is believed today that the accelerometer controlling the parachute's pyrotechnics was installed backwards, and would not have responded correctly to the deceleration levels the probe was experiencing (Fig. 9.118).

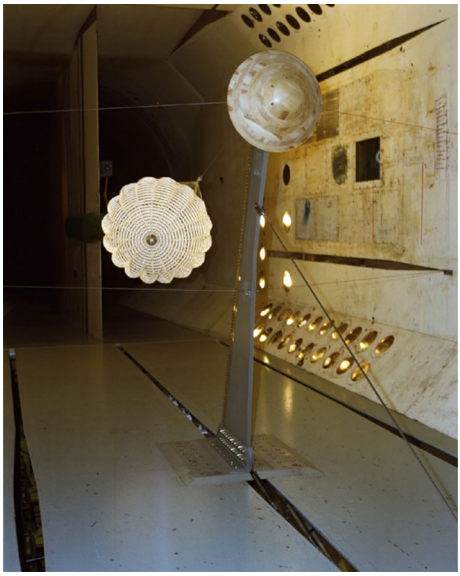
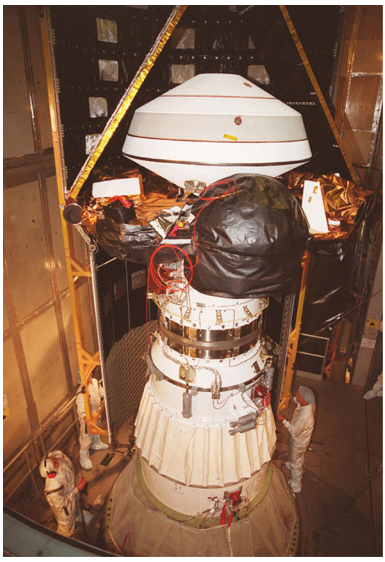


Fig. 9.118 Wind tunnel testing of the main parachute of the Galileo entry probe. Image: NASA

The same thing happened to the Genesis probe's sample return capsule when it returned to Earth in September 2004. Genesis had collected solar dust, stored the dust grains in a sample container and a small entry probe was to deliver the container back to Earth. The parachute failed to open, and the probe kept on tumbling and finally crashed in the Utah desert with a velocity of about 360 km/h. The capsule was largely destroyed, but fortunately the sample container more or less survived the impact. At least the mission was not a complete loss! Also in this case the cause of parachute-deployment failure was due to an accelerometer installed backwards.

These two instances are typical examples of Murphy's Law, attributed to rocket scientist Edward A. Murphy Jr. half a century ago. Although everybody knows this law in the popular formulation of *"If something can go wrong, it will go wrong"*, space historians point out that the original formulation of Murphy's Law was more limited: *"Every component that can be installed backward, eventually will be."* At the time, Murphy was testing human tolerance for g-forces during rapid deceleration. The tests used a rocket sled mounted on a railroad track with a series of hydraulic brakes at the end. Murphy had proposed using electronic strain gauges attached to the restraining clamps of the test-person's harness to measure the force exerted on them during the test.

Wired by his assistant, the sensors provided a zero reading; however, it became apparent that they had been installed incorrectly, with each sensor wired backwards. Of course Murphy blamed the poor assistant and made his pronouncement, despite being offered the time and chance to calibrate and test the sensor installation prior to the test. Although Murphy was at much at fault as his assistant, according to some he claimed: *"If that guy has any way of making a mistake, he will."* Others, for instance



(a) Genesis Spacecraft in fairing of the Delta Rocket, with on top the entry probe



(b) After the crash.

Fig. 9.119 Genesis entry probe. Images: NASA

Murphy's son Robert, say that the statement was along the lines of *"If there's more than one way to do a job, and one of those ways will result in disaster, then somebody will do it that way."* (Spark 2006).

9.7.4 Stardust

Stardust was a 300-kilogram robotic space mission launched by NASA on 7 February, 1999, with as primary mission to collect samples from the coma of comet Wild 2, coming within a distance of 149 km (Desai and Qualls 2010). En route to the comet it also observed asteroid 5535 Annefrank during its flyby at 3,079 km distance. After 6 years, 11 months, and 7 days the primary mission ended, when the Stardust spacecraft delivered the sample return capsule to Earth. This capsule entered the Earth's atmosphere on 15 January, 2006, and safely delivered the sample container with both interstellar and comet dust. The Stardust spacecraft continued its mission that was codenamed NExT. In February 2011 it intercepted comet Tempel 1, a small periodic Jupiter-family comet, which had previously been visited by the Deep-Impact mission (2005). It did so at only 181 km distance. After a total of 12 years, 1 month, and 17 days, Stardust ceased operations and was decommissioned on 24 March, 2011.

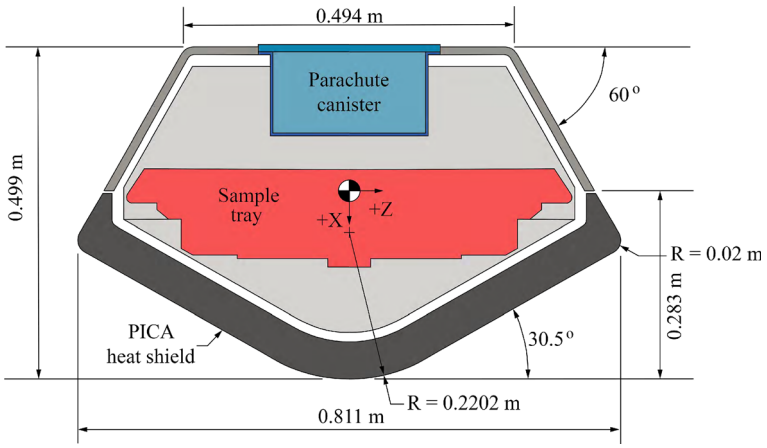


Fig. 9.120 The Stardust sample return capsule. Data taken from Desai et al. (2008)

Besides the scientific importance of the mission—in addition to the dust particles, scientists also found evidence for the presence of liquid water in comet Wild 2—the mission of the sample return capsule (SRC) was also a challenge.[44] With its inertial velocity of 12.799 km/s it was the entry of a man-made object with the highest velocity. Until then that honour had been reserved for the return of Apollo-10, which had an entry velocity of 11.069 km/s (Orloff 2001). For comparison: just over four years later (June 13, 2010), the Hayabusa sample return capsule made the second-fastest entry with an inertial velocity of 12.04 km/s, observed by NASA from a DC-8 flying observatory (Cassell et al. 2011). These high-velocity entries imply large thermo-mechanical loads, requiring a robust thermo-structural design. Also the sensitive payload would require a reliable parachute system to ensure soft landing.

The parachute system of the Stardust SRC consisted of two parachutes, i.e., a drogue and a main parachute (Witkowski 1999b), see Fig. 9.121. The drogue parachute was a disk-gap-band parachute, with a nominal diameter of 0.83 m (32.6 inch). It had been selected on the basis of required deployment at an altitude of approximately 32 km (105,000 ft) and a Mach number of $M = 1.4 \pm 0.3$. Its release was determined to be at approximately 2,438 m (8000 ft) and $M = 0.15 + 0.5$. The canopy was made of 1.1 oz nylon fabric, whereas the suspension lines were constructed of Kevlar cord with a tensile strength of 890 N (200 lbf). The drogue parachute was packed directly into the mortar sabot, eliminating the need for a deployment bag (and thus saving mass and volume). To protect the parachute from entry-heating damage, a thermal pad was included at the sabot mouth.

[44] The Stardust capsule has a diameter of about 0.811 m. Its forebody is a blunted sphere cone with a 60° half-angle. The afterbody is a 30° truncated cone (See Fig. 9.120). At entry, the capsule weighed 45.8 kg. Last, but not least, the capsule's thermal protection system was a phenolic-impregnated carbon ablator (PICA).

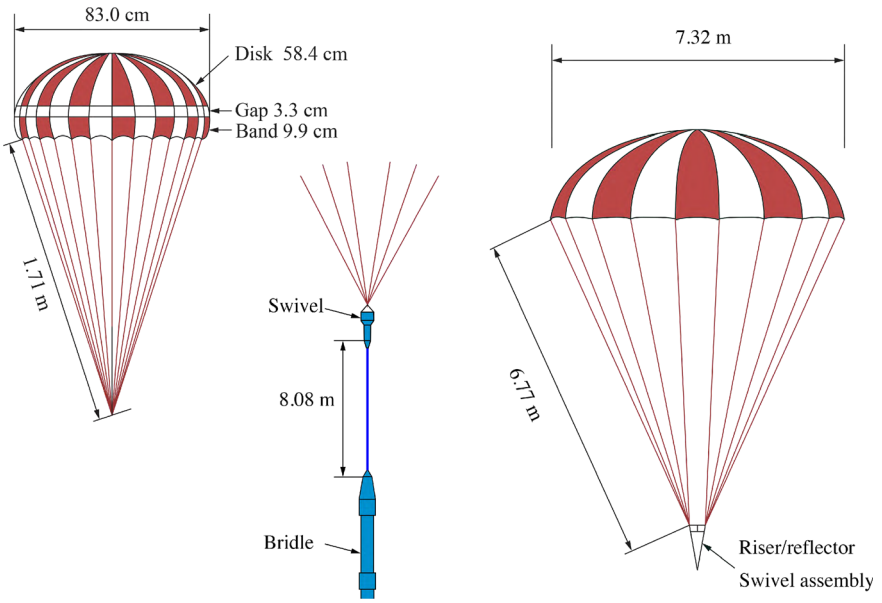


Fig. 9.121 Configuration of the Stardust probe drogue (left) and main (right) parachute. Data taken and converted from Witkowski (1999b)

The main parachute was selected as a best fit for meeting the descent rate under terminal conditions and the maximum allowable volume, which was limited due to the size of the capsule. The selected design was a triconical parachute with a diameter of 7.3 m (24 ft) and a total of 20 gores. As for the drogue parachute, also here the canopy fabric was 1.1 oz Nylon and the suspension lines were the same Kevlar cord, 6.8 m in length. In its packed state, the total weight of the main parachute was about 2 kg (4.5 lb). It was packed into an aluminised Kevlar deployment bag at a density of approximately 641 kg/m^3 (40 lb/ft^3). After deployment, the main parachute provided a predicted descent rate of 4.4 m/s (14.5 ft/s) at the elevation of the landing site.

The SRC entered the Earth's atmosphere at a very energetic velocity of 12.8 km/s. The high-altitude portion of the Stardust entry trajectory generated very high-energy, strongly non-equilibrium flow conditions. During its entry, measurements of the radiation emitted from the shock-heated flow were obtained with several instruments, installed on an aircraft some hundreds of kilometres away. Nonetheless, these data could be used to corroborate the aerothermodynamic design approach, discussed by Boyd et al. (2010) and Boyd and Jenniskens (2010).

For the aerodynamic analysis they applied both both continuum (computational-fluid dynamics solutions of the continuum Navier–Stokes equations) and particle (direct simulation Monte Carlo) methods to the capsule's forebody flow at altitudes of 81 and 71 km (near-continuum conditions). For both methods, an 11-species, 19-reaction, thermochemical non-equilibrium approach was used, but ablation processes

were not taken into account. It was found that at the selected (high) altitudes the continuum results were sensitive to the chemistry modelling and also non-continuum effects at the capsule's surface created significant differences in basic flow-field properties. Between the two methods. At lower altitudes the chemistry modelling became less of an issue for the continuum method, although the number densities of N_2 and electrons still gave significant differences.

The corroboration between the two flow solutions and the (external) measurements led to a modified approach for determining the temperature used by the CFD solver to control the dissociation and ionisation reactions. With the modified approach a much better agreement was achieved, and the importance of chemistry modelling for highly energetic flows was much better understood. Even though significant differences in temperatures near the capsule's wall and in the surface heat flux remain, a well-defined approach has been established for (future) high-velocity entries.

Having discussed the descent system, as well as aerothermodynamic design issues, let us now having a closer look at the entry and descent of the SRC (Fig. 9.122). On the morning of 16 January 2006, four hours before entry, the Stardust capsule was spun up to a nominal rate of 13.5 RPM and separated from the main spacecraft. This spin rate was required to maintain its inertial attitude during the coasting phase and guarantee the nominal zero angle of attack. Due to the lack of active guidance and control, the capsule relied on passive aerodynamic stabilisation during its entry flight through all aerodynamic regimes, from hypersonic all the way to supersonic parachute deployment.[45] This supersonic drogue parachute was required to stabilise the otherwise unstable probe in the transonic regime. The maximum *g-load* encountered was 32.89 g.

A deceleration switch, also known as *G-switch*, triggered a timer to release the drogue chute. Activated at 3 g, the timer sent a release pulse and the drogue chute was deployed at approximately Mach 1.23 and an altitude of 31.03 km. Video footage obtained from the Utah Test and Training Range (UTTR) radar tracking-stations was used to determined the deployment times of the drogue and main parachutes, and it was found that the former deployed 137.9 s after entry, which was 4.8 s later than had been determined during pre-mission analyses. Then, a second timer was initiated for triggering the main parachute deployment. That happened 331.3 s after drogue deployment, 19.3 s earlier than what had been planned. At that moment the capsule was at 3.428 km altitude (above mean sea level, but 1.8 km above ground). Due to the sustained winds, the landing footprint had shifted north, and the final landing location was 8.1 km north-northwest of the desired target.

Including 3σ-bounds on the atmospheric density, correlating this with balloon measurements of both density and winds, the actual entry of the Stardust SRC was very close to the predicted trajectory, and well within the 3σ-bounds of the performed Monte-Carlo analyses. More details about the Stardust entry reconstruction

[45] Because no accelerometers or gyroscopes were included in the capsule's design, the actual stability of the capsule could not readily be assessed. However, with the use of infrared video footage, as well as inspection of the heat shield attitude information could be derived. The charring patterns on the heat shield implied a symmetric heating and a uniform recession, and it was concluded that the capsule attitude must have been only a few degrees during hypersonic flight.

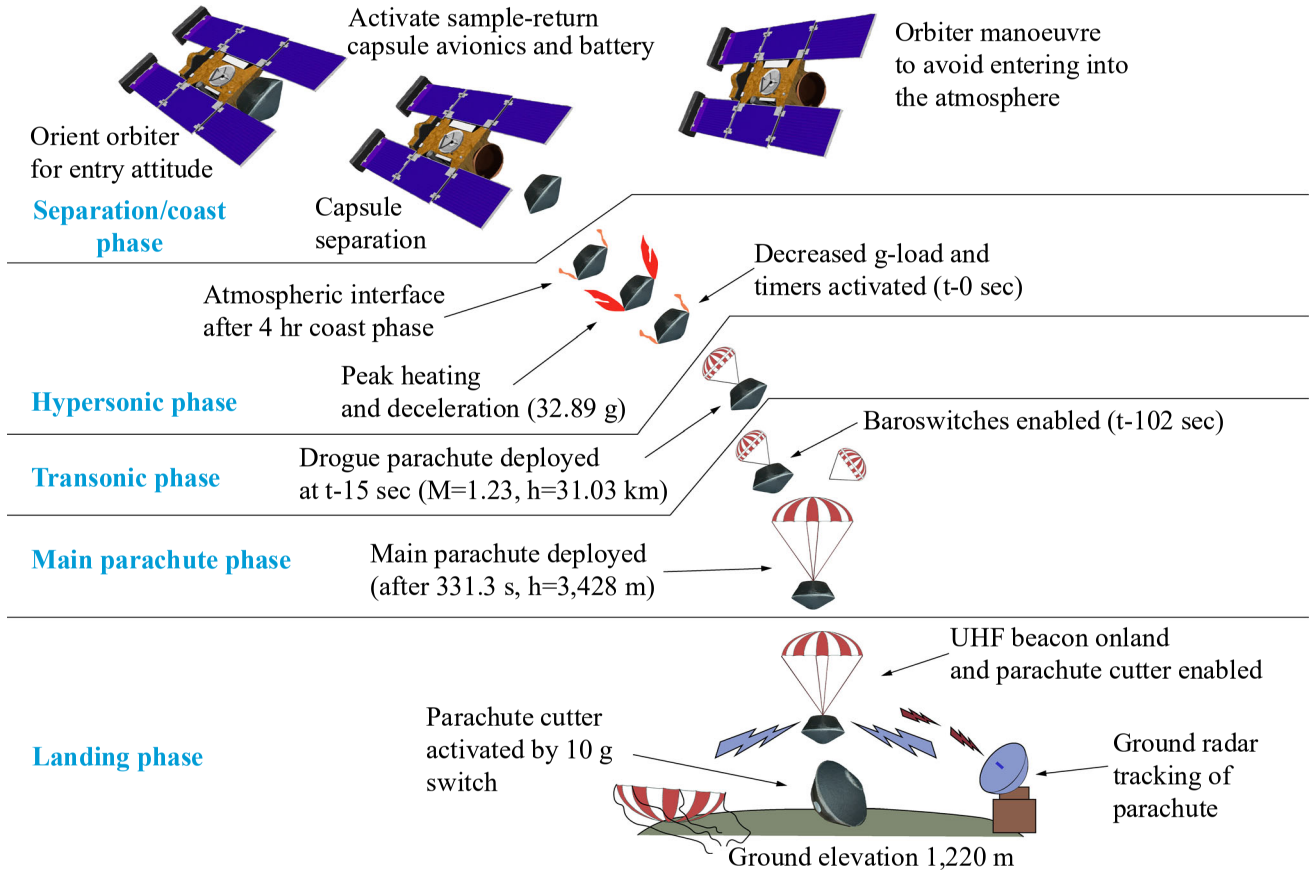


Fig. 9.122 Stardust sample-return capsule entry mission. Orbiter and capsule images: NASA, flight data taken from Desai and Qualls (2010)

is provided by Desai and Qualls (2010), whereas a detailed discussion on the mission design and entry, descent, and landing operations is given by Desai et al. (2008). There were no indications, established from the available data, that the capsule's aerodynamic performance deviated significantly from the nominal predictions. Thus, it was concluded that a sound methodology to analyse the flight dynamics, aerodynamics, and aerothermodynamics has been established.

9.7.5 MER-A and B

NASA's Mars Exploration Rover (MER) missions were robotic missions to mars to explore the Martian surface and geology. It can be seen as a direct follow up of the Mars Pathfinder mission,[46] using many of the same (or similar) sub-systems, as well as inheriting its entry, descent, and landing scenario. The first mission, designated MER-A and carrying the rover *Spirit*, was launched June 10, 2003, and landed on 4 January, 2004. The second mission, MER-B with on-board the rover *Opportunity*, took off from Kennedy Space Center on 7 July, 2003, and the rover was safely delivered to the surface on 25 January, 2004. The *Spirit* rover was targeted at the *Gusev crater* (14.59° south, 175.3° east), while *Opportunity* landed at the *Meridiani Planum* (1.98° south, 5.94° west). Compared with the *Sojourner* rover from NASA's earlier Mars Pathfinder mission, the size (about 1.5 m by 1 m) and capabilities of the MER rovers were considerably extended. The rovers were capable to transverse large distances and carried a larger suite of scientific instruments. Nominal mission duration of both rovers was 90 Martian solar days. However, the excellent state of the hardware allowed for several mission extensions. In the end, *Spirit* was active until March 22, 2010 (total mission duration of 2,269 days (2,208 sols) from landing to last contact) while MER-B Opportunity was active until June 10, 2018 (it was operational on Mars for 5,110 sols (5,250 days, or 14 years and 136 days).

The entry, descent, and landing scenario for both rovers was very similar to that of the Mars Pathfinder mission. The notable difference between the two mission's EDL was that the payload of MER was about 78% larger than Mars Pathfinder's payload (445 kg vs. 250 kg). Similar to Pathfinder, deceleration of the landers was achieved with the aid of an aeroshell, a supersonic disk-gap band parachute, solid propellant retro-rockets, and airbags. Consequently, the heritage of Pathfinder was used extensively to design the MER mission. However, the entry time, entry mass, and landing site altitude were all different. The difference between these principal EDL parameters required a significant redesign of the original Mars Pathfinder EDLS. For example, the increased entry mass required a more efficient aerodynamic decelerator. To sufficiently slow down the MER vehicle, the disk-gap band parachute area was

[46] Mars Pathfinder was an American robotic spacecraft that landed a base station with a rover *Sojourner* on Mars in 1997. Launched on December 4, 1996, it was NASA's first "faster, better, and cheaper" spacecraft, developed in a mere three years time at one fifteenth of the cost of the Viking program. See for more information on system and mission data the papers by Golombek (1997) and Witkowski (1999a).

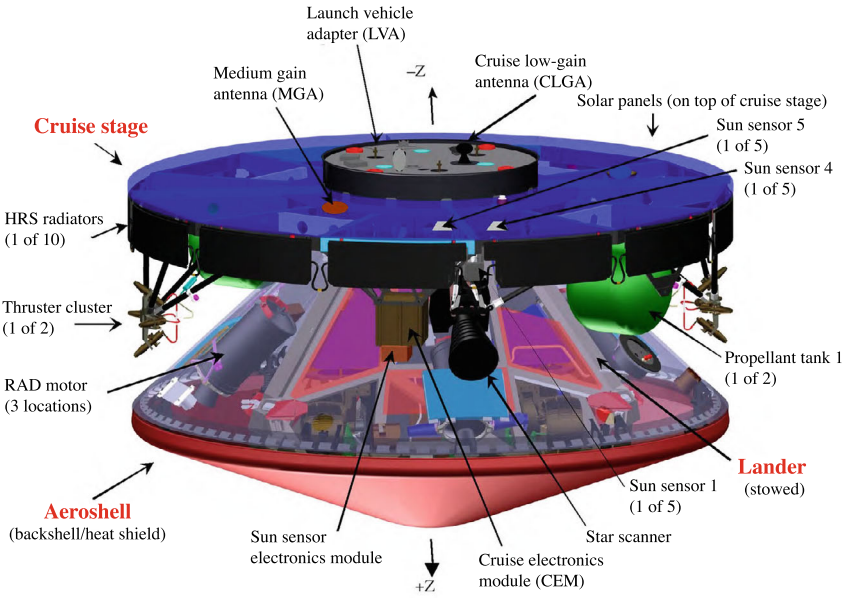


Fig. 9.123 Transfer configuration of the Mars Exploration Rover: cruise stage and entry capsule with lander stowed inside. Original image: NASA

increased by 40% and its overall construction was strengthened accordingly. The modification of the airbag system was also needed to satisfy the surface impact requirements.

The main two elements of Mars Exploration Rover were an orbital cruise stage, and an aeroshell that had the lander stowed inside and would protect it during the entry and descent (Fig. 9.123). With a total mass of 1,063 kg, the cruise stage would take 193 kg, the rover 185 kg, and the rest was taken up by the aeroshell and propellant (lander 348 kg, backshell and parachute 209 kg, heat shield 78 kg, instruments 5 kg, and the propellant 50 kg). As the focus of the current discussion is on the entry, descent, and landing, the relevant sub-systems are discussed now, taken information from the papers by Roncoli and Ludwinski (2002), Stelzner et al. (2005), and San Martin and Bailey (2005).

The design of the *aeroshell* was based on the Pathfinder architecture with minor modifications to increase volume. The shape of the forebody was a classical Viking 70° half-angle sphere cone. The thermal protection material was SLA-521. The heat-shield protection layer was downsized to 1.63 cm instead of Pathfinder's 1.91 cm due to less severe entry conditions (at the entry interface of at 125 km altitude, the (relative) velocity was considerably lower (5,630/5,700 m/s vs. 7,015 m/s) and the flight path was shallower (nominal inertial flight-path angle was −11.5° vs. −14.5°).

Due to combination of factors, such as higher entry mass, less dense atmosphere, and higher landing site altitude, a number of important changes to the *parachute sys-*

tem design had to be made. All events during entry and descent would take more time, and less time was available, thus a more efficient parachute deceleration had to be employed. The applied modifications involved a higher parachute deployment altitude, an adjusted parachute deployment algorithm and enlarged area of the parachute canopy. The deployment altitude was changed by setting the mean dynamic pressure of MER to a higher value of 725 Pa (for comparison, Pathfinder's value was about 600 Pa), which led to a deployment altitude of 9.5 km instead of 8.2 km. The deployment algorithm was triggered by the dynamic pressure and not, as was the case earlier, by the deceleration value. This approach provided more certainty if any off-nominal conditions were encountered. Finally, the parachute area was increased, i.e., the diameter went up from 12.74 to 14.1 m. The parachute drag coefficient also became slightly larger (C_D was 0.43 vs. 0.41) due to adjustments of the parachute band length.

After parachute deployment, it took the MER vehicle about 120 s to reach the Martian surface. However, the density variations could reduce this descent time to about 60 s or even less. Such reduction in parachute decelerator performance would result in the upper-bound terminal landing velocities of 85–90 m/s. As a consequence of this, thrust delivered by the retro-rockets had to be proportionally increased. Therefore, the impulse of the larger MER's retro-rockets was 70% higher than the corresponding Pathfinder value. The retro-rocket system, formally known as the *Rocket Assisted Decelerator* (RAD), consisted of three solid rockets, placed in a symmetric pattern along the circumference of the backshell, providing a total impulse of 95,000 Ns. The radar altimeter data was used during the descent to ignite the RAD system at approximately 160 m altitude. The deceleration rate was such that the vehicle was expected to dissipate all of its vertical velocity at 12 m above ground level. At that point, the bridle connecting the backshell and the rover was cut and the parachute/back-shell combination was removed by the residual RAD impulse.

The possibility of high, near-surface winds and wind shears required an additional set of steering rockets named the *Transverse Impulse Rocket System* (TIRS). This system was not included in Mars Pathfinder's entry, descent, and landing sequence. The TIRS consisted of three motors located around the backshell with 120° angular separation between the motors. The impulse delivered by each of these motors was 2,000 Ns. The TIRS motors were 85° canted with respect to the RAD thrust vector.

During the terminal phase before landing, a descent camera on MER took three pictures with 5 s separation time, such that the horizontal velocity could be estimated. The data supplied by the vehicle's Inertial Measurement Unit (IMU) and this horizontal velocity estimate were then processed by the control system to determine whether the TIRS had to be initiated. The performance of the TIRS proved to be very good, especially during the landing at the *Gusev crater*, reducing the horizontal velocity from an estimated 23 m/s to about 11 m/s. The latter value was well within the capabilities of the airbag system.

These *airbags* leaned heavily on Pathfinder heritage, as it had used the same airbag system. After the RAD system firing solution was determined, but prior to the actual firing of RAD or TIRS, the airbags were inflated at approximately 7.8 kPa (1.0 psig), significantly higher than the surface pressure that is in the range of 0.06–1.5 kPa.

The airbag consisted of many Vectran material layers, which protected the inner pressurised double-bladder system. This was different from Pathfinder's airbags that was designed as a single bladder system for a landed mass of 410 kg with an impact velocity of 26 m/s. The multi-layer airbags would allow for a landed mass of 550 kg, with the same impact velocity. In Fig. 9.102, the MER airbags system is shown during a test at NASA's Glenn Research Center in Ohio.

Both MER missions entered the atmosphere directly from interplanetary transfer trajectories, aided only by a cruise stage. After cruise-stage separation, the MER entry velocity was 5.63 km/s for MER-A and 5.70 km/s for MER-B (Desai and Knocke 2004), which is 20% less than that of Mars Pathfinder. The entry angle (the nominal value was −11.5°) was kept as small as possible to accommodate for a larger vehicle mass of 835 kg.

The general EDL sequence of MER is shown in Fig. 9.124. The parachute deployment was nominally targeted at the dynamic pressure of 700 Pa. This dynamic-pressure level was expected to occur approximately 244 s after the entry at the nominal altitude of 9.5 km.[47] The actual (reconstructed) dynamic pressures for parachute deployments occurred at nominal values of 729 and 765 Pa and Mach numbers from 1.71 to 1.94, but the already mentioned uncertainty in the atmospheric properties could have meant that the dynamic pressure could have been as high as 842 Pa and as low as 656 Pa (Witkowski and Kandis 2005). For sure, such a range of possible dynamic pressures would have impacted peak parachute loads, but these were difficult to reconstruct from the returned data. They were found to be within the limits suggested by pre-landing analyses, though.

Parachute deployment was followed by heat-shield release 20 s later. The lander was lowered by bridle 10 s after heat-shield jettisoning. Starting from an altitude of 2.4 km, the radar altimeter provided altitude measurements relative to the surface. The altimeter data allowed for determining the timing of airbag inflation, RAD system ignition and bridle-cut event. Nominally, airbags were inflated 4 s before the RAD firing. At about 15 m above ground level the bridle was cut and the lander was free to fall to the surface. The backshell and parachute were carried away by the RAD engines, which were still in operation.

To conclude this discussion, Fig. 9.125 shows the impact site of the heat-shield of MER-B. In this fantastic mosaic picture, taken by *Opportunity*, several elements and features can be discerned.[48] The main (inverted) part of the heat shield shows its metallic insulation layer. It is about 1 m tall and at about 13 m from the rover. The flat piece of debris near the centre of the image is about 14 m away. On the right side of the figure, the 2.8 m diameter crater made by the heat-shield's impact is clearly visible. The crater that is only about 5–10 cm deep, is about 6 m from Opportunity. Smaller fragments and debris can be seen all around the impact site.

[47] The actual (reconstructed) altitude was considerably lower at about 7.5 km (Desai and Knocke 2004). This was caused by a significant variation of the atmospheric density, due to a dust storm one month earlier.

[48] https://mars.nasa.gov/mer/gallery/press/opportunity/20050203a.html, date accessed August 8, 2021.

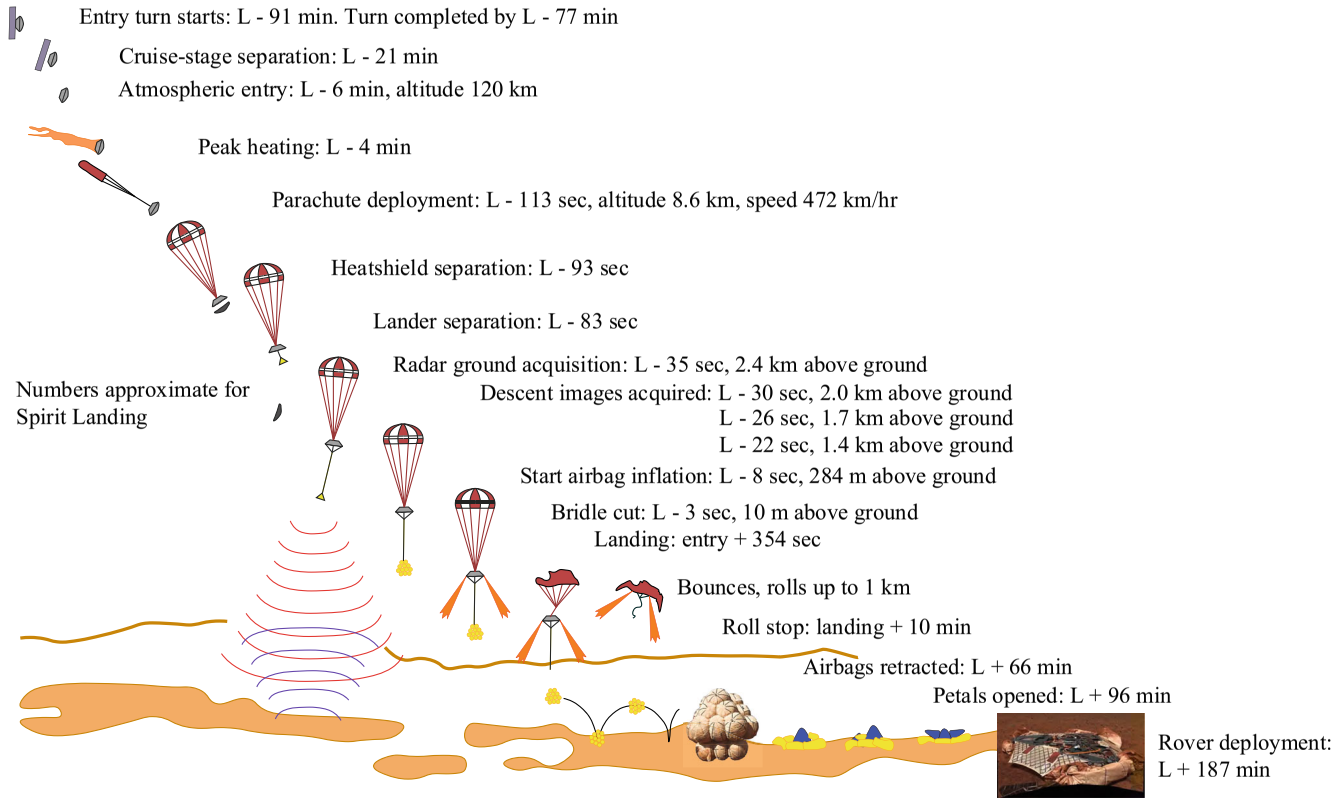


Fig. 9.124 Entry and descent mission profile of the Mars Exploration Rovers. Data taken from Desai and Knocke (2004)

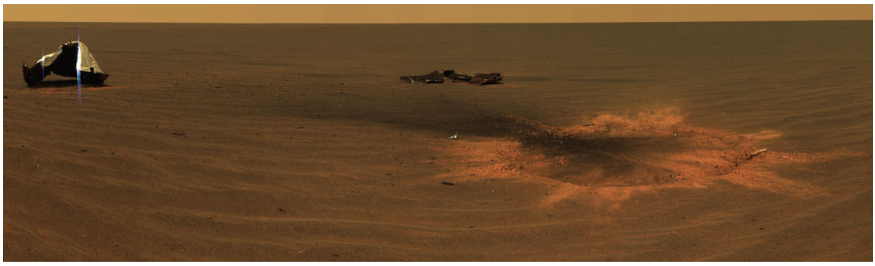


Fig. 9.125 The heat-shield impact site of Mars Exploration Rover *Opportunity*. Image: NASA/JPL/Cornell

The dark material, visible at the crater floor, have formed dark streaks towards the two larger debris parts, indicating the direction of impact and bouncing off.

9.7.6 Mars Science Laboratory

NASA's Mars Science Laboratory (MSL)[49] part of NASA's Mars Exploration Program, was launched on November 26, 2011, with a United Launch Alliance Atlas V launch vehicle, and it landed the rover *Curiosity* on 5 August, 2012, in the *Gale crater*. *Curiosity* was designed to assess whether Mars ever had an environment able to sustain small life forms (microbes). To design and execute the mission, two major challenges lay ahead, the first one being how to deliver such a heavy payload (the rover) of about 900 kg, almost five times the weight of its predecessors, MER-A and MER-B. The second challenge was to reduce the landing ellipse, because the scientifically most interesting regions to land are usually also the most inaccessible, in terms of having a region without too many rocks, slopes and craters. Therefore, precision-landing techniques needed to be developed, of which Knocke et al. (2004) stated that it would require robustness of real-time guidance algorithms to cope with navigation uncertainties, variations in atmospheric properties, and aerodynamic modelling inaccuracies.

Until MSL, all Martian entries had been ballistic and unguided, which explains the large landing ellipses of hundred(s) of kilometres of the earlier missions (see also Fig. 9.67). Even though the Viking entry capsules generated some lift, due to the offset in the centre-of-mass location, this was not actively used in its guidance and the semi-ballistic entry was effectively a ballistic one. To have the capability to

[49] The successor mission, Mars-2020, resembles MSL in most aspects, i.e., in its use of a guided entry, its large parachute, and the use of a sky crane. Because the Mars-2020 rover, *Perseverance*, is in the same weight class, we can focus in this section on MSL, it being the break-through mission. Of course, the inclusion of a helicopter in Mars-2020 that has successfully (and repeatedly) taken off and landed on the Martian surface is a unique and fantastic achievement, but does not really fit in the scope of this chapter.

guide *Curiosity* to its designated landing area, it was required to have lift, an active guidance system, and sensors to feed this system with relevant data. A sequence of complex autonomous operations decreased the landing ellipse to a relatively small target area of 20 × 6 km.

Due to the inherently thin nature of the Martian atmosphere, parachutes have to be rather large and should be deployed as early in the descent as possible, meaning at supersonic speeds. Even for earlier missions (and lower landing masses), the landing had to be supported by retro rockets and/or landing bags. With the large aeroshell mass at deployment of $m = 2{,}925.0$ kg (and a payload mass of about 900 kg), it had reached its limit in the amount of mass that can be delivered, and a different concept of putting the payload on the Martian surface had to be conceived: the sky crane. Both parachute and sky crane created major challenges in development and testing.

The MSL parachute was a disk-gap-band parachute with a nominal diameter of 21.35 m and a disk diameter close to 16 m, it had 80 suspension lines, and projected to a total length of more than 50 m (see Table 9.10 for more details on the parachute's geometry). It could deploy at Mach 2.3 with a maximum design opening force of up to 65,000 pounds, which equals a staggering 289 kN (the actual opening force peaked at 153.8 kN, though). Despite the fact that the disk-gap band is a familiar parachute, its sheer dimensions required extensive testing. Amongst others, its inflation needed to be confirmed, but this could not be done at the operational velocity. However, in 2009 flight-qualification tests at 36 m/s were executed inside the world's largest wind tunnel at NASA Ames Research Center. This wind tunnel stands 24 m tall and is 37 m wide, large enough to house a Boeing 737. Figure 9.126 shows the MSL parachute in its deployed state, with, for reference, in the insert shown in relation to two persons. Most of the orange and white fabric is nylon, though a small disk of heavier polyester is used near the vent in the apex of the canopy due to the higher stresses there.

Cruz et al. (2014) discuss the reconstruction of the MSL parachute performance, using the measurement data of the two inertial measurement units, mounted on the descent stage, providing both filtered and unfiltered data at 200 Hz. From the reconstructed events, it was found that mortar fire to deploy the main parachute started at an altitude of $h = 7{,}118$ m, at a Mach number of $M = 1.75$ with dynamic pressure of $\bar{q} = 493.6$ Pa. The flight-path angle was $\gamma = -22.4°$, i.e., a velocity orientation that is still very much horizontal. With a best time-estimate of 1.16 s later, the suspension lines were stretched ($h = 6{,}938$ m, $M = 1.737$, $\bar{q} = 493.2$ Pa, $\gamma = -22.9°$). Peak opening force happened at $t = 1.78$ s, with $h = 6{,}837$ m, $M = 1.695$, $\bar{q} = 474.0$ Pa, $\gamma = -23.2°$. About 18 s after this ($t = 19.749$ s), the heat shield was separated ($h = 4{,}934$ m, $M = 0.618$, $\bar{q} = 74.9$ Pa, $\gamma = -37.0°$). The parachute phase ended with descent-stage separation at 116.638 s after mortar fire, with conditions $h = -3{,}259$ m (below the reference ellipsoid), $M = 0.324$, $\bar{q} = 42.2$ Pa, and $\gamma = -85.5°$.

Reaching the proper deployment conditions, as well as the right flight path that would intersect with the landing ellipse, was the task of MSL's on-board guidance system. A guided entry into the Martian atmosphere had never been done before, so this justifies some discussion. The interested reader can find in-depth discussions in

Table 9.10 Key geometry parameters of Mars Science Laboratory's disk-gap-band parachute (Cruz et al. 2014)

Parameter	Value
Nominal (reference) diameter D_0, m	21.348
Disk diameter D_D, m	15.447
Band diameter D_B, m	15.608
Vent diameter (unconstrained) D_{VU}, m	1.576
Vent diameter (constrained) D_{VC}, m[a]	1.524
Gap height H_G, m	0.904
Band height H_B, m	2.580
Suspension-line length L_{SL}, m	36.578
Projected diameter D_P, m[b]	14.944
Nominal (reference) area S_0, m^2	357.94
Disk area S_D, m^2	187.40
Gap area S_G, m^2	44.01
Band area S_B, m^2	126.52
Vent area (unconstrained) S_{VU}, m^2	1.95
Vent area (constrained) S_{VC}, m^2	1.82
Geometric porosity (unconstrained vent) λ_{GU}, %	12.84
Geometric porosity (constrained vent) λ_{GC}, %	12.80
Number of gores (suspension lines), N_{gores}	80

[a]Constraint on the vent diameter is due to the vent cords
[b]Value of D_P is an estimate based on the assumption that $D_P/D_0 = 0.700$

the papers by San Martin et al. (2015), Mendeck and Craig (2011) and Way et al. (2013). Similar to Mars Pathfinder and the Mars Exploration Rovers, amongst others, MSL performed a direct entry. Delivered in the proximity of Mars by its cruise stage, the entry, descent, and landing GNC was activated one minute after separation from the cruise stage. The aeroshell was de-spun from the orbital 2 RPM by its reaction-control system and reoriented for the required attitude at entry. Cruise balance masses were jettisoned to create a shift in centre-of-mass location, inducing an L/D of 0.24 at $M = 24$. Travelling at 5.845 km/s, 9 min after GNC activation the spacecraft entered the Martian atmosphere at an entry flight path angle of $-15.474°$, 125 km above the MOLA reference surface. At that moment, the spacecraft was located 631.979 km downrange and 7.869 km cross range of the landing mark. Figure 9.127 gives a visualisation of the complete entry, descent, and landing mission. Details on the reconstructed entry and descent trajectory are provided by Dutta and Braun (2014) and Cruz et al. (2014). A comprehensive table that compares many reconstructed flight parameters to pre-flight simulation predictions is provided by Way et al. (2013).

The angle of attack at entry was maintained through a major part of the entry flight, such that the landing ellipse error could be reduced and the altitude at which the parachute was deployed could be increased. The best-estimate maximum heat-

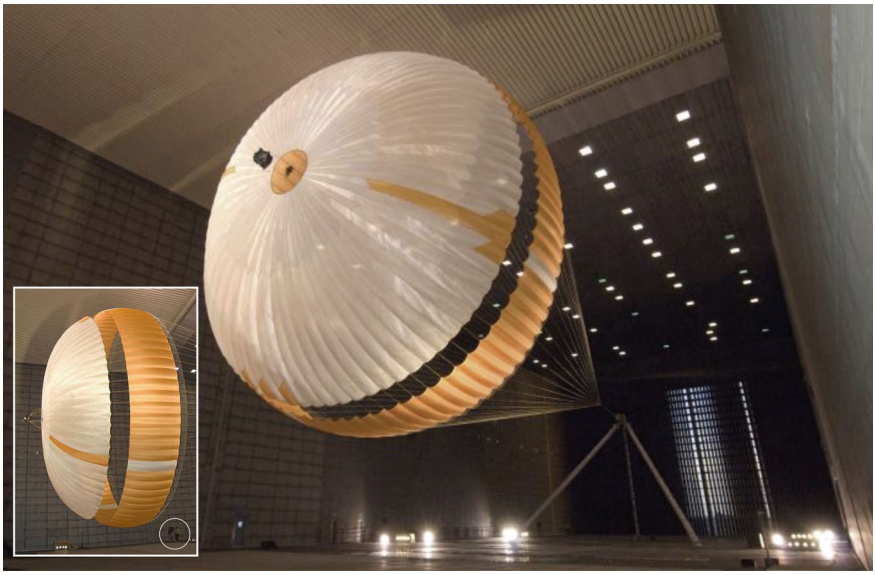


Fig. 9.126 Full-scale testing of the MSL main parachute in NASA's largest subsonic wind tunnel. Images: NASA/JPL-Caltech

flux peak occurred at $h = 25.8$ km with $q_c = 2.14$ MW/m^2 (Edquist et al. 2014). Maximum deceleration peaked at 12.61 Earth g.

The MSL entry guidance algorithm was based on the one used by the Apollo in the 1960s, which also served as the basis for the Shuttle guidance (See Sect. 7.6.2). Therefore, it should not come as a surprise that the main phases of the guidance are similar: (1) pre-bank phase, (2) range-control phase, and (3) heading-alignment phase. Obviously a certain level of aerodynamic force is required for the guidance to be effective, so an acceleration-trigger limit was set to activate range control. This occurred after sensors detected that acceleration had exceeded 0.2 g. Until that moment, in the pre-bank the bank-angle was kept at a zero command.

Range-to-go was continuously calculated and with the estimated drag deceleration and lift-to-drag ratio, the commanded (vertical) lift-to-drag ratio could be calculated, from which the commanded bank angle was derived. The bank angle was commanded in such a way that the predicted downrange error was minimised at parachute deployment (Mendeck and Craig 2011). The error in cross-range was preserved by managing the heading error in relation with a manageable dead-band limit. At a relative velocity of 900 m/s the heading alignment was initiated for the minimisation of the residual of the cross-range error. San Martin et al. (2015) report three executed bank reversals during MSL's entry flight. At the end of the heading-alignment phase parachute deployment was prepared by bringing the vehicle's angle of attack back to zero by shifting balance masses.

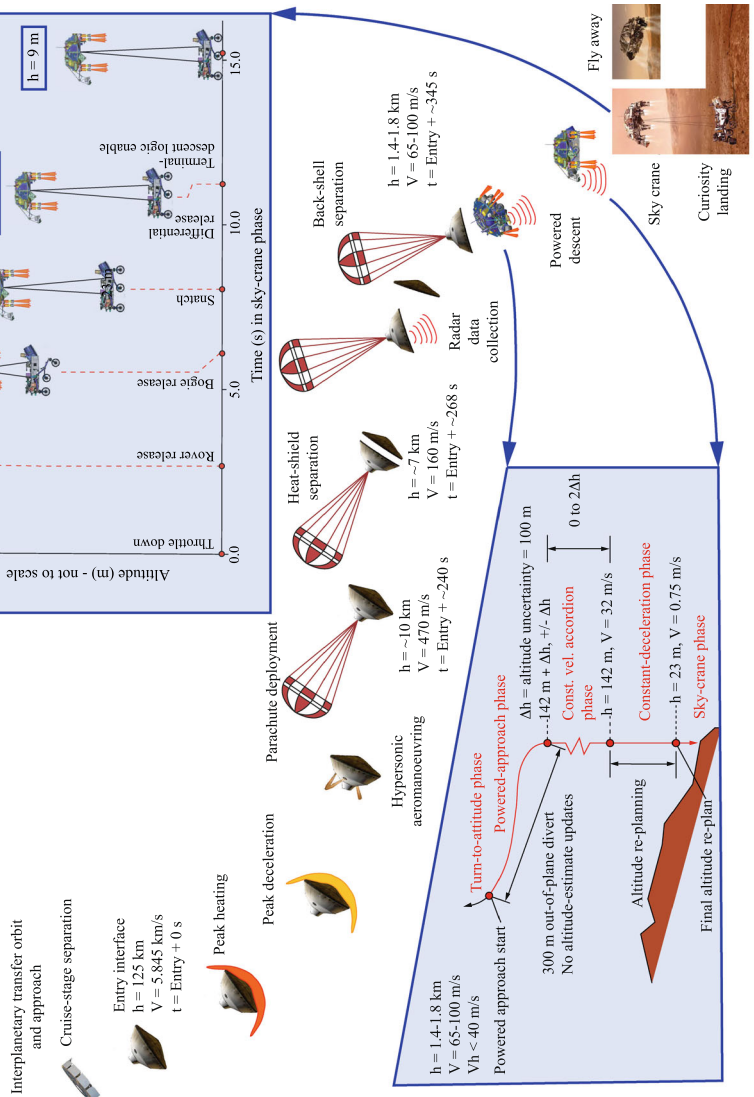


Fig. 9.127 Entry and descent mission profile of the Mars Science Laboratory. Source images of systems: NASA. Data taken from several references quoted in the text

Following the successful parachute phase that ended with back-shell (and thus parachute) separation as discussed above, the powered descent phase started, a phase that was basically very similar to previous soft-landers (e.g., Viking). The powered-descent guidance (PDG) logic distinguished four phases: (1) powered approach, (2) constant velocity profile, (3) constant deceleration profile, and (4) down throttling and transition to the sky-crane phase at 23 m altitude. In each phase, a polynomial function provided a reference trajectory (position, attitude, acceleration magnitude), which were to be followed closed loop, managed by the power-descent translational controllers (San Martin et al. 2015). During powered approach, the vertical velocity of the descent stage was smoothly reduced to 32 m/s, while the horizontal velocity was nulled simultaneously. Moreover, a diverting manoeuvre shifted the lateral position of the descent stage by 300 m to avoid the parachute/backshell from re-contacting.

Subsequently, to accommodate for any errors in altitude, the vehicle is driven into a constant velocity profile at an altitude of 142 m. An altitude-error margin due to radar error and terrain relief features of 100 m was considered. To accommodate the altitude error, a period of constant vertical velocity is used to "fly out" the altitude error. This is termed the *constant velocity accordion*. At the end of this phase, an altitude-velocity combination of 142 m and 32 m/s was achieved. During the constant deceleration phase, the descent stage's vertical velocity is further reduced to 0.75 m/s, while keeping the horizontal velocity at zero. This occurs at a throttling rate of 90% and eventually ends at an altitude of 23 m above ground. At this moment in time half of the total fuel has already been used.

At 23 m above the surface, the sky-crane phase begins. At that moment, when the vertical velocity is 0.75 m/s four of the the engines are throttled down, as more deceleration is not required (fuel usage during this complete final phase is 4 kg/s). To guarantee success of the sky crane, MSL was equipped with a so-called terminal descent sensor (TDS), a Ka-band (35.75 GHz) pulse-pair Doppler radar design (see Fig. 7.6 for a schematic overview of the TDS, as well as all other sensors and actuators). This sensor measures ground-relative velocity and slant-range along the boresight of each of its six narrow-beam antennas with a beam width of about 3.5°. One of the six antennas was nadir pointing to allow for maximum altitude measurement accuracy during vertical flight before to touchdown, three were oriented at 20°, and two were placed at 50° angles to guarantee that at least those two beams would not experience multi-path interference from the rover during the sky crane phase (San Martin et al. 2015). The measurements, i.e., a velocity/slant-range pair for a single antenna beam, were taken at 20 Hz.

Rover separation was initiated after 2.5 s, to provide sufficient time that the transient effects of the down throttling had died down. Wheels were released, and the rover was lowered until line snatch at $t = 9.5$ s. At that moment the descent stage had lowered down to 16 m. A buffer of 2 s was included, the so-called post-deploy settling phase. At the end of this phase the terminal-descent logic was enabled, and the touch-down phase was entered at an altitude of 14.5 m, with the rover about 5 m above the surface. Lowering the rover to the surface took another 4 s. Guidance in this last phase was linked to the throttle command. When the engine-throttle command was reduced to about 50%, a value that was consistent with the weight of the

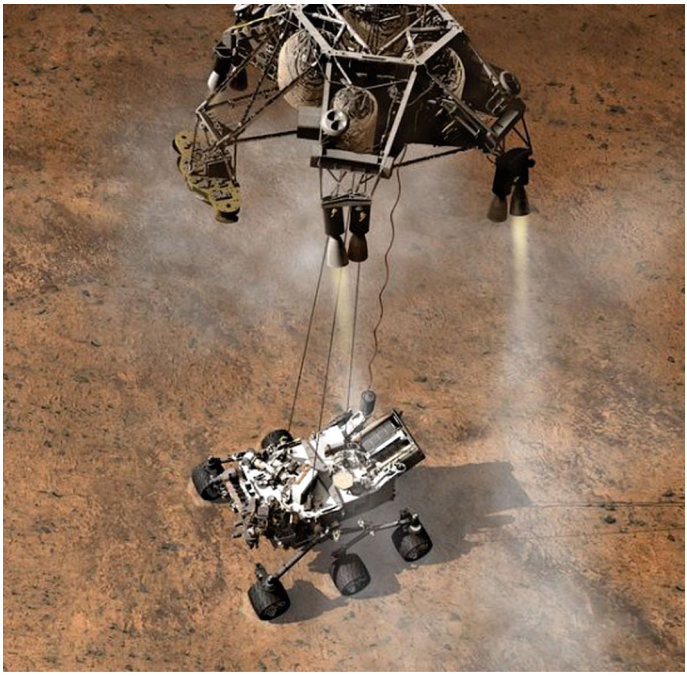


Fig. 9.128 Artist impression of the landing of *Curiosity* with the MSL sky crane. Image: NASA/JPL-Caltech

descent stage only (roughly equal to that of the rover), then touchdown could be declared. This occurred nominally about 15.5 s after the sky-crane phase had started. Figure 9.128 shows an artist impression of this moment.

9.8 Case Study: Huygens

In Chap. 5, Sect. 5.3.4, the first part of this case study of Huygens has been introduced, focusing on the ballistic entry phase. There, also some background information on the mission can be fond. Subsequently, in Chap. 6, Sect. 6.5.4, the characteristic motion and the influence of spin on the stability of the Huygens probe were analysed. In the current section, we continue the case study with the focus on the parachute-descent phase. A complete description of the mass, geometry and aerodynamic properties is provided in Appendix B, Sect. B.3. The discussion in this section is based on mission analysis done by Mooij (1992), and the reader is referred to that report for more details and background.

This case study will provide a brief introduction to the (descent) mission and the related hardware elements involved (Sect. 9.8.1). Section 9.8.2 summarises the

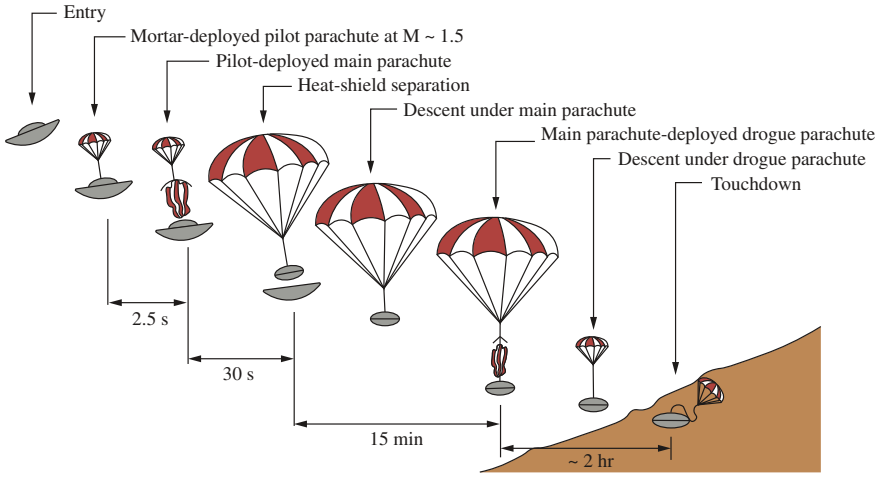


Fig. 9.129 Mission profile of the Huygens probe

simulation model that was used, and some results of the mission analysis are given in Sect. 9.8.3.

9.8.1 Mission and System

In Fig. 9.129 the mission profile of the Huygens probe is shown, covering the entry, descent, and landing. After separating from Cassini, Huygens entered the atmosphere at the reference altitude of 1250 km,[50] where the flight-path angle was within a permissible range of $-90°$ to $-60°$. This resulted in a steep entry that kept the probe within line of sight with the orbiter for the duration of the probe mission. The entry velocity relative to the surface had a value between 5.72 and 6.47 km/s. Because of low atmospheric density, drag is low at these high altitudes. Gravitational forces accelerated Huygens until the drag was high enough to initiate deceleration at an altitude of 600 km (note that the Titan atmosphere is much denser than the one of Earth, which explains this high altitude). During the entry, the descent module, which contained the experiments, was protected by a heat shield/aerodynamic decelerator and aft cover.

The descent control sub-system (DCSS) was initiated when an axial deceleration level of 8.25 m/s^2 was encountered, This event marked the end of the entry phase (for $\gamma_0 = -60°$ and $V_0 = 5.8$ km/s, the duration is about 4.5 min). By then the altitude was about 180 km. The first event since DCSS initiation was the deployment of the pilot chute in the supersonic regime (at a Mach number of $M \approx 1.5$) and

[50] The listed numerical values for the initial state are the design considerations; the actual initial state was (slightly) different, but is not relevant for the current discussion.

release of the aft cover as soon as the pilot chute was fully inflated. The aft cover pulled out the main parachute that decelerated the probe through Mach 1. When the probe was stabilised in the subsonic regime (30 s after pilot-chute deployment), the aerodynamic decelerator was jettisoned. After a further time, the main parachute was released, which in its turn deployed a stabilising drogue chute.

Originally, there were two design criteria for stabilising-drogue deployment. The first one assumed the release of the main parachute after 10 min at an altitude of about 142 km and subsequent drogue-chute deployment, and for the second one, the main chute was retained for 2,550 s until a lower altitude (about 90 km), where a minimum-sized stabilising drogue, required to maintain probe stability, would be deployed. For each of the scenarios, the descent phase could not last too long (orbiter data-relay link constraint).[51] The descent duration, starting from parachute deployment, was about 2 h and 15 min. Ground contact occurred with a velocity of about 7 m/s. Three minutes later, the mission of Huygens was officially ended, but, as mentioned earlier, the surface mission lasted about 1.5 h.

As is clear from the above mission description, the parachute system was designed with four main requirements in mind:

1. To remove the protective aeroshell,
2. To stabilise the probe in the transonic flight regime,
3. To control the descent time, which should not exceed 3 h due to line-of-sight constriction with the Cassini orbiter, and
4. To have no interference with the probe spin (before separation from Cassini, Huygens was spun up to 5 RPM. This spin rate was to be maintained in a certain range to allow for the horizon camera to have 360° vision).

The prime contractor for the DCSS was the United-Kingdom based company Martin-Baker Aircraft Company Ltd. The American company Airborne Systems Group, and in particular one of the companies Irvin Industries, provided design, analysis and manufacturing support, and manufactured all of the development, qualification and flight models for the program. The three parachutes, i.e., the pilot, main and drogue parachute had the following characteristics (Underwood 2005):

Pilot-chute design:

- Disk-Gap-Band, with a nominal diameter of $d_0 = 2.59$ m and a reference area of $S_0 = 5.29$ m^2, with Taschengurts (quick opening bands to improve the opening characteristics of the canopy) to stabilise the band
- Geometric porosity: 13.1%
- Number of suspension lines: 12
- Line length: $2d_0$
- Materials: Nylon broadloom fabric, and Kevlar lines and webbing
- Larger than required minimum size.

[51] The flight scenario was that 15 min after heat-shield jettisoning the main parachute was released, so the shorter of the two scenarios.

Main parachute design:

- Disk-Gap-Band, with a nominal diameter of $d_0 = 8.3$ m and a reference area of $S_0 = 54.1$ m^2
- Geometric porosity: 22.4%
- Number of suspension lines: 24
- Materials: same as for pilot chute.

Stabilising Drogue Design:

- Disk-Gap-Band, with a nominal diameter of $d_0 = 3.03$ m and a reference area of $S_0 = 7.23$ m^2
- Other properties similar to main parachute (see above).

To support the design team, of course many flight-simulation studies were done, both by the prime contractor as well as ESA in the form of a shadow-engineering activity. One of those shadow-engineering studies will be detailed below. It is noted that the results discussed below were obtained with the data provided in Appendix B.3, which represented the baseline at the time of the study. This means that detailed conclusions for the flight system cannot be drawn, but since the used data are not so much different, all general conclusions hold.

9.8.2 Simulation Model

From the above mission description, the simulation model for the parachute descent phase should be able to deploy a pilot chute, extract the main parachute and after a specified time release this chute and deploy a stabilising drogue chute. Each of these parachutes are single parachutes, so parachute clusters do not have to be considered. Since the descent module needs to have a certain spin rate—maintained by the spin vanes (see Fig. 7.41b)—this aspect will have to be simulated as well. And, since Titan has a severe wind environment the influence of wind on the stability of the parachute/payload system is one of the study goals.

For the simulations involved a dedicated simulation model was developed for the parachute phase. Initial conditions for this phase were obtained with a generic model that could simulate both three and six degrees-of-freedom, i.e., either only translational motion, or the combination of translational and rotational motion. For the entry phase and the stability analysis, both simulation models were used.

At the time of the study, one of the requirements was that the simulator should be simple and use as little CPU time as possible. The descent-simulation model includes seven degrees-of-freedom, and has been discussed in Sect. 9.2.6. Basic properties of the model are that both parachute and entry vehicle are modelled as rigid bodies, and connected to each other by a rigid bar. That means that the complete system is considered to be a single body that translates and rotates about the global centre of mass. Due to the rotation, the flow conditions, i.e., velocity magnitude and direction

(angle of attack and sideslip), are different for the individual bodies. Also, when the two bodies are not perfectly aligned, both gravitational and aerodynamic moments are introduced. The seventh degree of freedom is caused by a swivel in the rigid connection between parachute and payload, and allows for a free relative rotation along the spin axis, albeit that the swivel is not ideal, but has a friction moment proportional to the spin rate. Finally, added mass under the parachute is taken into account, which provides additional inertia of the system. Gravity, though, does not act on the added mass.

A limitation of this model is that only the so-called pendulum motion can be studied. The more complex scissors motion, i.e., the asymmetric rotation of parachute and payload, cannot be simulated (see also Fig. 9.35). Moreover, any elasticity in the suspension lines is also ignored, or, in fact, the complete geometry of the lines that may induce asymmetric loads on the parachute canopy. Even though parachute reefing can be simulated, a detailed calculation of the opening loads is not foreseen.

Specifically for Huygens, a spin-vane model was included. Spin vanes are small airfoils mounted on the vehicle wall, and induce a roll moment. This roll moment was essential for Huygens to guarantee a certain spin rate of the descent module, such that the horizon camera could keep scanning the horizon in a 360° motion. Details about the (simplified) spin-van model can also be found in Sect. 9.2.6, and numerical values for the nominal 24-vane model are given in Appendix B.3.

As mentioned earlier, the current case study is a summary of the mission analysis done by Mooij (1992). Not only does that report provide many more results, but also the used simulation model is described in a lot more detail.

9.8.3 Results

For the simulation study of the descent phase a number of study objectives was defined:

1. To determine the flight conditions at heat-shield separation to see whether interference with the descent module is possible,
2. To study the nominal descent for the two stabilising-drogue scenarios,
3. To establish the total descent time (which may not exceed 3 h),
4. To obtain the spin-rate profile for the two stabilising-drogue scenarios: at higher altitudes (i.e., above 10 km) it should be between 1 and 15 RPM, whereas below 10 km, it should be between 2 and 4 RPM, and
5. To get insight in the influence of a steady-state wind and horizontal wind gusts on the drift over the surface and the stability of the parachute-payload system.

Due to space limitations, the first objective will not be addressed here.

Because of the large uncertainty about the Titan atmosphere at the time of the design of Huygens, a range of atmospheric properties was assumed for the simulation study. Lellouch and Hunten (1987) had defined the so-called *nominal* atmosphere,

which would serve as the baseline for all engineering studies (see also Sect. 2.3.4). To allow for a sensitivity study on the nominal atmospheric properties, they also defined the *minimum* and *maximum* atmosphere. Minimum and maximum, in this case, mean a cold and hot atmosphere, respectively. The nominal atmosphere was believed to adequately represent Titan's atmosphere as it was at the time of the Huygens probe release, while the extreme profiles were conservative.[52]

Nominal Descent

The initial conditions at the beginning of parachute descent are, of course, the final conditions of the entry phase (and represent an axial (drag) deceleration of 8.25 m/s^2). For the so-called nominal dimensioning case that includes the definition of the nominal atmosphere, the altitude for parachute deployment is around $h = 176$ km, with a Mach number of $M = 1.5$. For the nominal descent two baseline scenarios have been simulated, i.e., deployment of a stabilising drogue parachute after 600 s from initiation of the DCSS, and deployment of the stabilising drogue after 2,550 s. For both scenarios, there are some differences with the actual mission scenario.

In the first place, initiation of the DCSS means the deployment of a pilot chute, which in its turn releases the after cover and extracts the main parachute after 2 s. This pilot chute is not included in our descent scenario, which will, of course, contribute to the differences compared with industrial results. In the second place, by misinterpretation of the event data, the release of the main parachute and the deployment of the stabilising drogue takes place 600 s (2,550 s) after heat-shield jettisoning, instead of after initiation of DCSS (a difference of 30 s). This results in another discrepancy compared with industry. Despite these errors, however, qualitatively speaking the results are in line with those presented by industry.

In Fig. 9.130 a set of images shows us what we would have seen if we had been present during the descent. As a function of altitude, the view of Huygens's descent imager/spectral radiometer is shown for the four main directions north, east, south and west.

In Fig. 9.131, left, the velocity-altitude profile is given for the 600-s scenario. After main chute deployment (which is simulated to be instantaneous), we see a strong deceleration from 400 m/s down to 60 m/s. Of course, this is due to the increased drag of the system. A velocity of 60 m/s is the equilibrium velocity at the corresponding altitude ($h \approx 165$ km). From that moment on, the velocity steadily decreases due to the increasing atmospheric density. After 600 s, corresponding with an altitude of $h \approx 142$ km, the drag area of the parachute is suddenly decreased due to the deployment of the much smaller stabilising drogue. The equilibrium velocity for this reduced drag area is of course higher, which can be seen in the graph. The velocity

[52] It is noted that currently more recent models are available and updated to the most current insight, most notably those by Yelle et al. (1997). However, at the time of the study these models did not exist.

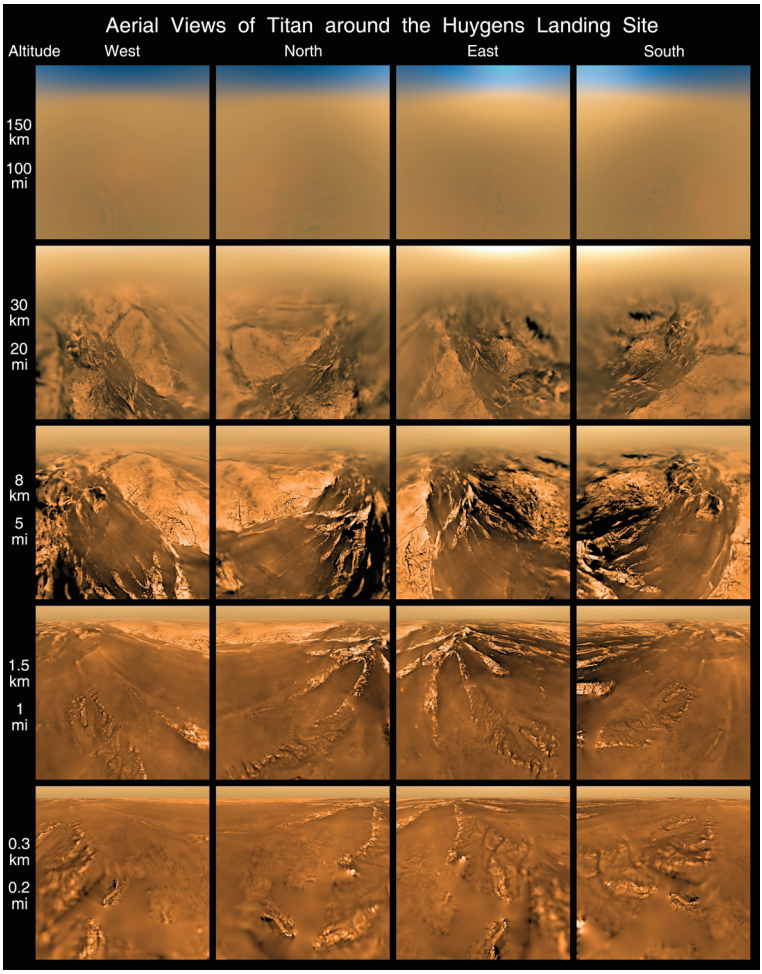


Fig. 9.130 Set of images acquired by the Huygens probe descent imager/spectral radiometer, in the four cardinal directions (north, east, south and west), at five different altitudes above Titan's surface. The images were taken on Jan. 14, 2005. Image credit: ESA/NASA/JPL/University of Arizona

increases to a value of about 120 m/s. From that moment on, the parachute system is steadily decelerated (increasing density) down to an impact velocity of $V \approx 6.5$ m/s.

Figure 9.132, left, shows the corresponding deceleration profile (height as a function of *g-load*; note that the *g-load* is referenced to the acceleration at Titan sea level[53]). The first peak—a positive *g-load* corresponds with a deceleration—is due to main-parachute deployment. The maximum amplitude of this peak is about 30 g.

[53] g_0, i.e., the acceleration at sea level, has a value of 1.354 m/s^2 for Titan, whereas the Earth value is 9.80665 m/s^2. This means, that to get the acceleration in m/s^2, the *g-load* should be multiplied

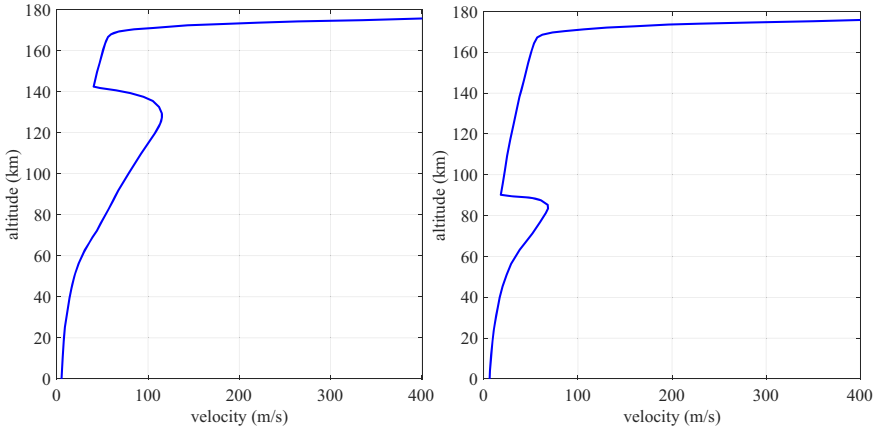


Fig. 9.131 Nominal descent for two stabilising-drogue scenarios (left $t = 600$ s and right $t = 2{,}550$ s): altitude versus velocity

At $h \approx 142$ km, we see a negative *g-load*, corresponding with the acceleration mentioned before (deployment stabilising drogue). The (absolute) peak value should indeed be lower than 1 g, since the acceleration can never be larger than the "free-fall" acceleration of 1 g. In this case, however, due to the drag the *g-load* is smaller than 1 g. Once the parachute system has reached equilibrium velocity again, the *g-load* is, although decreasing, always positive (continuous deceleration). While the velocity is getting smaller and smaller, hereby reducing the dynamic pressure (and the drag), the higher density is not sufficient to increase the drag of the parachute system.

In Figs. 9.131 and 9.132, right, the corresponding curves for the 2,550-s scenario are presented. The general shape of the curves is exactly the same as the ones for the 600-s scenario. Because the descent under the main parachute lasts longer, the velocity decreases to a lower value (just below 20 m/s for stabilising-drogue deployment at $h \approx 90$ km). Again, due to the smaller drag area, the velocity increases again. The new equilibrium value is lower, because of the lower altitude (higher density). Note that the smaller drag area of this particular drogue chute results in a higher equilibrium velocity as would have been the case for the drogue chute of the 600-s scenario. This can be verified when we look at the impact velocity which is higher for the 2,550-s scenario.

Descent-Time Analysis

In Table 9.11, the descent times for both drogue-deployment scenarios are shown. It is noted that some discrepancies with industrial values (shorter descent time) were

by 1.354. To express the *g-load* in Earth g, the Titan-based *g-load* should be multiplied by 0.138. For the project, it was customary to express the *g-load* in Titan g's.

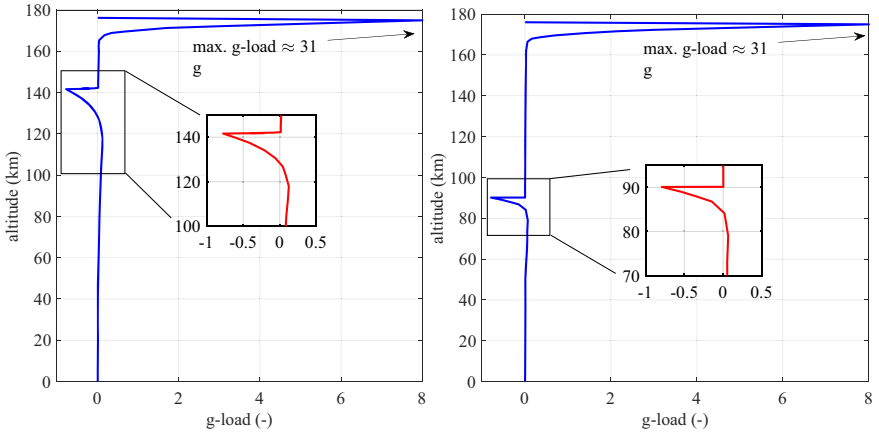


Fig. 9.132 Nominal descent for two stabilising-drogue scenarios (left $t = 600$ s and right $t = 2{,}550$ s): *g-load* versus altitude

Table 9.11 Descent times for two drogue-deployment scenarios

Drogue-deployment scenario (s)	Minimum atmosphere (s)	Nominal atmosphere (s)	Maximum atmosphere (s)
$t = 600$	7,342	8,144	8,997
$t = 2{,}550$	7,642	8,359	9,128

observed,[54] but at least three sources, which introduced the difference, can be identified. First, the deployment of the stabilising drogue took place 600 s (2,550 s) after heat-shield jettisoning, instead after initiation of the DCSS. As a result, the altitude at which the main chute was released was higher in our case, and the corresponding velocity lower. Both parameters increase the total descent time. Second, industry simulated the wake behind the descent module, which was neglected in this study. This means that the performance of its parachute was worse than the one used here, and this reduced drag area resulted in a higher velocity and thus shorter descent time. Third, alternate definitions of the minimum and maximum atmosphere also contribute to the differences. Unfortunately, it was not possible to check the influence of the mentioned aspects. Note that in all cases the descent time remains well below the 3-h maximum.

[54] The reader is referred to Mooij (1992) for a more detailed comparison (including references) with the industrial results of Martin-Baker.

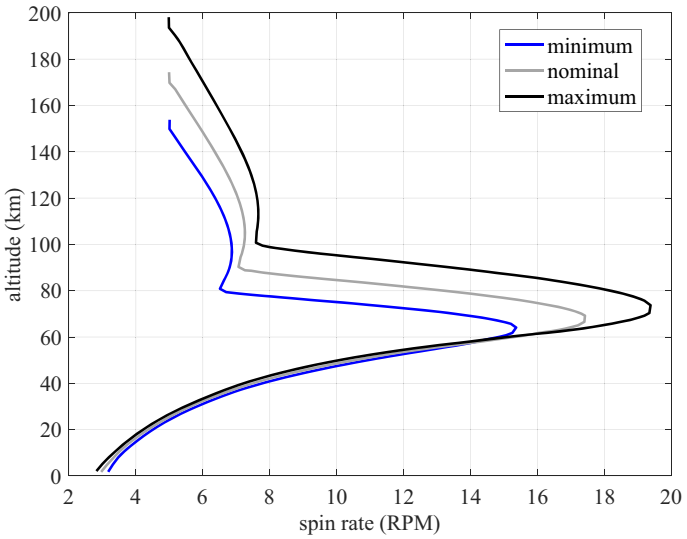


Fig. 9.133 The spin-rate profile for the minimum, nominal and maximum atmosphere. The stabilising drogue deploys after $t = 2{,}550$ s. Note that the initial conditions were set at the corresponding end conditions of the entry phase

Spin-Rate Analysis

Figure 9.133 shows the spin-rate profiles for the 2,550-s scenarios of stabilising drogue deployment and the three atmosphere models, with the nominal spin-vane configuration of 24 vanes. Starting at an initial value of 5 RPM, the spin rate p first decreases. This is due to the friction moment of the swivel. Because the heat shield has not yet been jettisoned, there is no spin-inducing moment due to the spin vanes. Once the heat shield has been separated, p increases rapidly. It should be noted, that for the descent module the aerodynamic roll damping is not zero, although the effect on the spin rate is small. It can be seen, that the peak values are too high for each of the curves. The conclusion is, that the nominal spin-vane configuration will not meet the mission requirements. The end values of p, on the other hand, do meet the specifications.

Compared with the 600-s scenario, the 2,550-s scenario results in a lower p: because of the longer descent under the main parachute, $\bar{q}$ will be lower, hereby decreasing the driving moment. During the main-chute descent the driving moment is not even sufficient to increase p. Once the stabilising drogue has been deployed, V increases and thus also $\bar{q}$, which results in a higher spin-vane-induced moment. However, compared with the 600-s scenario, $\bar{q}$ will not reach the same value, so p_{max} is smaller. Since $\bar{q}_{max}$ occurs at a lower altitude, also p_{max} is found at a lower altitude. The end values of p are higher for the 2,550-s scenario, because the descent time left from peak value down to impact is smaller, so the integrated effect of the damping is smaller.

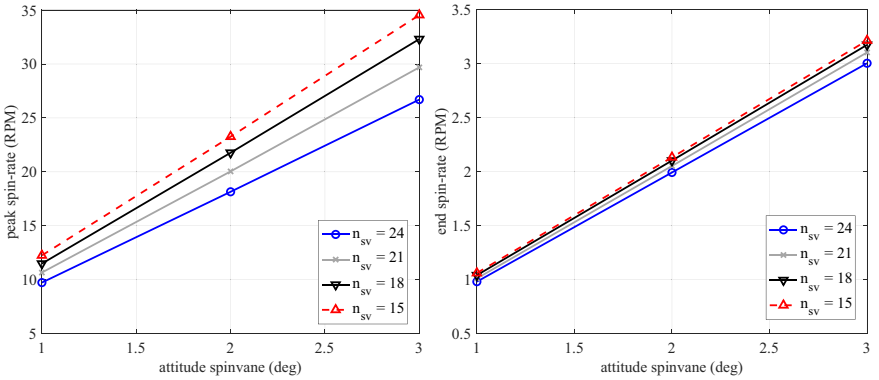


Fig. 9.134 Spin-vane design: the influence of the attitude of the spin vane on the spin rate, for a varying number of spin vanes

The spin-vane design analysis was performed for the 600-s scenario only, because it appeared to be the most demanding scenario. The earlier results show that for the 2,550-s scenario the maximum spin rate is smaller and the end spin rate is larger, which brings us closer to both requirements. So the choice of descent scenario has an important effect on the spin-rate profile.

Concluding, the spin rate did not meet the mission requirements, i.e., a value between 1 and 15 RPM above 10 km, and an end value between 2 and 4 RPM below. To see the influence of varying some of the design parameters, we did a sensitivity analysis with respect to the number of spin vanes, the attitude angle and the aspect ratio of the spin vanes. The obtained results serve a qualitative purpose only, because of the simplified model used in the simulations. Note that only the nominal dimensioning case with a nominal atmosphere was studied, and the 600-s scenario with respect to drogue deployment was selected.

Figure 9.134 shows the influence of the number of spin vanes and the attitude of the spin vane on the peak and end value of the spin rate. Comparing both figures shows us that we are dealing with two conflicting requirements. Concerning the peak spin rate, the attitude of the vanes should be between 1.2° and 1.6°, depending on the number of vanes. In principle, the number of vanes should be as low as possible to arrive at peak values below 15 RPM. These results are not surprising, since the nominal vane configuration ($\theta_{sv} = 2°$ and $n_{sv} = 24$) resulted in a large overshoot of the maximum allowable peak value (see Fig. 9.133 for the $t = 2{,}550$-s scenario; the comparing simulation for the $t = 600$-s scenario had a peak spin rate of 22 RPM). The smaller θ_{sv}, the smaller the aerodynamic forces acting on the vane, resulting in a smaller driving torque. Since the torque is an integrated effect of all the spin vanes, decreasing the number of vanes would lead to lower torques.

However, having a peak value which is within the allowable range, leads to a conflict with respect to the end spin rate. To end just above the lower boundary of 2 RPM, θ_{sv} should be between 1.9° and 2.1°, again depending on n_{sv} To end up in the middle of the required interval, i.e., 3 RPM, would demand an attitude of $\theta_{sv} = 3°$.

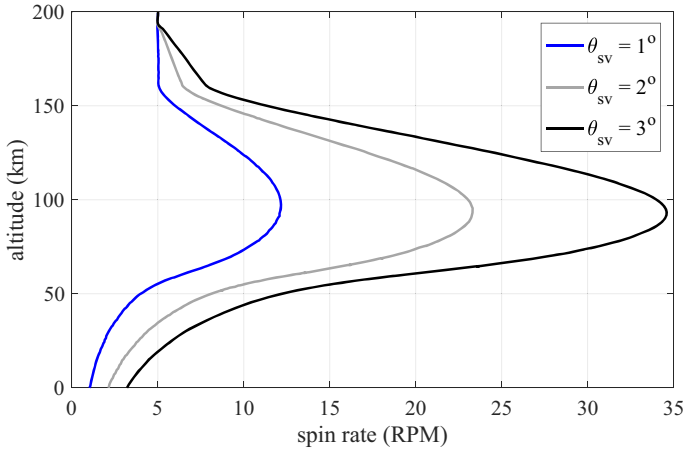


Fig. 9.135 Spin-vane design: the spin-rate profile, for a varying attitude of the spin vanes. The number of spin vanes is $n_{sv} = 24$

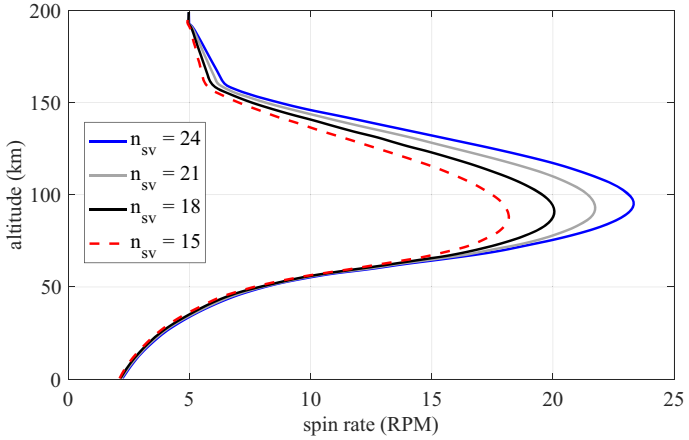


Fig. 9.136 Spin-vane design: the spin-rate profile for a varying number of spin vanes. The attitude of the spin vane is $\theta_{sv} = 2°$

To see the conflict, we have added Fig. 9.135, where the spin-rate profiles for three attitude angles is plotted, for 24 vanes. $\theta_{sv} = 3°$ gives an end spin rate of p 3 RPM, but the peak value is well above 34 RPM. And decreasing the number of vanes does not result in an acceptable solution, as can be derived from Fig. 9.136. For a fixed attitude of $\theta_{sv} = 2°$ (and an end spin rate at the lower part of the allowable range), it is not possible to get the peak below 15 RPM if we decrease n_{sv} from 24 to 15. So at least fewer vanes are required.

On the other hand, we can also influence the induced torque by varying the aspect ratio of the vane. As a (preliminary) final result, we looked at a more or less optimised

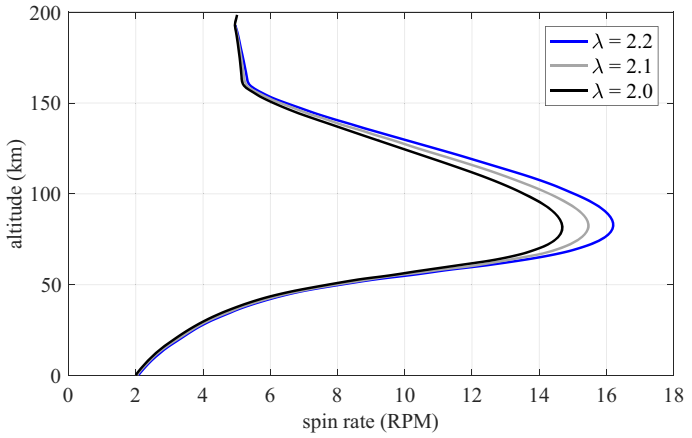


Fig. 9.137 Spin-vane design: the spin-rate profile, for a varying aspect ratio of the spin vanes. The number of vanes is $n_{sv} = 9$ and the attitude is $\theta_{sv} = 2°$

spin-vane configuration. For $\theta_{sv} = 2.2°$ and $n_{sv} = 9$, we studied three aspect ratios, i.e., $\lambda = 2.0$, $\lambda = 2.1$ and $\lambda = 2.2$. The resulting spin-vane profiles are depicted in Fig. 9.137. As we can see, the curves can be "fine-tuned" using the aspect ratio. The peak value can come below the 15-RPM level, but the end spin rate is still a point of concern.

The general conclusion drawn from the (short) sensitivity analysis is, that with the current modelling it is very difficult, let us not say impossible, to meet both mission requirements in an acceptable way. However, it must be stressed that the results are obtained with a simple model, but at least it should be clear that there is definitely a conflict. A variable number of vanes, or variable-pitch vanes, should at least be considered as a possible solution to this problem.[55]

[55] From post-flight analysis it was derived that Huygens, in fact, spun in the wrong direction. A recent study in the subsonic wind tunnel at the PRISME Laboratory at the University of Oréans, France, has confirmed the main cause. The flight spin-vane configuration consisted of 36 angled vanes. However, two of the descent module's main appendages, the Separation Subsystem (SEPS) and the Radar Altimeter (RA) antennae, produced an unexpected torque opposite to that generated by the vanes. This negative torque became even larger as the vanes altered the gas flow around the descent module, until it exceeded the influence of the vanes. Fortunately, the adverse spin rate was of the same order of magnitude as planned pre-flight, such that the measurements were not affected. However, this shows that a very detailed analysis, both with CFD and wind tunnels, is required to design an efficient spin-vane configuration. See also https://www.esa.int/Science_Exploration/Space_Science/Cassini-Huygens/Huygens_landing_spin_mystery_solved. Date accessed: August 1, 2021.

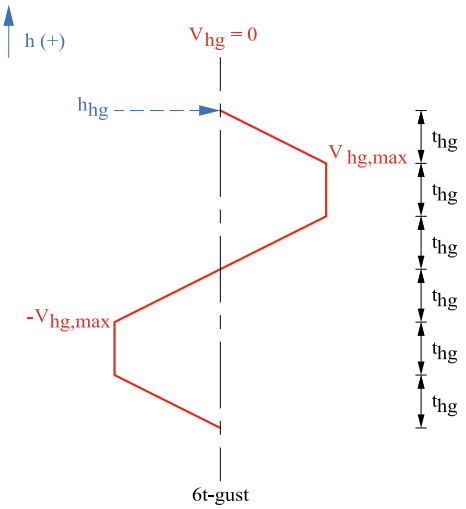


Fig. 9.138 The 6*t* horizontal wind-gust model

Wind Analysis

The wind climate on Titan is severe, as became obvious from, amongst others, Voyager data. Lunine et al. (1991) developed an engineering model for a steady-state zonal wind that became known as the Flasar model. This model results in $V_{wind} \approx$ 180 m/s for $h \approx 180$ km, and a wind velocity at the surface of 2 m/s. It has been found that this steady state wind does not have a noticeable effect on the descent time. However, for an equatorial entry large drift values as compared to the no-wind case were found:

(1) $\Delta\tau = 9.97°$ (or $\Delta R = 448.1$ km) for the 600-s scenario
(2) $\Delta\tau = 12.92°$ (or $\Delta R = 580.7$ km) for the 2,550-s scenario

Note that in case an orbiter would be moving in the direction of the drift, the duration of data contact will increase.

According to the system requirements, the parachute system should be stable enough to withstand any possible (horizontal) wind gusts in the atmosphere of Titan. The maximum pendulum oscillation must be less than $\pm 10°$. This criterion may be exceeded for periods up to 10 s (no data could be found on the maximum allowable peak values). The maximum pitch rate must be below 6°/s (no data were available for the yaw rate). To study the stability of the parachute system, the influence of three high-level gusts at $h_{hg} = 160$, 140 and 120 km, and four low-level gusts at $h_{hg} =$ 90, 60, 30 and 2 km for each of the two descent scenarios was simulated. The gusts have the so-called 6*t*-shape (Fig. 9.138), with a maximum velocity of 25 m/s (6 m/s) for the high (low)-level gust and a thickness t_{hg} of 250 m (60 m).

While studying the pitch-rate profile it became obvious that there was an overshoot of the maximum value of 6°, but for each of the occurrences the duration of the overshoot was less than 5 s. The conclusion drawn from the results was that the

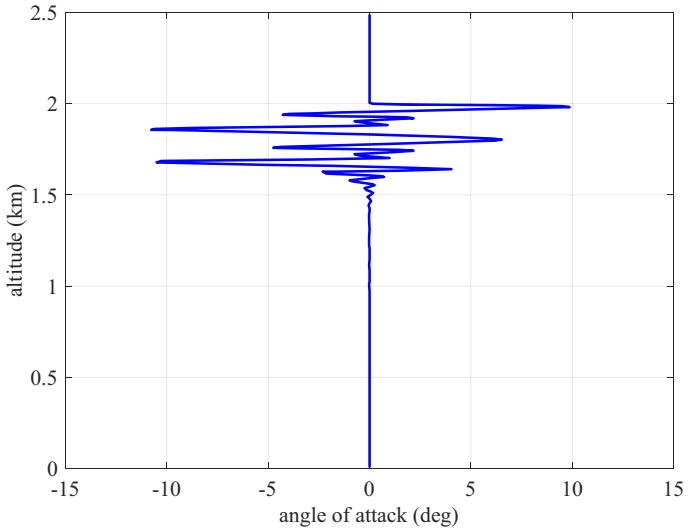


Fig. 9.139 The angle-of-attack profile for a low-level gust at $h = 2$ km. The 600-s deployment scenario for the stabilising drogue was used

parachute system is sufficiently stable. The major oscillations occur when the gust is active; once the wind velocity is back to zero again, the oscillations are damped in a very short time (as an example, in Fig. 9.139 the angle-of-attack profile for a low-level gust at $h = 2$ km has been depicted). The peak values of the total angle of attack have been compared with similar results obtained by Martin-Baker. However, only results due to a gust of 6 m/s were available, so the comparison for the high-level gust could only be qualitative. For $h = 160$ km, Martin-Baker found a peak value of 2° ($V_w =$ 6 m/s). The simulated value of just over 8° ($V_w = 25$ m/s) seems to be consistent with this. Martin-Baker found $\Delta\alpha_{tot} = 2.5°$ for $h = 100$ km ($V_w = 6$ m/s), whereas we obtained $\Delta\alpha_{tot} = 3°$ for $h = 90$ km ($V_w = 6$ m/s), again a good similarity. A difference was found for $h = 2$ km: 22° (Martin-Baker) versus 16°. However, the six-degrees-of-freedom parachute model showed some limitations, which should first be removed before this difference can be explained.

Problems

Judge whether the following statements are TRUE or FALSE (Problems 9.1 to 9.15).

9.1. During parachute reefing, the highest load usually occurs during the last reefing step.

9.2. Parachute systems for planetary landing are often supplemented or complemented by additional deceleration technologies, such as retrorockets or airbags, to ensure a safe and successful landing.
9.3. Parachute clustering is never used for small payload masses.
9.4. Parachute reefing is only used for systems when there is no pilot chute.
9.5. A descent guidance system can decrease the landing ellipse for a planetary landing.
9.6. Parachute systems for planetary landing are susceptible to environmental factors such as wind shear and atmospheric turbulence, which can affect descent stability and accuracy.
9.7. E-guidance is a two-point boundary-value problem.
9.8. For a Mars landing, a high (planetary) surface pressure is beneficial.
9.9. Precision landing requires the use of an HDA system.
9.10. An HDA capable system also requires precision landing capabilities.
9.11. The gravity turn is a two-point boundary-value problem.
9.12. Today, a landing ellipse of 10×5 km on Mars would be considered small.
9.13. Parachute systems for planetary landing are only suitable for relatively small and lightweight spacecraft due to the limited aerodynamic forces available for deceleration.
9.14. Parachute deployment altitude and timing are crucial factors in ensuring a successful landing, as they must be precisely coordinated to achieve the desired descent profile.
9.15. TAN can only be performed at high altitudes, and is therefore mainly used in the early phase of descent.

Parachute-system design (Problems 9.16 to 9.20).

As an aerospace engineer you are tasked to design a parachute system for a planetary lander (mass = 400 kg) on Titan. Given data: Titan: $\rho_0 = 5.44$ kg/m^3, $g_0 = 1.354$ m/s^2. Mars: $\rho_0 = 0.016$ kg/m^3, $g_0 = 3.727$ m/s^2. $C_{D_{para}} = 0.512$.

9.16. Assuming vertical equilibrium and a maximum impact velocity of 5 m/s, what will be the parachute diameter for landing on Titan? A circular parachute planform may be assumed, and the drag of the lander compared to the parachute is negligible. Derive any equation that you use.
9.17. For redundancy reasons, it is decided to equip the lander with a cluster of two parachutes. Given the same impact velocity, what will be the parachute diameter in that case?
9.18. Suppose we use the same design as under (b) for a Mars landing. For the same impact velocity, what would be the required parachute diameter?
9.19. Do you think that is feasible? If not, what would be the alternative?
9.20. For a parachute diameter of 6 m, the impact velocity on Titan would be 3.7 m/s. What would be the impact velocity on Mars using this parachute? Will parachute reefing help to reduce the velocity? Motivate your answer.

Powered descent (Problems 9.21 to 9.26).

The Mars Science Laboratory uses a combination of parachutes and landing engines to decrease its velocity. Close to the surface, it operates as a sky crane to lower a rover while it hovers above the surface. In this question we consider a similar vehicle and mission configuration. Data that can be used for this series of questions: the gravitational acceleration can be assumed to be constant and is $g_{Mars} = 3.69$ m/s^2 and $g_{Earth} = 9.81$ m/s^2. The initial vehicle mass is 2,196 kg, whereas the 8 engines provide a thrust of 3,000 N each and have a specific impulse of 200 s (Earth referenced).

9.21. After hypersonic descent and sequential parachute phase the velocity of the vehicle has decreased to 100 m/s at 1.8 km altitude. At that moment the parachute is released. For a vertical flight, set up the equations of motion for the successive powered flight.
9.22. In case the landing engines fail to ignite for 5 s, what will be the velocity at the beginning of the powered phase? One may assume an average (=constant) vehicle drag of 2,500 N.
9.23. What is the corresponding altitude (at the beginning of the powered phase)?
9.24. To lower the rover to the surface of Mars, the descent vehicle is hovering for 13 s. What would be the throttle setting if all 8 engines are used simultaneously? How much fuel is used during that timespan? Assume a (constant) vehicle mass at the beginning of the hovering phase of $m = 1{,}200$ kg.
9.25. What can you say about the maximum acceleration error that you make by having assumed a constant mass? Quantify the error.
9.26. What happens with the vehicle if the engines are not throttled to compensate for this error?

References

Açikmeşe B, Ploen SR (2007) Convex programming approach to powered descent guidance for Mars landing. J Guida Control Dyn 30(5):1353–1366
Ball AJ, Garry JRC, Lorenz RD, Kerzhanovich VV (2007) Planetary landers and entry probes. Cambridge University Press
Bennett FV (1970) Apollo Lunar descent and ascent trajectories, AIAA-1970-0025. In: AIAA/AAS astrodynamics specialist conference and exhibit
Bennett FV (1972) Apollo experience report - mission planning for lunar module descent and ascent. NASA-TN-D-6846, NASA
Benson HE (1966) Water impact of the Apollo spacecraft. J Spacecr Rockets 3(8):1282–1284
Blanchard RC, Desai PN (2011) Mars phoenix entry, descent, and landing trajectory and atmosphere reconstruction. J Spacecr Rockets 48(5):809–821
Boyd ID, Trumble KA, Wright MJ (2010) Modeling of Stardust entry at high altitude, part 1: flowfield analysis. J Spacecr Rockets 47(5):708–717
Boyd ID, Jenniskens P (2010) Modeling of Stardust entry at high altitude, part 2: radiation analysis. J Spacecr Rockets 47(6):901–909
Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press
Brandeau E, Clark IG, Ginn JM, Tanner C (2016) Ballutes for supersonic deceleration at Mars. In: IEEE Aerospace conference

Braun RD, Spencer DA, Kallemeyn PH, Vaughan RM (1999) Mars Pathfinder atmospheric entry navigation operations. J Spacecr Rockets 36(3):348–356
Bryan C, Strasburger W (1965) Lunar module structures handout LM-5 (For training purposes only; in support of LM-5 structures course), LSG 770-154-10 - LM-5, NASA
Câmara F, Rogata P, Di Sotto E, Caramagno A, Manuel J, Rebordão BC, Duarte P, Mancuso S (2005) Hazard avoidance techniques for vision based landing. In: 6th international ESA conference on guidance, navigation and control systems
Carpenter JD, Fisackerly R, De Rosa D, Houdou B (2012) Scientific preparations for lunar exploration with the European lunar lander. Planet Space Sci 74(1):208–223
Carson III JM, Hirsh RL, Roback VE, Villalpando CY, Busa JL, Pierrottet DF, Trawny N, Martin KE, Hines GD (2015) Interfacing and verifying ALHAT safe precision landing systems with the Morpheus vehicle, AIAA-2015-0325. In: AIAA SciTech forum, guidance, navigation, and control conference, Kissimmee, FL, 5–9 Jan
Cassell AM, Winter MW, Allen GA, Grinstead JH, Antimisiaris ME, Albers J, Jenniskens P (2011) Hayabusa re-entry: trajectory analysis and observation mission design, AIAA-2011-3330. In: 42nd AIAA thermophysics conference, Honolulu, Hawaii, 27–30 June
Çengal YA, Cimbala JM, Turner RH (2016) Fundamentals of thermal-fluid sciences, 5th edn. McGraw-Hill Companies, Boston
Cheng RK (1964) Surveyor terminal guidance. NASA-CR-57550, Jet Propulsion Laboratory
Cheng Y, Johnson A, Matthies L (2005) MER-DIMES: a planetary landing application of computer vision. In: IEEE computer society conference on computer vision and pattern recognition, pp 806–813
Cherkasov II, Mikheev VV, Smorodinov MI (1986) 20 years of Soviet investigations of lunar soils. Soil Mech Found Eng 23(6):241–244
Cherry GW (1964a) A general, explicit, optimizing guidance law for rocket-propelled spaceflight, AIAA-1964-0638. In: AIAA astrodynamics guidance and control conference, Los Angeles, CA, 24–26 Aug
Cherry GW (1964b) A class of unified explicit methods for steering throttleable and fixed-thrust rockets. Technical report, Massachusetts Institute of Technology, Boston
Citron SJ, Dunin SE, Meissinger HF (1964) A terminal guidance technique for lunar landing. AIAA J 2(3)
Clark IG, Adler M, Manning R (2015) Summary of the first high-altitude, supersonic flight dynamics test for the low-density supersonic decelerator project, AIAA 2015-2100. In: 23rd AIAA aerodynamic decelerator systems technology conference and seminar, Daytona Beach, FL, 30 March–2 April
Cohanim BE, Dhillon S, Ho JA (2012) Statistical hazard detection using shadows for small robotic landers/hoppers. In: IEEE aerospace conference, Big Sky, MT, 3–10 March
Cooley CG, Lewis GJ (1977) Viking 75 project: Viking lander system primary mission performance report, NASA-CR-145148, NASA
Cooper M, Sinclair R, Sanders J (2005) Design and testing of an airbag landing attenuator system for a generic crew return vehicle, AIAA-2005-1616. In: 18th AIAA aerodynamic decelerator systems technology conference and seminar, Munich, Germany, 23–26 May
Cruz JR, Kandis M, Witkowski A (2003) Opening loads analyses for various disk-gap-band parachutes, AIAA-2003-2131. In: 17th AIAA aerodynamic decelerator systems technology conference and seminar, Monterey, CA, 19–22 May
Cruz JR, Lingard JS (2006) Aerodynamic decelerators for planetary exploration: past, present, and future, AIAA-2006-6792. In: AIAA guidance, navigation, and control conference and exhibit, Keystone, CO, 21–24 Aug
Cruz JR, Way DW, Shidner JD, Davis JL, Adams DS, Kipp DM (2014) Reconstruction of the Mars Science Laboratory parachute performance. J Spacecr Rockets 51(4):1185–1196
Cuseo J, Dallas C (1989) Planetary terminal descent simulation with autonomous hazard avoidance, AIAA-1989-0515. In: 27th Aerospace science meeting, Reno, NV, 9–12 Jan

De Ridder K, Mooij E (2017) Convex guidance, navigation, and control for pin-point lunar landing. In: 10th international ESA conference on guidance, navigation & control systems, Salzburg, Austria, 29 May–2 June

De Rosa D, Fisackerly R, Pradier A, Philippe C, Houdou B, Carpenter J, Gardini B (2011) ESA lunar lander mission, 8th international ESA conference on guidance, navigation & control systems, Carlsbad, Czech Republic, 5–10 June

Desai PN, Knocke PC (2004) Mars Exploration Rovers entry, descent, and landing trajectory analysis, AIAA-2004-5092. In: AIAA/AAS astrodynamics specialist conference and exhibit, Providence, Rhode Island, 16–19 Aug

Desai PN, Lyons DT, Tooley J, Kangas J (2008) Entry, descent, and landing operations analysis for the Stardust entry capsule. J Spacec Rockets 45(6):1262–1268

Desai PN, Qualls GD (2010) Stardust entry reconstruction. J Spacecr Rockets 47(5):736–740

Dillman RA, DiNonno JM, Bodkin RJ, Gsell V, Miller NA, Olds AD, Bruce WE (2013) Flight performance of the inflatable reentry vehicle experiment 3. In: 10th international planetary probe workshop (IPPW-10), San Jose

Do S, De Weck O (2012) A personal airbag system for the Orion crew exploration vehicle. Acta Astronaut 81:239–255

Doengi F, Burnage ST, Cottard H, Roumeas R (1998) Lander shock-alleviation techniques. In: ESA bulletin, Nr. 93, European Space Agency

D'Souza CN (1997) An optimal guidance law for planetary landing, AIAA-1997-3709. In: AIAA guidance, navigation, and control conference, New Orleans, LA, 11–13 Aug

Dutta S, Braun RD (2014) Statistical entry, descent, and landing performance reconstruction of the Mars Science Laboratory. J Spacecr Rockets 51(4):1048–1061

Edquist KT, Hollis BR, Johnston CO, White TR, Mahzari M (2014) Mars Science Laboratory heat shield aerothermodynamics: design and reconstruction. J Spacecr Rockets 51(4):1106–1124

Epp CD, Robertson EA, Carson III JM (2014) Real-time hazard detection and avoidance demonstration for a planetary lander, AIAA-2014-4312. In: AIAA SPACE conference, San Diego, CA, 4–7 Aug

Epp CD, Smith TB (2007b) Autonomous precision landing and hazard detection and avoidance technology (ALHAT). In: IEEEAC paper 1379, IEEE aerospace conference, Big Sky, MT, 3–10 March

Euler EA, Adams GL, Hopper FW (1978) Design and reconstruction of the Viking lander descent trajectories. J Guid Control 1(5):372–378

Ewing EG, Bixby HW, Knacke TW (1978) Recovery system design guide. Irving Industries, California

Fallon II EJ (1995) Supersonic stabilization and deceleration - ballutes revisited, AIAA-1995-1584. In: 13th aerodynamic decelerator systems technology conference, 15–18 May

French KE (1968) The initial phase of parachute inflation. In: Proceedings of the 1969 AIAA aerodynamic deceleration systems conference, Air Force Report FTC-69-11

Gelly G, Vernis P (2009) Neural networks as a guidance solution for soft-landing and aerocapture, AIAA-2009-5664. In: AIAA guidance, navigation, and control conference, Chicago, IL, 10–13 Aug

Gerth I, Mooij E (2014) Guidance for autonomous precision landing on atmosphereless bodies, AIAA-2014-0088. In: AIAA guidance, navigation, and control conference, National Harbor, MD, 13–17 Jan

Golombek MP (1997) The Mars Pathfinder mission. J Geophys Res 102(E2):3953–3965

Goodrick TF (1975) Theoretical study of the longitudinal stability of high-performance gliding airdrop systems, AIAA-1975-1394. In: AIAA 5th aerodynamic decelerator systems conference, Albuquerque, NM, 17–19 Nov

Grimwood JM (1963) Project Mercury, NASA-SP-4001. In: National Aeronautics and Space Administration, Washington, D.C

Grimwood JM, Hacker BC, Vorzimmer PJ (1969) Project Gemini: technology and operations, a chronology, NASA-SP-4002. In: National Aeronautics and Space Administration, Manned Spaceflight Center, Houston, TX

Grumman Aerospace Corporation (1971) Apollo operations handbook. Lunar module LM 10 and subsequent. Volume I - subsystems data. Technical report NAS-9-1100, Exhibit E - Paragraph 10.4

Hall JL (2000) A review of ballute technology for planetary aerocapture, IAA-L-1112. In: 4th IAA conference on low cost planetary missions, Laurel, MD, 2–5 May

Harding J (2006) Flying's strangest moments: extraordinary but true stories from over 1100 years of aviation history. Strangest series. Anova Books

Harvey B (2007a) Russian planetary exploration. History, development, legacy, prospects. Springer-Praxis Books in Space Exploration

Harvey B (2007b) Soviet and Russian lunar exploration. Springer-Praxis Books in Space Exploration

Helmut GH, Lawrence WR (1965) Dynamic stability of a parachute point-mass load system. Technical report, FDL-TDR-64-126

Hirano Y, Miura K (1970) Water impact accelerations of axially symmetric bodies. J Spacecr Rockets 7(6):762–764

Holmberg NA, Faust RP, Holt HM (1980) Viking '75 spacecraft design and test summary. Volume I - lander design. NASA Reference Publication 1027

Howard AM, Jones BM, Serrano N (2011) Integrated sensing for entry, descent, and landing of a robotic spacecraft. IEEE Trans Aerosp Electr Syst 47(1):295–304

Hughes SJ, Dillman RA, Starr BR, Stephan RA, Lindell MC, Player CJ, Cheatwood FM (2005) Inflatable re-entry vehicle experiment (IRVE) design overview, AIAA-2005-1636. In: 18th AIAA aerodynamic decelerator systems technology conference and seminar, Munich, Germany, 23–26 May

Hume RG (1973) A two-dimensional mathematical model of a parachute in steady descent, C.P. No. 1260. Aeronautical research council (Ministry of Defence), London

Ibrahim SK, Engdahl AA (1974) Parachute dynamics and stability analysis. Prepared under NASA contract no. NAS 8-28607, Honeywell Inc., Minneapolis

Ingoldby RN (1978) Guidance and control system design of the viking planetary lander. J Guid Control 1(3):189–196

Ingoldby RN, Michel FC, Flaherty TM, Doty MG, Preston B, Villyard KW, Steele RD (1976) Entry data analysis for Viking landers 1 and 2. Final Report, NASA-CR-159388, NASA

Johnson CJ, Hixson RA (2008a) Orion vehicle descent, landing, and recovery system level trades, AIAA-2008-7745. In: AIAA Space 2008 conference & exposition, 9–11 Sept

Johnson AE, Montgomery JF (2008b) Overview of terrain relative navigation approaches for precise lunar landing. In: IEEEAC paper 1657, IEEE aerospace conference, Big Sky, MT, 1–8 March

JPL (1966) Surveyor I mission report. Part I: mission description and performance, TR-32-1023. In: Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

JPL Special Review Board (2000) Report on the loss of the Mars polar lander and deep space 2 missions. JPL D-18709, 22 March

Ke P, Yang CX, Sun X, Yang X (2009) A novel algorithm for simulating the general parachute-payload system: theory and validation. J Aircr 46(1):189–197

Klumpp AR (1971) Apollo guidance navigation and control. TR-R-695. Massachusetts Institute of Technology, Boston

Knacke TW (1987) Parachute system technology lectures: design of a parachute recovery system, Munich

Knacke TW (1991) Parachute recovery systems design manual, NWC-TP-6575. In: Naval Weapons Center, China Lake, CA

Knocke PC, Wawrzyniak GG, Kennedy BM, Desai PN, Parker TJ, Golombek MP, Duxbur TC, Kass DM (2004) Mars Exploration Rovers landing dispersion analysis, AIAA-2004-5093. In: Astrodynamics specialist conference and exhibit, Providence, Rhode Island, 16–19 Aug

Lafontaine Jde, Neveu D, Lebel K (2004) Autonomous planetary landing with obstacle avoidance: the quartic guidance revisited, AAS 04-274. In: 14th AAS/AIAA space flight mechanics conference, Maui, Hawaii, 8–12 Feb

Lellouch E, Hunten DM (1987) Titan atmosphere engineering model. Report ESLAB 87/199, ESLAB

Li S, Jiang X, Tao T (2016) Guidance summary and assessment of the Chang'e-3 powered descent and landing. J Spacecr Rockets 53(2):258–277

Litton DK, Bose DM, Cheatwood FM, Hughes S, Wright HS, Lindell MC, Derry SD, Olds A (2011) Inflatable re-entry vehicle experiment (IRVE) - 4 overview, AIAA-2011-2580. In: 21st AIAA aerodynamic decelerator systems technology conference and seminar, Dublin, Ireland, 23–26 May

Lobo MS, Vandenberghe L, Boyd S, Lebret H (1998) Applications of second-order cone programming. Linear Algebra Appl 284(1–3):193–228

Lorenz RD (2006) Spin of planetary probes in atmospheric flight. J Br Interplanet Soc 59(8):273–282 August

Lorenz RD (2011) Planetary penetrators: their origins, history and future. Adv Space Res 48:403–431

Lundstrom RR, Raper JL, Bendura RJ, Shields EW (1974) Flight tests of Viking parachute system in three Mach number regimes, NASA-TN-D-7692, NASA

Lunine Jl, Flasar FM, Allison M (1991) Huygens probe wind drift: science issues and recommendations- a report to the Huygens project. Final draft 3 Oct

Maass B, Woicke S, Oliveira WM, Razgus B, Krüger H (2020) Crater navigation system for autonomous precision landing on the Moon. J Guid Control Dyn 43(8):1414–1431

Mammarella M, Rodrigálvarez MA, Pizzichini A, Sánchez Montero AM (2011) Advanced optical terrain absolute navigation for pinpoint lunar landing. In: Advances in aerospace guidance, navigation and control, Springer, pp 419–430

Marraffa L, Kassing D, Baglioni P, Wilde D, Walther S, Pitchkhadze K, Finchenko V (2000) Inflatable re-entry technologies: flight demonstration and future prospects. ESA Bulletin (103):78–85. European Space Agency

Martin LJ, Zurek RW (1993) An analysis of the history of dust activity on Mars. J Geophys Res 98(E2):3221–3246

Mazaheri A (2013) High-energy atmospheric reentry test aerothermodynamic analysis. J Spacecr Rockets 50(2):270–281

McCullough JE, Stafford FA, Benson HE (1964) Attenuation of landing impact for manned spacecraft, NASA-CR-53291. In: Presented to the parachute symposium on technology and evaluation, El Centro, California, 7–9 April

McEwan AJ (1970) An investigation of parachute opening loads, and a new engineering method for their determination, AIAA-1970-1168. In: AlAA aerodynamic decelerator systems conference, Dayton, OH, 14–16 Sept

McFarland RK (1964) The development of metal honeycomb energy-absorbing elements. Technical Report No. 32-639, Jet Propulsion Laboratory

McGehee JR, Hathaway ME, Vaughan VL (1959) Water-landing characteristics of a reentry capsule, NASA-MEMO-5-23-59L

McShera JT (1963) Aerodynamic drag and stability characteristics of towed inflatable decelerators at supersonic speeds, NASA-TN-D-1601. In: National Aeronautics and Space Administration

Mendeck GF, Craig L (2011) Entry guidance for the 2011 Mars Science Laboratory mission, AIAA-2011-6639. In: AIAA atmospheric flight mechanics conference, Portland, Oregon, 8–11 Aug

Milo T (1991) On the initial-stage slamming of a rigid sphere in a vertical water entry. Appl Ocean Res 13(1):43–48

Milwitzky B (1952) Generalized theory for seaplane impact. NACA-Report 1103

Moog, R.D. and Michel, F.C., "Balloon Launched Viking Decelerator Test Program Summary Report, NASA-CR-112288 (Martin-Marietta Technical Report TR-3720359), March 1973

Moog RD, Sheppard JD, Kross DA (1979) Space Shuttle Solid Rocket Booster decelerator subsystem drop test results, AIAA-1979-0463. In: 6th Aerodynamic decelerator and balloon technology conference, Houston TX, 5–7 March

Mooij E (1992) ESA Huygens probe entry and descent analysis. ESTEC Working Paper EWP 1679, ESTEC, Noordwijk

Mooij E, Wijnands QGJ, Schat B (2003) 9 DOF parafoil/payload simulator development and validation, AIAA-2003-5459. In: AIAA modeling and simulation technologies conference, Austin, TX, 11–14 Aug

Mooij E (2016) Stability analysis of multibody systems for Mars descent and landing, AIAA-2016-0022. In: AIAA atmospheric flight mechanics conference, San Diego, CA, 4–8 Jan

Murrow HN, McFall JC (1969) Some test results from the NASA planetary entry parachute program. J Spacecr Rockets 6(5):621–623

NASA (1970) Space vehicle design criteria (structures). Landing impact attenuation for non-surface-planing landers, NASA-SP-8046. National Aeronautics and Space Administration, Washington, D.C

National Space Science Data Center (2012) Luna 9. http://nssdc.gsfc.nasa.gov/nmc/spacecraftDisplay.do?id=1960-010A Accessed 9 Oct 2012

Nebiker FR (1961) Feasibility study of an inflatable type stabilization and deceleration system for high-altitude and high-speed recovery. WADD Technical Report 60–182. Goodyear Aircraft Corporation, Akron, Ohio

Ohlmeyer EJ (2003) Control of terminal engagement geometry using generalized vector explicit guidance. In: Proceedings of the 2003 IEEE American control conference, Denver, CO, 4–6 June 2003, pp 396–401

O'Keefe SA, Bose DM (2010) IRVE-II post-flight trajectory reconstruction, AIAA-2010-7515. In: AIAA atmospheric flight mechanics conference, Toronto, Ontario, Canada, 2–5 Aug

Orloff RW (2001) Apollo by the numbers, NASA-SP-2000-4029, NASA History Series

Parreira B, Vasconcelos J, Montaño J, Ramón J, Peñin LF (2013) Hazard detection and avoidance in ESA lunar lander: concept and performance, AIAA-2013-5020. In: AIAA guidance, navigation and control conference, Boston, MA, 19–22 Aug

Pien H (1991) Autonomous Hazard Detection and avoidance for Mars exploration, AIAA-1991-3731-CP. In: AIAA 8th computing in aerospace conference. MD, Baltimore, pp 174–181

Pleasants JE (1974) Parachute mortar design. J Spacecr Rockets 11(4):246–251

Poynter D (2007) Parachuting: the skydiver's handbook, 10th edn. Para Publishing

Raiszadeh B, Queen E (2004) Mars Exploration Rover terminal descent mission modeling and simulation, AAS 04-271. In: AIAA/AAS space flight mechanics meeting

Ray HA, Burns FT (1967) Development and qualification of Gemini escape system. NASA-TN-D-4031, National Aeronautics and Space Administration

Relchenau DEA (1972) Aerodynamic characteristics of disk-gap-band parachutes in the wake of Viking entry forebodies at Mach numbers from 0.2 to 2.8. AEDC-TR-72-78

Rodier RW, Thuss RC, Terhune JE (1981) Parachute design for the Galileo Jupiter entry probe, AIAA-81-1951. In: AlAA 7th aerodynamic decelerator and balloon technology conference, San Diego, CA, 21–23 Oct

Rogers WF (1972) Apollo experience report - lunar module landing gear subsystem. NASA-TN-D-6850, NASA

Rohrschneider RR, Braun RD (2007) Survey of ballute technology for aerocapture. J Spacecr Rockets 44(1)

Roncoli RB, Ludwinski JM (2002) Mission design overview for the Mars Exploration Rover mission, AIAA-2002-4823. In: AIAA/AAS astrodynamics specialist conference and exhibit, Monterey, California, 5–8 Aug

San Martin AM, Bailey E (2005) The Mars Exploration Rover (MER) transverse impulse rocket system (TIRS). In: AAS guidance and control conference

San Martin AM, Mendeck GF, Brugarolas PB, Singh G, Serricchio F, Lee SW, Wong EC, Essmiller JC (2015) In-flight experience of the Mars Science Laboratory guidance, navigation, and control system for entry, descent, and landing. CEAS Space J 7(2):119–142
Schilling DL (1957) A method for determining parachute opening shock forces. Lockheed Aircraft Corp. Report 12543
Schroeder K, Bayandor J, Samare J (2018) Structural sizing and synthesis of Venera-class landers. J Spacecr Rockets 55(3):561–574
Schröder S, Grimm CD, Witte L (2020) A crushable shell for small body landers. CEAS Space J 12:65–72
Smith BP, Tanner CL, Mahzari M, Clark IG, Braun RD (2010) A historical review of inflatable aerodynamic decelerator technology development, IEEEAC paper 1276. In: IEEE aerospace conference, Big Sky, MT
Smrekar S et al (1999) Deep Space 2: the Mars microprobe mission. J Geophys Res 104(E11):27,013–27,030
Sostaric R, Rea J (2005) Powered descent guidance methods for the Moon and Mars, AIAA-2005-6287. In: AIAA guidance, navigation, and control conference, San Francisco, CA, 15–18 Aug
Spark NT (2006) A history of Murphy's law. Periscope Film
Spencer DA, Blanchard RC, Braun RD, Kallemeyn PH, Thurman SW (1999) Mars Pathfinder entry, descent, and landing reconstruction. J Spacecr Rockets 36(3):357–366
Steinfeldt BA, Grant MJ, Matz DM, Braun RD, Barton GH (2008) Guidance, navigation, and control technology system trades for Mars pinpoint landing, AIAA-2008-6216. In: AIAA atmospheric flight mechanics conference and exhibit, Honolulu, Hawaii, 18–21 Aug
Stelzner A, Desai PN, Lee W, Bruno R (2003) The Mars Exploration Rovers entry descent and landing and the use of aerodynamic decelerators, AIAA-2003-2125. In: 17th AIAA aerodynamic decelerator systems technology conference and seminar, Monterey, California, 19–22 May
Strandmoe S, Jean-Marius T, Trinh S (1999) Toward a vision based autonomous planetary lander, AIAA-1999-4154. In: AIAA guidance, navigation, and control conference and exhibit, Portland, OR, 9–11 Aug
Stubbs SM, Hathaway ME (1969) Effects of bottom-structure flexibility on water landing loads of Apollo spacecraft models. NASA-TN-D-5108
Surkov YA, Kremnev RS (1998) Mars-96 mission: Mars exploration with the use of penetrators. Planet Space Sci 46(11):1689–1696
Surveyor Project Staff (1968) Surveyor VI mission report. Part I: mission description and performance. Technical Report 32-1262, NASA
Thurman SW (2004) Surveyor spacecraft automatic landing system, AAS-04-062. In: 27th annual AAS guidance and control conference, Breckenridge, CO, 4–8 Feb
Turner MJL (2006) Rocket and spacecraft propulsion. principles, practice and new developments, 2nd edn. Springer
Underwood J (2005) Parachute system design case study - Huygens, parachute seminar. In: 3rd international planetary probe workshop, Anavyssos, Greece, June 27–July 1
Von Kármán T (1929) The impact of seaplane floats during landing. NACA Technical Note 321, Washington, D.C
Wagner H (1932) Über Stoss und Gleitvorgänge an der Oberfläche von Flüssigkeiten (in German: About impact and sliding processes on the surface of liquids). Zeitschrift für Angewandte Mathematik und Mechanik 12:193–215
Walton WC, Herr RW, Leonard HW (1964) Studies of touchdown stability for lunar landing vehicles. J Spacecr Rockets 1(5):552–556
Wang JT, Lyle KH (2007) Simulating space capsule water landing with explicit finite element method, AIAA-2007-1779. In: 48th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, Honolulu, Hawaii, 23–26 April
Way DW, Davis JL, Shidner JD (2013) Assessment of the Mars Science Laboratory entry, descent, and landing simulation, AAS-13-420. In: 23rd AAS/AIAA space flight mechanics meeting, vol 148, pp 563–581, Kauai, Hawaii

White L (1968) The invention of the parachute. Technol Culture 9(3):462–467
Willey CE, Sandy C, Welch J, Timmers R (2007) Impact attenuating airbags for Earth and planetary landing systems, AIAA-2007-6172. In: AIAA SPACE 2007 conference & exposition, Long Beach, CA, 18–20 Sept
Witkowski A (1999a) Mars Pathfinder parachute system performance, AIAA-1999-1701. In: 15th aerodynamic decelerator systems technology conference, Toulouse, France, 8–11 June
Witkowski A (1999b) The Stardust sample return capsule parachute recovery system, AIAA-1999-1741. In: 15th aerodynamic decelerator systems technology conference, Toulouse, France, 8–11 June
Witkowski A, Kandis M (2005) Mars Exploration Rover parachute system performance, AIAA-2005-1605. In: 18th AIAA aerodynamic decelerator systems technology conference and seminar, Munich, Germany, 23–26 May
Witkowski A, Kandis M, Kipp DM, Buecher D (2019) Mars InSight parachute system performance, AIAA-2019-3481. In: AIAA aviation forum, Dallas, TX, 17–21 June
Witte L, Schroeder S, Kempe H, Van Zoest T, Roll R, Ulamec S, Biele J, Block J (2014) Experimental investigations of the comet lander Philae touchdown dynamics. J Spacecr Rockets 51(6):1885–1894
Witte L (2015) Touchdown dynamics and the probability of terrain related failure of planetary landing systems - a contribution to the landing safety assessment process. PhD Thesis, University of Bremen
Wittenburg J (2008) Dynamics of multibody systems. Springer, Berlin
Woicke S (2014) Stereo vision for hazard detection. development, verification, and analysis. MSc Thesis, Delft University of Technology, Faculty of Aerospace Engineering,
Woicke S, Mooij E (2014) Stereo-vision algorithm for hazard detection during planetary landings, AIAA-2014-0272. In: AIAA guidance, navigation and control conference, National Harbour, MD, 13–17 Jan
Woicke S, Mooij E (2016) A stereo-vision hazard-detection algorithm to increase planetary lander autonomy. Acta Astronaut 122:42–62
Woicke S, Mooij E (2017) Terrain relative navigation for planetary landing using stereo vision measurements obtained from hazard mapping. In: Advances in aerospace guidance, navigation and control. Selected papers of the fourth CEAS specialist conference on guidance, navigation and control held in Warsaw, Poland, in April 2017, Springer
Woicke S (2018) Hazard relative navigation. Towards safe autonomous planetary landings in unknown Hazardous Terrain. PhD Thesis, Delft University of Technology, Delft, The Netherlands
Wolf DF (1971) Dynamic stability of a nonrigid parachute and payload system. J Aircraft 8(8):603–609
Wolf DF, Spahr HR (1975) A parachute cluster dynamic analysis, AIAA-1975-1398. In: AIAA 5th aerodynamic decelerations systems conference, Albuquerque, New Mexico, 17-19 Nov
Wolf DF (2005) Parachute materials and stress analysis. In: Parachute seminar, 3rd international planetary probe workshop, Anavyssos, Attiki, Greece, June 27–July 1
Wong EC, Singh G, Masciarelli JP (2006) Guidance and control design for hazard avoidance and safe landing on Mars. J Spacecr Rockets 43(2):378–384
Yan R, Cao Z, Zhu L, Fang Z (2013) Hazard avoidance via descent images for safe landing. In: Eighth international symposium on multispectral image processing and pattern recognition. International Society for Optics and Photonics, vol. 8918, pp. 1–6

Chapter 10
Thermal Protection Systems



During the course of the text, many different issues related to the entry and descent into a planetary atmosphere have been discussed. The critical nature of entry conditions have been made clear, and the relation with the aero-thermal environment has been established. Earlier, in Sect. 3.6 it was estimated that 1 kg of something in a 200 km orbit around Earth contains about 32 MJ of energy. During re-entry, this energy has to be removed from the system, and without taking any flow conditions around the vehicle into account, it became clear that this would lead to a huge temperature increase. Therefore, thermal protection of the entry vehicles is a crucial necessity, as the accident with the Space Shuttle *Columbia* back in 2003 made painfully clear. During launch, a piece of the insulation foam had broken off from the external tank and struck the thermal protection system tiles on the Orbiter's left wing. Subsequently, during entry hot atmospheric gases locally penetrated the heat shield and destroyed the internal wing structure, including its control lines to the control surfaces. This caused the Orbiter to become unstable and break apart due to the excessive thermo-mechanical loads on the vehicle.

To conclude from this that the thermal protection system (TPS) is the most critical sub-system of a re-entry vehicle may on one hand be true, but on the other hand it is "only" one of the important sub-systems. Each one of them should closely work together with the others, and the failure of one may cause a problem for the other. In the case of *Columbia*, it was failure of the TPS that destroyed part of the structure, which in turn affected the control of the vehicle. Severe rotational motion was the result and the interaction with the flow field pushed the loads on the vehicle to unacceptable levels, far beyond the limit loads, and the structure could no longer withstand them.

In this chapter, several aspects of thermal protection systems will be considered. To begin with, in Sect. 10.1 a historical overview of TPS development is presented, starting from the early days of re-entry flight with the Mercury, Gemini, and Apollo capsules, through the X-planes and SR-71 to the Space Shuttle and planetary probes. Conceptually, this overview will not show so much variation, albeit that the materials

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_10

development progressed well during those 60 years. Section 10.2 will introduce different aspects that should be taken into account, while starting the design process: the main requirements to which the TPS has to be developed, as well as a recap of some the primary physical aspects that make meeting these requirements very difficult. In Sect. 10.3 an overview of the most commonly used materials is given, applied in both ablative and re-usable heat shields. How these materials can be applied in TPS design is covered in Sect. 10.4, where the different options are presented. Section 10.5 introduces several (simplified) models that can be used for conceptual design and analysis of a variety of TPS systems. Section 10.6, finally, concludes this chapter with a case study: the optimisation of an entry capsule, while combining the aerodynamic shape, the trajectory, and the TPS design in the optimisation process. Due to the vast nature of the topic, involving many aspects of aerothermodynamics, materials science, structural dynamics and analysis, the setup of this chapter is in part that of a literature overview. The aim is to provide the reader a solid starting point for their own research.

10.1 Historical Development

10.1.1 The Early Days

The space age started with the development of capsules, both unmanned and manned. Because less information is available on TPS concepts and materials that have been used in Russian (and other) spacecraft, the focus will be on the American development, with some side steps to Europe, where appropriate. However, it is safe to say that especially in the early years, the design solutions have been quite similar. The capsules form one branch in the development of predominantly ablative, non-reusable heat shields, although over the years the design process has much improved. Within this family, we will look at Mercury, Gemini, and Apollo, which all had the same kind of TPS. With the Dragon (Space-X) and Orion (NASA) capsules, a new era was entered, but these will be included in this discussion. The second branch is formed by the experimental vehicles, and aircraft that were not part of the space-system development but provided a jump in technology level nonetheless. Subsequently, the X-15, X-20, SR-71, and X-33 will be discussed. The development of the X-20 provided a treasure of experience for the development of the Space Shuttle, which, because of it being a reusable system, will be the third branch on its own. Finally, within the fourth branch, some planetary probes will be briefly discussed, although they could technically also be included in the first branch.

Early in the space program, metals were the primary structural materials for both aircraft and spacecraft. Naturally, metals were one of the first options considered for TPS as well. However, for the relatively short-duration re-entry with Mercury, Gemini, and Apollo flying a nearly ballistic trajectory, a huge amount of energy had to be absorbed. Consequently, the windward surface heating was much too high to

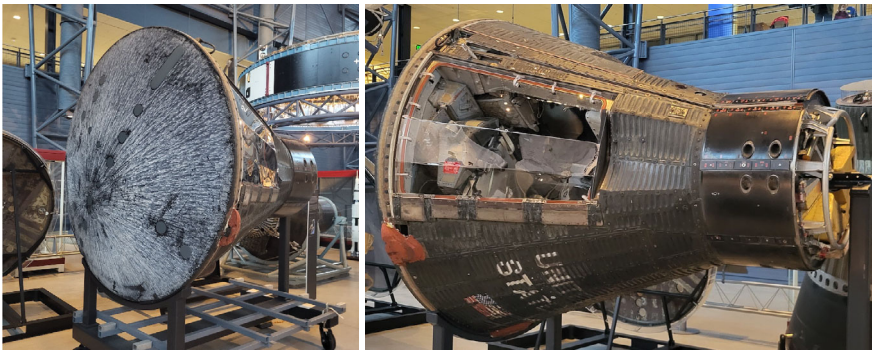


Fig. 10.1 On the left: the ablative heat shield of the (unmanned) Mercury entry capsule, launched on September 9, 1959, and nicknamed *Big Joe*; on the right its metal aft structure. On display at the smithsonian air and space museum Steven F. Udvar-Hazy center. Images: author

use metallic heat shields, and one had to change focus. This choice was already made early in the Mercury program: engineers debated about the trade off between heat sinks[1] and ablation (Swenson et al. 1989). In the end, the choice was easy: the high heat loads would not be sustained by heat sinks, so the development of ablative heat shields took form.

Figure 10.1 shows the ablative heat shield of the unmanned Mercury capsule nicknamed *Big Joe*. Launched on September 9, 1959, it was sent into a sub-orbital flight, heavily instrumented, amongst others to gather information about the thermo-mechanical environment. The heat shield shows the traces of the flow lines, all originating from the stagnation point. It is clear that this point is not centred on the capsule's heart line, which means that the capsule flew with a non-zero angle of attack. As discussed previously for Apollo, this can be achieved through an offset of the centre of mass. The same figure also shows the aft structure of Mercury. Metallic heat shields could be used on these leeward surfaces of both the Mercury and Gemini. For Mercury the structure was made of an Inconel alloy, whereas the heat shield was made of phenolic resins and fibre glass.

The Gemini project was set up with several objectives in mind. Apart from carrying two astronauts and gaining experience with on-orbit rendezvous and docking procedures, amongst others, it also greatly increased the knowledge about atmospheric entry and the heating environment. This helped to improve the design of the Apollo capsule, which would enter the atmosphere with a much larger entry velocity coming from a lunar return flight. To shield the Apollo Command Module from the extreme heat of re-entry, NASA was forced to use an ablative heat shield for the windward side of the capsule. However, because of uncertainties in the aerodynamic

[1] A heat sink is a heat dissipator that radiates it back to the surrounding air, before the heat can transfer into the underlying structure. A popular use of the heat sink can be found in computers, where a large block of metal with many extending fins is mounted on the CPU to radiate the produced heat away from the electronics.

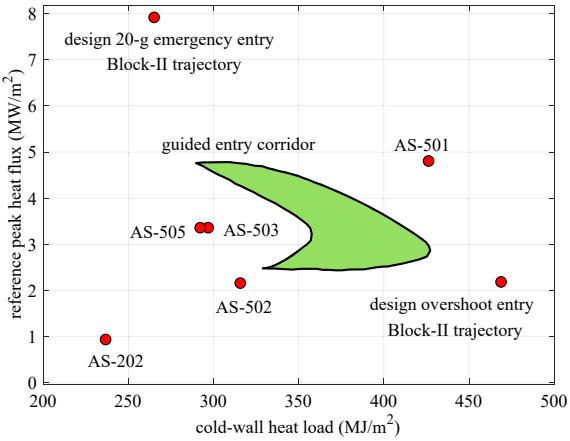


Fig. 10.2 Apollo heat-load entry corridor, with several of the pre-operational flights indicating the heat-load test conditions. Data taken from Erb et al. (1970)

heating on the leeward surface, also there ablative protection was the design choice, rather than metallic heat shields, which could otherwise have been overloaded. The earliest test flights of the Apollo command modules were primarily focussed on getting as much information about the TPS performance as possible.

Erb et al. (1970) discuss the Apollo TPS development, starting with the so-called Block-I designs, unmanned capsules (apart from Apollo-1, had it flown) to be used for early Low Earth Orbit test flights only. This design was extended by including a docking hatch and incorporating weight reduction and lessons learned, such that this Block-II design was lunar capable.[2] The many test flights gave valuable input to the design process. Figure 10.2 shows the guided entry corridor, in this plot shown as the peak heat flux (at the reference location) as a function of the integrated cold-wall heat load.[3]

Many potential thermal-protection materials were tested in the early 1960s, varying from the phenolic-glass system similar to the Mercury heat shield, phenolic nylon, phenolic melamine, the "Thermo-lag" materials developed by Emerson Electric, the Century Series material, and the Avcoat 5026-22. The latter material, an epoxy novolac resin with special additives, was finally chosen, and was initially to be applied as moulded solid tiles of about 30×30 cm, bonded to a shell of stainless steel honeycomb sandwich. However, doubts arose about the attachment of the tiles, and since Gemini with its filled honeycomb had successfully flown in the meantime,

[2] An overview of the different Apollo command modules is given on https://en.wikipedia.org/wiki/Apollo_command_and_service_module. Date accessed: March 29, 2023. A total of 35 capsules were built, of which 19 were launched. Of the 35, 13 were the so-called Block-I modules, the last one (CSM-020) launched on April 4, 1968, on Apollo-6. The others were the Block-II types with CSM-107 being the Apollo-11 Command Module. The last operational one was CSM-118, launched on November 16, 1973, as Skylab 4, the third crewed Skylab mission.

[3] The cold-wall heat load gives a conservative input to the TPS design, in contrast to the hot-wall model (see Sect. 3.6). It is noted that for the high-speed lunar return flight, also radiative heating is an important source of heat. Therefore, this should be included in the heat balance, as it will more than use up the margin of using a cold-wall model.

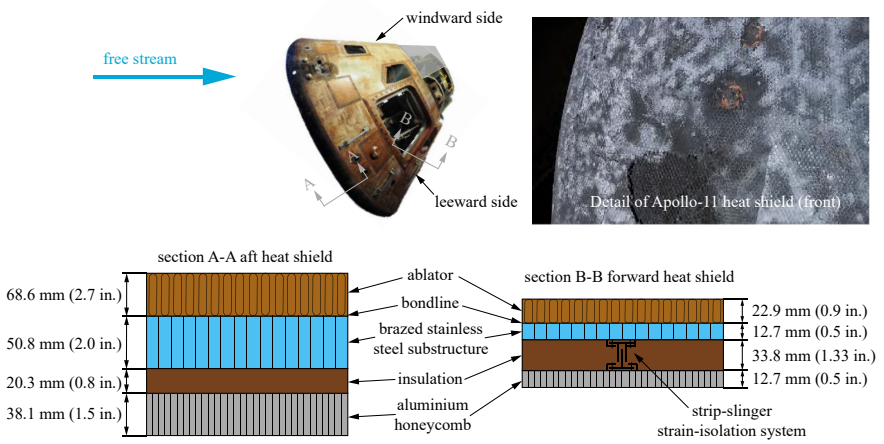


Fig. 10.3 Cross-section of the Apollo heat shield, with the insert showing the ablated condition and the cell structure of the integrated honeycomb. Image credits Apollo-11 command module (on display at the Smithsonian National Air and Space Museum at The Mall): capsule, author, and heat shield, National Air and Space Museum, open access

it was decided to use the same approach for Apollo: an upgraded Avcoat 5026-39 was applied by tamping it in mastic form into an open, phenolic-glass honeycomb. This honeycomb was, as for the earlier design, bonded to the stainless-steel honeycomb sandwich. The latter was not the primary, load-carrying structure, though. This was taken care of by a lower shell of aluminium honeycomb attached to the inner structure of the capsule, and separated from the top shell by an insulation layer to reduce heat conduction from the heat shield to the cabin. To accommodate differential expansions due to thermal gradients between the two shells, a slip-stringer strain isolation system was bonded to both honeycomb structures.

Design requirement for the bond line was a maximum temperature of 315 °C (600 °F), whereas requirements for the outer shell included, amongst others, to withstand micrometeoroid impacts, vacuum (zero pressure and potentially very low temperatures), and exposure to ultraviolet radiation. As mentioned, the many test flights provided crucial input to the TPS design process, and the many (small) design changes were aimed at robustness and mass savings. As indicated in Fig. 10.3, the ablative layer of the windward heat shield ("aft") had a thickness of about 6.9 cm (2.7 in.), whereas this was 2.3 cm (0.9 in.) in the crew compartment and forward portions. The total weight of the shield including substructure varied significantly in between preliminary design and the Block-I capsules. The latter had a (measured) TPS mass of about 771 kg. Several design changes lowered the mass to about 680 kg for the Block-II design.

The process of assembling the ablative layer of different sections of the command module was a tedious process (Pavlovsky and StLeger 1974). The three sections consisted of a stainless-steel subassembly to which the phenolic-glass honeycomb was bonded. The assembly was vacuum bagged, and for one hour the adhesive was

oven-cured at 163 °C (325 °F). A nondestructive ultrasonic transmission evaluation was used to test the bonding, and any deficiencies were repaired. Next, the Avcoat 5026-39 had to be injected into each cell of the honeycomb with the aid of a special gun developed for that purpose. This was done manually, for each of the about 370,000 cells of the honeycomb. After this part of the process had finished, the assembly was yet again vacuum bagged and oven-cured for 16 h at 93 °C (200 °F); after this it was post-cured for an additional 16 h at 121 °C (250 °F). To detect any faults, the assembly would be x-rayed for voids, and repairs would be made. Seals (for the doors), inserts and other details would be bonded in place. As one of the last steps, the main ablator was checked for moisture content. To ensure sealing of the porous material, a layer of thin, epoxy-based pore sealer and a moisture-protective plastic coating was then applied to the surface. Certification for the manned flights was achieved by many ground tests and the earlier mentioned unmanned test flights.

10.1.2 Winged Vehicles

The X-15 was one of the many experimental aircraft constructed in the 1960s to further technology development (see also Sect. 1.3.3 and Fig. 1.37). Constructed by a joint team of USAF and NASA, its objectives were to study the hypersonic flight regime up to Mach 7, including the structures, propulsion, and control systems required for such a flight. In total, three different models were designed and built, all of which performed sub-orbital flights and quite a few pilots earned their astronaut wings. The program provided quality data that is even useful till date. The fuselage of the X-15 was a slender cylinder, centred around a titanium frame and an Inconel-X heat-sink. This material had been selected, due to of its strength at temperatures over 650 °C. As heavy machined bar stock, it was used on the wing and vertical-stabiliser leading edges, whereas the leading edge of horizontal stabilisers were formed from Inconel-X sheets; they were built up internally with weld material to form a heat sink. To allow for thermal expansion, the wing leading edges were divided into segments with expansion joints. Further, the X-15 integrated a large, man-rated throttleable rocket engine and an advanced adaptive control system (Miller 2001; Jenkins 2007). The maximum recorded speed was Mach 6.7 (about 2.2 km/s), and still stands as the speed record for a piloted, rocket-powered aircraft. This record was achieved by a reconstructed X-15 model (X-15A-2), after the original aircraft was partially damaged in an accident.

This new configuration was modified to handle the higher heat by including an ablative coating. The nose of the vehicle (one of the stagnation points) sustained temperatures to 1,000 °C (Fig. 10.4), whereas temperatures on the wing leading edges (a stagnation line) could get as high as 1,315 °C. Temperatures at the fuselage bottom would decrease from the nose temperature to values just below 600 °C. Temperatures at the control surfaces, though, could be well over 1,000 °C (vertical stabiliser) or close to that (horizontal stabiliser). It is clear that the ablative coating was a necessity, especially if one considers that the plan was to increase the maximum flight Mach

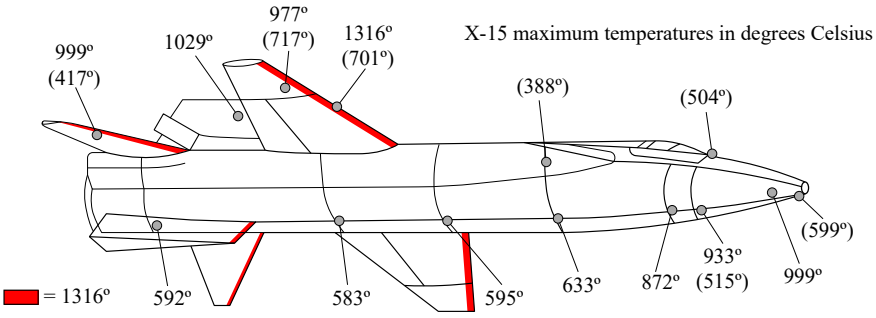


Fig. 10.4 Temperature distribution on the X-15A-2, with the values in brackets holding for the standard X-15 flights. Maximum speed $V_{\max} = 2{,}438\,\text{m/s}$ @ altitude $h = 30.5\,\text{km}$. Data converted from the original data in the book by Miller (2001)

number from 6 to 8. For comparison, the total heat load at $M = 6$ is 26 MJ/m^2, but more than 148 MJ/m^2 at Mach 8. Similarly, the wing leading edge absorbed 108 MJ/m^2 at Mach 6, but this increased to 312 MJ/m^2 at Mach 8 (Jenkins 2007).

The SR-71 *Blackbird* was a long-range, high-altitude, Mach 3+ strategic reconnaissance aircraft, developed and manufactured by the Lockheed's Advanced Development Company, commonly known as the *Skunk Works*. Its maiden flight was on 22 December, 1964 (Merlin 2009); the project came to an official end in October 1989, but was revived in the 1990s with limited success. The US airforce finally retired the *Blackbird* in 1998, although NASA operated the two last airworthy *Blackbirds* until 1999, to do high-speed flight testing. Amongst these flight tests, boundary-layer flow effects, digital integrated controls, heat transfer, and drag effects were studied. NASA Dryden engineers scheduled the aircraft for three Space Shuttle Orbiter landing approach simulation flights, paving the way for the approach and landing tests done with the Space Shuttle *Enterprise*.

The maximum design Mach number was 3.3, which subjected some parts of the aircraft to temperatures as high as 565 °C, with average surface temperatures varying between 240 and 330 °C. Because of the high thermal loading, it was thus not possible to use aluminium as a basic structural material, and the Skunk Works team turned instead to titanium, stainless steel, and other advanced alloys, as well as high-temperature plastics. The preferred material was titanium, though, because it is as strong as stainless steel (with the exception of the highest grade of stainless steel), but at the same time it is light-weight and durable at high temperatures. The disadvantage is that it is a sensitive alloy and brittle in nature. The use of titanium was not common and the team was pioneering with its use, and had to develop many new tooling and manufacturing techniques. To reduce the radar signature, large sections of the leading and trailing edges, vertical stabilisers, chines, and inlet spikes were made of plastic laminates of phenyl silane, silicone-asbestos, and fibreglass.

Concerning thermal protection, apart from the use of titanium, major portions of the surface panels were made as so-called corrugated webs, although these were not part of the standard airframe version, see Fig. 10.5. These corrugated face plates

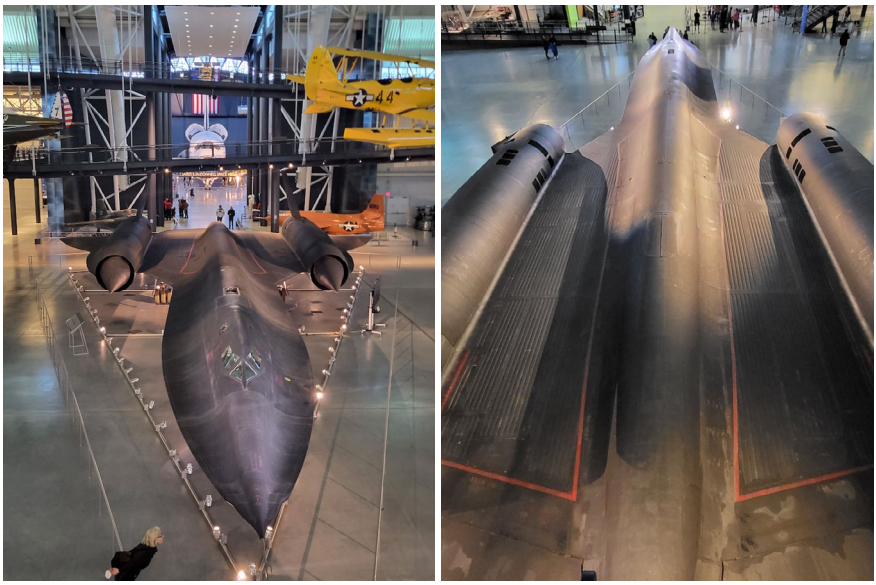


Fig. 10.5 The SR-71 *Blackbird*, on display at the Smithsonian Air and Space Museum Steven F. Udvar-Hazy Center. The figure on the right clearly shows the corrugated web panels, part of the vehicle's TPS. Images: author

would allow for thermal expansion, due to its sinusoidal/trapezoidal cross-section, and thus significantly reduce the thermal stresses. Finally, to decrease the wall temperature and consequently lowering the thermal stresses on the airframe, the complete aircraft was coated with a high-emissivity black paint, earning the aircraft its nickname. Despite all measures of heat rejection, after landing, the vehicle had to be allowed to cool down for over half an hour before the ground crew could handle it.

The X-20 (see Sect. 1.3.3 and Fig. 5.58), nicknamed Dyna-Soar (abbreviated from *dynamic soaring*), was designed to provide a piloted manoeuvrable vehicle to do a variety of military missions, amongst others: aerial reconnaissance, bombing, space rescue, satellite maintenance, and as a space interceptor to sabotage enemy satellites. To broaden the mission profile, it could also carry equipment for conducting experiments in the hypersonic and orbital flight regime (Miller 2001). Developed in the 1950s and early 1960s, the X-20 came in different configurations. Cancellation in December 1963 came only eight months before drop tests from a B-52 bomber aircraft would commence, and a first manned flight in 1966. The reasons being the large cost and the absence of a clear mission objective.

Focussing on the induced temperatures and the selected materials, in Fig. 10.6 the temperatures that could be sustained by the materials on different parts of the vehicle have been indicated. From the change of values, the material development over the four year period can be clearly seen. For example, the preliminary design of the primary structure and leading edge could sustain temperatures up to 316 °C, but

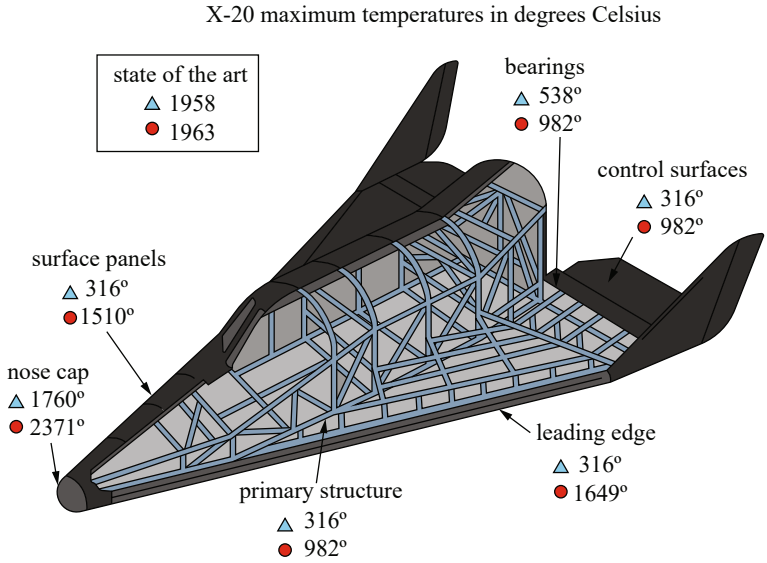
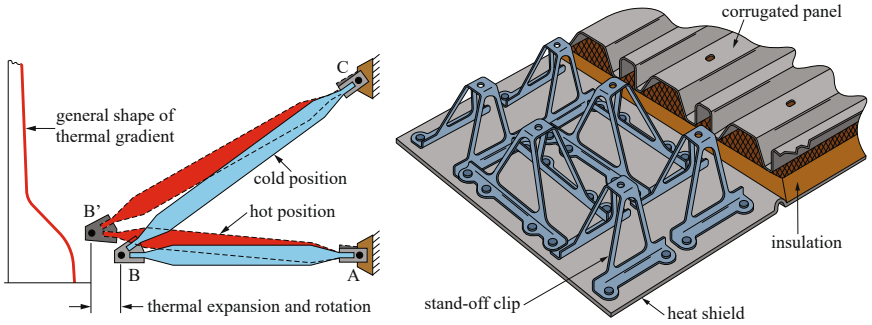


Fig. 10.6 Temperature distribution on elements of the X-20. Geometry and original temperature data obtained from Rotelli (1965)

with the new materials and designs developed in between 1958 and 1963, a major improvement was obtained, i.e., the related parts could sustain temperatures up to 982 and 1649 °C, respectively.

Whereas the X-15 had introduced a heat sink made out of Inconel-X (a nickel steel), material development in the Dyna-Soar program went considerably further, developing radiation-cooled insulated structures manufactured from René-41 (a nickel-base superalloy) and refractory materials. The X-20 was the first re-entry vehicle designed to have a fully metallic TPS. Trusses with rotating joints were chosen to allow for reorientation to the thermally induced shape without causing excessive induced stresses. Corrugated René-41 panels were used after reinforcement with a thin-gauge made from René-41 on the outer side and with insulation or refractory alloys on the inner side of the panel (Fig. 10.7). Also, the nose cap of this vehicle was made of carbon and zirconia (maximum sustainable temperature of 2,371 °C) (Rotelli 1965). Active cooling was also proposed for reducing temperatures within the crew compartment and two equipment bays. An absorbent material between sheet metal panels filled with water and gel would be used. The purpose of the gel was to decelerate the fluid flow, while the absorbent wicking kept it uniformly distributed to prevent hot spots (Hargis 1964). The windward side was covered with a TPS made from molybdenum (max. 1,510 °C) and the leading edges were coated with TZM molybdenum (max. 1,649 °C) (Rotelli 1965). This TPS configuration for the X-20 vehicle was expected to be reused for a minimum of four re-entry flights (Hargis 1964).



(a) Truss (b) Panel

Fig. 10.7 Typical X-20 TPS elements. Geometry inspired by NASA (Hargis 1964)

The TPS design was far ahead of its time, and despite its smaller size, the related spaceplane research, most notably the advanced materials and active cooling concepts, influenced the design flow of the much larger Space Shuttle.

Moving on to the Space Shuttle, it is most likely the most visually recognisable launcher of all times, with a total of 135 launches (also now we refer the reader to go back to the introductory chapter, Sect. 1.3.3 and Fig. 1.42), an operational span from 1981 to 2011 and a total of more than 1322 d in-orbit mission time. As it has been an important space-launch system, both for manned space flight in general, and the associated technology development in particular, it will be discussed in some more detail. The initial development of the Space Shuttle began in the late 1960s, and derived from experience acquired in the X-20 program, and the many spaceplane studies that were performed in the 1960s.[4] Key requirements were a reduction of launch cost, reusability, and a short turn-around time in between missions. Focussing on the TPS, it seemed logical that the starting point of the engineers was the ablative and metal heat-shield developments of the capsules and X-20, as most of their experience came from those earlier programs.

All throughout the 1960s many reusable launcher concepts were studied, but it was in 1969 that NASA awarded contracts to four main aerospace contractors for what later turned out to be the Phase-A studies of the Shuttle program. The end of this phase, which had included over 200 man-years of work, and many wind-tunnel tests and materials evaluations, had resulted in four baseline vehicles, i.e., straight-wing, delta-wing, and stowed-wing/variable-geometry wing vehicles, as well as lifting-bodies. In the mid 1970s two contracts were awarded for a Phase-B study, one to a team of McDonnell Douglas and Martin-Marietta, and one to North American Rockwell. The latter came up with a proposal of an orbiter atop a booster. This orbiter came in two versions, one with a high cross-range and one with a low one, the latter of which is shown in Fig. 10.8, and the interesting part is that most of the

[4] The book by Jenkins (2001) gives an excellent overview of all the different concepts that were conceived and studied in those early years.

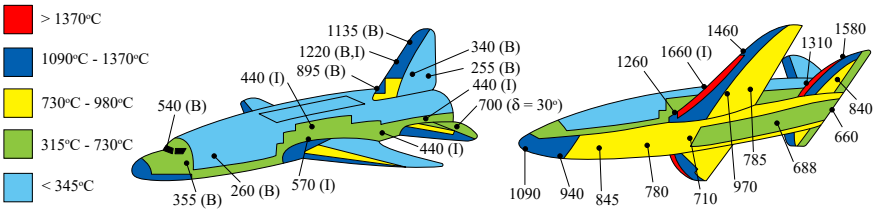


Fig. 10.8 North American Rockwell low cross-range orbiter, with temperature zones and corresponding peak values. The labels (B) stand for maximum temperature during boost, (I) interference (shock impingement), whereas the other temperatures are for a 200 nautical-mile entry (370.4 km). Data taken from Scoville and Gorsuch (1970)

proposed TPS was metal-based. As this fits nicely with discussions later on in this chapter, this concept will be briefly reviewed.

The figure shows the maximum temperatures and indicates whether the peak surface temperatures result from the boost flight, interference aerodynamics between booster and orbiter (shock impingement), or entry heating. Apart from the wing leading edges, which were made of reinforced pyrolised plastics (RPP), the vehicle design was entirely made of metal. Of course, for the different temperature zones, a variety of metals was used. These have been listed in Table 10.1, together with the primary selection criteria. The selected materials were, in part, heritage of the X-20 program discussed earlier. It is noted that the high cross-range orbiter, a delta-winged vehicle with vertical fins mounted at the wing tip, endured a higher heat load during entry because of its extended flight time. The maximum temperature would stay just below the value of the low cross-range orbiter. The region where columbium (Cl29Y) was applied was slightly larger, and for the lower temperature range TD-NiCr (thoria-dispersed nickel-chromium alloy) was selected.

Also the McDonnell Douglas/Martin-Marietta team considered a metallic TPS for the Shuttle Orbiter (Groninger 1971). This particular study focused on the structural performance, weight and cost of the heat shield and supports, with little attention paid to the thermal performance, insulation, and other TPS-related criteria. Three edge-supported tile configurations were studied, i.e., a skin with a corrugated web below, a skin with stringers, and a honeycomb, as well as two point-supported tiles, i.e., a honeycomb sandwich panel and a so-called isogrid. It was found that the use of point-supported honeycomb sandwich panels would generally result in the lightest metallic heat shield.

It was not just metal heat shields that were considered. A concept to actively cool the vehicle wall such that metal TPS could be more easily used, was studied by (Niblock et al. 1972), which would work, but under certain conditions. It could potentially lead to weight reduction and cost savings (the design cost would go up, but the operational cost would go down), but obviously the design would be more complex and had to be integrated in the remainder of the TPS design. Dulak and Gecka (1970) and Seifert (1973) studied a variety of ablative solutions for the Shuttle's TPS. Both replaceability and cost reductions were the main drivers for

Table 10.1 Material selection for North American Rockwell low cross-range orbiter (Scoville and Gorsuch 1970)

Temperature (°C)	Baseline material	Primary selection criteria
>1340	Reinforced	○ High-temperature properties
	Pyrolised plastics	○ Reserve margin
1090–1370	Columbium (Cl29Y)	○ High-temperature properties
	Silicide coated	○ Availability
730–980	Haynes 188	○ High-temperature properties ○ Oxidation resistance ○ Availability
315–730	Inconel 718	○ High strength to 730 °C ○ Fabricability
<345	Titanium 6AL-4V	○ High strength/density ○ Fabricability ○ Cost

this research, where for cost reductions the use of different materials, tooling and methods that provide for ease of fabrication and require fewer man-hours of labour, were considered. As material a phenolic/nylon/micro-balloon powder system and a silicone elastomer/micro-balloon mixture were considered, in combination with an underlying honeycomb. Significant progress was made, but ultimately discarded in favour of the final design choice, a passive system mainly consisting of the many black ceramic tiles.

The top-level design requirements for the Space-Shuttle TPS were to protect the structure from maximum temperatures of 1,540 °C, to be light-weight, cost effective, and reusable for 100 missions. Related to the reusability—to guarantee that only through inspection and maintenance—and thus the cost-effectiveness, many things can be said, but unfortunately here is not the space for that. The focus will be on the technical aspects, starting with the thermal environment. Once the mission profile had converged, the thermal loading became clear. Figure 10.9 shows the temperatures over the surface of the Orbiter, which supported the materials-selection process for a passive system, i.e., no active cooling elements have been used.[5]

The materials, selected for stability at high temperatures and weight efficiency, are shown in Fig. 10.10, with details on the covered areas and the total mass per material listed in Table 10.2. Some details on the materials are (Curry 1993): (1) Reinforced carbon-carbon (RCC) is used on the wing leading edges, the nose cap, including an area near the nose cap on the lower surface (the chin area between the nose cap and nose landing gear doors), the arrowhead aft of the nose landing gear door, and

[5] It is noted that even when the choice for a ceramic TPS had been made, NASA Langley Research Center continued a technology development program to improve metallic TPS, which has been documented by, for instance, Eidenoff and Rose (1974), Johnson and Killpatrick (1976), Black (1977), and Varisco et al. (1978).

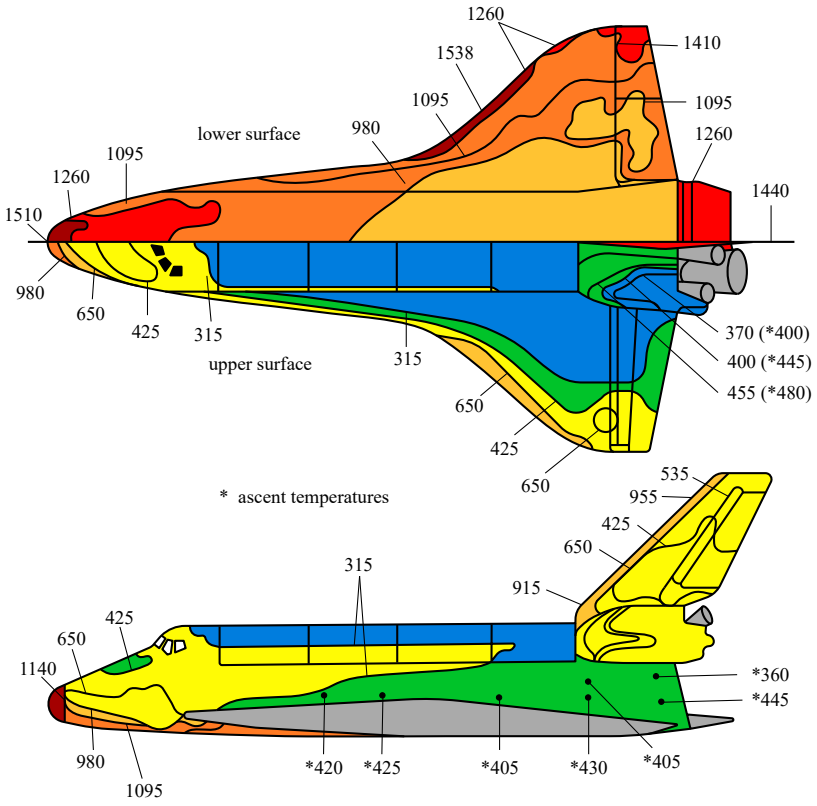


Fig. 10.9 Temperature distribution (in °C) over the space shuttle orbiter surface TPS elements. Data taken from Curry (1993)

an area around the forward orbiter/external tank structural attachment. RCC is used for the hottest surface areas, where temperatures will exceed 1,260 °C during the entry flight; (2) Black High-Temperature Reusable Surface Insulation (HRSI) tiles, made of coated LI-900 silica ceramics. They are used on many areas of the orbiter where the temperature is lower than 1,260 °C, but quite high nonetheless. The HRSI tiles amount to almost 20,000, and are the heaviest part of the TPS (over 4,400 kg); (3) Black tiles called Fibrous Refractory Composite Insulation (FRCI), which were developed later in the design process. As they showed improved strength, durability, resistance to coating cracking and weight reduction, they were used to replace some of the HRSI tiles on the hotter surface areas, mostly on the base of the orbiter. A total of about 3,000 tiles of this type were used; (4) White Low-Temperature Reusable Surface Insulation (LRSI) tiles, formerly used on the upper fuselage, upper wing and vertical tail, but later replaced by (5) Flexible Insulation Blankets (FIB), also known as Advanced Flexible Reusable Surface Insulation (AFRSI) blankets. This is a quilted, flexible blanket-like surface insulation, covering areas where the re-

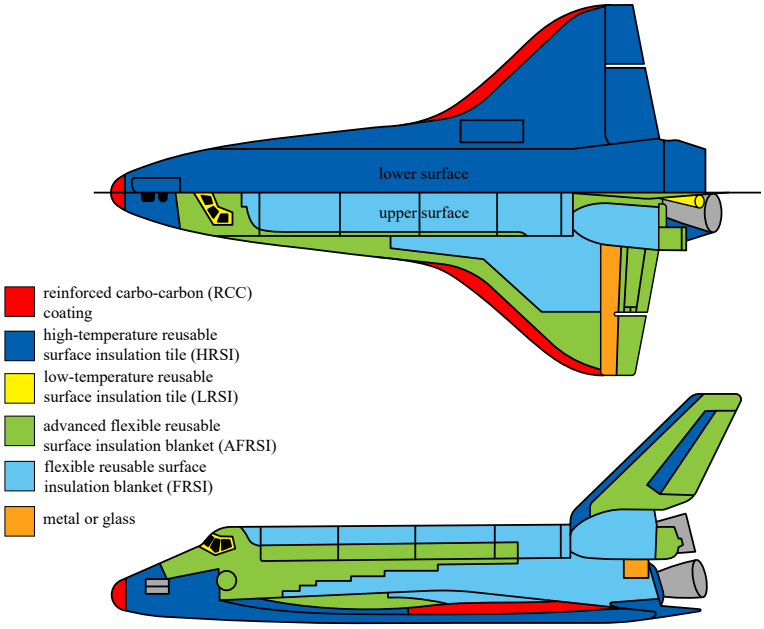


Fig. 10.10 Overview of the space shuttle thermal protection system elements. Material distribution taken from Curry (1993)

Table 10.2 TPS elements space shuttle

Element	Area (m^2)	Mass (kg)
FRSI	332.7	532.1
LRSI	254.6	1014.2
HRSI	479.7	4412.6
RCC	38.0	1679.3
Miscellaneous		918.5
Total	1105.0	8574.7

*Includes bulk insulation, thermal barriers, and close-outs

entry temperature was below 649 °C; (6) White blankets made of coated Nomex Felt Reusable Surface Insulation (FRSI) that were used on the upper payload bay doors, portions of the upper wing surface, and a portion of the OMS/RCS pods (OMS = orbital manoeuvring system; RCS = reaction control system). The maximum operational temperature does not exceed 371 °C for these areas. Later on in the program, in 1996, Toughened Unipiece Fibrous Insulation (TUFI) tiles became available, which are stronger, tougher tiles that were used in high and low temperature areas.

To conclude the discussion on the Space Shuttle Orbiter TPS, some words on the tiles that were used (see also Fig. 10.11). Each tile has a unique identification number with batch and location information, stored in a computer database. This

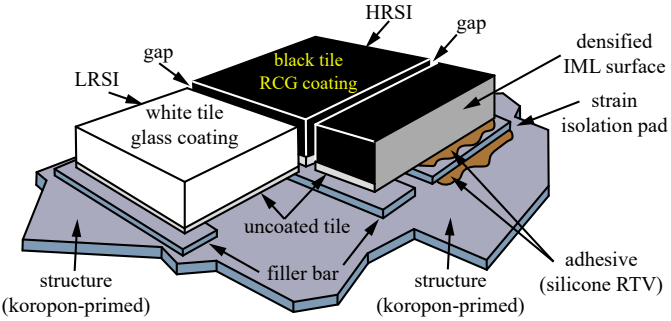


Fig. 10.11 Space shuttle orbiter TPS tiles. Abbreviations: LTRSI = Low temperature reusable surface insulation, HTRSI = High temperature reusable surface insulation, IML = Inner-mould line, RTV = Room-temperature vulcanising, and RCG = Reaction-cured gas. Data taken from Johnson (2016)

facilitates inspection, maintenance and replacement procedures. The outer portion of an HLSI tile is covered with a black-glazed coating of borosilicate. About 95% of the incoming heat is thus shedded, and only 5% will enter the tile and is absorbed by its interior. The tiles will prevent the heat flow from reaching the aluminium substructure. The tiles are composed mostly of air (90%) and only 10% are silica fibres a few millimeters thick (hence the light weight). The silica fibres are derived from high-quality sand. Finally, a silicon-rubber glue similar to common bathtub caulking was used to bond a tile to a felt pad, that is in turn bonded to the orbiter's skin. The felt is flexible and absorbed the stresses of airframe structural bending that could otherwise damage the tiles. It is noted that the gaps in between the tiles are filled with a thermal conductor, Nomex, a woven ceramic cloth arc. They were used in regions where the surface pressure gradient is high. This pressure gradient would otherwise cause excessive tile-to-tile gap heating, creating local hot spots. These could damage the tile adhesive bonds, the nylon felt, or the Orbiter's structure. Due to their flexible design, they allowed for tile movement without causing damage to the tile side walls or resulting in excessive sidewall loads (Dotts et al. 1982). This paper also compares the pre-flight thermal-response predictions with measurements taken during STS-1, the first Shuttle flight. Their conclusion was that the RSI subsystem thermal performance met the design requirements.

The tiles and other elements of the Orbiter's TPS can clearly be seen in Figs. 10.12 and 10.13. The Space Shuttle *Discovery*, on display at the Smithsonian Air and Space Museum Steven F. Udvar-Hazy Center provides a unique opportunity to closely inspect the TPS. In the front view, the large RCC nose cap is prominent. Around the nose, the HRSI tiles are used, and also on the nose-wheel bay cover. The leading edge of the delta wings are also made of RCC, whereas the sides of the wing tips are covered with tiles, as well as parts of the elevons. Since the body flap is heavily thermally loaded during re-entry, also there the HRSI tiles have been applied. The



Fig. 10.12 Space shuttle *Discovery* thermal protection system, front view. On display at the Smithsonian Air and Space Museum Steven F. Udvar-Hazy center. Images: author

(white) sides of the Orbiter are covered with the (advanced) flexible reusable surface insulation blankets.

The X-33, shown earlier in Fig. 7.29, was the demonstrator design of a future, reusable launch vehicle proposed by Lockheed Martin, called *VentureStar*. This vehicle concept belongs to the lifting-body family of vehicles (Sect. 1.3.3), supposed to be launched vertically and after completion of its mission to perform a horizontal landing. The operational vehicle, a single-stage-to-orbit configuration, was meant to be fully reusable, and was expected to have a seven-day turnaround period between flights. Key systems, which were part of a major technology development effort, were the unique linear aerospike engines, composite liquid hydrogen tanks, and a metallic thermal protection system (Miller 2001). The X-33 was a wedge-shaped sub-orbital vehicle to serve as the technology demonstrator providing useful information for development of the full-scale model. However, due to budget issues and some technological setbacks the project was cancelled before the demonstrator could be built.



Fig. 10.13 Space shuttle *Discovery* thermal protection system, rear view. On display at the Smithsonian Air and Space Museum Steven F. Udvar-Hazy center. Images: author

The sub-orbital mission profile of the X-33 depended on the configuration studied. For instance, the Rockwell configuration showed a maximum Mach number up to 14.72, which was successively used for a TPS sizing study for that particular winged concept (Olynick 1998). With the aeroheating conditions generated, he created several one-dimensional TPS stack-ups, applying different combinations of materials, that worked well given the procedure. One of the main conclusions was that to select the materials, it appeared to be critical to predict the transition to turbulence and turbulent heating.

For the winning lifting-body concept of Lockheed-Martin, different trajectories were designed to support the evaluation of materials and technology for the *VentureStar*, abiding by the constraints enforced on aerodynamics, aeroheating and structural loading. Hill et al. (1998) analyse five different trajectories targeted for landing at Michael Army Airfield, including short-range and long-range trajectories, and one with maximum thermal loading. The latter would have a maximum stagnation heat flux of $q_c = 325\,\mathrm{kW/m^2}$ (resulting in an equilibrium temperature of 1,610 °C, calculated for a cold wall, with an emissivity of $\varepsilon = 0.85$). These trajectories served

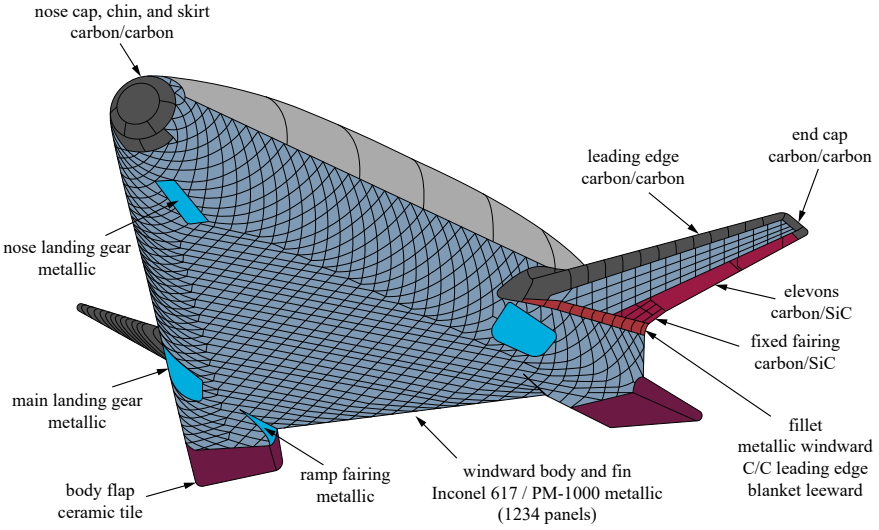


Fig. 10.14 Conceptual design of the X-33 thermal protection system. TPS overview obtained from Palmer et al. (1999)

as input for more detailed aero-thermal CFD analyses (Prabhu et al. 1998a, b; Hollis et al. 2001), which, in turn, provided input to the TPS design.

The X-33 technology demonstrator was foreseen to have a metallic TPS, apart from the peak-heating areas such as the nose and leading edges; these were protected with RCC, similar to the Space Shuttle TPS. According to representatives from Rohr (the subcontractor building the TPS), "*the metal is more durable and lighter than normal shuttle insulation, but not as good an insulator*" (Gonzales et al. 1995). The lifting-body design of the X-33 allowed for re-entry with a less severe heat-loading profile than the Space Shuttle, affording the use of the metallic TPS. In fact, temperatures on some of the bottom fuselage surfaces were predicted to be lower than 700 °C and could be protected by titanium. Higher-temperature areas would have a combination of Inconel 617 and PM-1000.

The TPS layout on the windward surface of the X-33 is shown in Fig. 10.14. This surface consists of 1234 honeycomb panels, with face material of either Inconel 617 or PM-1000, the latter to be used for areas with higher thermal loading. Each TPS panel consisted of outer and inner metallic face sheets with a metallic honeycomb core in between. A layer of insulation contained within an Inconel foilbag, which was attached to the back of the honeycomb panels, can be found under the inner face sheet. The insulation was sized to keep the vehicle's substructure and tank insulation below 723 °C (450 K). Standoffs at each corner of the panel were used to attach the TPS panels to the substructure. The honeycomb construction and geometry is typical of that found on existing commercial and military aircraft; however, the material gauges were somewhat smaller to minimise weight.

Part of the TPS design process was a correlation study of computational aeroheating predictions with experimental data (Hollis et al. 2001). Both laminar and turbulent aeroheating predictions were generated and experimentally checked in the windtunnel at Mach = 6 conditions. The former prediction appeared to be within 10% accuracy, whereas the latter was found to be within 25%, thus allowing the use of Navier-Stokes solutions to be used for the conceptual design and analysis of the TPS.

10.1.3 Planetary Probes

Having covered the development of thermal protection for more advanced flight systems, a return to capsules is justified here. Capsules are being used extensively for planetary exploration and sample-return missions, and, after retirement of the Space Shuttle, Space-X's Dragon capsule is used for ferrying astronauts to and from the International Space Station, and the return of astronauts to the Moon is centred around the Orion capsule. The unmanned Artemis-1 mission launched on November 16, 2022, was the first orbiting-Moon mission since the cancellation of the Apollo program.

One of the first planetary targets beyond our moon was Venus, the primary target of the Sovjet Union from the 1960s to the mid 1980s with their Veneras. The US followed with the Pioneer Venus mission. Not much is known about the structural design of the entry vehicle of Venera, other than that for most of them the lander was encapsulated in a spherical insulated aeroshell, to withstand the large thermal loads, see, for instance, the paper by Keldysh (1977). The entry velocity was typically on the order of 11 km/s, resulting in very large convective and radiative heat fluxes. However, most information to be found is on the thermal protection of the instruments on the lander, although Ball et al. (2007) mention that the forward aeroshell is made of asbestos composite over honeycomb, with a total mass of about 900 kg out of a vehicle mass of about 1,500 kg (note that, due to the unknown nature of the Venusian atmosphere at the time, other than that it would be very hot and dense, the Veneras were very much over-designed).

Peterson and Nicolet (2001) discuss two options to be used for Venus missions, in support to the Pioneer-Venus program. Their two approaches to ablative thermal protection design were a typical carbonaceous charring ablator and a dielectric reflective ablator, with the main difference in their optical properties. The charring material of the carbon-phenolic composite is highly absorptive to the gas-cap radiation, whereas the dielectric is highly reflective. It followed from the study that mass loss was not a key issue for the sizing of the heat shield. The carbon phenolic behaved more like a heat sink, and lost very little mass, such that the bond-line temperature determined the sizing. The Teflon absorbed little to no energy, but rejected the heat by a much larger mass loss. However, with sufficient material present it can reflect the energy for the duration of the radiation pulse.

Comparing the different material behaviours it was found that the Teflon heat shield would be the lighter solution. For an Apollo-shape capsule and a steep entry, the mass saving was computed to be 13%, compared to the carbon phenolic heat shield on a hemisphere cone and a shallow entry. Nonetheless, the carbon-phenolic design was chosen, as it was determined with further study that the convective flux would be dominating the heat balance compared to the flux radiating from the gas (Brewer 1975). The reflective properties of Teflon would thus be quite diminished, and the weight saving would be negated by the thicker required material layer.

To aid the understanding of TPS performance, the selected carbon phenolic heat shield had thermocouples installed to enable a comparison between the predictions and the measurements. Pitts and Wakefield (1980) describe the results of this comparison. With an entry velocity of 11.54 km/s the heating environment was severe. Nonetheless, the heat shield performed well, even better than predicted. Their conclusion based on the preliminary analysis of the measurements was that the design methods used to calculate the heat-shield ablation provided conservative results.

The situation for a Mars entry would be a completely different one, given the much thinner and colder atmosphere, as well as the lower entry velocities. To prepare for landing missions on Mars (with Viking being the first one), Martin-Marietta was tasked by NASA to come up with a thermal-protection design for this particular environment. Strauss (1967) discusses the development of these new super-light ablators (SLA), starting from two conflicting requirements: the heat shield must (1) provide efficient and reliable thermal protection and (2) allow radio-frequency (RF) signal transmission before, after, and possibly during ablation. Solutions would be to have a highly efficient ablator, but with windows to allow for the RF transmission, or to settle for a less-efficient ablator, which would be transparent for RF after ablation. The second option could be achieved at the expense of a thicker ablation layer, thus a weight penalty. Martin-Marietta developed two highly efficient ablators, i.e., the highly filled elastomeric silicone ablators SLA-741 and SLA-561, of which a comparison in terms of density, thermal conductivity, and specific heat values with those of conventional low-density charring ablators can be found in the cited paper. Explored materials for the RF-transparent ablators included both non-charring materials (i.e., porous Teflon) and compositions that form silica chars.

Two mission scenarios for two capsules were considered, with differences in ballistic parameter, $K = 63\,\mathrm{kg/m^2}$ ($47\,\mathrm{kg/m^2}$), entry velocity, $V_E = 3{,}658\,\mathrm{m/s}$ ($4{,}877\,\mathrm{m/s}$), entry angle, $\gamma_E = -10°(-18°)$, maximum dynamic pressure, $\bar{q}_{\max} = 1{,}963\,\mathrm{N/m^2}$ ($10{,}534\,\mathrm{N/m^2}$), maximum stagnation flux, $q_{c,\max} = 182\,\mathrm{kW/m^2}$ ($772\,\mathrm{kW/m^2}$) with a pulse duration of 250 s (50 s), and total integrated heat load, $Q = 23\,\mathrm{MJ/m^2}$ ($15\,\mathrm{MJ/m^2}$). It was found that the SLAs, composed of a silicone resin with cork, phenolic microballoon, silica microballoon, and refractory fibre fillers (with densities varying between 190 and 250 $\mathrm{kg/m^3}$) were the best-performing candidates. Moreover, they had a thermal conductivity of about 40% lower and an effective heat capacity of about 50% higher compared to the conventional low-density charring ablators, such as filled or foamed silicone, and filled epoxy or nylon-phenolic. Tests in an arc-jet plasma facility also showed an RF transmission efficiency of better than 90% when tested in the charred condition.

For the Viking entry capsules the SLA-561 was selected for their heat shield, along with an RF-transparent material known as SLA-220. Because the Viking Lander performed an entry from Mars orbit, the entry conditions were less stringent than tested for by Martin-Marietta, and the maximum design heating rates were "only" about 300 kW/m^2. This changed when the US went back to Mars, 20 years after Viking, with Mars Pathfinder, which had to perform a direct entry into the atmosphere (Willcockson 1999a). But, despite the much more stringent conditions, the material developed for Viking still outperformed the more recently developed ceramic derivatives, as in their late-1960s study Martin-Marietta had actually tested SLA-561 for the higher-energy application of a direct Mars entry. With new arc-jet tests performed, the SLA-561V showed an excellent performance for heat rates as high as 2.4 MW/m^2. This would allow for margins in the aerothermal computational accuracy, angle-of-attack effects, material variability, as well as basic manufacturing uncertainty. The design analyses showed that an SLA-561V thickness of 1.52 cm would suffice, but in the end it was recommended to go up to 1.78 cm.

With the development of the Stardust mission in the 1980s, with one of its mission objective to collect dust samples from the coma of comet Wild 2, as well as samples of cosmic dust, and return these to Earth for analysis, the need for a new type of TPS that could survive the very demanding heating environment arose.[6] Even though the common solution could have been a carbon-phenolic ablator, it would have made the sample return capsule (SRC) too heavy. NASA Ames Research Center was researching and developing a new material, centred around impregnating a preform of carbon foam with a phenolic resole resin to produce favourable pyrolysis and ablation characteristics. This development led to the Phenolic Impregnated Carbon Ablator (PICA).

The main advantages of PICA are a much lower density than other carbon/phenolic systems, with an average density of about 270 versus 561 kg/m^3 for Avcoat 5026-39 used for Apollo, and an efficient ablative capability at high heat fluxes. This material showed great promise, although it was, in essence, an R&D material that had never flown before and, thus, represented some risk for Stardust flight qualification. Nonetheless, it was selected and it all worked out very well. After the successful entry of the Stardust SRC, it became the material of choice for the Martian entry missions Mars Science Laboratory and Mars 2020, and it became also the material of choice for Orion. A variant of this material, PICA-X, is currently used on Space-X's Dragon capsule.[7] Additional phenolic-based composite materials have been used (Sawant et al. 2017) or are under development. Note that Agrawal et al. (2016) discuss the arc-jet tests of multiple PICA samples, exposed to very high heat fluxes and pressures to study the operational limits of the material. The test conditions were an estimated cold-wall heat flux and stagnation pressure of 18 MW/m^2 and 130 kPa, respectively, and, as it turned out, all samples survived without any failure either during or after

[6] To date, the 2006 Earth entry of the Stardust sample-return capsule with a velocity of $V_E = 12.799$ km/s is the fastest one recorded. It led to a peak heat flux of 9.42 MW/m^2.

[7] https://www.nasa.gov/centers/ames/thermal-protection-materials/tps-materials-development/low-density-ablators.html. Date accessed: March 31, 2023.

the exposure. This proves that PICA can be used for future missions with operational conditions well beyond the current mission profiles.

Returning to Stardust, Willcockson (1999b) reviews the capsule design experience, whereas Trumble et al. (2010) discuss the post-flight aerothermal analysis (for a post-entry picture of the Stardust SRC, see Fig. 5.5, and the reader is referred to Sect. 9.7.4 for an overview of the sample-return entry and descent mission). The Stardust SRC has a diameter of about 0.811 m. Its forebody is a blunted sphere cone with a 60° half-angle, with a rounded nose of radius 0.2202 m around the stagnation point. The afterbody is a 30° truncated cone (Fig. 9.120).

The forebody heat shield of the SRC was manufactured as a single piece of PICA to withstand the high heat flux. The heat shield was constructed around the rounded nose with a pre-flight thickness of the heat shield of 6.45 cm. The pre-entry, detailed predictions showed that the surface could recede by as much as 0.65 cm (Beerman et al. 2009). The TPS requirements for the afterbody were a lot more relaxed, but nonetheless still relatively high, with a peak flux of 900 kW/m^2. Meeting with this thermal loading was the super light ablator SLA-561V (Johnson et al. 1997). It had also successfully been used on Viking and Mars Pathfinder, as well as on the Space Shuttle external tank, and had been qualified for high thermal loading, i.e., 24 MW/m^2. Both TPS systems were bonded to the carbon composite structure that formed the inner shell of the entry capsule. Due to the critical nature of the entry mission, extensive arc-jet testing was performed on both the PICA and the SLA-561V.

Being a sample return capsule, this allowed for an extensive post-flight analysis of the TPS and the first-time use of PICA. Stackpoole et al. (2008) discuss this in their paper, using three core samples taken from different locations: near the stagnation point, at a flank, and a segment taken from the shoulder of the heat shield. The evaluation and analysis focussed on density profiles, recession determination, a thermal analysis profile, PICA bond-line examination, strength assessment of the remaining virgin PICA, an emissivity profile, a chemical analysis profile, and a micro-structural analysis. Overall, their findings were very positive, as the examined PICA was in good condition and mostly intact. The observed degradation was only due to the entry heat load, without any signs of impact damage. Also, a substantial amount of virgin PICA was present in all three cores. Last but not least, the observed values were also compared with the pre-flight predictions by NASA's Fully Implicit Ablation and Thermal response code, FIAT.

Another challenging entry mission for which detailed analysis of the entry environment and its interaction with the heat shield was required, was the Galileo entry-probe mission (see for the details Sect. 9.7.3). Launched on October 18, 1989, it reached Jupiter on December 7, 1995, which also marked the date of the entry-probe mission: a small capsule, 86 cm high and having a mass of 339 kg, entering the Jovian atmosphere at tremendous speed and encountering the most severe aerothermal environment to date. However, to fully appreciate the significance of this, we need to go back in time. The project had been initiated much earlier (1977) and at that time there were no facilities to test the extreme heat environment of Jupiter. Such a facility only became available after 1980, when NASA built a special laboratory, the Giant Planet Facility, to simulate the heat load (Meltzer 2007).

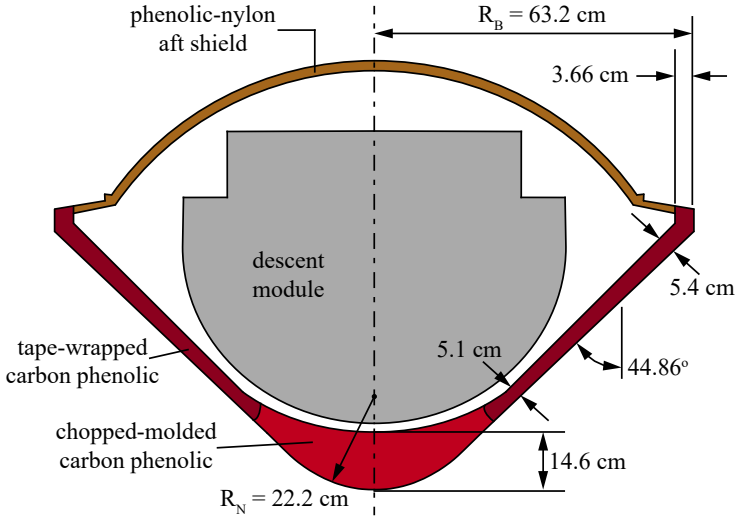


Fig. 10.15 Schematic layout of the thermal protection system of the Galileo-probe deceleration module. Data taken from Milos (1997)

Upon entry, the probe decelerated from Mach 50 (at 47.4 km/s), the forebody heat shield had to survive peak heating rate and heat load on the order of 300 MW/m^2 (!) and 3,000 MJ/m^2, respectively, with a maximum deceleration of 250 Earth g under normal conditions. The temperature would go up to 16,000 °C during this flight of terror. Milos (1997) describe the Galileo-probe heat-shield design and the ablation experiment, i.e., the installation of a sensor suite to obtain measurements of the heating environment and heat-shield performance. Milos et al. (1999) present the analysis of this experiment. To begin with the design, apart from surviving the very high wall temperature, the heat shield also had to maintain the structural bond-line temperature below 371 °C prior to parachute deployment, which released the heat shield from the descent module.

The forebody exterior shape was an axisymmetric sphere-cone with a nose radius of 22.2 cm and a cone half-angle of 44.86° (see Fig. 10.15). The nose cap was made out of chopped-moulded carbon phenolic, the frustum was tape-wrapped carbon phenolic, and the aft heat shield was phenolic nylon. The forebody heat-shield thickness varied in between 14.6 cm at the centreline and 5.1 cm at the front of the frustum. It then increased again to 5.4 cm at the rear of the frustum. These thicknesses included a 50% safety margin to account for the effect of atmospheric uncertainties on the aerothermodynamic modelling. The ablator experiment was meant to give insight in the total mass loss, the surface recession, and by inspecting the final ablated state to determine the drag coefficient to within 5%. A total of ten sensors was installed. After analysing the transmitted data, it was concluded that the measured recession was 3.63 $\pm$ 0.15 cm, whereas the estimated peak heat flux and heat load were

$134 \pm 5\,\mathrm{MW/m^2}$ and $1{,}610 \pm 60\,\mathrm{MJ/m^2}$, respectively (Milos et al. 1999). This turned out to be only about half the load that was assumed in the design phase.

Before moving back to the US Mars entry capsules, some words on the TPS of the European Huygens probe, designed for entry in the Titan atmosphere (see Sect. 9.8 for a mission and system overview). Being the European part of the Cassini/Huygens mission meant that the selected TPS design (including the materials) was European, too. The front shield was made of a sandwich structure (aluminium honeycomb + carbon-fibre reinforced-plastic (CFRP) skins) and of two ablative thermal-protection materials developed and produced by EADS-ST in France, i.e., AQ60/I silica-fibre tiles that are bonded on the front face's structure—a dome-shaped carbon-fibre honeycomb structure—and jointed by a silicone glue, and for the rear face, where the heat flux would be moderate, Prosial-1000, implemented using a spraying process. The back cover was made of an aluminum shell covered with Prosial (Bouilly 2004).

Labaste and Mignot (2003) and Bouilly (2006) give a summary of the material properties of AQ60/I and Prosial-1000. AQ60 is a felt made of short silica fibres, and is realised by vacuum processing of an aqueous suspension of silica fibres. This is followed up by reinforcing this moulded form by an impregnation of phenolic resin to get AQ60/I. To extract tiles to the required shape and tolerances it is then machined. The final density of the reinforced felt is about $280\,\mathrm{kg/m^3}$, similar to that of PICA, and the total porosity is 84%. Prosial-1000 is made up of a silicone elastomer with great thermal properties and of silica hollow spheres, making it possible to decrease density down to values of 600–$540\,\mathrm{kg/m^3}$. Prosial-1000 is directly sprayed onto the surface to be protected and can thus accommodate very complex shapes.

Originally, for Mars Science Laboratory the TPS was based on the use of SLA-561V, as it had been used on all previous Mars missions. However, due to the different vehicle and mission parameters, the heating environment would be much more extreme. The larger entry mass and speed resulted in boundary-layer transition to turbulent flow before peak heating, thus significantly increasing the peak heat flux (Edquist et al. 2009). They found peak values of $q_{c,\max} = 1.97\,\mathrm{M/m^2}$ and a maximum pressure of 0.371 Earth atm, with a total heat load of $Q = 54.77\,\mathrm{MJ/m^2}$. Even though for Mars Pathfinder the SLA-561V had been successfully tested up to $q_{c,\max} = 2.4$ $\mathrm{MW/m^2}$, and the material had been selected under the (confirmed) assumption of a laminar boundary layer, the margin was considered to be too small. Mars Pathfinder, with a diameter of 2.65 m (vs. 4.5 m for MSL) and a mass of 585 kg (MSL: 3,380 kg), encountered a peak heat flux of $1.18\,\mathrm{MW/m^2}$. However, the problem was even more complicated than that. The boundary-layer transition took place on the leeward side of the heat shield (MSL flew with an angle of attack of about 16°), which meant away from the stagnation point. The peak heat flux occurred towards the shoulder of the heat shield, and with the large variation of temperature in the shield combined with the growing boundary layer, also the (thermal) shear stresses were large. For comparison, for Mars Pathfinder these were about $150\,\mathrm{N/m^2}$, whereas for MSL they were about three times as large. All these aspects combined required new qualification testing. It was during these tests that the material failed, and new design choices had to be made (Beck et al. 2009; Szalai et al. 2014). At the time, new material was being tested for Orion, albeit it had been known for a longer period, i.e., the PICA material

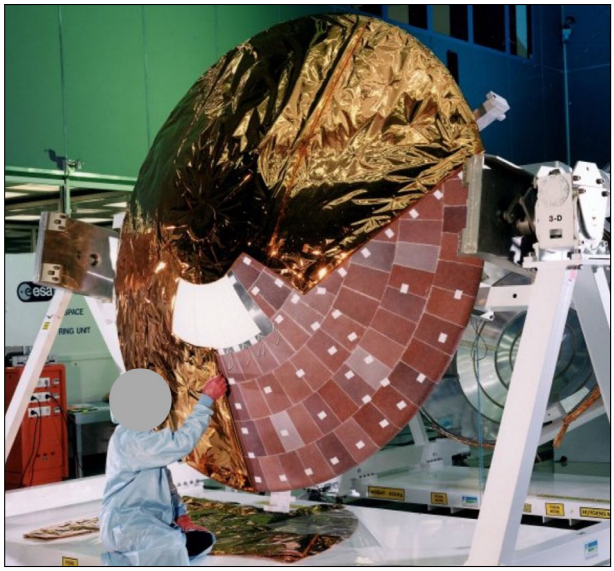


Fig. 10.16 The Huygens heat shield, with the multi-layer insulation (MLI) being installed for protection during the interplanetary transfer. Image: ESA

used for Stardust. With the good test results of PICA, the decision was made to step away from SLA-561V and use PICA instead. However, with the expected launch date known, there were only 18 months left to design, fabricate, test, and deliver the heat shield.

The dimensions of the Orion heat shield (at 5 m diameter slightly larger than the one of MSL with 4.5 m) made it necessary to design the TPS as a tiled system with some type of gap filler, as due to manufacturing limitations the size of a single piece of PICA was limited. Due to the size similarity, for MSL the same approach had to be followed. However, to change from a monolithic design based on SLA-561V to a tiled design using PICA proved to be a challenging design effort. Beck et al. (2009) state six key requirements for the design to meet, amongst others that each tile is limited by the vendor standard PICA block size, which is 60.96 × 106.68 cm (24 by 42 in.) with three different thicknesses of 15.24, 20.32 or 25.40 cm (6, 8, or 10 in.), and that the design shall not only minimise the total number of tiles, but also the number of differently-sized tiles. Also, the tiles shall be designed with a uniform thickness of 31.75 mm (1.25 in.), except where necessary to maintain the outer mould line (transition and shoulder tiles). The final tile layout, together with the actual flight hardware, is shown in Fig. 10.17. The colour coding and the numbers indicate the different groups of tiles, with their individual types. The two centre tiles (–141 and –142) are known as the spherical dome tiles for the nose, the tiles in the concentric

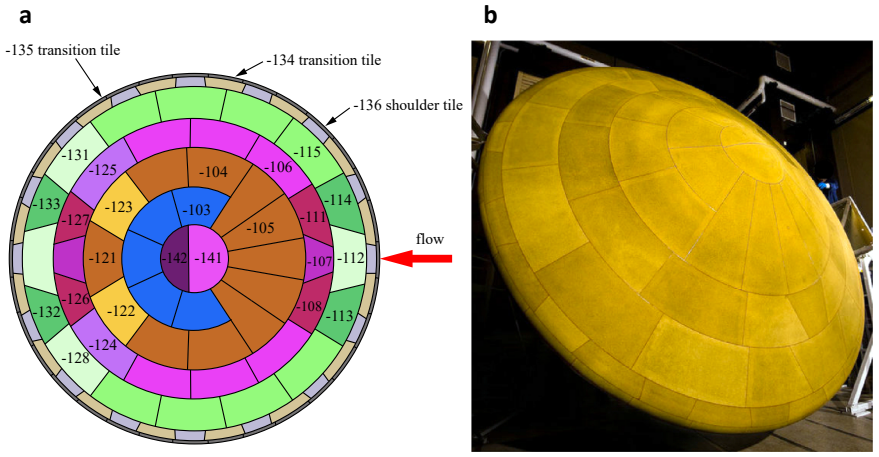


Fig. 10.17 Mars science Laboratory thermal protection system. **a** Tile layout, data taken from Beck et al. (2009). **b** Flight heat shield. Image: NASA

rings up to coding –112 are conical "field" tiles, and the two rings at the edge contain the transition and the shoulder tiles.

A uniform thickness was used for all tiles, except for the transition and shoulder tiles, such that the number of different tiles was minimised. In the end, 27 separate tile designs were used out of a total of 113 tiles. The tiles were bonded directly to the aeroshell, a stiff composite design, with aluminum honeycomb and M55J-BCTy-1 graphite cloth face sheets. As an adhesive, HT424, an epoxy-based material, was used. Extensive testing throughout the design, manufacturing, and integration phases had to guarantee the proper functioning of the heat shield.

To check the flight performance, the heat shield was instrumented with a comprehensive set of pressure (7) and temperature (7) sensors. Based on the heat-shield flight temperature data, Mahzari et al. (2015) performed an inverse estimation of the MSL entry aeroheating and heat-shield response. The thermocouple flight data indicated that the fully implicit ablation and thermal response predictions obtained by FIAT agreed with the true boundary condition and material properties. The heating profiles on the leeward side, estimated from the temperature measurements, had been well-predicted by CFD, although the heating increase due to the turbulent boundary layer was slightly lower than predictions. For the stagnation region, the heat-rate profile appeared to be slightly higher than the predictions, which has been partially attributed to radiative heating. Finally, the nominal PICA model consistently over-predicted the temperature profile at one windward and one leeward location, which led to the conclusion of having inaccurate data on the virgin properties of PICA.

The Mars 2020 TPS design closely followed that of MSL, as the capsule shapes and mission parameters were very similar. Because of these similarities, the Mars 2020 design will not be further discussed. Instead, the interested reader is referred to the paper by Mahzari et al. (2022) for specific details.

Above, it was already mentioned that for the Orion capsule a materials program was well on its way to test the PICA material, as part of the materials trade-off to select the heat-shield material (a total of eight candidate materials were considered). For a return flight from the Moon, the aerothermal heating will be severe, and the heat shield and the backshell will reach outer-surface temperatures over 3,300 and 1,500 °C, respectively. Requirement for Orion was that the ablative and insulative materials on the heat shield and backshell will be sized to maintain bondline temperatures below 290 °C. Stewart et al. (2018) describe how TPS experience from both the Apollo and Space Shuttle program have been used to reduce program risk and improve affordability to meet NASA's future manned exploration missions. Given the same capsule shape and operational environment, the Apollo heritage led to cost and schedule reduction concerning the design and interfaces. The Shuttle experience was more elaborate, and led to many savings related to materials manufacturing and testing, as well as tile development, and the use of facilities and training of personnel.

After an extensive study and intense test campaigns, NASA selected the Avcoat ablator system over PICA, the other final candidate, for the Orion crew module.[8] Contrary to the process used for Apollo, i.e., manually impregnating the Avcoat into the thousands of honeycomb cells, a total of 180 (unique) large blocks of the ablative material were manufactured in a new process, which were then bonded to the heat shield's surface. The blocks varied in thickness between 2.5 and 7.6 cm (1–3 in.). The small seams between the blocks were filled with a mixture that took some time to harden, after which a coat of white epoxy paint was applied to the surface. Finally, after the painted surface had dried, aluminised tape was applied to provide surface resistivity, and improved absorption of solar heat and infrared emissions. The completed heat shield was subsequently bolted to Orion's structure with 68 bolts.[9] A second ablator material was also used in certain locations on Orion. This material, the 3-Dimensional Multifunctional Ablative Thermal Protection System (3DMAT), is made of woven threads of quartz in resin, and was developed at NASA Ames Research Center. Stronger than Avcoat, it was used to strengthen connection points along the spacecraft.[10] After a successful launch and mission, Artemis-1 landed in the Pacific on December 11, 2022. The heat shield performed well, but at the moment of writing detailed inspection should provide more insight in the actual performance.

[8] NASA release 09–080: https://www.nasa.gov/home/hqnews/2009/apr/HQ_09-080_Orion_Heat_Shield.html. Date accessed: April 10, 2023.

[9] https://www.nasa.gov/feature/heat-shield-install-brings-orion-spacecraft-closer-to-space. Date accessed: April 10, 2023.

[10] https://www.nasa.gov/feature/ames/after-15-years-1000-tests-orion-s-heat-shield-ready-to-take-the-heat. Date accessed: April 10, 2023.

10.2 Design Aspects

This section will address some aspects that are relevant for the design of thermal protection systems. As this chapter is only an introductory text, there is only space for a global overview. Nonetheless, the introduced aspects and the cited references will be sufficient to make the reader aware of the importance of these aspects. Any design process begins with the formulation of requirements. Or, when requirements are stated, with a critical review of those requirements, on one hand to check whether they are not unrealistically strict, and on the other hand to derive sub-system requirements that will aid the designer with lower-level design decisions. Some of the most important top-level requirements will be presented and briefly reviewed in Sect. 10.2.1. Next, a brief recap of the aerothermal environment will be provided, as well as the main components of the heat balance (Sect. 10.2.2).

10.2.1 TPS Requirements

The requirements for designing a thermal protection system can vary depending on the specific application and operating conditions, but as a bare minimum it should protect the vehicle and its payload against the harsh thermal environment and should preferably be weight-minimised. As we know from the elaborate discussions in Chap. 5, the total vehicle mass has an effect on the trajectory and its thermo-mechanical loads. However, without going into too much detail some more general considerations and requirements can be stated. They may include:

1. *Temperature range*: The TPS should be designed to withstand the expected temperature range of the operating environment, including both steady-state and transient conditions. The steady-state conditions will determine the maximum allowable heat flux and thus the maximum wall temperature, whereas the transient effects may give insight in the origin of thermal stresses in the structure. In case there is a change in temperature there will be a lag in the heat flow into the structure. Depending on, for instance, thickness variations of the structure, thermal stresses will occur when thinner material will heat up faster than thicker parts. Differences in the stresses will be more prominent the quicker the temperature changes. It is noted that when optimising an entry trajectory it is good practise to not only include the maximum heat flux as a path constraint, but also the gradient.
2. *Conductivity*: The TPS should have a low thermal conductivity to reduce the transfer of heat to the underlying structure or components. This will be especially useful when the peak heat flux is sharp and reduces again after not too long a time. With low conductivity, the structure would need "more time" to heat up. However, with a higher conductivity a structure may heat up more uniformly and in that case thermal stresses will be reduced.

3. *Emissivity*: The TPS should have a high emissivity to efficiently radiate heat away from the surface and prevent the buildup of heat. This will be discussed at length in the next section(s). With a maximum value of $\varepsilon = 1$ for a black body, materials with a high emissivity can be found in the range between 0.8 and 0.9. For materials with a lower emissivity, coatings can be applied to increase this value.
4. *Mass*: The TPS should be lightweight to minimise the impact on the overall mass and performance of the system. The relation between flight time and integrated heat load has been brought up at an earlier stage, in the sense that a minimum total integrated heat load will minimise the mass of the TPS. To achieve minimum flight duration it is advised to fly along the maximum heat-flux constraint for as long as possible. Obviously this will put a demand on the selected material, as it has to withstand a higher temperature in case the maximum flux is set to a high value.
5. *Durability*: The TPS should be designed to withstand the expected operating conditions and potential impacts or stresses that may be encountered during use. Even though durability is not synonymous with reusability, the two are, of course, linked. As an example we mention the ceramic tiles of the Space Shuttle TPS. They were durable when it came to surviving the thermal environment. But, they appeared to be very brittle when it came to the impact of, for instance, micro-meteorites or small particles of space debris. That meant that the overall durability was not as high as one had hoped for, and intense inspection and maintenance was required in between flights.
6. *Thermal expansion*: The TPS should be able to accommodate thermal expansion and contraction without losing its effectiveness or damaging the underlying structure. Differential expansion between parts of the TPS should be avoided, as it will induce potential large stress concentrations. A typical example to account for this is the use of corrugated panels, as discussed in the introduction about the X-20 and the SR-71 *Blackbird*. Using a wave pattern for the face plates allows for a differential expansion, like a harmonica. Another example of preparing sandwich panels with a different material for facings and core (and thus different expansion coefficients) will come up for discussion later.
7. *Ease of manufacturing*: The TPS should be designed for ease of manufacturing, installation, and maintenance. Some materials may be very hard to machine, and then to have complex (three-dimensional) geometries would not be a good idea. Also, the manual impregnation of the honeycomb cells, as was done for Apollo, should be avoided, and whenever possible, be automated.
8. *Cost-effectiveness*: The TPS should be cost-effective and within the budget constraints of the project. Of course, cost is always an issue. But it should not be at the expense of technical performance and safety. It is a good criterion in a trade-off study when multiple solutions to the problem exist, but when only a single choice can be made, then the cost are more or less fixed. However, when the cost seems to be too high, it may be necessary to go back to the

top-level mission and system requirements to see whether a reduced scope will allow for cost reduction.

9. *Compatibility*: The TPS should be compatible with the other materials and components of the system and not negatively impact their performance. An example may be the release of a parachute system. Depending on the location of different TPS elements, it could even be as simple as a flexible blanket, one should, for instance, consider the anchor points of the bridle lines. Ripping through the TPS blanket is not the most reliable way for deployment. Similarly, the release hatch from where the parachute pack will be ejected, should be properly sealed for thermal protection, but at the same time allow for a smooth opening.
10. *Regulatory compliance*: The TPS should comply with any relevant regulatory requirements or standards for safety, environmental impact, or performance. This does not only concern the environment, but also personnel involved. Working with hazardous (toxic) materials should be minimised, or better yet, avoided. However, whenever that is necessary, special precautions should be taken. The same would apply to tooling, which should always be operated in a safe manner.

10.2.2 Aerothermodynamics

10.2.2.1 General

Aerothermodynamics is an important field to provide input to the design process of a TPS. Once the vehicle enters the atmosphere at high speed, the large amount of kinetic energy has to be somehow absorbed to reduce the speed to acceptable levels for a safe landing. Typically, the only mechanism available is to convert this energy to heat, but it is not trivial to determine how much of the energy will, in the end, enter the vehicle. The design of the TPS has a lot to do with this, of course, and depends on design solutions and materials choices. However, also the physical flow effects play an important role. In Sect. 3.6, several aspects of aerodynamic heating have been discussed. To be able to put the TPS design in perspective and to understand the relation between the structural aspects and the flow, a brief recap of the most important elements of the earlier section will be given, extended with some aspects that are particularly relevant for the current discussion.

As a result of varying density of the (Earth's) atmosphere relative to the vehicle's altitude above the planetary surface, a vehicle entering the atmosphere may encounter fundamentally different flow regimes (Sect. 3.1.5 and Fig. 3.3). The vehicle will first encounter widely scattered particles in what is called the free-molecular flow regime. Energy flux in this regime is relatively low, because of the low density. When density increases, slowly a shockwave starts to form. Viscous forces in the flow become stronger and the corresponding velocity gradients behind the shock are limited to a region adjacent to the vehicle surface called a boundary layer. When the flow in

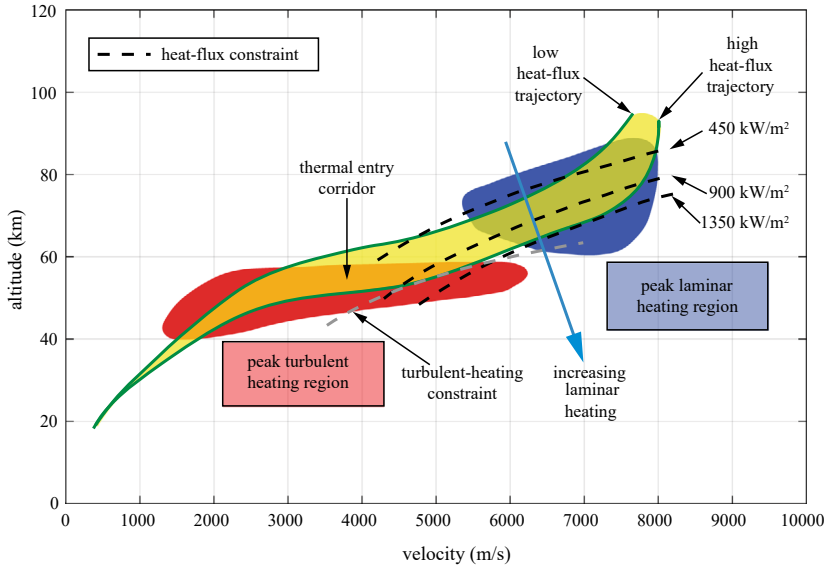


Fig. 10.18 Schematic areas of constant heat flux

the boundary layer is parallel to the vehicle surface the boundary layer is laminar, with layers at different velocities and temperatures. However, the laminar flow can become unstable and velocities normal to the surface appear that mixes the layers into turbulent flow. A turbulent boundary layer can have convective heat-transfer rates to the surface that are several times higher than the corresponding laminar boundary layer. Vehicle designers, therefore, will try and develop vehicles where this transition to turbulent flow is delayed as much as possible. If this cannot be guaranteed, though, the TPS designer needs to take the increased heat flux for a turbulent boundary layer as the nominal design load.

Figure 10.18 shows the "thermal entry corridor", in the form of maximum heat-flux constraints. These constraints are very much affected by the nature of the boundary layer, as indicated. Whereas the heat flux serves more as a local boundary, i.e., it should be considered in relation with the maximum allowable temperature of both the outer wall and the underlying structure, the duration of a certain flux acting on the vehicle wall will determine the total integrated heat load. For trajectories with a low heat-flux constraint, for instance, a metallic TPS solution can be found. For a higher heat flux—as long as the heat load is not too large—a ceramic solution could be chosen. For the combination of high heat flux and high integrated heat load, typically an ablative TPS is the only solution. Thus, one can also see a difference in entry trajectories: with the high-temperature capabilities and heat-rejection (i.e., insulative) nature of ceramic TPS, a steeper entry trajectory with short flight time (equals high heat flux and low heat load) is enforced, whereas for the metallic TPS a

low heat-flux, but higher heat-load trajectory is allowed to stay within the temperature limits. Once the boundary layer transitions from laminar to turbulent, a much higher heat flux can be expected. This shall be taken into account by the mission designers by adjusting the thermal path constraints, forcing the vehicle to fly a shallower trajectory. To include the heat flux as a path constraint has been discussed in detail in Sect. 5.4.3. There, the cold-wall Chapman equation was used to define an altitude-velocity profile, given a maximum heat flux, $q_{c,\max}$:

$$q_{c,\max} = c^* \frac{1}{R_N^n} \left(\frac{\rho}{\rho_0}\right)^{1-n} \left(\frac{V}{V_c}\right)^m \tag{10.1}$$

with $c^* = 1.1097 \cdot 10^8 \sqrt{\text{m}}$, $\rho_0 = 1.225\,\text{kg/m}^3$ and $V_c = 7{,}905.4\,\text{m/s}$. Commonly applied values for n are $n = 0.5$ for a laminar boundary layer, and $n = 0.2$ for a turbulent one; R_N is a characteristic radius (nose, leading edge, ...), and m is usually chosen to be 3 for an Earth entry. Given a velocity range, the above equation will provide the density, ρ, and through the atmosphere model, the altitude, h.

As an example, the case where a significant temperature rise was found due to transition from a laminar to a turbulent boundary layer is shown in Figs. 10.19 and 10.20. Edquist et al. (2009) did the aerothermodynamic design of the Mars Science Laboratory heat shield by doing Navier-Stokes flow-field simulations for a variety of entry trajectories. Part of their assessment was to study boundary-layer transition, for which they defined a criterion based on a critical laminar boundary-layer momentum-thickness Reynolds number,[11] conservatively defined as $Re_\theta > 200$.

Figure 10.19 shows that Re_θ exceeds the transition criterion well before—more than 20 s—the maximum heat flux occurs. This happens on the leeward side of the heat shield, well away from gthe stagnation point. The highest values of Re_θ is found 5 s after peak heat flux, on a location very close to the leeward shoulder of the heat shield, where the boundary-layer running length is longest. Checking the corresponding curves for maximum heat flux, shear stress and pressure in Fig. 10.20 confirms that this moment corresponds to the one for maximum shear stress. What this plot also shows is that the stagnation-point heat flux is much smaller than the actual peak value, obviously because the boundary layer is still laminar. This analysis led to the decision to always treat the boundary layer as turbulent to provide the input

[11] The momentum-thickness Reynolds number is given by

$$Re_\theta = \frac{\rho_e u_e \theta}{\mu_e}$$

where the momentum thickness of the boundary layer is defined as

$$\theta = \int_0^e \frac{\rho\mu}{\rho_e \mu_e} \left(1 - \frac{u}{u_e}\right) \mathrm{d}\eta$$

Here, the subscript e indicates the edge of the boundary layer. Further, ρ is the density, μ the dynamic viscosity, and u the flow speed.

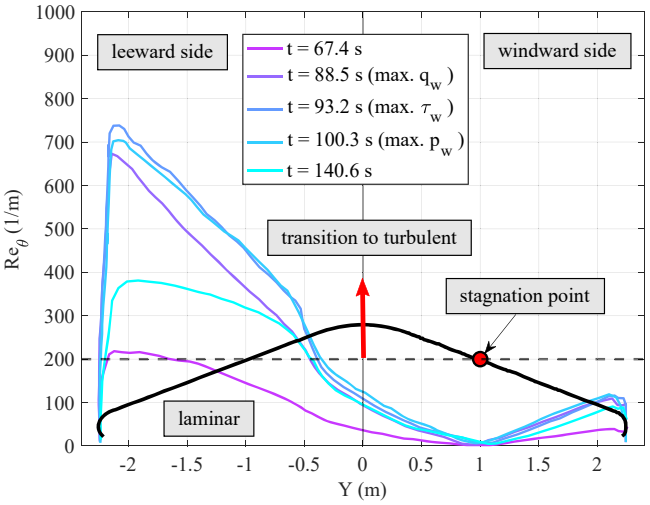


Fig. 10.19 Reynolds-number history and boundary-layer transition on the MSL heat shield. Data taken from Edquist et al. (2009)

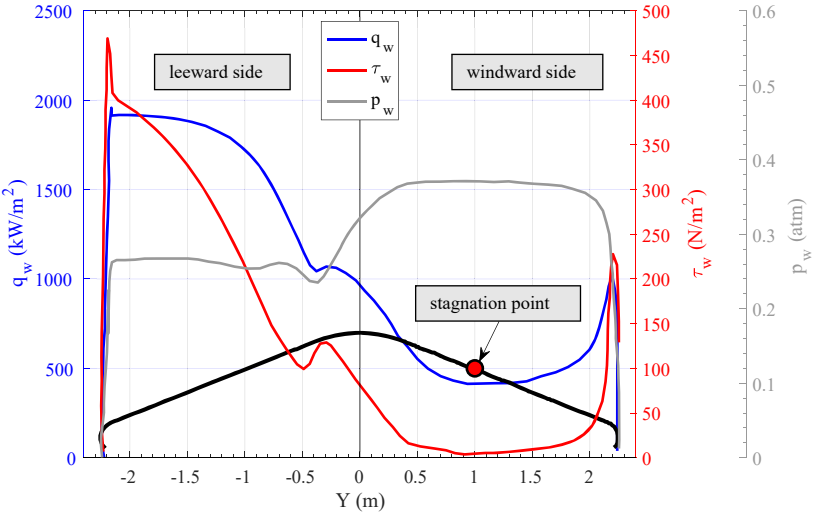


Fig. 10.20 Peak loads on the MSL heat shield. Data taken from Edquist et al. (2009)

for the TPS design. It is noted that MSL (as well as the more recent Mars 2020) was the first Mars probe for which this had to be done; for all other Mars mission the boundary layer stayed laminar.

This example shows that it is not trivial (and quite often also not correct, depending on the application, of course) to make assumptions about the heating environment. Often it is simply assumed that the boundary layer is laminar and that the highest heat flux occurs at the stagnation point. But it is not just the difference between laminar and turbulent boundary layers. While passing through the shockwave, an almost instantaneous increase of energy state of the gas particles occurs. There may also be additional chemical processes, such as dissociation and ionisation that might affect the thermodynamic behaviour of the gas dramatically. Not all of these changes are instantaneous, though; in case the reaction times are longer than the time required for the gas to flow past the vehicle, the gas will not reach an equilibrium state, and the flow is then known as frozen flow. Finally, under some conditions, there is a significant increase in radiation emitted by the gas in the non-equilibrium region. That means, that on top of a convective influx there will be a radiative influx to the vehicle wall, increasing the heat load on the TPS. Figure 3.83, based on the analysis by Gnoffo et al. (1999), shows this heat load for several entry missions. The relation between the entry speed and the magnitude of the heat load is evident: the higher the entry speed, the larger this load.

It will go far beyond the nature of this chapter to dive into all specifics of the aerothermal environment. The discussion will be focussing on the high-level heat balance that can help preparing input to some simplified TPS analysis methods.

There are multiple modes through which heat can be transferred *through* a medium or *between* different media; for the purpose of the current discussion, the focus will be on the most common modes, i.e., convection, radiation, and conduction. Even though the first two have been discussed at length in Sect. 3.6, a brief summary is given, because a good grasp of the different modes of heat transfer will be useful in setting up a thermal energy balance.

10.2.2.2 Convection

Convection is a form of heat transfer generally seen in fluids and gases: it is the physical movement of molecules that leads to an energy flow. There are two main types of convection, either the motion of the fluid is natural or it is forced. In general, it is difficult to calculate the convective heat transfer, since it depends on various factors, such as flow speed, surface condition, geometry and many more. Convection is generally defined in terms of a convective heat-transfer coefficient, α, that describes the heat transferred to or from a fluid per unit temperature difference between the fluid and the surrounding surface. The convective heat transfer is calculated with:

$$q_c = \alpha(T_{fluid} - T_w) \tag{10.2}$$

Other forms of this equation exist, e.g., Chapman's equation that has been extensively used in Chap. 5. This equation can be used to calculate the stagnation-point heat flux, where the wall can be considered being either hot or cold, the latter of which is given by Eq. (10.1). Specific details about this can be found in Sect. 3.6.

10.2.2.3 Radiation

Radiation heat transfer occurs via electromagnetic waves. Unlike conduction and convection, radiation does not need a medium for transmission, i.e., it also occurs in vacuum. For TPS applications the radiation is typically emitted by the (solid) structure, either outwards into the atmosphere, or inwards, depending on the structural layout. However, specifically related to re-entry, the hot gasses of the plasma around the vehicle also radiate heat to the vehicle wall, and should thus be considered an influx in the heat balance. Radiative heat flux (per unit area, so with a unit W/m^2) can be simply estimated by using the Stefan-Boltzmann law:

$$q_{rad} = \varepsilon \sigma T_w^4 \tag{10.3}$$

with $\sigma = 1.380649 \cdot 10^{-23}$ m^2 kg/(s^2 K) is the Stefan-Boltzmann constant, ε is the emissivity of the material, and T_w is the wall temperature (with wall being either the outer surface or any internal structural part). The emissivity is an important material property, as it is an indication of how efficiently a body absorbs or radiates heat. It is a dimensionless value between 0 and 1. A black body is one which radiates and absorbs heat perfectly, i.e., $\varepsilon = 1$. For re-entry applications, the emissivity of the outer wall should be as high as possible; the material and surface conditions are important aspects in determining the emissivity.

The emissivity of a body is equal to its absorptance, and in reality is a function of the wavelength. For wavelengths corresponding to the average temperature of a Lunar entry, the Apollo Command Module has $\epsilon \approx 0.8$ (Robinson et al. 2009). In theory, also the incoming radiative heat as discussed in Sect. 3.6.3 should be multiplied by an emissivity/absorptance; nevertheless, this radiation occurs at a variety of wavelengths. To be conservative, it is safer to let the corresponding emissivity/absorptance be equal to 1. When the heat-shield is ablative, a large portion of the heat is dissipated in the ablation process. Therefore, the wall temperature would, in theory, be smaller.

In case two or more radiative surfaces are facing each other, the above expression for radiative heat flux is modified by a factor known as the view factor or geometric form factor, indicated by F (Rohsenow et al. 1998; Janna 2009). It can be seen as a correction or scaling factor to account for the geometric view of one surface from the other (typical values can be found in the cited books and references therein). The relation between $F_{1,2}$ and $F_{2,1}$ is commonly given as $F_{1,2}A_1 = F_{2,1}A_2$ (the reciprocity theorem for view factors), where A_1 and A_2 are the areas of surface 1 and 2, respectively. This means that for two surfaces 1 and 2, $F_{1,2}$ can be different from $F_{2,1}$ in case the surface areas are not the same. It also implies that given the surface

areas and one view factor, one can calculate the other one. When no reflection of the radiation is considered the radiative heat flux between two surfaces with areas A_1 and A_2 can be derived as follows.

Independently from each other the two surfaces would radiate the following energy:

$$q_{rad,1}A_1 = \varepsilon_1 A_1 \sigma T_1^4 \quad \text{and} \quad q_{rad,2}A_2 = \varepsilon_2 A_2 \sigma T_2^4 \tag{10.4}$$

The energy radiated from area A_1 that reaches A_2 is then

$$q_{rad,1-2} = \sigma \varepsilon_1 T_1^4 A_1 F_{1,2} \tag{10.5}$$

and following the same logic, for A_2 to A_1

$$q_{rad,2-1} = \sigma \varepsilon_2 T_2^4 A_2 F_{2,1} \tag{10.6}$$

That means that the net radiative heat transfer, $\Delta q_{rad,1-2}$, is

$$\Delta q_{rad,1-2} = \sigma \varepsilon_1 T_1^4 A_1 F_{1,2} - \sigma \varepsilon_2 T_2^4 A_2 F_{2,1} \tag{10.7}$$

With the reciprocity equation substituted,

$$\Delta q_{rad,1-2} = \sigma \varepsilon_1 T_1^4 A_1 F_{1,2} - \sigma \varepsilon_2 T_2^4 A_1 F_{1,2} \tag{10.8}$$

or

$$\Delta q_{rad,1-2} = \sigma \left(\varepsilon_1 T_1^4 - \varepsilon_2 T_2^4\right) A_1 F_{1,2} \tag{10.9}$$

However, in a more realistic setting there may be reflections on one or both surfaces, i.e., part of the energy radiated by the one will be reflected by the other surface. In that case the body is considered a grey body.[12] To analyse this, consider the configuration of two parallel flat plates, as shown in Fig. 10.21. The flat plates are considered to be "two-dimensional", i.e., they are "infinitely" wide. The transmittance is considered to be zero, so only emittance (ε), absorption (α), and reflection (ρ) take place. Since the surfaces are considered grey, Kirchhoff's Law states that for an arbitrary body emitting and absorbing thermal radiation in thermodynamic equilibrium, the emissivity is equal to the absorptivity, so $\varepsilon = \alpha$, and also that $\alpha + \rho = 1$. They each have a certain temperature T_1 and T_2 and emit energy $E_1 = q_{rad,1}$ and $E_2 = q_{rad,2}$.

Surface 1 emits energy E_1, which is absorbed by Surface 2 as $\alpha_2 E_1$. It reflects $\rho_2 E_1 = (1 - \alpha_2)E_1$. The reflected energy reaches Surface 1, which absorbs

[12] A black body absorbs all colours of light and emits so-called black-body radiation (with $\varepsilon = 1$). In contrast, a white body is one with a rough surface that reflects all incident rays completely and uniformly in all directions. In between, there are grey bodies that partially reflect and partially absorb the radiation.

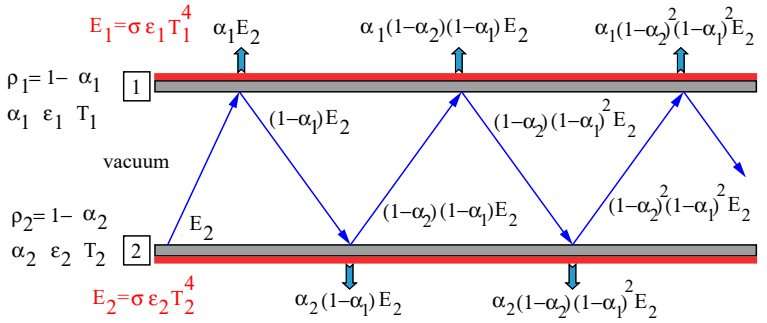


Fig. 10.21 Exchange of radiation energy between two parallel, flat grey surfaces. Here, the emittance is defined by ε, absorption by α, reflection by $\rho = 1 - \alpha$, and temperature by T

$\alpha_1(1-\alpha_2)E_1$ and following the same logic, it will reflect $(1-\alpha_1)(1-\alpha_2)E_1$. Continuing, Surface 2 will absorb $\alpha_2(1-\alpha_1)(1-\alpha_2)E_1$ of this reflected energy, and reflects $(1-\alpha_2)(1-\alpha_1)(1-\alpha_2)E_1$. This process continues indefinitely. A similar process can be followed for Surface 2. It begins with the emittance of E_2, of which $\alpha_1 E_2$ is absorbed by Surface 1. This surface reflects $(1-\alpha_1)E_2$, which reaches Surface 2, and $\alpha_2(1-\alpha_1)E_2$ is absorbed and $(1-\alpha_2)(1-\alpha_1)E_2$ is reflected. And so on.

All energy components E_1 that are absorbed by Surface 1 can be added, and the same can be done for Surface 2. So, the energy absorbed by Surface 1 is

$$E_{absorbed,1} = \alpha_1(1-\alpha_2)E_1 + \alpha_1(1-\alpha_1)(1-\alpha_2)E_1 + \cdots \tag{10.10}$$

Using the short-hand notation $a = (1-\alpha_1)(1-\alpha_2)$ the above absorbed energy can also be written as

$$E_{absorbed,1} = \alpha_1(1-\alpha_2)(1 + a + a^2 + \cdots)E_1 \tag{10.11}$$

Realising that $1 + a + a^2 + ...$ is, in fact, a Taylor series, i.e.,

$$\frac{1}{1-a} = (1-a)^{-1} = 1 + a + a^2 + \cdots \tag{10.12}$$

which means that $E_{absorbed,1}$ can be written as

$$E_{absorbed,1} = \frac{\alpha_1(1-\alpha_2)}{1-a}E_1 = \frac{\alpha_1(1-\alpha_2)}{1-(1-\alpha_1)(1-\alpha_2)}E_1 \tag{10.13}$$

Similarly, it follows for all energy components E_2 and absorbed there as well that

$$E_{absorbed,2} = \frac{\alpha_2(1-\alpha_1)}{1-a}E_2 = \frac{\alpha_2(1-\alpha_1)}{1-(1-\alpha_1)(1-\alpha_2)}E_2 \tag{10.14}$$

So, for Surface 2 that emitted E_2 but absorbed again $E_{absorbed,2}$, the net energy absorbed by surface 1 is thus the difference, i.e.,

$$\begin{aligned}\Delta E_{absorbed,1} = E_2 - E_{absorbed,2} = & \; E_2 - \frac{\alpha_2(1-\alpha_1)}{1-(1-\alpha_1)(1-\alpha_2)} E_2 \\ = & \; \frac{\alpha_1}{1-(1-\alpha_1)(1-\alpha_2)} E_2\end{aligned} \tag{10.15}$$

and similarly for surface 1:

$$\begin{aligned}\Delta E_{absorbed,1} = E_1 - E_{absorbed,1} = & \; E_1 - \frac{\alpha_1(1-\alpha_2)}{1-(1-\alpha_1)(1-\alpha_2)} E_1 \\ = & \; \frac{\alpha_2}{1-(1-\alpha_1)(1-\alpha_2)} E_1\end{aligned} \tag{10.16}$$

The net radiative heat flux from surface 1 to 2 is thus

$$\begin{aligned}q_{rad,1\to 2} = & \; E_1 - \frac{\alpha_1(1-\alpha_2)}{1-(1-\alpha_1)(1-\alpha_2)} E_1 - \frac{\alpha_2}{1-(1-\alpha_1)(1-\alpha_2)} E_2 \\ = & \; \frac{E_1 - (1-\alpha_1-\alpha_2+\alpha_1\alpha_2)E_1 - \alpha_1 E_1 + \alpha_1\alpha_2 E_1 - \alpha_2 E_2}{1-(1-\alpha_1-\alpha_2+\alpha_1\alpha_2)}\end{aligned} \tag{10.17}$$

which yields

$$q_{rad,1\to 2} = \frac{\alpha_2 E_1 - \alpha_1 E_2}{\alpha_1 + \alpha_2 - \alpha_1\alpha_2} \tag{10.18}$$

With $\alpha = \varepsilon$ and $E = q_{rad} = \varepsilon\sigma T^4$ the above equation can also be written as

$$q_{rad,1\to 2} = \frac{\varepsilon_2\varepsilon_1\sigma T_1^4 - \varepsilon_1\varepsilon_2\sigma T_2^4}{\varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2} \tag{10.19}$$

or

$$q_{rad,1\to 2} = \varepsilon_{1,2}\sigma(T_1^4 - T_2^4) \quad \text{with} \quad \varepsilon_{1,2} = \frac{\varepsilon_2\varepsilon_1}{\varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2} \tag{10.20}$$

In literature, the above equation is often rewritten as:

$$q_{rad,1\to 2} = \frac{\sigma(T_1^4 - T_2^4)}{\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2} - 1} \tag{10.21}$$

Of course, the above approach could be extended to multiple surfaces with finite areas and different view factors, but this is considered to be beyond the scope of the current discussion. The interested reader is referred to standard texts, such as the books by Rohsenow et al. (1998) or Siegel and Howell (1992). One particular method is mentioned, though. For grey bodies the radiative heat flux can also be calculated using the Gebhart method (Çengel and Ghajar 2015). By this method, the radiative heat flux from Surface 1 to Surface 2 is given by

$$\Delta q_{rad,1-2} = \sigma \varepsilon_1 A_1 B_{1,2} \left(T_1^4 - T_2^4\right) \tag{10.22}$$

with $B_{1,2}$ being the so-called Gebhart factor, defined as the fraction of radiation emitted by one surface to that absorbed by the other surface, including reflections from other k surfaces. This factor can be estimated by iteratively solving

$$B_{1,2} = F_{1,2}\varepsilon_2 + \sum_k F_{1,k}(1 - \varepsilon_k) B_{k,2} \tag{10.23}$$

It is noted that $\varepsilon_i A_i B_{ij} = \varepsilon_j A_j B_{ji}$ and $\sum_{j=1}^{n_s} B_{ij} = 1$, with n_s the number of surfaces in view of Surface i.

10.2.2.4 Conduction

The rate at which the heat flows from a hot region to a cold region is determined by means of the temperature difference between the two regions and the thermal conductivity of the material. Heat conduction can be estimated by means of the Fourier law, which states that the negative gradient of temperature and the time rate of heat transfer is proportional to the area at right angles of that gradient through which the heat flows:

$$\mathbf{q_{cond}} = -k \nabla T \tag{10.24}$$

Often, one considers conduction in a single direction, so

$$q_{cond} = -k \frac{\partial T}{\partial y} \tag{10.25}$$

with k being the material's thermal-conductivity coefficient in W/(m K) for a conductive flux in W/m^2. For the heat flux due to conduction through a material in one

dimension, a simple formulation can often be used. Assume a plate with thickness Δy and a temperature difference between Points 1 and 2 at either end of $\Delta T = T_1 - T_2$, Fourier's Law is rewritten as

$$q_{cond} = \frac{k}{\Delta y}(T_1 - T_2) \tag{10.26}$$

where the heat flows from the higher temperature to the lower one. This formulation will be used in Sect. 10.5.5, where an engineering model for TPS design is presented. Note that for a stacked TPS, and assuming that Layer 1 is on the outside, temperature $T_1 > T_2$. If the direction of y is positive from 1 to 2, it explains why the minus sign no longer appears in Eq. (10.26).

For the estimation of 2-D and 3-D heat transfer by conduction, complex models are required, which are beyond the scope of the material presented in this chapter. The interested reader is referred to the books by, for instance, Rohsenow et al. (1998), Janna (2009), and Çengel and Ghajar (2015) for more information.

10.3 Materials

Early in the space program, metals were the primary structural materials for both aircraft and spacecraft, as has been extensively discussed at the beginning of this chapter. It was initially also used as candidate material for TPS design, but the severe heat environment and associated challenges in TPS design necessitated the application and sometimes the development, of new materials. For the extreme heat environment of the early capsules, the use of metals was out of the question, as were most other materials. Single-use ablators were developed to accommodate the high heating rates, although for a lower-loaded parts of the vehicles other design solutions could be pursued, such as the metallic heat shields that were used on the leeward surfaces of both the Mercury and Gemini re-entry capsules. With the advent of winged vehicles, the encountered heating rates were substantially lower, and combinations of ceramic materials and metals could be considered for the new vehicle's TPS. Also in the category of ablative materials substantial development was made, and it is safe to say that current-day ablative heat shield are definitely not the same as they were in the 1960s.

Before we move on to provide an overview of materials that are typically used in heat-shield design, as well as to address thermal characteristics that are key input to the TPS design process, a simple example to show the severity of the re-entry heat environment is in place. To put the heat that needs to be absorbed during a re-entry into a "worldly" perspective, the example of evaporating a block of 1 kg solid iron will be considered.[13] Subsequently, the transition of a solid to a liquid to a gas will be addressed. The specific heat capacity of iron (Fe) is 450 J/(kg K) at room

[13] All related properties of iron that are used in this example can be found at https://webbook.nist.gov/cgi/inchi/InChI%3D1S/Fe. Date accessed: March 3, 2023.

temperature.[14] So, to heat up iron from room temperature (20 °C) to its melting point ($T_{melt} = 1536\,°C$ requires $\Delta Q_{heating,1} = 682$ kJ of energy. Then, to fully melt the iron at this temperature (a process called fusion), will take $\Delta Q_{melting} = 33$ kJ/kg (at constant pressure).[15] After all iron has melted, more energy is added to boil the iron (up to a temperature of $T_{boil} = 2{,}861\,°C$), which requires another $\Delta Q_{heating,2} = 1193$ kJ. Note that any specific heat capacity is temperature dependent; for high temperatures, the specific heat capacity for iron is at least twice the value of that at room temperature. Without focussing too much on the actual value (which is also not constant during the process of raising the temperature), an average value of 900 J/(kg K) has been used. To evaporate boiling iron (at $T = 2{,}861\,°C$), i.e., to go from a liquid to a gaseous state, requires 6088 kJ/kg (the value commonly provided for the so-called enthalpy of vaporisation, also known as the (latent) heat of vaporisation or heat of evaporation, is 340 kJ/mol, with 1 kg/mole = 17.90670606142). So, to sum up, to evaporate one 1 kg of iron requires

$$Q = \Delta Q_{heating,1} + \Delta Q_{melting} + \Delta Q_{heating,2} + \Delta Q_{evaporisation}$$

is 682 + 33 + 1193 + 6,088 = 7,996 kJ, so about 8 MJ/kg. This is only 25% of the energy content of a re-entry vehicle, and no matter which flow conditions were neglected, it may be obvious that some form of thermal protection is required to have the vehicle survive the harsh entry environment.

Of course, it is not just the heat capacity of the material that is important. Many aspects come into play when designing hot structures, but the same (or similar) aspects would hold for the design of cold structures as well. In the next few sections, potential candidates for TPS materials will be introduced, including their particular properties. To begin with, in Sect. 10.3.1 an overview of relevant materials will be provided. Then in Sect. 10.3.2 several of the key material characteristics will be discussed in more detail. Primary focus will be on the thermal properties of the materials, as these are more relevant for the thermal analysis of TPS covered later in this chapter.

[14] The specific heat capacity of water is 4184 J/(kg K) at 20 °C, which makes water the most suitable substance to store heat. This is a property that makes water an excellent candidate for an actively cooled thermal protection system, as will be discussed later on in this chapter.

[15] Note that all numbers provided in this example will change with the slightest change in composition, or deviation from purity. For instance, white cast iron, a type of carbon-iron alloy that contains carbon content greater than 2% in the form of cementite, will require a fusion heat of 138 kJ/kg, rather than the 33 kJ/kg for (pure) iron.

10.3.1 Material Types

From a technical point of view, the materials used to design a TPS with good functionality should have a low thermal conductivity, or, in other words, a high thermal resistance, high emissivity to reject or radiate out as much heat as possible, and have a good thermal stability (refractoriness). To be more specific, a refractory material is one that will (almost) not decompose when subjected to severe conditions of heat, pressure, or chemical attack, and it will keep its strength and form at high temperatures. There are other aspects to consider as well, of course, such as the material's density, which will determine the total TPS mass, its cost, manufacturability, and so on. An overview of these requirements has been given earlier in this chapter, Sect. 10.2.1. What is also important to realise is that given the fact that the flow field around an entry vehicle is not the same everywhere, there will be a strong variation in thermal loading, which allows for the selection of different materials for different areas of the vehicle. The design of a TPS is not simply a local optimisation problem for the hottest regions, but a global problem with several localised optima.

10.3.1.1 Ceramic Composites

Ceramic composites are important materials for re-entry technology, as they can withstand higher temperatures and thermal stresses encountered during atmospheric entry than other (reusable) materials. Their role is crucial in the design and performance of re-entry vehicles, by providing lightweight, durable, and thermally resistant materials that can withstand the extreme conditions at critical locations of the vehicle. Some of the most important ceramic composites used in re-entry technology are:

1. Carbon/carbon composites: These are composite materials made of carbon fibres in a carbon matrix, which are known for their high strength, stiffness, and thermal resistance. As is the case for all composites, the purpose is to combine the advantage of high specific strength and stiffness of the fibres with the refractory properties of the matrix. The fibres are laid in near-net shapes with multi-directional reinforcements, resulting in an ideal high temperature structure with high specific thermal conductivity and low thermal expansion (Devi and Rao 1993). Carbon/carbon composites are used in applications such as parts of the TPS, rocket nozzles, and aircraft brake discs. They are sometimes good alternatives for some of the applications currently using a heavier metallic material. At higher temperatures it is prone to oxidation, so its use should be combined with protective coatings. Also, its manufacturability is a slow process that should be taken into account.
2. Silicon carbide (SiC) composites: SiC is a ceramic material that is known for its high temperature strength, oxidation resistance, and thermal shock resistance. Overall, it has a high strength-to-weight ratio. It is a particular type of ceramic matrix composite (CMC), with ceramic fibres or particles that lie in a silicon-carbide matrix phase. Since SiC-SiC composites have a relatively high thermal

conductivity and can operate at very high temperatures due to their inherently high creep and oxidation resistance, they are very suitable as TPS material for wing leading edges and other parts of heat shields, and rocket nozzles.
3. Carbon/silicon carbide composites: These are composite materials that combine the properties of carbon and SiC, such as high strength, stiffness, and thermal conductivity. They are used in applications such as thermal protection systems, nose cones, and wing leading edges.
4. Zirconia-based composites: Zirconia (ZrO_2) is a ceramic material that is known for its high thermal conductivity, low thermal expansion, and chemical stability. Besides heat shields, Zirconia-based composites are used in applications such as insulators and bearings.

10.3.1.2 Metallic Alloys

Metallic alloys are also important materials for re-entry technology, as they can provide high strength and toughness in addition to good thermal and electrical conductivity. The following are some of the most commonly-used metallic alloys:

1. Titanium-based alloys: Titanium alloys are lightweight, strong, and corrosion-resistant, and have seen many aerospace applications. For instance, Ti-6Al-4V is a widely used alloy for its high strength-to-weight ratio and high operating temperature properties, although it has shown to ignite in oxygen-rich and nitrogen-rich test streams. When introduced to the monatomic oxygen rich and high enthalpy re-entry environment, X-link components from the Space Shuttle *Columbia* that were made out of Ti-6Al-4V, showed an intensified oxidation and combustion behaviour. This has been studied in detail by Buckner (2016), and although the mechanism is well understood, no ready-made solution is available to prevent this. However, titanium alloys remain to be used in several applications, such as low-to-medium loaded heat shields, structural components, and rocket motor cases.
2. Aluminum-based alloys: Aluminum alloys are lightweight, easily fabricated, and have good thermal and electrical conductivity. They find similar fields of application as those of titanium alloys, albeit at lower temperatures. Due to its light weight, it is a commonly applied material in satellites.
3. Intermetallics: Gamma-titanium aluminide (γ-TiAl) is a relatively new material that is applicable at temperatures up to some 850–900 °C (Fischer et al. 2003). Much of its development was done in the United States, in the 1980s, in the framework of the National AeroSpace Plane (NASP) programme, with its X-30 demonstrator. Continuing development has made the alloy competitive, with a significantly higher specific yield strength than nickel-based alloys. Tests indicated a very low catalycity and an emissivity of about 0.8. Its good characteristics and high specific strength have made γ-TiAl a very attractive material in the temperature range in between 250 and 800–950 °C. It can thus be applied to a large part of the surface of many concept reusable launch vehicles.

4. Nickel-based superalloys: These are high-strength alloys that are resistant to high-temperature deformation, corrosion, and oxidation. They are used in applications such as turbine blades, engine components, and heat exchangers. The Inconel-family of nickel-chromium-based alloys are well-known for their aerospace applications, as they are oxidation and corrosion resistant, and well-suited for service in extreme environments subjected to high mechanical loads. With increasing temperature, Inconel forms a thick and stable passivating oxide layer protecting the surface from further exposure. Compared to aluminium and steel, which would give in to creep as a result of thermally-induced crystal vacancies, Inconel retains its strength over a wide temperature range. This high-temperature strength is developed by solid-solution strengthening or precipitation strengthening, depending on the alloy.
5. Refractory metal alloys: Refractory metals, such as molybdenum, tungsten, niobium (also called columbium), and tantalum, have very high melting points and can thus be used in environments with high temperatures, but they are also very heavy, too sensitive to oxidation, and due to hardness or abrasiveness they present difficulty in maintaining close tolerances. Refractory metal alloys can be used for specific parts of heat shields, but also in nozzles, and other re-entry vehicle components. But, due to their manufacturing issues and large weight, they are thus not suitable for TPS covering large areas of a vehicle.
6. Copper-based alloys: Copper alloys are used in re-entry technology for their high thermal conductivity and electrical conductivity. They are used in applications, such as heat shields (but then more as thermal shields in a cryogenic environment), thermal management systems, and electrical connectors. Due to their high thermal conductivity, it is possible to achieve a low temperature drop across a thermal shield with a relatively thin shell, which favours them over aluminium, except when weight reduction is of primary importance.

10.3.1.3 Oxide-Dispersion-Strengthened Materials

Oxide-Dispersion-Strengthened (ODS) material is a type of advanced material that is designed to have superior mechanical and high-temperature properties. It is a composite material that consists of a metallic matrix (usually a high-temperature alloy, such as nickel-based or iron-based alloys), and a dispersion of small oxide particles (typically yttria or zirconia) that are uniformly distributed throughout the matrix. The oxide particles act as a strengthening agent by inhibiting the movement of dislocations in the metallic matrix, thereby increasing its strength, hardness, and creep resistance. ODS materials also exhibit good thermal stability, and resistance to radiation damage and high-temperature oxidation, making them suitable for use in a variety of high-temperature and high-radiation environments encountered during re-entry, as well as gas turbines and nuclear reactors.

The fabrication of ODS materials involves a complex process of powder metallurgy, which includes mechanical alloying, powder consolidation, and sintering. The resulting material has a fine grain structure and a high density, which contributes to

its superior mechanical properties. Due to their unique properties, ODS materials are being actively researched and developed for use in advanced energy systems, such as next-generation nuclear reactors and fusion reactors, as well as in high-temperature components for gas turbines and other industrial applications.

The following are some of the most important ODS alloys used in re-entry technology, as well as applications, such as combustor liners, heat exchangers, and turbine blades. The materials can be roughly divided into two main groups of alloys, i.e.,

1. FeCrAlY alloys: FeCrAlY alloys are iron-based ODS alloys that are known for their high-temperature strength, corrosion resistance, and oxidation resistance.
2. NiCrAlY alloys: NiCrAlY alloys are nickel-based ODS alloys that are similar to FeCrAlY alloys, but have higher thermal stability and oxidation resistance.

with individual materials given by

1. INCOLOY® alloy MA954: MA954 is a nickel-chromium superalloy, produced by mechanical alloying. Its yttrium oxide dispersoid provides exceptional high-temperature strength and creep resistance to the highly corrosion-resistant nickel-chromium matrix. Its base composition is ≈78.0% Ni, ≈20.0% Cr, ≤1.0% Fe, ≤0.3% Al, ≤0.5% Ti, ≤0.05% C and ≈0.6% Y_2O_3 (yttrium oxide), but it is noted that on the manufacturer's website the element percentages are given as a range.[16] Dedicated testing is thus always required to establish the exact thermo-mechanical properties. In terms of composition and thermo-mechanical properties, MA754 is similar to the no-longer produced PM-1000 material (75.6% Ni, 20.0% Cr, 3.0% Fe, 0.3% Al, 0.5% Ti, and 0.6% Y_2O_3).
2. INCOLOY® alloy MA956: MA956 is an iron-based ODS alloy that contains yttria particles in a Fe-Cr-Al matrix. It is known for its high-temperature strength, thermal stability, and radiation resistance. The alloy has excellent oxidation resistance up to 1,300 °C and significant creep strength up to its unusually high melting point of 1,482 °C. In terms of composition and thermo-mechanical properties, it is similar to PM-2000.
3. PM-2000 alloy: as mentioned, also PM-2000 is an iron-based material, with a composition of 74.5% Fe, 19.0% Cr, 5.5% Al, with 0.5% Ti and 0.5% Y_2O_3.

Overall, ODS alloys play an important role in the design and performance of re-entry vehicles, by providing materials that can withstand the extreme conditions encountered during atmospheric entry. By tailoring the composition, microstructure, and processing of ODS alloys, researchers can develop new strategies for enhancing the efficiency, reliability, and safety of re-entry technologies. ODS materials have been considered as TPS for the Space Shuttle as well. For instance, Johnson and Killpatrick (1976) have evaluated the application of a dispersion strengthened nickel-base alloy as heat-shield material. Included in their investigation were cyclic load, temperature, and pressure effects on sheet material residual strength, the effects of

[16] see, for instance, https://www.specialmetals.com/ and https://www.aeether.com/. Date accessed: May 7, 2023.

braze reinforcement in improving the efficiency of spot-welded joints, and radiative properties. Besides material-sample tests, also full-scale heat-shield panels were developed to test, amongst others, meteoroid impact and a simulated entry heating environment. Despite excellent technical performance, the final TPS choice was based on more than just that: cost, mass, reusability, manufacturability, but also in the course of the program the mission parameters changed. However, after many flights of the Shuttle and careful review of the TPS's performance it was concluded that thermal protection tiles with a metallic outer protecting casing would be most promising, which was the main reason that this new technology was applied in the conceptual design of the X-33 (see Fig. 10.14).

From the early 1990s to the mid 2000s, a dedicated hot-structures programme was run in The Netherlands by the AEOLUS consortium. A dedicated materials program was part of that, centred around the ODS materials PM-1000 and PM-2000. Mooij et al. (2006a, b) document some of the achievements, both in terms of material testing, manufacturing technologies, sub-scale and full-scale samples, and final products. The materials' database included static mechanical properties, fatigue and fracture mechanics, thermal properties, chemical stability, and corrosion resistance, each for a range of temperatures with a maximum as high as 1,200 °C. The mastered manufacturing technologies for ODS TPS tiles included cutting and machining (including inserts, bolts/rivets manufacturing), bending and forming, high-temperature brazing, laser welding, non-destructive inspection, and pre-oxidation. The samples produced were web clips, corrugated webs, honeycomb cut, brazed sandwich panels with inserts and edge members (sub-scale and full-scale). Finally, the rudders of the X-38, the demonstrator in anticipation of the development of the Crew Return Vehicle, were originally designed to be hot structures, i.e., made out of metal (PM-1000) the structure was meant to endure both thermal and structural loads. The same applied to the outer skin of the European Experimental re-entry Testbed (EXPERT), a re-entry vehicle developed by ESA to do aerothermodynamic research on boundary-layer transition, catalycity and oxidation, real-gas effects on shock-wave boundary-layer interactions, and micro-aerothermodynamics (Muylaert et al. 2003). Unfortunately, despite being built, the vehicle was never launched due to lack of funding for the launch and operations.

10.3.1.4 Insulation Materials

A TPS system does not only consist of an outer wall, which rejects all heat coming to the vehicle. A proper TPS design will be stacked in nature, having as a minimum an outer wall and some form of internal insulation that adds an extra layer of protection for the inner structure. Therefore, insulation materials are important in re-entry technology. Even though in principle any TPS material could be called an insulator, some of them, for instance ablative materials, will be covered elsewhere. Some of the more important insulation materials used in re-entry technology are:

1. Ceramic tiles, which are lightweight and low-cost, are highly efficient insulators. They are made from materials such as silica, alumina, and zirconia, and are/were widely used on, for instance, the Space Shuttle.
2. Flexible blankets, made from materials such as silica, alumina, and carbon fibres, and are not just used in applications such as TPSs, but are also used as insulation blankets, and flexible joints.
3. Aerogels are highly porous materials with low thermal conductivity, making them excellent insulators. They are made from materials such as silica, alumina, and carbon, and besides being used in TPS, they can also serve as general thermal insulation and acoustic insulation.
4. Foams made from materials such as polyurethane, polystyrene, and silicone are lightweight, low-cost, and have good thermal insulation properties. They may be used as thermal, structural, and acoustic insulation.

Typical examples of complete insulation systems are the multi-layered insulation (MLI) used as a protective system on satellites—it also happens to be the most effective insulation design for cryogenic systems—as well as the reusable surface insulation (RSI), as used on, for instance, the Space Shuttle Orbiter. MLIs have known a long history starting from the 1950s. Figure 10.22 shows how multiple foils made of Kapton or Mylar coated with a highly reflective metal are placed parallel to each other, with low thermal conductive spacers arranged in between the foils to avoid direct contact with the foils. The significant modes of heat transfer are radiation, solid conduction and gas conduction. The principle is the same as has been discussed in Sect. 10.2.2, most notably the part on radiation. In summary, the external radiation heat hits the first reflective layer where a part of it is reflected back to the environment. The remaining radiation energy heats up the first layer of spacer. The temperature of this layer will increase, and thus solid conduction, gas conduction and radiation takes place through the spacer material to the next foil. There, the process will repeat

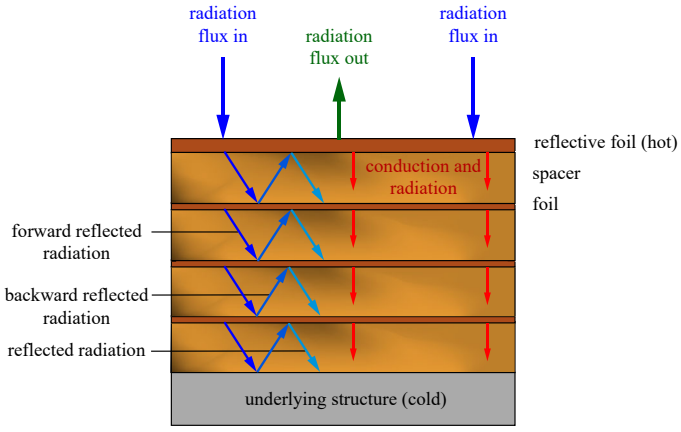


Fig. 10.22 Schematic layout of multi-layered insulation

itself: the second foil temperature will increase, some radiation will be reflected back through the spacer to the first foil, whereas the remaining energy transfers to the third foil. This process continues down to the bottom layer.

An overview of the development and characteristics of MLI is given by Sutheesh and Chollackal (2018). They also list some of the materials used by other researchers, albeit only for satellite applications: double-face aluminum-coated terylene film with fibre paper spacer, aluminised Mylar foil with "vegetal" super-thin fibrous spacer, double aluminised Mylar with polyester spacer, and aluminium single-layer foil. A much more extensive guideline on selecting materials and adhesives, finishing techniques, and the impact of the space environment is provided by Finckenor and Dooling (1999). For re-entry applications no materials are mentioned, but it would all depend on the aerothermodynamic environment, and, therefore, many combinations of facing, foil, and spacer materials are possible. As an example, Fig. 10.16 shows the MLI being applied to the Huygens heat shield, although this served primarily as protection during the interplanetary and coast phases. The MLI was made out of Kapton sheets with an aluminium coating on the back side. Kapton is a polyimide film that remains stable at temperatures as low as −269 up to 400 °C. It has excellent thermal properties to radiate heat away from the surface.

Apart from application on satellites, MLI has also been proposed for a planetary probe entering Titan's atmosphere and an Earth based space transportation system (Keller et al. 1992). They discuss the design of a stacked system with reflective screens and ceramic spacers of a few millimeters thick; depending on the application (and thus thermal environment), their solution could include up to 200 layers. For both systems they achieved minimum insulation mass, by optimising the layer's thickness using mostly analytical models. Similar results were obtained by Daryabeigi (2000, 2001), who used numerical models and confirmed the results with a test setup over a temperature range from 20 to 1,000 °C. A comparison was made between a metal MLI (consisting of gold-coated foils with fibrous insulation spacers made of high purity polycrystalline alumina in between) and a fibrous insulation sample with the same thickness (but which required a higher density such that the temperature of the modelled vehicle's inner structure would not be exceeded). The research showed that insulation mass-per-unit-area savings of 22% may be achieved with a 76.2 mm thick MLI using 16 foils compared to the all-fibrous solution. Kumar and Mahulikar (2016) developed a model of an ablative system with an underlying insulation layer to evaluate the selection of materials and design of a multilayer TPS, and provide a list with insulation materials. These materials and some of their properties are listed in Table 10.3.

A variation on the MLI concept is the water-cooled approach that will be discussed in Sect. 10.4.4. The TPS consists of an outer skin, an air gap, and a layer with porous material that holds a certain amount of water. Even though those concepts will only be discussed later, it is important that the porous material to be used will be introduced here. Buursink (2005) considered several options, based on a set of essential characteristics. The material should

Table 10.3 Thermophysical properties of insulation materials (Kumar and Mahulikar 2016)

Insulation material	Density (kg/m^3)	Specific heat (J/kg K)	Max. service temp. (K)	Thermal conductivity (W/m K)
Saffil	50	942–1340	1273–1975	0.036–0.544
Q-fibre	56	841–1190	1533	0.030–0.242
Carbon foam	51	1210–1310	3650–3800	0.04–0.11
Graphite foam	70	750–800	2850–2960	0.07–0.20
Glass-wool	26	1000	873	0.03256

1. be able to contain a large amount of water,
2. and retain this water under high g-forces;
3. be able to resupply water to the surface by capillary action,
4. have a high emissivity, for good radiation coupling,
5. have a low density,
6. have low irreducible saturation, and
7. be chemically inert under operational conditions.

It was found that the most suitable, porous material is an alumina insulation material, with the trade name ZAL-15.[17] This material consists of 85% Al_2O_3 and 15% SiO_2, with a porosity of 93% and a density of 24 kg/m^3. When dry, it has an emissivity of $\varepsilon = 0.81$, which increases to $\varepsilon = 0.91$ when wet. The coolant of choice is water, and the great advantage of ZAL-15 is that it can retain a large amount of water: its porosity is 93%, which is the theoretical maximum volume percentage of water it can contain. In reality, this may be slightly lower, as small bubbles of air may remain in the material, lowering the actual saturation level. Figure 10.23 shows a sample of ZAL-15.

Apart from the reinforced carbon-carbon nose cap and wing leading edges, the high-temperature reusable surface insulation made of coated LI-900 silica ceramics, and the fibrous refractory composite insulation tiles (see Fig. 10.10), for the areas with lower temperatures so-called thermal blankets were used as insulator. The reusable surface insulation on the Space Shuttle Orbiter came in two versions, i.e., the flexible insulation blankets (FIBs), a quilted, flexible blanket-like surface insulation, and felt reusable surface insulation (FRSI). The FRSI was used where temperatures stayed below 371 °C, and was found, amongst others, on the aft fuselage sides and a portion of the OMS/RCS pods, see Fig. 10.24.

The FRSI blankets were originally made from a Nomex felt (a so-called meta-aramid felt). A white silicone elastomer coating provided the required thermo-optical properties and the water-proofing. Using a silicone-based adhesive, the blankets were bonded directly to the aluminium airframe. During a major update of the TPS, a majority of the low-temperature tiles were replaced by FIBs (in full known as Advanced Flexible Reusable Surface Insulation, AFRSI). These were blankets with

[17] https://zircarceramics.com/product/zal-15/. Date accessed: May 15, 2023.



Fig. 10.23 Sample of the porous fibre material ZAL-15. Image: author

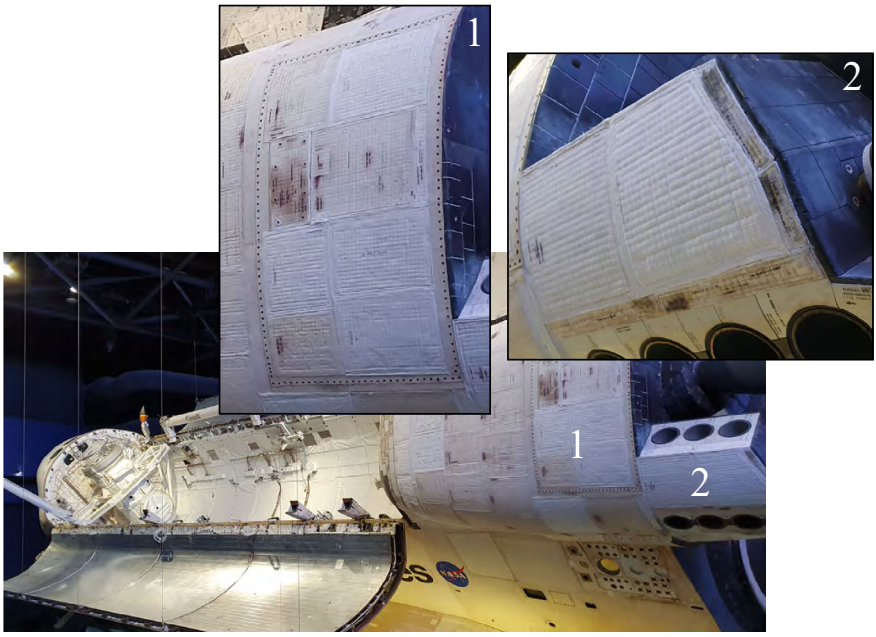


Fig. 10.24 Sideview of Space Shuttle *Atlantis*, on display at Kennedy Space Center: the use of Nomex Felt Reusable Surface Insulation (FRSI) at the back and a portion of the OMS/RCS pods is clearly visible (see inserts). Images: author

a maximum size of 76.2 × 76.2 cm (30 × 30 in.), and consisted of layered, pure silica felt sandwiched between a layer of silica fabric on the outer, hot side, and a layer of S-glass fabric on the other. The blankets were through-stitched with pure silica thread in a 2.54 cm (1 in.) grid pattern, treated to provide water repellency, and coated with a ceramic coating to provide durability and resistance to damage. The AFRSI was of low density, with a value of 176 kg/m^3 (Curry 1993).

Another felt type is the Zirconia felt (main constituent zirconium oxide, ZrO_2), also known as zirconia ceramic fibre felts or zirconia ceramic fibre insulation. They have excellent thermal resistance and low thermal conductivity and work by limiting heat conduction and providing a thermal barrier. They can be used at temperatures up to 1,000 °C. Further properties of ZrO_2 are its chemical inertness, resistance to molten metals, wear resistance, a high fracture toughness, and high hardness.[18]

10.3.1.5 Coatings

When the (surface) properties of materials are not completely meeting all design requirements, coatings can be applied to fill in the missing parts. Coatings are used in re-entry technology for a variety of reasons, including:

1. Thermal protection, where coatings can provide an additional layer of protection to the re-entry vehicle, reducing heat transfer to its interior and protecting the vehicle and its occupants from high temperatures,
2. Protection against corrosion, caused by exposure to either the atomic-oxygen environment of low Earth orbit, or the atmosphere during entry,
3. Ablation resistance, where the coatings either resist ablation or delay the process,
4. Radiation shielding, where the coatings can protect the vehicle and its occupants from the harmful effects of ionising radiation encountered during entry, and
5. Surface stability, with the coating preventing surface erosion, especially due to hot atomic oxygen.

Selecting the right coating is not a trivial matter, and sometimes a very dedicated one has to be "designed". Nonetheless, there are many coatings available. On their website, Aremco[19] provide a list with coatings that they manufacture, for instance, ceramic-based, black-pigmented coating for ceramic fibre modules, light-weight refractory board, and dense refractories, to be used either up to 1,093 °C or up to 1,316 °C, and a variation just for dense refractories and refractory metals up to 1,093 °C. Another series of ceramic-based, black pigmented coatings can be used for carbon and stainless steel, either up to 1,093 °C or up to 1,316 °C, and yet another, a silicone-ceramic, black pigmented coating to be used on aluminium, copper, carbon and stainless steel up to 593 °C.

[18] https://precision-ceramics.com/materials/zirconia/. Date accessed: May 16, 2023.

[19] https://www.aremco.com/. Date accessed: May 14, 2023.

For their application, Groh et al. (1994) required a material with a high emittance and a surface that was durable in an elevated-temperature, high-vacuum environment for an extended period of time. As base material, Haynes 188 was selected and 35 samples were exposed to 14 different types of surface-modification techniques. Out of these 14, seven were found to meet with the durability requirement, but two of them stood out: alumina-titania (AlTi) coatings ($\varepsilon = 0.85$ after 2695 h of heat-treatment in vacuum) and zirconia titania-yttria coatings ($\varepsilon = 0.86$ after 2024.3 h). This example shows the relevance for dedicated testing, once a specific application has been identified.

A particular category of coatings include the self-healing chromia, silica, and alumina scales, a type of protective coating used in high-temperature applications, such as in gas-turbine engines, power generation, and aerospace. Chromia (Cr_2O_3), silica (SiO_2) and alumina (Al_2O_3) are all oxide materials that form naturally on the surface of certain metals, such as stainless steels, superalloys, and intermetallics (MX, where M is Ti, Fe, Co or Ni, and X denotes Al, Si, or Cr), when they are exposed to high temperatures in air (Sloof 2007). These oxides form a protective layer, or scale, on the metal surface, which helps to prevent further oxidation and degradation of the material. In self-healing chromia and alumina scales, the oxide layer is engineered to have certain properties that allow it to repair itself when damaged. For example, the coating may contain microcapsules filled with a healing agent, which can be released to repair cracks or other damage in the oxide layer. Alternatively, the coating may contain particles that can migrate to areas of damage and help to repair the oxide layer. Self-healing chromia, silica, and alumina scales are important, because they help to improve the durability and longevity of high-temperature materials. By providing a self-repairing layer of protection, these coatings can help to reduce maintenance costs, increase equipment reliability, and extend the life of critical components.

A specific example of additional protection can be found in the TPS of the Space Shuttle. Johnson (2016) presents the combination of a coating and surface treatment. The coating, reaction cured glass (RCG), is a black coating consisting of tetraboronsilicide and low porosity borosilicate glass. It was applied to the Orbiter's top and sides to protect the porous silica. On silica-based tiles, RCG is very effective at temperatures up to 1,920 °C. A modified version, RCG-M, has an even higher temperature capability with a maximum of just above 2,000 °C. Its application was widespread, as it was applied to most Shuttle tiles, as well as many of the X-37b tiles, currently in operation.

As surface treatment, the earlier introduced Toughened Unipiece Fibrous Insulation (TUFI) is mentioned, which consists of borosilicate glass ($B_2O_3.SiO_2$), siliconboride (BxSi), and molybdenum disilicide ($MoSi_2$), which produces a stronger, tougher silica tile. The standard version of TUFI, the black variant, was used on the tiles located at the Shuttle Orbiter's underside. White variants with higher impact resistance and conductivity were used on the upper body. A second surface treatment that was developed from TUFI and has been flown on several X-37b missions is the High Efficiency Tantalum-based Composite, or HETC, which includes tantalum disilicide ($TaSi_2$) besides the standard TUFI components. HETC was designed

to operate at higher temperatures than TUFI and to accommodate higher thermal expansion differences between the substrate and coating.

Building on the developments described above, a more recent development for the X-37 is that of Toughened Unipiece Fibrous Reusable Oxidation Resistant Ceramic (TUFROC), which is lightweight, has a high total hemispherical emissivity of $\varepsilon = 0.9$, and a low catalytic efficiency.

10.3.1.6 Ablative Materials

Ablative materials are crucial for highly thermally loaded re-entry systems, and are mostly the only possible solution in those situations. The applicable materials are designed to sacrificially ablate, or erode, in response to the extreme temperatures, forming a protective charred or gas layer that insulates the underlying structure from further heat transfer. Ablative materials typically consist of a resinous matrix embedded with reinforcing fibres or particles, such as carbon or ceramic. The moment the vehicle encounters the extreme heat, the ablative material undergoes a controlled degradation, absorbing and dissipating the thermal energy through a combination of phase change, chemical decomposition, and gas evolution.

Quite a few ablative materials have already been mentioned in previous sections, but nonetheless they are summarised here.

1. Phenolic impregnated carbon ablator (PICA) is a composite material consisting of carbon fibres and phenolic resin, with a much lower density than other carbon/phenolic systems (i.e., about 270 kg/m^3). Moreover, it has an efficient ablative capability at high heat fluxes, confirmed by the qualification heat flux of 24 MW/m^2. When it is highly thermally loaded, PICA forms a protective char layer that helps insulate the underlying structure and prevents further heat transfer, the so-called *blocking effect*. Further, PICA has low thermal conductivity, limiting heat transfer through the material, and exhibits good mechanical strength and structural stability even at high temperatures. It retains its integrity during re-entry and can withstand the aerodynamic forces and vibrations experienced during the descent. Finally, PICA heat shields can be reused, as after multiple re-entry cycles there is no significant degradation or loss of performance.
2. Carbon phenolic is a composite material consisting of low-density carbon fibres and phenolic resin, with excellent thermal and ablative properties. From the huge volume of literature available on this particular material, it is clear that a large variation in properties is found, depending on the exact composition, also leading to different densities. For instance, Duffa (2013) presents test results of a material with a density well above 1,000 kg/m^3, whereas Pulci et al. (2010) discuss new developments with a density of 500 kg/m^3; a follow-up study has reduced that even further to 300 kg/m^3 (Paglia et al. 2021). The research of the former focussed on the application of two different carbon-based reinforcements, i.e., a graphitic foam (Grafoam®) and a rigid graphite felt (Sigratherm®)

tested for a heat-flux range between 2 and 9 MW/m^2, representative for a Moon return flight and entry. They found that applying Sigratherm® could potentially lead to very effective insulator up to temperatures of about 2,000 °C. Paglia et al. (2021) tested their samples under heat-flux conditions of 6 and 13 MW/m^2 and found, for the lower end, a difference in temperature of more than 2,400 °C from the exposed surface to the back of the samples. Sample thickness was 4 cm and the exposure time about 30 s. Their results matched very well with comparable tests applied on PICA samples.

3. Avcoat is a composite material consisting of an epoxy novolac resin with special additives in a fibreglass honeycomb matrix. In the Apollo era the Avcoat 5026-22 version had a density of 1,060 kg/m^3, which reduced to 530 kg/m^3 for the final version 5026-39. This was achieved by adding micro-balloons of phenolic resin to the initial composition. Avcoat has excellent thermal properties and is highly effective at dissipating heat. In terms of thermal loading, one of the most severe environments was met for Apollo-4; having a heat shield made of Avcoat 5026-39G, it encountered a maximum dynamic pressure of 80 kPa and a total heat flux of 5.3 MW/m^2, of which about 35% was a radiative flux (Duffa 2013).
4. SLA-561V is a composite material consisting of silica fibres and an ablative resin of low density (224.3 kg/m^3) and low conductivity that can either be used as ablator or as insulator.[20] The material can be bonded with a silicone adhesive using vacuum bag pressure, and is available in different forms: as moulded panels with or without honeycomb support, as an ablative barrier layer, which is integrally moulded against flat or contoured substrates, or even as a three-component spray mix. SLA-561V begins significant ablation at a heat flux of approximately 1.1 MW/m^2 and will work well up to heat fluxes of 3 MW/m^2, above which it will fail. The emissivity of a virgin ablator is $\varepsilon = 0.58$, whereas it is 0.97 for the charred ablator. That means that for a cold-wall flux of $q_c = 1.1$ MW/m^2, the equilibrium temperature of the virgin ablator is $T_w \approx 2{,}132$ °C, whereas for the high-end flux and a charred ablator it is $T_w \approx 2{,}445$ °C. Extensive testing was done within the Viking programme and many material properties were published by Strauss (1967).
5. SIRCA (Silicone Impregnated Refractory Ceramic Ablator) is a lightweight composite material consisting of a ceramic matrix and a silicone binder (Sepka and Samareh 2015), used for ceramic substrates on, amongst others, the Viking capsules and the Space Shuttle. SIRCA typically has a density between 200–400 kg/m^3 and can withstand heat fluxes up to 3 MW/m^2. With an emissivity of $\varepsilon = 0.85$ (Kumar and Mahulikar 2016), this would yield a cold-wall equilibrium wall-temperature of 2,535 °C. Finally, it is easily machined to custom shapes, which makes it a good candidate for curved surfaces.

[20] https://www.lockheedmartin.com/content/dam/lockheed-martin/space/documents/thermal/SLA-561%20Product%20Information.pdf. Date accessed: May 17, 2023.

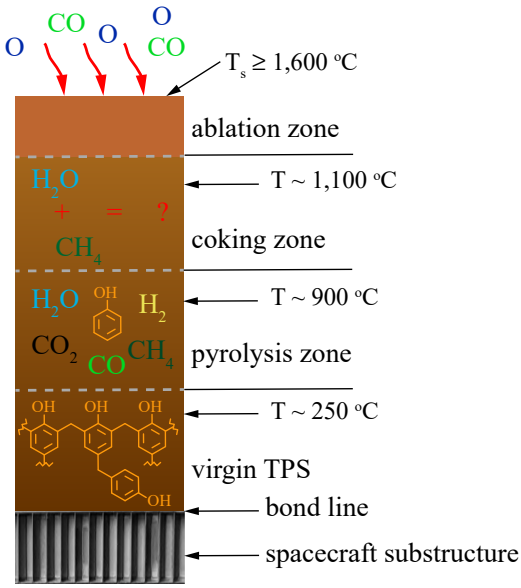


Fig. 10.25 Thermal decomposition of PICA, with the dominating decomposition products shown. Data taken from Bessire and Minton (2017)

To discuss a little more the ablation process, using PICA as an example, Bessire and Minton (2017) investigated the pyrolysis of PICA by measuring the molar yields of the volatile decomposition products as a function of temperature. They applied four different heating rates of 3.1, 6.1, 12.7, and 25 °C/s for temperatures between 100 and 1,200 °C. It was found that four products, H_2, CH_4, H_2O, and CO, dominated the molar yields, albeit not in terms of yield. That was reserved for phenol (C_6H_5OH) and its methylated derivatives, cresol (a group of aromatic organic compounds) and dimethyl phenol ($C_8H_{10}O$). Figure 10.25 shows the different layers of the PICA TPS, which are present during an atmospheric entry. All zones are present simultaneously with the thermal gradient progressing from the surface of the TPS toward the bond line between the TPS and the spacecraft substructure. In the pyrolysis zone, the TPS is heated sufficiently that the polymeric resin decomposes and small gaseous molecules are released. Further up, in the coking zone (coking is the precipitation of carbon), a large fraction of the gaseous products of thermal decomposition permeate through a hot char layer. The remainder may also coke the char layer or react between themselves. In the ablation zone, gaseous product molecules enter the boundary layer. There they may either pass through or react with partially dissociated atmospheric gas components that can be found in the boundary layer. *Vice versa*, the boundary-layer gases can also ablate the hot outer char layer.

The chemical reactions in the different layers occur at different temperatures, as indicated in the figure. This makes it difficult to come up with an estimate for the mass loss at a given (wall) temperature.

In the next Sect. 10.3.2 the most important TPS-material characteristics will be introduced in some more detail.

10.3.2 Characteristics

In this part, several material characteristics and interactions with the hot plasma around the re-entry vehicle will be presented. It is not meant to be exhaustive, but serves as an indicator which aspects may be of importance for a certain application. In any case, the reader is referred to books by Anderson (2019) for chemistry effects in the flow, Duffa (2013) for the interaction of ablative material and the boundary layer, Thornton (1996) for the thermal behaviour of aerospace structures, and Wijker (2008) for the design of space structures. These books may give a good starting point for further research.

Many aspects are of importance for the operational use of potential TPS materials:

1. thermal properties as a function of temperature, related to thermal expansion, heat transfer (convection, conduction, and radiation) and the material's heat capacity,
2. optical properties, such as the emissivity and absorptivity, not just of the outer surface, but for layered TPS also for the inner layers,
3. chemical behaviour, such as oxidation resistance and resistance to other chemicals, like salt, yet again for a range of temperatures,
4. strength parameters, such as the limit strength, ultimate tensile strength, and the stress-strain relations for a range of temperatures, and
5. fatigue properties, including mechanical fatigue, thermal fatigue and fatigue-related chemical effects.

As has been stated earlier, the focus will not be as much the structural design and strength analysis for high-temperature applications, but more the thermal aspects. Subsequently, the discussion will zoom in on emissivity, hypersonic-flow chemistry and wall catalycity, and the protection against hot oxygen. Towards the end, some remarks will be made about strength and stiffness at elevated temperatures, as well as the issue of thermal stresses. It is noted that the mechanics of heat transfer have been discussed earlier (Sect. 10.2.2). Thermal properties of materials, such as specific heat capacity and the coefficient of thermal conduction, will only be summarised at the end of this section; the detailed data tables will be presented separately, as these are an important input to the engineering models for TPS design that will be introduced later (Sect. 10.5).

10.3.2.1 Emissivity

Emissivity is an important material property that influences the behaviour of materials. It is a measure of a material's ability to emit thermal radiation, which is the energy that is emitted by an object in the form of electromagnetic waves due to its temperature. The emissivity of a material is typically given as a value between 0 and 1, where a value of 0 represents a perfect reflector (no emission), and a value of 1 represents a perfect emitter (total emission). Materials with high emissivity tend

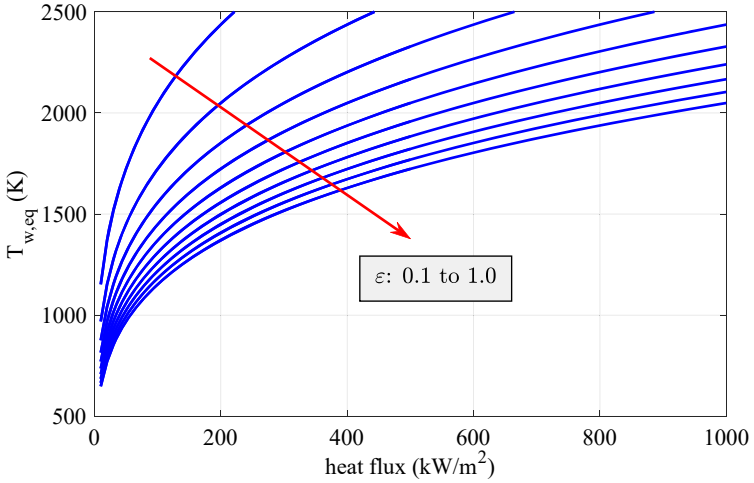


Fig. 10.26 The equilibrium wall temperature as a function of the flux-equilibrium value ($q_c = q_{rad}$) and different values for the emissivity, ε. Calculated under the assumption of a cold wall

to absorb and radiate thermal energy more efficiently than those with low emissivity. For example, a material with a high emissivity will heat up more quickly when exposed to thermal radiation and will also radiate more heat back to its surroundings. In general, for TPS materials a high emissivity is wanted.

As an example, consider a bowl of (pure) water. Pure water has an emissivity of $\varepsilon = 0.96$. While leaving the bowl with water in the Sun, it will warm up quickly. Then, when putting the warm bowl in a colder environment, it will also cool down quite quickly. Of course, the cooling down will progress slower when the environment is warmer. In case the influx of heat is equal to the emitted heat, it is said that *radiation equilibrium* is reached and the temperature of the water will remain constant. The same principle can be applied to the wall of a re-entry vehicle, as we have discussed in Sect. 3.6. For radiation equilibrium the convective flux (inwards) is equated to the radiative flux (outwards) for equilibrium, and from this the equilibrium wall temperature can be derived. The larger the emissivity, the lower this equilibrium temperature. To show this, Fig. 10.26 shows the equilibrium wall temperature for the condition that the incoming convective heat flux is matched by the outgoing radiative heat flux. From Eq. (10.4) it follows that[21]

[21] In case a significant portion of the incident heat comes from radiation (of the plasma), this influx, $q_{rad,air}$, should be included in the heat balance. And, to take the loss of outgoing radiation into account would mean that the actual wall temperature, $T_{w,eq}$, is a bit lower. When more accurate analysis is required, the equation to be used is then:

$$q_c \left(1 - \frac{T_{w,eq}}{T_{w,ad}}\right) + q_{rad,air} = \epsilon \sigma T_{w,eq}^4$$

.

$$q_c = q_{rad} = \varepsilon\sigma T_w^4 \quad \Rightarrow \quad T_w = \left(\frac{q_c}{\varepsilon\sigma}\right)^{0.25} \tag{10.27}$$

The (inverse-proportional) effect of emissivity is clearly seen, as the difference in equilibrium temperature is significant.

Some examples of emissivity values for different types of TPS materials will now be presented, but it is noted that this is by no means an exhaustive discussion. It serves the purpose of introducing the variation between several materials and the difficulty of obtaining consistent test data. Carbon-carbon (C-C) is a non-oxide engineering ceramic that is commonly used for TPS applications due to their high strength, low weight and thermal stability. The emissivity of C-C composites can vary depending on their composition, surface finish, wavelength of the incoming radiation, and temperature. Studying literature confirms that it is not easy to find uniform and consistent numbers. It more or less confirms that for any application of high-temperature resistant materials, dedicated thermal tests have to be done to establish the appropriate values of material properties. Nonetheless, some numbers will be given here, mainly to show the ranges and sensitivities to some test conditions.

Balat-Pichelin et al. (2006) discuss emissivity measurements on carbon-carbon composites at high temperature under high vacuum, as carbon-carbon composite was intended for the thermal shield of the *Solar Probe* space mission. They tested a variety of samples with different properties, e.g., a density in between 1,500 and 1,810 kg/m^3, a porosity between 0 and 17%, and a variety of pore sizes. Some samples were tested "as received", while others underwent a heating treatment (at the maximum test temperature of 2,400 K). A large group of the samples had an emissivity of $\varepsilon = 0.42-0.50$ at $T = 1{,}500$ K, and increasing to $\varepsilon = 0.48-0.56$ at $T = 2{,}250$ K. Two particular samples had much higher values measured, i.e., $\varepsilon = 0.65$ ($T = 1{,}500$ K) and 0.68 ($T = 2{,}250$ K), and $\varepsilon = 0.76$ ($T = 1,\ 500$ K) and 0.77 ($T = 2{,}250$ K). Both samples had received a heat treatment at very high temperatures, i.e., 3,000 and 2,800 K, respectively. For specific details, the reader is referred to the cited paper, although the main conclusion is that both material properties and condition, as well as pre-test heating treatments are important aspects to consider.

The nose and wing leading edges of the Space Shuttle were made of reinforced carbon-carbon, i.e., carbon-carbon with a SiC coating applied. Coatings are typically applied to change (improve) the surface properties, emissivity being one of them. Caram et al. (1993) conducted emittance tests on both virgin and flown samples of RCC, both at room temperature, and at elevated temperatures between 1,200 and 2,500 K. It was found that the total normal emittance for the virgin material decreased from 0.89 to 0.63 over this temperature range. Measuring at room temperature—and using directional data to obtain the so-called total hemispherical emittance—it was observed that the data varied from 0.87 to 0.613. The RCC flown samples were also thermally tested, albeit for total normal emittance at lower temperatures only. The emittance values only decreased to 0.75. For the total hemispherical emissivity at the higher temperatures, it was only reduced to 0.72. Despite the limited test data, their conclusion was that the entry flight environments has affected the

emittance characteristics of the RCC/SiC coating. The highly thermally-loaded Shuttle at peak-heating conditions can be compared with the earlier mentioned pre-test heating treatment.

Staying with the Shuttle, the main portion of its TPS consisted of ceramic tiles. They are a common type of TPS material, which are used for their high-temperature resistance and low thermal conductivity. There are two types of these reusable surface insulation (RSI) tiles, both black and white ones, necessitated by the requirement to control different temperature regimes (see Fig. 10.9). The black tiles, known as high-temperature RSI (HRSI) were on surfaces that were highly thermally loaded, and were optimised for maximum emissivity such that they would lose their heat faster. The white tiles, known as low-temperature RSI (LRSI), were situated on the upper surface of the Orbiter and had higher thermal reflectivity. This means that they tended to absorb little heat. They were pointed towards the Sun to minimise solar gain when the orbiter was on the illuminated part of the orbit. The HRSI material is a high-purity amorphous silica fibre, ceramic bonded and fabricated into blocks or tiles (Throckmorton 1982). They were manufactured in two densities, i.e., 144 and 352 kg/m^3, indicated by LI-900 and LI-2200.[22] The tile's surface was coated with a layer of reaction-cured glass (RCG) with a black pigment, to provide the most efficient radiation of thermal energy. The LRSI material was the same as the low-density HRSI material, but now a white pigment was added to provide the earlier mentioned low solar absorptance. To complete the family of RSI, there was also a flexible RSI (FRSI), which was a Nomex felt. The material was coated with a white silicon elastomeric film to provide waterproofing and low solar absorptance.

Throckmorton (1982) discusses two sets of dissimilar sets of HRSI surface emissivity information, obtained from experimental measurements by NASA Ames Research Center (total hemispherical data based on spectral measurements at room temperature, which were then converted to temperature-dependent emissivity data), and NASA Langley Research Center. They measured the normal spectral emittance of the material at elevated temperatures. The Ames data showed $\varepsilon \approx 0.9$ with minimal variation with temperature. The Langley data, on the other hand, showed $\varepsilon \approx 0.87$ at $T = 120\,^{\circ}\mathrm{C}$ and $\varepsilon \approx 0.62$ at $T = 1{,}330\,^{\circ}\mathrm{C}$. The author mentions that the difference in results at the higher end of the temperature range is not understood.

Emissivity values for both LRSI and FRSI surfaces have also been determined using the same spectral-emittance methods, and compared with data from other sources. A similar trend as for the HRSI was observed, but it was also mentioned that having accurate quantitative values was less important, as these two materials operate at much lower temperatures.

With respect to metals, Del Campo et al. (2010) discuss the emissivity test results of three Ni- and Co-based aeronautical alloys, most notably Inconel 718 (planned to have been used for the Shuttle design with metal TPS), René 41 (planned to have been

[22] A tile is made from 99.9% pure silica glass fibres, and is 94% air by volume. An LI-900 tile can be heated to 1,204 °C and then immediately dipped into cold water and suffer no damage. Information taken from NASAFacts "Space Shuttle Thermal Protection System", https://www.nasa.gov/centers/kennedy/pdf/167473main_TPS-06rev.pdf. Date accessed: April 13. 2023.

used for the X-20), and Haynes 25, each with three different surface roughnesses, i.e., brushed, sandblasted, and wire-cut by electrical discharge machining (EDM). They conclude that the emissivity is nearly the same for the three studied alloys. In general terms, the emissivity decreases with increasing wavelength, and it is nearly independent of the temperature in the analysed range of about 180–640 °C. The values found for this temperature range are between 0.45 and 0.6 for a wavelength of $\lambda = 2\,\mu$m, and decrease to values just below 0.2 for $\lambda = 22\,\mu$m. Also some oxidised samples were tested, and the emissivity would increase significantly for the lower wavelengths, going up to almost $\varepsilon = 0.8$. For the higher wavelengths, the emissivity was still larger for the oxidised samples, but the differences were not so prominent.

Some other metal alloys considered for metal TPS are Molybdenum, Haynes 188, Inconel 617, Inconel-X, and PM-2000. Molybdenum (proposed for the X-20) has a low emissivity of 0.06 ($T = 38\,°$C) and going up to 0.18 ($T = 1{,}093\,°$C),[23] which was one of the reasons to apply a coating on the wing leading edges. Haynes 188 is a cobalt-nickel-chromium-tungsten alloy, which has excellent high-temperature strength combined with very good resistance to oxidising environments up to 1,095 °C) for prolonged exposures. The emissivity of pristine material is quite low ($\varepsilon = 0.11$), but Groh et al. (1994) report that with surface treatments it can be increased to values between 0.28 and 0.86, depending on the treatment applied. Out of the treatments tried, two were selected for their excellent vacuum heat-treatment durability: alumina-titania coatings ($\varepsilon = 0.85$ after 2,695 h) and zirconia titania-yttria coatings ($\varepsilon = 0.86$ after 2024.3 h). Even though these were studied for heat receiver surfaces for solar dynamic power systems, they may have similar potential as re-entry-TPS coating.

Inconel, and more specifically Inconel-X, and Inconel 718 and 617 have been considered as metal TPS candidates on several occasions. Amongst others, Inconel-X was used as a heat sink on the X-15, Inconel 718 was proposed for the X-20, and Inconel 617 was considered for the bottom side of the X-33 (Fig. 10.14). Inconel, having been around since 1932, has come a long way with its applications in the automotive and aerospace industry. Currently a registered trademark of Special Metals Corporation, it is a family of austenitic nickel-chromium-based superalloys.[24] Inconel-X, once oxidised, has an emissivity coefficient of $\varepsilon = 0.895$ at $T = 315\,°$C and $\varepsilon = 0.925$ at $T = 1{,}095\,°$C.[25] It can be used at maximum temperatures in excess of 650 °C, as applied on the X-15. Within the Inconel family, it is a steel-nickel-chromium alloy, with 5–9% iron, and it is also known under the name Inconel-X-750.

[23] https://www.omega.co.uk/literature/transactions/volume1/emissivitya.html. Date accessed: April 13, 2023.

[24] https://www.specialmetals.com/documents/technical-bulletins/inconel/. Date accessed: April 14, 2023.

[25] https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-x-750.pdf. Date accessed: April 14, 2023.

Inconel 718[26] is a high-strength, corrosion-resistant material to be used at temperatures between $T = -253$ to $+705\,^{\circ}\text{C}$. It has various applications, ranging from cryogenic storage tanks, for instance, to the other side of the temperature spectrum, where it is a prime candidate material for constructing heat resistant turbines. It was selected as one of the TPS materials for the predecessor of the Space Shuttle, because of its high strength to $730\,^{\circ}\text{C}$ and its fabricability. In terms of emissivity, it is not the best of materials. Keller et al. (2015) tested the material on emissivity, for a potential application in very-high-temperature gas reactors. They found that ε ranged from 0.21 to 0.28 in the temperature interval from 487 to 1,202 °C for the untreated samples. However, with an oxidation process in air at $800\,^{\circ}\text{C}$ ε increased to 0.52 at $T = 927\,^{\circ}\text{C}$. At the lower temperature of $T = 487\,^{\circ}\text{C}$, no change was detected.

For the windward body and fin of the *VentureStar* demonstrator X-33, a combination of Inconel 617 and PM-1000 TPS panels was considered,[27] elaborately discussed by Blosser (2000) and expanded on by Blosser et al. (2004). Inconel 617 is a solid-solution strengthened (nickel-chromium-cobalt-molybdenum), high-temperature strength, corrosion and oxidation resistant, high workability and weldability material.[28] Its operational temperature ranges to at least $980\,^{\circ}\text{C}$. Cunnington et al. (1983) evaluated the effects of simulated re-entry exposure conditions on the thermal emittance properties of statically oxidised Inconel 617, and found that the total normal emittance of the material remained above 0.80 at $987\,^{\circ}\text{C}$.

We may conclude from the above descriptions that the emissivity of metallic TPS materials can vary depending on factors such as temperature, surface finish, and composition. For example, at high temperatures (above 1,000 K), the emissivity of titanium alloys can range from 0.3 to 0.6, while nickel-based alloys can have emissivity values between 0.3 and 0.8. It is important to note that the emissivity values of metallic TPS materials can also depend on the wavelength of the thermal radiation being emitted, and some materials may have higher emissivity at certain wavelengths than others. Additionally, the emissivity can be affected by factors such as oxidation and surface roughness, and may not be well defined. This could be, for instance, in regions where the gas is highly ionised and the plasma properties dominate the radiative behaviour.

---

[26] https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-718.pdf. Date accessed: April 14, 2023.

[27] PM-1000 was an oxyde-dispersed strengthened (ODS) material, a nickel-chromium alloy, developed and produced by the Austrian company Plansee. Despite its favourable properties at high temperatures, the company ceased production of this material. Another of their ODS materials, PM-2000, an Iron-based alloy, is still in production, and is also a good candidate for use in metallic TPS. Within the Inconel family, there are two alternatives for PM-1000 and PM-2000, i.e., MA-754 and MA-956, respectively. With basically the same composition, the properties are very similar, although with a variety of heat treatments these properties can be affected. In the framework of the Dutch AEOLUS programme the emissivity of both PM-1000 and PM-2000 have been measured, amongst many other properties (Mooij et al. 2006a, b). For PM-1000, $\varepsilon = 0.78$ @ $T = 670\,^{\circ}\text{C}$ and $\varepsilon = 0.86$ @ $T = 1,300\,^{\circ}\text{C}$, whereas for PM-2000 it is a bit lower, about 0.7.

[28] https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-617.pdf. Date accessed: April 14, 2023.

Of course, the above summary is far from complete. There are many more candidate materials that can potentially serve as a high-emittance TPS material. However, that is not the only criterion that should be considered. Operational temperature range, maximum yield stress at elevated temperatures, conduction properties, and thermal elongation, are others to name just but a few. These criteria are often more important, and then, for a particular design, material thicknesses/thickness variations and the total mass of the TPS come into play. And finally, the reader should also not forget that when other properties are more favourable over the emissivity, coatings can be applied to increase it.

10.3.2.2 Chemistry and Wall Catalycity

Chemistry plays a critical role in hypersonic flow, as at these high speeds, the gas molecules in the flow are subjected to high temperatures and pressures, which can cause chemical reactions to occur. One of the primary reactions is the phenomenon of chemical dissociation, which means that at high temperatures, the molecules in the gas flow can break apart into their constituent atoms. It may also lead to the formation of reactive species, such as free radicals and ions. These species can then react with other species in the plasma, leading to the production of new species and the transfer of energy between different modes of motion. The dissociation of molecules can have a significant effect on the thermodynamic properties of the gas, such as its specific heat capacity, which in turn affects the aerodynamic properties of the flow. In addition, chemical reactions between the gas molecules and the surrounding surfaces can lead to the production of species that can alter the properties of the flow, such as heat transfer, mass transfer, and chemical kinetics. For example, the interaction of oxygen molecules with the surface of a spacecraft during re-entry can lead to the production of atomic oxygen, which can react with the spacecraft material and cause erosion. Understanding the chemical reactions is essential for optimising the performance and safety of re-entry vehicles.

In hypersonic flow, chemical reactions can either release or absorb energy depending on the specific chemical reaction, the composition and temperature of the plasma, and the thermodynamic conditions of the flow. In general, endothermic chemical reactions are those that absorb energy from the surroundings in the form of heat, and they could potentially cause a cooling effect in the hypersonic flow. Some examples of endothermic chemical reactions that may occur include dissociation of molecular oxygen, nitrogen, and carbon dioxide, as well as the decomposition of hydrocarbons such as methane or propane. Exothermic chemical reactions release energy in the form of heat, and such reactions can thus cause an increase in temperature and heat release. An example of an exothermic reaction in hypersonic flow is the recombination of dissociated molecular species, such as that of atomic oxygen to form molecular oxygen.

The net effect of the chemical reactions on the plasma energy depends on the balance between the energy gained and lost through the various reaction pathways. Therefore, the amount of energy taken away from the plasma due to chemistry can

vary widely, depending on the specific conditions of the hypersonic flow and the reactions taking place. Detailed computational simulations and experimental studies are often needed to accurately quantify the energy transfer due to chemistry in hypersonic flows. The different types of chemical reactions are very well covered by Anderson (2019).

Chemistry in hypersonic flow and wall catalycity (catalysis) are linked, because catalytic reactions can play an significant role in modifying the chemistry of the gas flow by providing a surface for chemical reactions to occur. This can promote certain desired reactions while suppressing others. For example, in a hypersonic engine, a catalyst coating on the inside of the engine nozzle can help to break down exhaust gases into simpler compounds, reducing the amount of pollutants emitted into the atmosphere. The catalyst coating can also help to reduce the temperature of the exhaust gases, which can reduce thermal stress on the engine components and improve the efficiency of the engine. The catalytic efficiency of TPS defines the finite catalytic reaction rates of gases in chemical non-equilibrium on the materials' surface. During re-entry the airflow molecules (oxygen and nitrogen) passing through the bow shock become dissociated. The atoms can then recombine at different rates (depending on catalytic efficiency of the material) to molecules on the TPS surface or in the gas phase. As the recombination energy is typically transferred to the material surface, the heat flux on the re-entry vehicle depends therefore on the catalytic efficiency of the TPS materials. Since the catalytic properties of these materials depend on temperature, the heat flux indirectly depends on the vehicle-surface temperature and, consequently, on the thermal emissivity of the TPS materials. Therefore, for the accurate prediction of thermal loads during re-entry and for the lightweight design of the TPS, it is important to know both the catalytic efficiency and the emissivity of the material used.

Materials with high catalytic efficiency are often used in TPS designs, because they can promote rapid recombination of dissociated air molecules, which can help dissipate heat more quickly and reduce the thermal load on the vehicle. For example, some TPS materials use catalysts such as platinum, to promote the recombination of oxygen and nitrogen atoms, which can help reduce the heating of the TPS surface. However, the use of highly catalytic materials can also have drawbacks. For example, these materials can be more susceptible to erosion and degradation due to the high temperatures and chemical reactions during re-entry. Additionally, the use of highly catalytic materials can increase the risk of unwanted chemical reactions, such as the formation of toxic compounds or the release of corrosive gases.

Two examples of the effect of catalycity on aerodynamic heating are provided by Gupta et al. (1984), who analyse the Space Shuttle heating due to a variation angle of attack and catalycity, and Inger and Baker (2005) who study non-equilibrium viscous shock-layer heat transfer with arbitrary surface catalycity. The latter authors have formulated an analytical method for predicting non-equilibrium dissociation/recombination effects on stagnation heat transfer to arbitrarily catalytic surfaces, which gives accurate results over a large range of hypersonic speeds and altitudes. These expressions could very well be used in conceptual design studies when rapid results are required.

10.3.2.3 Protection Against Hot Atomic Oxygen

The main natural atmospheric constituent in low Earth orbit is electrically neutral atomic oxygen, also known as hot atomic oxygen. By colliding with orbiting spacecraft surfaces it can be energised to 5 eV energy, and this is a critical consideration for spacecraft and satellites operating in that altitude range. Collisions with these charged particles have proven to be a major factor in the degradation of a wide range of candidate spacecraft materials. Significant material-interaction effects have been observed in detail, for instance, during the six years of operation of the Long Duration Exposure Facility (LDEF) mission, as well as three Evaluation of Oxygen Interaction with Materials (EOIM) experiments, flown on the Space Shuttle in the late 1980s and early 1990s (Linton et al. 1993). Whereas the Shuttle experiments were of short duration, the Materials International Space Station Experiment 2 (MISSE 2) ran for nearly 4 years, where 41 different polymer samples, collectively called the Polymer Erosion and Contamination Experiment (PEACE) Polymers, were exposed to the low Earth orbit (LEO) environment on the exterior of the International Space Station (Groh et al. 2008). The experiment yielded very valuable erosion yield data aiding future spacecraft design.[29]

Atomic oxygen originates from photo-dissociation of molecular oxygen in the upper atmosphere by solar radiation of wavelength less than or equal to 243 nm. Several processes can take place when oxygen atoms hit a spacecraft surface at orbital velocities. Amongst them are chemical reactions with surface atoms or adsorbed molecules, elastic scattering, scattering with partial or full thermal accommodation, recombination, or excitation of ram species (Banks et al. 2003). For many (unprotected) materials this may lead to corrosion, erosion, structure modification and surface roughening. This, in turn, can cause a degradation of the optical, thermal, electrical, and mechanical properties. The longer the exposure time, the more damaging the effects may be.

To protect against hot atomic oxygen, spacecraft designers and engineers use a variety of materials and coatings that are resistant to oxidation and can withstand the extreme thermal and mechanical stresses associated with exposure to hot atomic oxygen. For instance, among the Inconel familiy of materials, MA754 has excellent oxidation resistance. Thus, it can be used in high-temperature oxidising environments without the need for expensive special coatings, which is often required for conventional high-temperature alloys. The alloy X-750 has also proven to be resistant to oxidation up to 980 °C. Kim et al. (2009) evaluated the oxidation resistance of Inconel 617 and Haynes 230, and tested samples at 900 and 1,100 °C. They found that the oxidation resistance depends on the stability of the surface oxide layer, which

[29] Amongst others, ESA has a Low Earth Orbit Facility (LEOX), which fires a laser to generate "atomic oxygen" normally encountered only in low orbits. The facility can generate atomic oxygen at energy levels that are equivalent to orbital speed, to accurately reproduce the space environment. However, to save time and money for testing, it can also test at a higher flow. See: https://www.esa.int/ESA_Multimedia/Images/2017/05/Atomic_oxygen_generator. Date accessed: May 13, 2023.

can be improved by minor alloying elements such as Ti and Mn. The 1,100 °C test revealed that both alloys showed a double oxide layer consisting of an inner Cr_2O_3 and outer $MnCr_2O_4$ or TiO_2.

Overall, protection against hot atomic oxygen is a complex and ongoing process that involves careful material selection, testing, and monitoring to ensure the long-term viability and reliability of spacecraft and satellites operating in harsh environments. Even for entry systems that are not exposed to this environment for a long time it is important to select materials and coatings carefully, since degradation of the thermo-mechanical properties of the TPS could be disastrous.

10.3.2.4 Strength and Stiffness at Elevated Temperatures

The strength and stiffness of materials can change significantly at elevated temperatures. Generally, as the temperature increases, the strength and stiffness of most materials decrease due to thermal softening, which is the reduction in the elastic modulus of the material as a result of thermal expansion and atomic vibrations. Some specific factors that may lead to the reduction in strength and stiffness when temperature is rising include (i) a reduced bond strength, which means that at high temperatures, the strength of atomic bonds within a material decreases, (ii) an increased atomic mobility, that is, when temperature increases, atoms within a material have increased kinetic energy and are more likely to move, resulting in a decrease in stiffness, (iii) thermal expansion, which typically happens when a material is heated and can result in internal stresses and thus a reduction in strength, and (iv) recrystallisation, which some materials may undergo at elevated temperatures. This can lead to changes in grain size and orientation, affecting the material's strength and stiffness. It is noted, though, that some materials, such as ceramics and certain alloys, can exhibit increased strength at elevated temperatures due to changes in their microstructure or crystal structure.

The effects of elevated temperatures on strength and stiffness can be characterised through experiments, such as tensile and compression tests, which can be used to determine the material's response to stress and strain at different temperatures. The results of these experiments will be valuable input to designers who are responsible for the design of structures and components that are expected to operate at elevated temperatures.

In Fig. 10.27 a global overview of several high-temperature metals for different temperature ranges is shown. Of course, selecting the material for a high-temperature application from just a graph like this, will not do, as there are many more aspects to consider, such as cost, manufacturability, and weight. However, it is safe to say that for a given aerothermal environment the lightest alloy that can accommodate this will be a good starting point to use as structural material in the design process. At the lower end of the temperature range (up to about 320 °C), some advanced aluminium alloys and aluminium-matrix composites are attractive materials, as they may lead to a low structural weight (the density of these materials is about 2,770 kg/m^3). Titanium and titanium-matrix composites can be used to slightly higher temperatures (up to

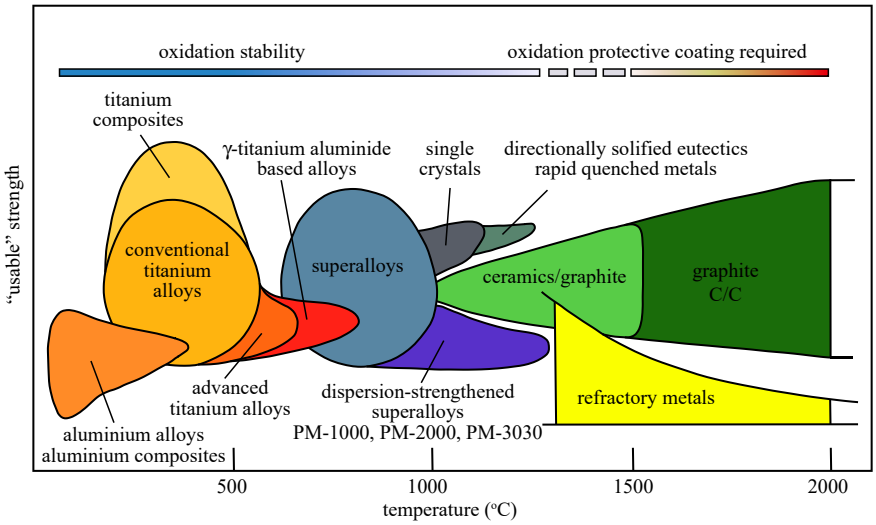


Fig. 10.27 Typical strength as a function of temperature, for a range of materials. Data taken from and image inspired by Blosser (2000)

about 600 °C) and, with a density of about 4,430 kg/m^3, they were considered to be good candidates for application at the leeward side of launch vehicles such as the X-33/*VentureStar*.

The maximum allowable temperature for iron and nickel-based superalloys is substantially higher (980−1, 040 °C), although they are relatively heavy, with a density of about 8,300 kg/m^3. A concentrated development of variations of titanium intermetallics has led to medium-weight materials with densities around 3,900 kg/m^3 that can be used up to 815 °C. Improved dispersion strengthened superalloys were developed in the late 1990s and early 2000s that could be used at temperatures as high as 1,100 °C. PM-1000 and PM-2000 (or the alternatives MA-754 and MA-756) can even be used up too about 1,300 °C.

A very special category is formed by so-called refractory metals. Despite their high melting points and good mechanical properties, their sensitivity to oxidation requires them to be coated to protect them, which usually works up to 1,650 °C. Lighter alternatives for high-temperature applications may be found in ceramics and carbon/carbon. Their sensitivity to oxidation requires coatings as well, and despite their high strength at elevated temperatures they are also impact sensitive, and costly to apply.

To provide a more detailed example of temperature effects on materials' strength, in Fig. 10.28 the yield stress (i.e., the maximum stress below which deformation occurs linearly, whereas above plastic deformation takes place) of several materials has been plotted as a function of temperature. It may be obvious that the yield stress reduces with increasing temperature, and at one point there is a significant drop. The same would apply to the tensile strength, so again it is important to take this behaviour

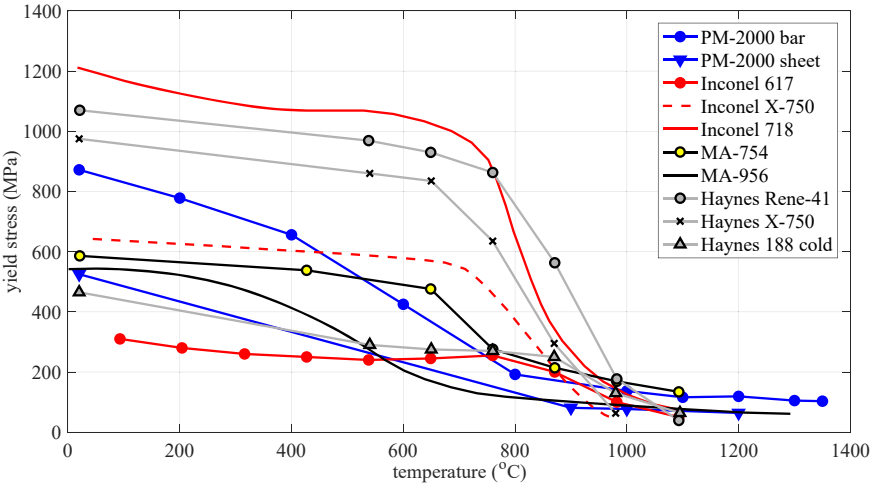


Fig. 10.28 Yield stress as a function of temperature, for a range of materials (The materials properties have been taken from the manufacturers' websites, most notably Hayes (https://www.haynesintl.com/alloys/alloy-portfolio_) and Special Metals (https://www.specialmetals.com/markets/aerospace.html)

into account while doing the structural design of any thermally loaded vehicle. At elevated temperatures the structure will weaken and will (plastically) deform when too severely loaded. While researching the data it was confirmed that the materials' properties are very much dependent on the geometry (plate, rod, sheet), treatment (for instance, the annealing temperature), and, of course, chemical composition (how much of the element percentages are deviating from the nominal values) and, for ODS materials, also the degree of crystallographic texture (i.e., the preferred crystal orientation) imparted during hot working and the grain size. For instance, for the data shown in Fig. 10.28 the PM-2000 data are given for a sheet of grain class 6, for Inconel 617 the data relate to an annealed hot finished rod (no further details provided), and for Inconel 718 they are provided for a 12.5 mm diameter hot-rolled round, annealed at 980 °C for 1 h and aged at 720 °C for 8 h, fast-cooled to 620 °C, and held at that temperature for a total ageing time of 18 h. For most of the other materials no details were provided. The conclusion is once more, that when materials such as these are used, a materials-testing programme should be included.

Concerning stiffness, defined to be the ability of a material to resist deformation in response to an applied force, or in other words, a measure of how much force is required to cause a given amount of deformation in a material, a similar overview as for the yield stress can be prepared. The stiffness of a material can typically be described by two different moduli, the so-called E and G moduli that are used to predict the amount of deformation a material will experience under a given stress

in a linear-elastic regime.[30] The E-modulus (also known as the elastic modulus or Young's modulus), is a measure of a material's resistance to linear or axial deformation in response to an applied tensile or compressive stress. It describes the ratio of the stress applied to a material to the resulting strain in the same direction. The G-modulus, also known as the shear modulus, is a measure of a material's resistance to shear deformation. This is a deformation, which occurs when forces are applied in parallel planes offset from each other. It describes the ratio of shear stress to shear strain in the material. The two moduli are determined by measuring the slope of the stress-strain curve for a material in the linear elastic region. The stress-strain curve is a graphical representation of how a material responds to an applied force, showing the relationship between the stress (force per unit area) and strain (deformation) of the material. While both the E and G moduli describe a material's stiffness, they measure different types of deformation and are related to different types of stress. Therefore, their values may be different for the same material. For example, some materials, such as rubber, have a low E-modulus but a high G-modulus, because they are relatively easy to stretch axially but resist deformation under shear.

Figure 10.29 shows the Young's modulus as a function of temperature for the same materials as before (Fig. 10.28). With increasing temperature, E shows an almost linear decrease. Over the full temperature range, the decrease is substantial and is something not to be ignored. Concerning the shear modulus, G, Fig. 10.30 gives an example comparison of the elasticity and shear moduli of Inconel 617 and Inconel 718. The first observation is that the shear modulus is significantly smaller than E—for these two materials over 50%—which dictates careful load-path analysis for the structural design. The second observation is that also G reduces linearly with increasing temperature, albeit at a slightly lower pace.

It is noted that even though a material's strength typically degrades at higher temperatures, one should also not forget the other end of the spectrum. Although at low temperatures the strength and stiffness of most materials increase, they tend to shrink in size and become rather brittle. In space, close to Earth, temperatures get as low as 10 °C (due to a combination of Solar flux and Earth albedo), although when in eclipse the temperatures tend to get a lot lower. One example of what can happen with materials at low temperatures is the O-ring failure on one of the Space Shuttle solid rocket boosters (SRBs) during the launch of *Challenger* on January 28, 1986. At the day of the launch, it was just above 0 °C and the O-ring material contracted. After ignition of the SRBs, an O-ring in the right SRB did not expand to prevent the seepage of hot gases. A hot gas plume from the failed joint could escape and impinged directly on the external tank and burning through it. This caused a structural failure and led to the fateful event of an exploding external tank, destroying the Orbiter.

[30] When a material is subjected to twist loading, the corresponding elasticity modulus is approximately equal to the shear modulus and is sometimes also called modulus of rigidity. There are basically two main methods to measure the moduli, i.e., (i) the static method, which is based on the measurement of the deformation induced in a material by the application of a known force, and (ii) the dynamic method, which implies measuring the ultrasonic body wave velocities. The relationship between the obtained moduli is a linear one.

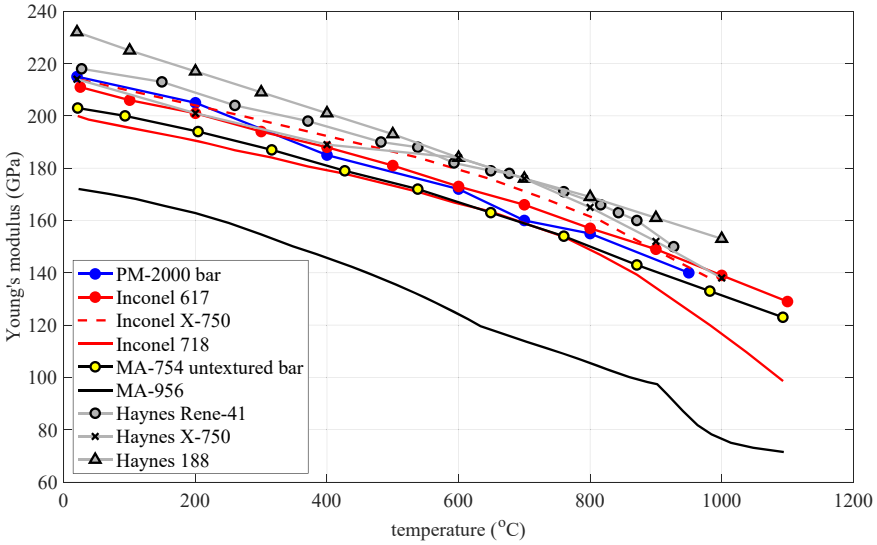


Fig. 10.29 Young's modulus as a function of temperature for several materials. Data taken from manufacturer's website (see references in the text)

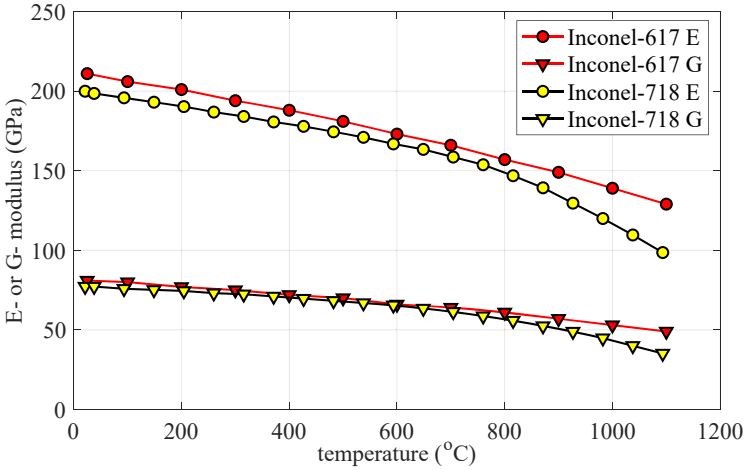


Fig. 10.30 The elasticity and shear moduli (E and G, respectively) as a function of temperature for Inconel 617 and Inconel 718. Data taken from manufacturer's website (see references in the text)

It makes the structural design of spacecraft even harder, especially if they have a mission in space and subsequently have to survive atmospheric entry.

It is noted that composite materials exhibit a different behaviour. Just like with other materials, each has its advantages and disadvantages. Carbon/carbon composites belong among the best of the composites, in the sense that they are one of the strongest and lightest high-temperature engineered materials (Forsberg et al. 2005). Not only can they operate at higher temperatures than any other material, they also keep their room-temperature strength at temperatures up to 2,225 °C. Further, Fitzer (1988) comment on some of their other properties, amongst others the excellent fatigue behaviour. After more than several million cycles, C/C retains in between 70 and 80% of the original modulus and strength. On top of this, it has a very low thermal expansion and extremely high thermal conductivity, such that the material's high thermal shock resistance makes it a very good candidate for high thermally loaded parts of the TPS. In terms of properties, the most severe disadvantage of C/C is its sensitivity against oxidation at temperatures above 600 °C. Up to 800 °C, inhibitors like phosphates, borates, and silicates will reduce the oxidation rate, but at higher temperatures coatings are required.

10.3.2.5 Thermal Stress

Thermal stress refers to the stress induced in a material due to a change in temperature. When the temperature of a material changes, its dimensions may also change due to thermal expansion or contraction. This change in dimensions can result in internal stress within the material, which can affect its mechanical properties and structural integrity. Thermal stress can be caused by both heating and cooling of a material and can lead to failure or deformation, when the stress exceeds its strength. It is an important consideration in the design and operation of structures and systems that are exposed to significant temperature changes.

The calculation of thermal stress involves understanding the thermal expansion and contraction properties of the material, as well as the temperature change to which the material is subjected. The following steps provide a general approach to calculating thermal stress:

1. Determine the coefficient of thermal expansion, α, of the material, which is a measure of how much the material will expand or contract, given a temperature change and given the thermal-equilibrium state of the material. This coefficient can usually be found in material data sheets or handbooks. In Fig. 10.31, the average coefficient of thermal expansion is shown for several materials. The data provided is based on the average value over a temperature range, e.g., the datum at 200 °C means that this is the average α for the temperature range starting at 20 °C up to 200 °C. What is clear from this graph is that the expansion will increase with rising temperature. Also, there is quite a bit of variation in between the materials, so having structural elements made from different materials will require special attention;

2. Determine the temperature change, ΔT, i.e., the difference in temperature between the initial and final temperatures of the material;
3. Calculate the thermal strain, ε_T, by multiplying ΔT with α;
4. Calculate the thermal stress, σ_T, by multiplying the thermal strain by the elastic modulus, E.

Thus, the formula for calculating thermal stress (for a constrained part) is as follows:

$$\sigma_T = -\alpha E \Delta T \tag{10.28}$$

where the minus-sign has been added to get the right sign of the stress: negative for compression (which will occur for a clamped system and a temperature rise) and positive for tension. Note that the thermal stress is only one of the stresses that a material can experience, and the total stress on a material is the sum of all the individual stresses. Additionally, the actual stress experienced by a material may be affected by other factors, such as its shape and geometry.

To illustrate this process, consider a rod of Haynes 188 that is clamped on one side, and can thus freely expand on the other. Suppose that the rod is $L_i = 1$ m long, and is at equilibrium (i.e., free of stresses) at room temperature (20 °C). The rod is heated up to 200 °C, so $\Delta T = 180$ °C. What is the elongation of the rod? From Fig. 10.31, we find $\alpha = 12.6 \cdot 10^{-6}$ per °C, valid over the temperature range of [20, 204] °C. Thus, the strain is $\varepsilon_T = \alpha \Delta T = 0.0023$. With ε_T defined by

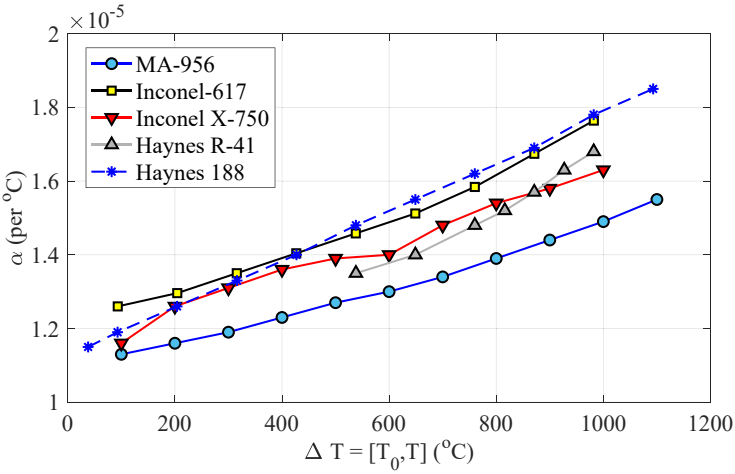


Fig. 10.31 The average coefficient of thermal expansion for a range of materials, taken from the manufacturers' websites (see earlier references). The data points are given for discrete values of the temperature range, and represent average values of α for the temperature range beginning from 20 °C up to that temperature

$$\varepsilon_T = \frac{L_f - L_i}{L_i} \quad \Rightarrow \quad L_f = (1 + \varepsilon_T) L_i \tag{10.29}$$

where L_f is the elongated length, it follows that $L_f = 1.0023\,\text{m}$. Thus, the rod has extended by 2.3 mm.

Suppose now that the rod is clamped at both sides, what is then the induced thermal stress? From Fig. 10.29, at $T = 200\,^\circ\text{C}$ Young's modulus is found to be $E = 217\,\text{GPa}$. Using Eq. (10.28) it follows that $\sigma_T = -492.2\,\text{MPa}$, which (obviously) means that the rod is under compression. To get a feeling for the reaction force at the clamped end, one could consider that the rod is part of a chained structural system, and the adjoining element is preventing the rod to expand. In line with the discussion above, and using the definition of Young's modulus as the ratio of stress divided by strain,

$$E = \frac{\sigma}{\varepsilon} \tag{10.30}$$

and force being equal to stress multiplied by the cross-sectional area, it follows that

$$F = \sigma A = \varepsilon E A = \frac{\Delta L}{L_i} E A \tag{10.31}$$

Assuming that the rod has a diameter of 1 cm, giving a cross-sectional area of $A = 0.25\pi d^2 = 7.854 \cdot 10^{-5}\,\text{m}^2$, and substituting all known numbers, it follows that the reaction force is $F = \sigma_T A = 38.7\,\text{kN}$.

For comparison, in case the rod were made out of aluminium, with $\alpha = 25 \cdot 10^{-6}$, the elongation would be almost doubled, about 4.5 mm. With $E = 70\,\text{GPa}$, the thermal compression stress would be $-315\,\text{MPa}$, and the reaction force $F = 24.7\,\text{kN}$. And finally, in case the rod would have a diameter of 10 cm ($A = 0.0079\,\text{m}^2$) then the force would be $F = 3{,}865\,\text{kN}$ ($F = 2{,}474\,\text{kN}$ for the aluminium rod). The conclusion may be obvious: design to accommodate thermal expansion! Fig. 10.32 summarises the above discussion with the different mechanisms of where thermal stress comes from.

As an example of how to use thermal stresses to ones advantage, in the framework of the Dutch AEOLUS programme Sudmeijer et al. (2001) discuss the development of an ODS sandwich structure for the stiffened load carrying skin of the rudder of a reusable launcher. The sandwich panel consisted of two PM-1000 face plates, and a PM-2000 honeycomb cell structure. The facing thickness was 0.2 mm, the core depth of the honeycomb, i.e., the distance between top and bottom facing, 10.0 mm, the cell-wall thickness 0.1 mm, and the cell size 8.0 mm. The honeycomb and facings were joined by vacuum brazing at a temperature of 1,238 °C and then slowly cooled down. At the brazing temperature the panel was stress-free, and because the coefficients of thermal expansion are different for PM-1000 and PM-2000, reaching room temperature meant that the sandwich panel was pre-stressed. This process was also reproduced with finite element analysis, and it was confirmed that the thermal stresses always remained below yield stress. The consequence of the pre-stress is easy to imagine: once heated up to conditions encountered during re-entry, effec-

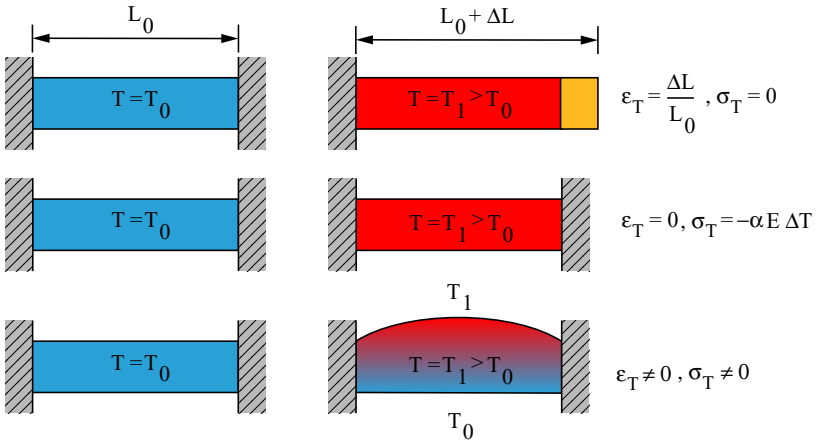


Fig. 10.32 Different mechanisms for the origin of thermal stress: the material's thermal expansion, a temperature differential, and structural or mechanical restraint of thermal growth

tively resulting in temperatures close to the brazing temperature, means that the panel will be virtually stress-free, as it was during the brazing! The panel was tested during 47 re-entry cycles and did not show any damage or degradation. Permanent deformation was negligible and also the braze joint remained in excellent condition.

In summary, with the above discussion and examples on thermal stress, the field of thermoelasticity has been touched upon, i.e., the studies of the response of a material to changes in temperature and mechanical stress. It deals with the coupling between thermal and mechanical effects, which means that changes in temperature can cause changes in the material's shape, and vice versa. The behaviour of a material is described by its elastic properties, which are related to its ability to resist deformation under (thermo-mechanical) load. The elastic properties of a material can be characterised by its Young's modulus, Poisson's ratio, and shear modulus. When a material is subjected to thermal loading, it experiences thermal expansion or contraction. This change in dimension can be calculated using the material's coefficient of thermal expansion, which is a measure of how much the material's length, width, or height changes per unit change in temperature. The mechanical behaviour of a material under thermal loading can be predicted using thermoelasticity theory, which relates the stress and strain in the material to changes in temperature. The theory is based on the assumption that the material's elastic properties do not change significantly with temperature. If that is not the case, though, this change has to be accounted for.

10.4 Design Options

10.4.1 Characterisation

From the introduction to this chapter it has become clear that there are several options to design and implement an operational TPS. The type of TPS that is chosen depends largely on the magnitude and duration of the surface heating. Form aerothermodynamics analysis it has become clear that the heating varies over the surface of a vehicle, which implies that for different parts of the vehicle, the TPS may vary from one concept to the other. It should also be noted that a particular solution for one vehicle and mission, may not work for another vehicle, even if flying the same mission. It is for that reason that active research and development continues to take place; even when a certain concept has been selected for a vehicle it will not stop the ongoing development. One example is that of the Space Shuttle; when the ceramic TPS had been selected as baseline TPS, NASA Langley Research Center coordinated further research to improve metallic TPS, see, for instance, the report by Varisco et al. (1978).

In general, the most commonly used TPS concepts can be grouped together as passive systems, semi-passive/active systems, and active systems. These groups have been visualised in Fig. 10.33 with a total of eight different concepts, albeit that variations on these concepts are possible:

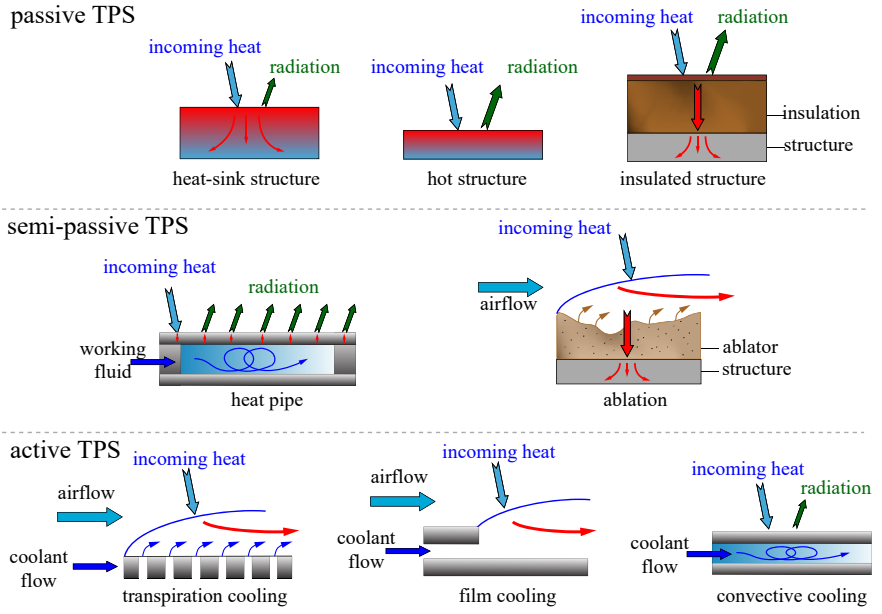


Fig. 10.33 Classification of different TPS concepts, divided over passive, semi-active, and active systems. Inspired by Blosser (2000)

1. Passive TPS: unlike active-cooling concepts, where a coolant flows through the TPS to remove the heat incident on the vehicle, a passive concept has no fluid flow. Instead, part of the heat is absorbed by an distributed through the structure, whereas another part is radiated out again from the surface. If the passive system is loaded not too close to the material's thermal limit, such a concept could be reused. Different implementations exist:
   (a) Heat-sink structure,
   (b) Hot structure, and
   (c) Insulated structure, occasionally called a *cold structure*, referring to the strict separation of the thermal function and the structural function.
2. Semi-passive TPS: logic dictates that this kind of concept is somewhere in between the passive and active TPS. It does make use of a coolant, but the flow is not controlled by any external mechanism. To main concepts can be distinguished:
   (a) Heat pipes, and
   (b) Ablators.
3. Active TPS: as mentioned, an active TPS requires a cooling liquid that is actively controlled. Three main categories are mentioned here:
   (a) Transpiration cooling,
   (b) Film cooling, and
   (c) Convective cooling. Within this (family of) concept(s), also the concepts of enhanced radiation cooling and nucleid boiling are considered.

In the next sections, each of the categories with the concepts will be briefly discussed.

10.4.2 Passive Systems

10.4.2.1 Heat-Sink Structure

A heat sink consists of a material with a high thermal conductivity that will easily absorb all the aerodynamic heat incident on the surface of the vehicle and uniformly distribute it around and away from the affected region. The higher the conductivity, the faster this will go. The total amount of heat that can be absorbed depends on several material properties, such as its temperature limit and specific heat capacity. Furthermore, the total heat is proportional to the structure's mass. Thus, to sustain a larger heat load one can simply increase the material layers or the mass of the TPS. However, this does not come for free and the heat-sink concept generally becomes prohibitively heavy for all but very brief heat pulses.

Because the heat sink operates well below the radiation equilibrium temperature, the outward surface radiation is limited. This means that indeed most energy coming

from the flow must be stored in the structure. The primary advantages of heat sink structures are simplicity and reliability, but because of the high thermal loads during entry pure heat-sink concepts are generally too heavy. Heat sink structures were used for noses of most early ICBMs, after-bodies of the Mercury and Gemini re-entry capsules, and parts of the structure of the X-15 experimental vehicle.

To get some feeling for the heat-sink concept and to understand the useful range of the heat sink approach, a simple calculation will allow us to calculate the amount of heat that can be stored per unit mass of a selected number of aerospace materials. The specific heat coefficient, c_p provides the information on how much energy per unit mass, Q, is required to increase the temperature by one degree, or, *vice versa*, how much of a temperature increase, ΔT, is the result after a given amount of energy has entered this unit mass. So:

$$\Delta Q = c_p \Delta T \tag{10.32}$$

In the appendix to this chapter, c_p has been given as a function of temperature for many materials.[31] Starting from a temperature of 25 °C, for each degree increment the current temperature will determine c_p and with the above equation the amount of energy required (or stored) can be calculated for this ΔT of one degree. Integration/summation over the complete temperature range will then give the total heat capacity. Figure 10.34 shows the result of this calculation. Nickel, chromium, titanium, and iron have nearly identical and relatively low heat capacities. Concerning the use as TPS material, these four families make up most of the high-temperature, oxidation resistant metallic alloys. Of the other materials shown, aluminium, graphite polyimide and carbon/carbon have similar heat capacities, but this cannot be seen decoupled from their (different) operating temperature limits. Beryllium is the material that can hold most heat, i.e., at 625 °C this is about 1,600 kJ/kg. To put this number in perspective, a typical integrated heat load for a point on the windward surface of an RLV may be in the range of 170–227 MJ/m^2 and even without having more details about the primary TPS configuration, once divided by the specific heat capacities it means that a heat-sink approach is much too heavy for consideration on an RLV.[32] Note, though, that for small portions away from stagnation points where the thermal loading is smaller, heat sinks can be applied. How and where this can

[31] Using a single specific-heat coefficient, thus assuming temperature is constant, may have a serious impact on the results, which is shown by an example given by Holland (1963). He measured the specific heat of titanium over a temperature range from 600 to 1345 K. The specific heat at 600 K is 720 J/(kg K) (±2.5%), and increases to 930 J/(kg K) (±6%) just below the transition temperature of 1155 K.

[32] Beryllium has a density of $\rho = 1,854.6\,\text{kg/m}^3$, which is on the very low end of metals. A 1-m^2 plate, with thickness of 1 m has thus a mass of 1,854.6 kg. With a specific heat capacity of 1,600 kJ/kg at $T = 625\,°\text{C}$, the total amount of heat stored is thus 2,967 MJ. For higher temperatures this gets even more: at $T =$ at 1,200 °C the heat stored is 6,198 MJ. However, for a plate of 1 cm thick this has reduced to about 62 MJ (with a TPS mass of 185 kg/m^2). In the worst-case scenario, with a total heat load of 227 MJ/m^2 this means that a thickness of almost 4 cm is required for each square meter of heat-sink TPS.

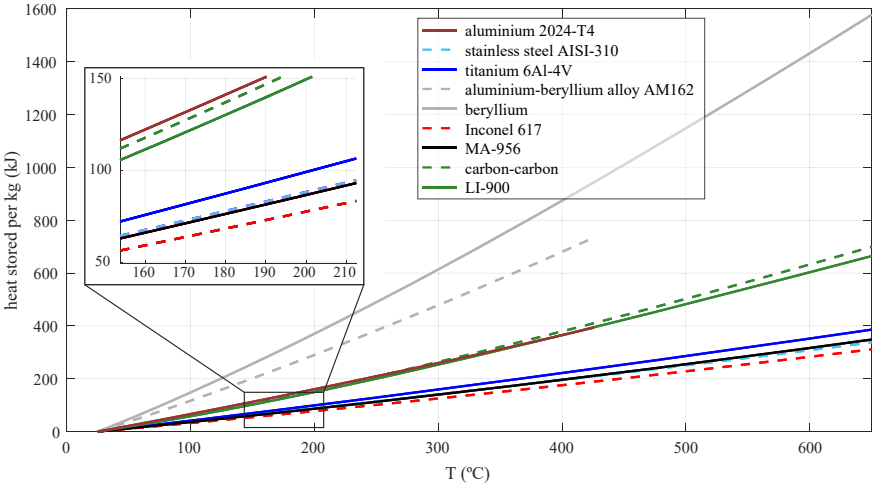


Fig. 10.34 The heat capacity of some aerospace materials in relation to temperature. The data and their sources are provided in the appendix to this chapter

be done is covered by, for instance, McDonnel Douglas (1972) where the structural active cooling and heat sink systems for the Space Shuttle have been studied.

In summary, the heat-sink concept can be used for conditions of moderate heat flux and short flight times at the higher end of the flux profile. Most of the heat will be absorbed by the structure, only a small portion is radiated away.

10.4.2.2 Hot Structures

A hot structure refers to a design concept where a structure is intentionally designed to operate at elevated temperatures. In this context, "hot" signifies that the structure's surface temperature is allowed to rise until the heat being radiated from the surface is equal to the incident heating. The concept is not limited by the duration of the heat pulse, but rather by the maximum acceptable surface temperature of the available materials. A hot-structure concept can be considered when there is a need to handle high heat loads. By allowing the structure to reach higher temperatures, the need for insulation or other cooling mechanisms can be reduced or eliminated. This can result in weight savings, increased structural robustness, and simplified thermal management.

However, designing and implementing hot structures come with various challenges. The selection of materials becomes crucial as they need to withstand the elevated temperatures without compromising their structural integrity. Materials capable of operating at high temperatures often have reduced structural efficiency compared to those used at lower temperatures. Moreover, the structural properties of materials tend to decrease as temperature increases. Non-uniform heating of the structure can

cause thermal stresses and distortions, requiring careful design and thermal analysis. Additionally, ensuring compatibility and thermal management between the hot structure and adjacent cooler components is important to prevent overheating and thermal stress-related issues.

The surface temperature of a hot structure approaches the radiation equilibrium temperature, as discussed in Sect. 10.3.2.1. The temperature of the surface is determined by Eq. (10.27), which considers the surface emittance of the material along with the heating rate (which may include surface catalysis). Figure 10.26 illustrates the relationship between surface temperature and heat flux for different values of surface emittance. By referring to this figure, one can identify suitable materials for a hot structure based on their surface emittance and maximum use temperature.

As an example of a hot structure the honeycomb metallic standoff TPS concept designed by BF Goodrich (formerly Rohr, Inc.) is briefly discussed. This concept was meant to be the TPS for the bottom side of the X-33 (Fig. 10.14). Several reasons led to this particular selection (Jackson and Dixon 1980), operational cost, durability, and maintainability being a few of them. Metallic panels possess innate durability surpassing that of ceramic insulation, as exemplified by their superior ruggedness compared to the ceramic tiles employed on the Space Shuttle. These panels are mechanically affixed, allowing for expedited installation when compared to the adhesive application required for ceramic tiles. Additionally, the design of these panels minimises the disturbance caused when replacing a single panel. Consequently, using metallic panels is anticipated to generate substantial cost savings over the lifespan of the vehicle.

Kontinos and Grant (1998) evaluated the panel bowing in a numerical manner, starting from the notion that a honeycomb panel of which the top plate is severely heated by the incidence flux, whereas the bottom plate under the honeycomb core is much cooler, the panel will bow outward due to the thermal stresses. The problem was amplified as the panel bowing resulted in significant temperature variations on the surface of the panel, inducing lateral temperature gradients that increase the thermal stress within the panel even more. Reusing such a TPS multiple times could potentially lead to thermal fatigue, even though it was found that panel bowing does not appreciably alter the trajectory integrated heat load.

However, in the previous section it was already mentioned that Sudmeijer et al. (2001) found a way to remove the thermal stresses from their sandwich panel altogether. Using PM-1000 face sheets and a PM-2000 core, the sandwich panel was brazed at high temperature and while being stress-free then, after cooling down the panel was pre-stressed due to differences in thermal expansion between the two materials. Any time that the panel was heated to re-entry thermal conditions, the panel was stress-free yet again, and thus no panel bowing occurred.

In conclusion, hot structures are used when the heat flux is moderate, but where the exposure time can be long. The material will be in a steady state of radiation equilibrium, where heat is radiated away. The remaining part of the heat is conducted inward. The structure is in a hot operating condition.

10.4.2.3 Insulated (Cold) Structures

An insulated structure combines the features of a heat sink and a hot structure. Sometimes also mistakenly indicated to be a "cold" structure, because the load-carrying capabilities have been divided over the TPS (the hot part, taking care of the thermal loading) and the underlying structure (the cold part, dealing with the mechanical loading). It is the latter that the term "cold structure" refers to. A schematic representation of the insulated structure can be seen in Fig. 10.35. The heat incident on the surface of the TPS material is rejected in two ways. The outer layer of the TPS is allowed to heat up and store the heat as a heat sink, and Fig. 10.34 can aid in selecting the right material for this to maximise this capacity. The heating up continues until it reaches the equilibrium temperature. Like a hot structure, heat is rejected by radiation, so also now Fig. 10.26 can aid in selecting a material (or a coating) to maximise the emissivity. Finally, the outer surface is insulated to prevent all but a small fraction of the incident heating from reaching the underlying structure. The Space Shuttle Orbiter, for instance, used ceramic tiles and blankets to insulate most of its surface area to limit the temperature of its structure to below 176 °C.

To size an insulating TPS to limit the heat reaching the structure is not a linear process, but requires an iterative, transient thermal analysis. Myers et al. (2000) describe this process of using an implicit, one-dimensional finite-element sizing code with models to account for coatings, fasteners, adhesives, and strain isolation pads, and applied it to an advanced metallic panel, ceramic blanket, and ceramic tile TPS. They verified their approach with results from detailed two-dimensional models and found a good match, confirming that their 1D approach was an accurate approximation.

In general, the insulating TPS does not carry any primary mechanical vehicle loads so that it can be made from low-density or thin-gauge materials to reduce weight. As

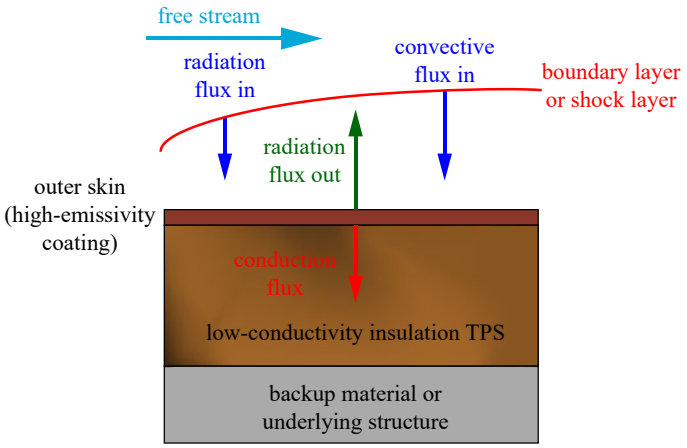


Fig. 10.35 Schematic overview of an insulated structure

a consequence, the outer surface of the TPS may be less robust than the underlying vehicle structure. Considering the insulation, internal multilayer insulation (IMI) can be applied (see Fig. 10.22), which is very similar in its operation as the external multilayer insulation (MLI) by blocking the heat transfer due to radiation, using layers of reflective foil. Compared to the insulation that was foreseen for the X-33 (4.4 kg/m^2), this is much lighter (1.4 kg/m^2) and can handle temperatures up to 1,700 °C.

In summary, an insulated structure is used under moderate heat-flux conditions with relatively short flight times. When the surface emissivity is maximised, most of the heat can be radiated away, whereas a minimum amount of heat is conducted inward. The underlying structure remains cool.

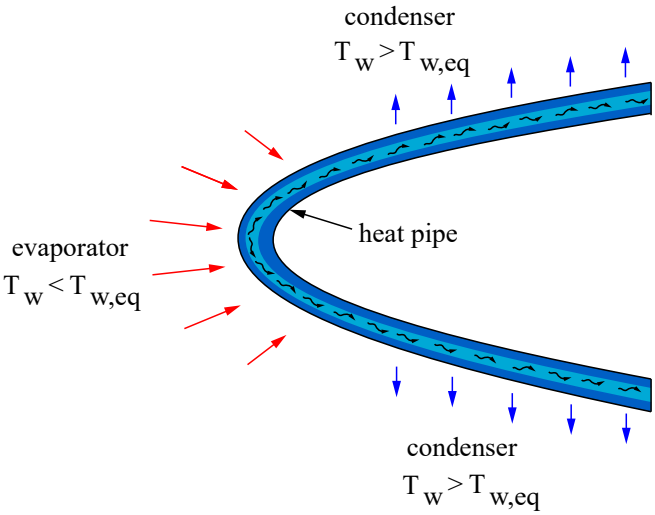
10.4.3 Semi-Passive Systems

10.4.3.1 Heat Pipes

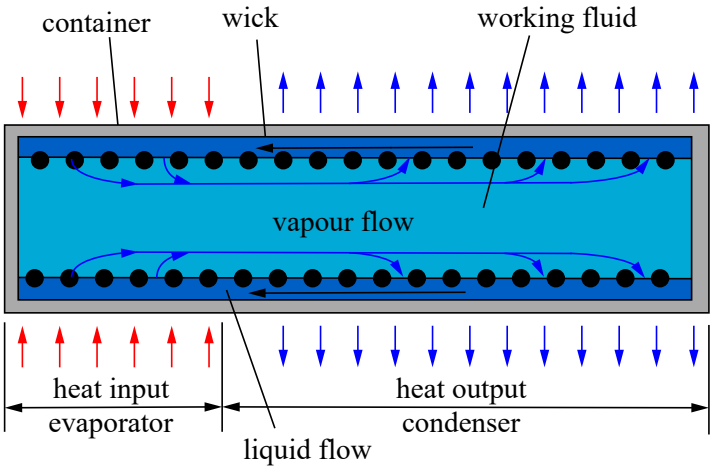
Heat pipes are mainly used for parts on the vehicle that have adjacent areas of high and low heating, for instance, the wing leading edge of a space plane. Heat will be absorbed in the areas where the incoming heat flux is high, to vaporise a working fluid that is sealed in a pipe. Subsequently, the vapour flows to regions with lower pressure and heating, where it condenses while rejecting the heat. Thus, the operating principle of a heat pipe is based on the principles of heat transfer, phase change, and capillary action. The sealed pipe or tube is made of a material that is compatible with the working fluid. Examples are copper pipes with water as the working fluid or aluminium for ammonia heat pipes, but many more combinations can be found in the overview article by Shukla (2015). After sealing the pipe, a vacuum pump is typically used to remove the air from the empty heat pipe. Figure 10.36 shows the schematics of a heat-pipe implementation in the leading edge of a wing (for instance, as foreseen for the X-30, the demonstrator for the National Aerospace Plane single-stage-to-orbit waverider concept conceived in the 1990s), as well as its operating principle.

A heat pipe consists of a sealed metal tube with an inner lining called a wick structure. The wick structure is typically made of a porous material such as sintered metal or grooved metal. The heat pipe is partially filled with a working fluid, usually a low-boiling-point liquid such as water or ammonia. One end of the heat pipe is exposed to a (high) thermal flux, which causes the working fluid in that section of the pipe to evaporate due to its low boiling point. This process of vaporisation absorbs quite a bit of the heat energy and part of the fluid transitions from liquid to vapour. The vapour flows towards the cooler end of the heat pipe, where it condenses back into a liquid phase, releasing the heat energy.

The main operational aspect of the heat pipe is the capillary action of the wick structure. It acts as a capillary pump, facilitating the movement of the liquid from the condenser section to the evaporator section. Capillary action helps to overcome the



(a) Schematic wing leading-edge implementation



(b) Operating principle

Fig. 10.36 Heat-pipe concept

gravitational forces and ensures the continuous circulation of the working fluid. When the liquid condenses in the cooler end, it creates a pressure gradient within the heat pipe. This pressure difference causes the liquid to flow back to the evaporator section, where it is again vaporised by absorbing heat, and the cycle repeats. Contributing to the heat pipe's efficiency is the high heat-transfer coefficient associated with the phase change of the working fluid and the large surface area available for heat exchange. Despite the moving working fluid, heat pipes are primarily considered to be passive devices, as they do not require an external power source. Yet, it can transfer heat over long distances with minimal temperature drop.

The interested reader is referred to the book by Reay and Kew (2006) that provides a detailed overview of how to design heat pipes, and how to analyse their performance. Shukla (2015) provides a(n) (historical) overview of the development and application of heat pipes for aerospace applications, amongst others those for hypersonic cruise vehicles. He also provides several references to heat-pipe experiments on quite a few Space Shuttle flights, as well as one on the International Space Station. Glass and Camarda (1992) discuss the thermal/structural analysis of a carbon-carbon/refractory-metal heat-pipe-cooled wing leading edge, which was meant for integration with the National Aerospace Plane. Their nominal thermal loading was set at 10.3 MW/m^2, and they analysed different aspects of their heat-pipe design: a range for the leading-edge radius in between 1.25 and 5.1 cm (0.5 and 2.0 in.), a heat-pipe length between 0.3048 and 1.2192 m (12 and 48 in.), a C/C wall thickness between 1 and 2 mm (0.04 and 0.08 in.), and a heat-pipe spacing between 0.5 and 2.5 cm (0.25 and 1.0 in.). As one of the results, they found that to keep the C/C leading-edge wall below 1,650 °C, the wall thickness should be below 1.65 mm (for a given leading-edge radius of 1.25 cm).

In summary, heat pipes are used for regions exposed to high heat flux and for a longer period of time. The heat is transferred by a working fluid, and it will condense in cooler areas. The heat will enter those areas and is radiated away. Overall, the heat-pipe structure operates in a hot condition.

10.4.3.2 Ablative Systems

In a way similar to the operating principle of heat pipes, where a phase change of the working fluid takes care of the heat transport, in the case of ablative systems a chemical reaction generates gasses from the solid material, which block a major part of the aerodynamic heating to the vehicle surface. Ablators belong to the category of earliest TPS designs, and was the main focus of development from the 1950s onwards. Applications were found in military re-entry bodies (ICBMs), and the Gemini, Mercury and Apollo capsules, when it was found that metallic heat shields would not be effective against the high thermal loads. Currently, this type of heat shield is still widely used on, for instance, Shenzou, Soyuz, Dragon, and planetary probes like Huygens, MER, MLS, etc.

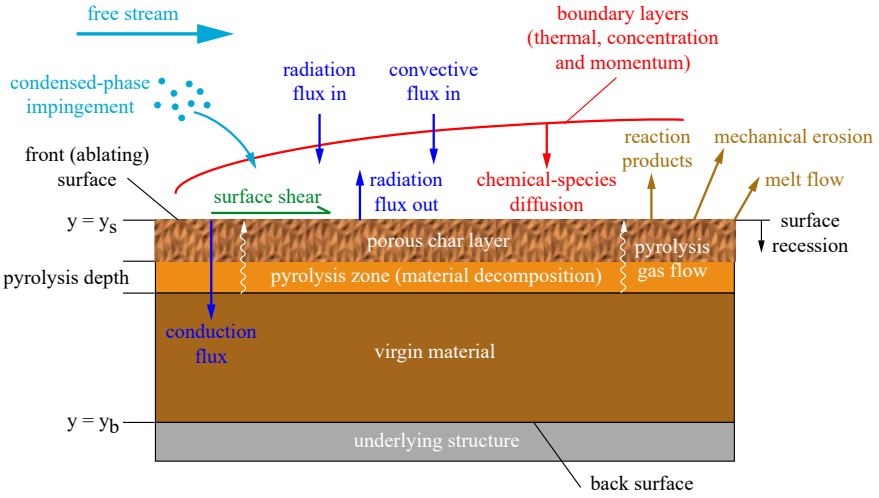


Fig. 10.37 The processes involved with ablation

A schematic representation of ablative TPS can be seen in Fig. 10.37. Hot gases in the boundary layer convectively heat the surface, and additional heat will reach through radiation from the shock layer. When the entry velocity is large, this will be the dominating heating mechanism. Part of the heat will be re-radiated, depending on the temperature and the emissivity of the surface. The remaining heat will be conducted into the surface. The polymer in the composite begins to decompose, a process that already absorbs energy, and pyrolysis gases are formed. Carbon remains and a char layer begins to form.

The thermal front moves through the material, resulting in more decomposition. Pyrolysis gases are forming throughout the composite, with temperatures that are highest near the surface. Thus, these gasses flow from deeper in the composite where they have a lower temperature towards the surface, thereby cooling the char layer. The charred surface reacts with the boundary layer through, for instance, oxidation and ablation, resulting in the removal of material. This causes recession of the surface. After the pyrolysis and gases have been formed at the surface, they penetrate the boundary, which is consequently thickened and, as a result, the convective heating is reduced. This is the so-called blocking effect.

The above is only a simple description of the very complex ablation process. Hurwicz et al. (1972) state that the environmental parameters associated with hypervelocity flight in planetary atmospheres affecting ablation mechanisms are: gas temperature, enthalpy, mode of energy transport, total heat load, peak heating rate, shape of the heat pulse, duration of the pulse, external and internal pressure forces,

aerodynamic shear, particle impingement,[33] mechanical and acoustical vibration, inertial and dynamic forces, and the chemical reactivity. However, to give a full description of all processes involved does not serve the spirit of this section. More information can be found in the many publications available on this topic, the book by Duffa (2013) would be a good starting point.

A modification of the standard ablator concept is the surface-protected ablator, where a layer of C-SiC is applied to the top. This layer protects the burnt and evaporated material from variation in dynamic pressure, while it is protected from excessive heating by the underlying ablator. The advantage of this system is that it can not only sustain higher heat rates, but also higher dynamic pressures. A considerable part of the aerothermal heat load incident to the protective surface layer is directly re-radiated to space, thus lowering the influx to the TPS. The remaining part is transported into the TPS by thermal conduction. The performance of a surface-protected ablator was successfully tested on the MIRKA capsule back in 1995. However, it was concluded that the system is neither reusable nor very cost-effective (Reich and Jahn 1996).

In summary, ablators can accommodate (very) high heat flux, albeit for a relatively short time since the ablator is consumed in the process. Therefore, it is meant for single use only, although the latest designs allow for a major refurbishment due to their tile-wise layout (see Fig. 10.17). The incidence heat is blocked by the ablation products, at the expense of regression of the ablative layer. As all heat is absorbed by this process, the underlying structure remains cool.

10.4.4 Active Systems

10.4.4.1 Transpiration Cooling

The underlying operational principle of both transpiration and film cooling, the latter to be discussed shortly, is similar to that of ablation: a coolant ejected from the surface blocks most of the aerodynamic heating from reaching the structure. Due to this blocking effect, transpiration cooling can handle large heat fluxes. This cooling principle involves the flow of a gas or fluid through a porous medium (integrated into the outermost layer of the TPS), normal to the surface in the outward direction, i.e., into the boundary layer (Fig. 10.38). Due to the inflow of the coolant, the boundary layer will grow by the push of the coolant layer. Initially, the coolant velocity will

[33] Particle impingement, also known as condensed phase impingement is a phenomenon that occurs when particles in a gas stream (or droplets in a fluid) collide with and stick to a surface, as it can happen in the ablation process. The gas stream carries particles that are intended to react with or transfer heat to a surface. However, some of these particles may collide with the surface and stick, forming a layer of material on the surface. This layer can then impede the desired process by reducing the effectiveness of the heat transfer or inhibiting the reaction. Condensed phase impingement can be mitigated through various methods, such as modifying the surface properties to reduce adhesion. Understanding the mechanisms of condensed phase impingement is important while designing, analysing, and optimising an ablative TPS.

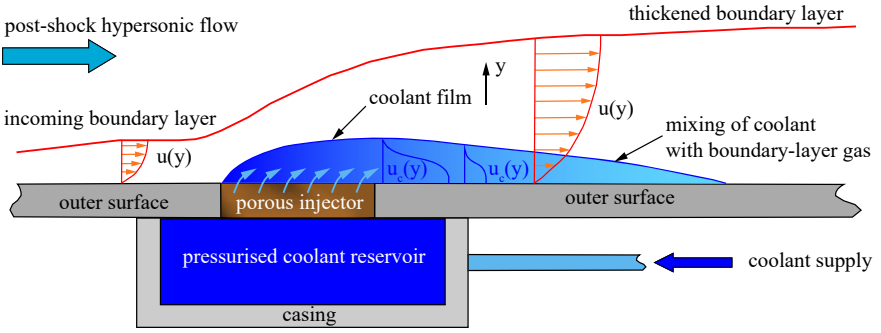


Fig. 10.38 The concept of transpiration cooling; $u(y)$ is the horizontal flow velocity in the boundary layer, whereas $u_c(y)$ is the horizontal velocity of the coolant

be large enough to progress downstream. Shear forces will slow down the coolant downstream, and the cooling effect diminishes, unless new coolant is inserted into the flow at a down-stream location.

The coolant flowing along the surface absorbs part of internal energy, thus affecting the convective heat flux going into the vehicle. Also, the coolant film along the wall has a blocking effect on the heat being transferred to the surface through conduction. Another advantage of the coolant film is that it can aids in reducing the oxidation of the material, for example, when nitrogen gas is used as a coolant for carbon-based materials, which is an important issue for such structures.

One of the advantages of this type of active cooling over other methods is that the cooling effect can be controlled by regulating the mass flow of the coolant through the surface. Böhrk et al. (2015) describe how DLR performed an in-flight transpiration-cooling experiment with the second SHarp Edge Flight EXperiment (SHEFEX II) mission, which proved the feasibility of this concept. SHEFEX II had a faceted nose-TPS consisting of tiles that could be replaced by experiments. One of these experiments, called AKTiV, consisted of a porous tile made of permeable C-C ceramic that was cooled by nitrogen passing through the porous sample. To prepare for the experiment they used their in-house tool HEATS that determines the transient wall heat flux to a transpiration-cooled parallel flat plate under laminar or turbulent flow conditions.[34] Comparing results with a laminar-flow experiment and a CFD solution for a trajectory point of SHEFEX II showed deviations of the wall temperature below 10% and deviations of heat flux below 18%, within an acceptable range to support the flight-experiment design. The predicted coolant mass-flow rate for the experiment was found to be between 0.1 and 1 g/s. The flight experiment confirmed a significant cooling effect.

Other experiments worth mentioning are those by Ifti et al. (2022a), who looked at laminar transpiration cooling in hypersonic flow (Mach = 7), using nitrogen and

[34] The paper by Böhrk et al. (2015) gives a clear description of the mathematical model used, and provides ample references for more information.

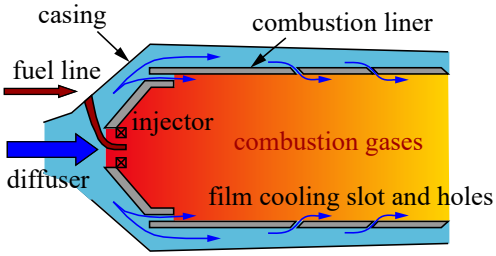
helium as the coolant gas. They confirmed that the cooling film eventually starts to decay monotonically in downstream direction, but observed also that in all experiments a heat-flux reduction was achieved. Finally, it provided insight how the cooling layer mixes with the boundary layer, and the results aided the validation process of their numerical and analytical models. With other experiments, Ifti et al. (2022b) studied the flow characterisation of porous ultra-high-temperature ceramics (UHTC) for transpiration cooling of hypersonic vehicles. The UHTC was confirmed to be an excellent material, because of its high melting point above 3,300 °C. The outflow velocity showed good flow uniformity, and the results validated the so-called Darcy-Forchheimer model for flow in porous media.

In summary, transpiration cooling can be used when high heat fluxes for long duration are expected. The coolant is injected into flow through a porous structure, which then decreases the heat. The remaining flux to the structure will heat it up, so the structure operates in a hot condition. The additional hardware required for this cooling concept is an external pumping system and a reservoir, which, in addition to the (expendable) coolant, adds mass penalties to the TPS. For larger surface areas the additional mass will be too large, thus its use is restricted to smaller, but highly heated areas of the vehicle (for instance, nose tips or small propulsion components).

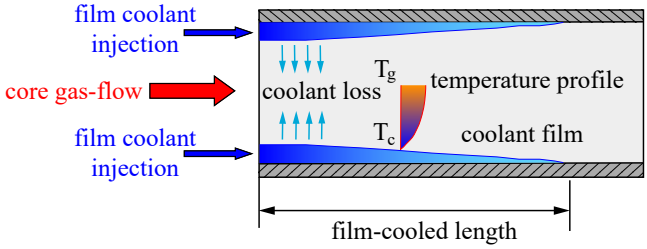
10.4.4.2 Film Cooling

With film cooling, a coolant is injected directly into the flow, a concept that is often used for cooling gas turbines. The injected coolant creates a thin, cool layer on the walls of the vehicle. However, turbulent mixing and heat conduction from the hot gas flow are the cause of a slow decay of this layer, and as a consequence no uniform wall temperature is possible. At sufficient distance away from the injector, the wall temperature will become equal again to the hot-gas temperature, unless at regular distance along the wall more coolant is injected. In Fig. 10.39 a film-cooling configuration for a jet engine is shown, as well as for a rocket engine; one can easily imagine that a similar layout would suffice for cooling the wall of an entry vehicle.

Because of its importance as a cooling mechanism, both inside and outside the aerospace industry, film cooling has been intensively studied. One relevant example is provided by Sahoo et al. (2005), who studied film cooling effectiveness on a large-angle blunt cone flying at hypersonic speed. They performed multiple simulations and experiments, and could indeed confirm a decrease in the heat flux in the cooled region. Using helium as a coolant gas, about 30–45% overall reduction in surface heat transfer was observed (except for the stagnation region). Using other coolants, this reduction in surface heat-transfer rate was between 10 and 25%. On the downside, they found a major drawback for this system as well. The coolant, which is injected in the flow, affects the boundary layer and has a (negative) influence on the aerodynamic force. The drag coefficient increase by 12% with helium injection, whereas with other gases the increase even higher, i.e., about 27%. This increase in drag will obviously reduce the lift-to-drag ratio and affect the flight range and manoeuvrability of the vehicle.



(a) Jet engine



(b) Rocket engine

Fig. 10.39 Film-cooling applications

In summary, film cooling can be used for applications where a high heat flux can be expected, which can occur for long times. The coolant is injected into flow to provide a thin, cool, "insulating" blanket. Despite the cooling effect, the structure operates in hot condition.

10.4.4.3 Convective Cooling

Convective cooling is based on the forced convection of liquid or gaseous coolants to transport the heat from thermally critical surface regions and hot spots to cooler regions. The coolant flows flows through small channels below or within the wall. Since the aerodynamic heating must be conducted through the outer skin, convective cooling cannot accommodate heat fluxes as large as those that can be endured using transpiration and film cooling. McConarty and Anthony (1971) distinguish two types of convective cooling systems: direct and indirect. Direct systems transport the coolant directly to the critical area, from where it is then carried away to other areas of the vehicle. Indirect systems consist of a closed heat transport which transfers heat from the areas to be cooled to heat exchangers. The complexity for indirect systems is higher than that for direct systems. Both systems have been tested and have proven to be effective in reducing the heat load. A disadvantage of the system is that after the cooling the coolant needs to be transported outside the vehicle.

In summary, convectively cooled systems can be used for long time, even when the heat flux is high. The heat is transferred into coolant that heats up, and carries the heat away. Because the heat conducts through the outer wall, the structure operates hot and is thus considered a hot structure. Applications can be found in studies of airframe structures for NASP; as cooled engine nozzles they were used on the X-33 and as such also proposed for the *VentureStar* (Kelly and Blosser 1994). It was also considered in system studies on active TPS for the sub-orbital passenger transport vehicle, SpaceLiner (Schwanekamp et al. 2014). The study by McConarty and Anthony (1971) is the first in a series of three NASA contractor reports on the design of a convective cooling system for a Mach 6 hypersonic transport airframe and wing structures, with the second and third part by Helenbrook et al. (1971) and Helenbrook and Anthony (1971). They provide a lot of insight in the design and analysis of such active TPS.

10.4.4.4 Variants of Active Cooling

Radiation Shielding

In case the surface area to be cooled is large and, thus, the total heat load drives the amount of required coolant to large values, a combination of an insulated structure and active cooling may be considered. Because of the low thermal conductivity of gases the most efficient methods of insulation involve the use of a blanket of air with arrangements for minimising heat transmission by convection and radiation. Under the outer skin a filler material is inserted that is porous, such that in the pores gas can accumulate. While describing the concept, McConarty and Anthony (1971) state that when the pressure is low the conductivity is reduced, and when the insulation material can break up the air space into small cells also the convection is reduced. They further describe an attractive alternative of using the low conductivity of air, i.e., to omit the fill material, such that weight can be saved. This so-called air gap/radiation shield system provides the two parallel walls bounding the air space with highly reflective (low emittance) surfaces to minimise radiative heat transfer, see Fig. 10.40. However, such a system will only be effective when convection is minimal.

In the figure, $\varepsilon_{w,c}$ is the effective emissivity of a group of n radiation shields. For $n = 0$, this means that there is only a gap between the outer heat-shield surface and the inner actively cooled structure, without any radiation shields. The corresponding heat flux from the outer to inner wall is then given by Eq. (10.21), and the effective emissivity is then equal to Eq. (10.20):

$$\varepsilon_{w,c} = \frac{\varepsilon_w \varepsilon}{\varepsilon_w + \varepsilon - \varepsilon_w \varepsilon} \tag{10.33}$$

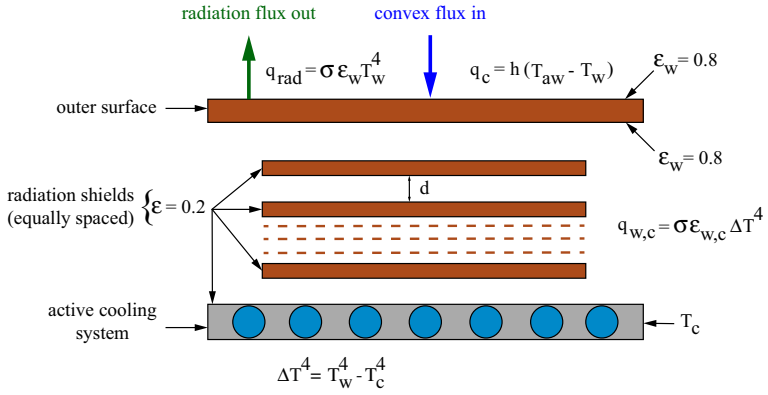


Fig. 10.40 Radiation shielding in combination with an active cooling system. The wall and radiation-shield emissivities have been given as an example; $\varepsilon_{w,c}$ is the effective emissivity taking into account all layers in between the wall and the cooling system

For the cases of $n = 1$ and $n > 1$, McConarty and Anthony (1971) give:

$$n = 1 \quad \Rightarrow \quad \varepsilon_{w,c} = \frac{c_1 c_3}{c_1 + c_3} \tag{10.34}$$

$$n > 1 \quad \Rightarrow \quad \varepsilon_{w,c} = \frac{c_1 c_2 c_3}{c_1 c_3 + c_2 c_3 + c_1 c_2} \tag{10.35}$$

with

$$c_1 = \frac{\varepsilon}{2 - \varepsilon} \qquad c_2 = \frac{\varepsilon}{2 - \varepsilon}(n - 1) \qquad c_3 = \frac{\varepsilon_w \varepsilon}{\varepsilon_w + \varepsilon - \varepsilon_w \varepsilon} \tag{10.36}$$

Nucleate Boiling

Nucleate boiling is a process in which bubbles of vapour form on a heated surface due to the rapid evaporation of a liquid. These bubbles, called nucleation sites, are created when the surface temperature exceeds the saturation temperature of the liquid. As the liquid comes into contact with the hot surface, it undergoes phase change from liquid to vapour, leading to the formation and release of bubbles. Nucleate boiling plays a crucial role in various heat-transfer applications, such as cooling systems and boiling water for cooking. As an active cooling mechanism, nucleate boiling is the process where a coolant is placed directly under the wall of the vehicle. Due to the high temperature of the wall the coolant will start to boil. The bubbles of nucleate will dissipate heat efficiently and cool down the surface.

The vapour bubbles act as tiny heat carriers, effectively removing heat from the surface through the process of evaporation. As the bubbles rise and reach cooler regions, they condense back into liquid, releasing the absorbed heat. This cycle of

bubble formation, detachment, rise, and condensation aids in the efficient removal of heat from the surface, enhancing the cooling process (Gogonin 2023).

Nucleate boiling offers advantages in cooling systems due to its high heat-transfer coefficient, meaning it can transfer heat more effectively compared to other cooling methods. However, it has not been tested yet in aerospace applications. In theory, the method could provide the necessary cooling for the wall of the vehicle, although there are some disadvantages to this system with application in the hypersonic regime. First, the low boiling temperature (for water, for instance, this is 100 °C) and a low temperature at, for instance, the nose results in low radiation cooling. However, this could be solved by applying a coating on the inside of the outer skin to increase the temperature. Second, at the stagnation point superheated pressurised bubbles could form, which could locally increase the temperature. Third, bubbles could merge and form one large bubble. During the re-entry, due to the loads acting on the vehicle the large bubble could shift and cool a different region than the desired one.

Enhanced Radiation Cooling

Enhanced radiation cooling is a relatively new concept in the category of cooled metallic TPS, which was developed at Delft University of Technology (Sudmeijer et al. 2002), was patented by Van Baten and Buursink (2002) in 2001, and extensively tested by Buursink (2005). Whereas traditional TPS only radiates heat outwards to provide insulation to the inner layers, enhanced radiation cooling allows the heat to be radiated inwards as well. As the heat radiates inwards, a coolant evaporates, effectively cooling the TPS. This additional heat-dissipation mechanism increases the allowable heat load for a given TPS temperature, or enables reaching a lower temperature for a specific heat load. The evaporated coolant is expelled from the system, effectively expending the inwardly radiated heat energy. The supply of coolant to the system may be active or passive, depending on the cooling requirements.

Figure 10.41 shows the layout and operating principle of the concept. Several layers can be discerned: the outer surface layer (vehicle wall), which is mainly thermally loaded by the aerodynamic heat loads through convection,[35] an "empty" layer through which the coolant vapour will flow, a porous layer with the coolant (most likely water, due to its high specific heat capacity), and finally the inner structure.

During flight, the outer wall of the TPS heats up and it radiates thermal energy both outwards and inwards. The underlying porous layer, filled with coolant, experiences a temperature rise due to the heat transferred from the outer wall. Once the outer surface of the porous layer reaches the boiling temperature of the coolant, evaporation begins. Simultaneously, heat is conducted through the porous layer until the entire layer, along with the coolant, reaches the boiling temperature. The evaporated coolant is expelled from the vehicle, and capillary action pushes more coolant to the

[35] If in addition to a convective heat flux, also an inward radiative flux is present, because of the high speed of the vehicle, enhanced radiation cooling will most likely not be able to protect the vehicle. The heat load will be too large for the system to carry all heat away. A typical TPS for such cases will most likely be an ablative heat shield.

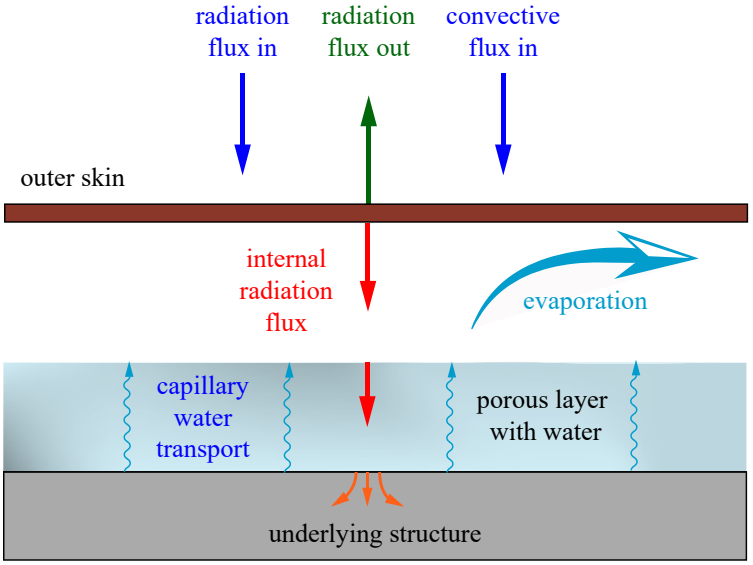


Fig. 10.41 Schematic of enhanced radiation cooling (not to scale)

outer surface of the porous layer. If the coolant is not actively supplied, capillary action continues until the porous layer reaches its critical saturation level. At this point, the cohesive forces of the water in the capillaries exceed the capillary force, inhibiting further transport. As the evaporation plane descends into the porous layer, the temperature of the layer rises above the boiling point, completely drying it out and eventually reaching a temperature equivalent to the outer wall. Consequently, the heat radiated inwards becomes negligible, posing a risk of damage to the outer skin if the excessive heating causes material melting. In such a case, the porous layer directly contacts the environment and serves as insulation. Therefore, once the critical saturation level of the porous layer is reached, the cooling system's useful operation ends. This implies that sufficient coolant should be available for the entire re-entry flight, either through a sufficiently thick porous layer capable of carrying the required coolant mass or via a tank connected by a pipe to replenish the porous layer. The amount of water needed depends on the evaporation rate, which is directly influenced by the temperature difference between the outer wall and the porous layer. A higher outer wall temperature increases the evaporation rate and subsequently the rate of capillary transport.

10.5 TPS Modelling

During the design process, starting from a conceptual level, it is crucial that the interrelation between the sub-systems is carefully analysed, demanding a highly integrated, multi-disciplinary design process. For instance, one assumes a particular vehicle shape, or range of shapes, such that the aerodynamic properties can be established. This, of course, all in relation with the (preliminary) mission requirements. In combination with the system requirements, a preliminary trajectory can be designed, taking path constraints on thermo-mechanical loads into account. At that moment, at best an initial guess of the vehicle mass is known, so if more information about the TPS could be available—a significant fraction of the total vehicle mass—the better it is. Without going to a complicated combination of aerothermal CFD and structural analysis, a first selection of the concept and assessment of the performance in relation to the thermal loads in the stagnation-point is considered. With some scaling of that heat load for other parts of the vehicle a conceptual, yet global evaluation of the TPS can be made. Once a better understanding of the vehicle shape and mission parameters is obtained, more detailed analysis (most notably CFD calculations to get more detail on the actual heat distribution over the complete vehicle) can be performed.

Depending on the experience of the design team, as well as the available software and computational resources some of the above steps may be skipped, so in the next Sect. 10.5.1 the general approach to TPS design is discussed, assuming the availability of (high-fidelity) software. In case this software is not available, it may be sufficient to use so-called TPS mass estimating relationships that have been developed by extensive analysis, see Sect. 10.5.2. In anticipation to the development of an engineering model for TPS design, the following two sections discuss two specific aspects of TPS modelling: Sect. 10.5.3 introduces the specifics of ablation modelling, whereas Sect. 10.5.4 covers modelling aspects for active cooling. Section 10.5.5, finally, combines these two concepts with the standard convection and conduction models to set up a one-dimensional, finite-difference model for quick assessment of TPS designs. Besides the mathematical model, also a detailed verification and validation process is included.

10.5.1 General Approach

The TPS of a re-entry vehicle is a challenging system to model accurately. To approach this problem, CFD software can be utilised to create a three-dimensional discretisation of both the vehicle and the surrounding fluid along its trajectory. While the temperature profiles obtained through this approach are generally reliable, the process requires a significant computational load for each individual vehicle, trajectory, and attitude. If the model also considers non-steady-state thermo-chemical reactions such as material decomposition, charring, pyrolysis, and mass injection effects, the problem becomes even more complex. The next step would then be use the pressure and thermal loading as input to a finite-element (FEM) software to do a coupled thermo-mechanical analysis, taking the proper boundary conditions of wall and underlying structure into account.

To give some insight in this approach, let us briefly review the approach taken by NASA to develop the (ablative) heat shield for their planetary (Mars) landers. This design approach has most recently been applied to the TPS design of Mars Science Laboratory (Wright et al. 2014) and Mars 2020 (Mahzari et al. 2022), and has been fine-tuned over the many years of Mars-mission development. In Fig. 10.42, the flowchart illustrates the process used for overall TPS sizing and margin determination. There is a clear separation between the input preparation and the actual TPS design. To prepare for the latter, nominal trajectories for various design cases are computed, extended with Monte Carlo simulations to evaluate the impact of trajectory, aerodynamic and atmospheric dispersions on these nominal trajectories. Such an analysis is inherent to the mission-design process, but in this particular case the focus is on establishing the bounds on heat-pulse duration and heat-load trajectories for use in the TPS analysis. The heat-pulse duration, in this context, is the time interval when the heat flux starts deviating from the minimum value (typically zero) after entry, reaching the peak value and then reducing again to the minimum value. So, in other words, it is the duration of non-zero heat flux. Note that a worst-case loading is obtained when the boundary layer is considered to be turbulent.

The 3σ bounds provided are used to calculate the design environments with associated uncertainties noted as load factor of safety, at each TPS sizing point. In the end, these bounds will determine the TPS thickness. Edquist et al. (2009) describe this aerothermodynamic design process for MSL, where they simulated the flow field with a Navier-Stokes solver on shallow (maximum total heat load) and steep (maximum heat flux, shear stress, and pressure) entry trajectories. Boundary-layer transitions were included, whereas the effects of distributed surface roughness on turbulent heat flux and shear-stress peaks were considered using empirical correlations. Output of the analysis were the time time-varying conditions at a number of pre-defined heat-shield locations, which were subsequently pushed further down the line for the TPS design and analysis, and in the case of MSL, also for flight-instrumentation development.

The input to the TPS design and sizing analysis can be divided into three parts: (1) the nominal heating is used in combination with the nominal bond-line temperature limit; this is considered the nominal load case; (2) the aerothermal margins (3σ bounds) are considered, but still in combination with the nominal bond-line temperature limit (the so-called *aerothermal margin* or the load factor of safety, FoS); and (3) the nominal heating is considered after the bond-line temperature limit has been lowered by a margin (typically with a 2σ value). For this case, a Monte-Carlo process is used to calculate the impact of material variability and uncertainties in the material response model at each TPS location. Sepka and Wright (2009) describe the developed tool for this process in more detail (the Fully Implicit Ablation and Thermal Response program, FIAT). Many aeroshell heat-shield design considerations can be addressed, such as the maximum bond-line temperature, recession, isotherm depth, surface temperature, and char depth.[36] The reduction in the allowable bond-

[36] The authors present a case study for the MSL back shell and found that the most influential parameters on bond-line temperature dispersion were virgin material thermal density (34%), initial material temperature (20%), virgin material conductivity (17%), and char density (13%).

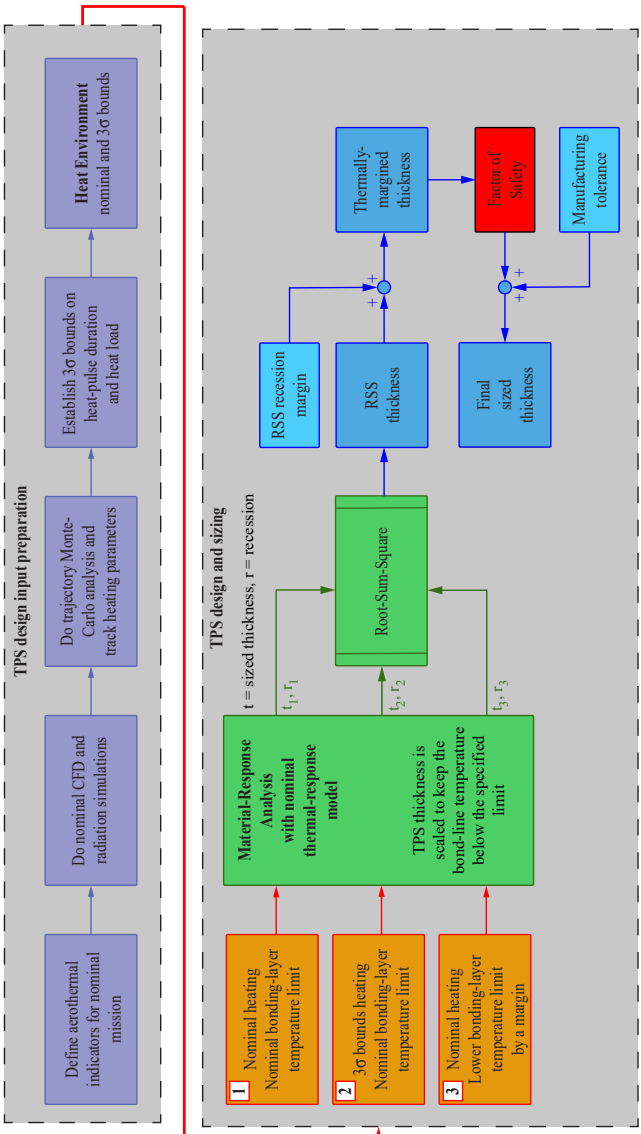


Fig. 10.42 TPS design approach. Adapted from Mahzari et al. (2022)

line temperature that is found with this Monte-Carlo process is referred to as *thermal margin*.

Output from the material-response analysis will be a sized thickness and the amount of recession. Next, the sized thicknesses from the three branches are combined in a Root-Sum-Squared (RSS) formula:

$$t_{RSS} = t_1 + \sqrt{(t_2 - t_1)^2 + (t_3 - t_1)^2}$$

with which the impact of the aerothermal uncertainties and the material variability (thermal margin) on the final TPS thickness can be assessed. Any limitations in the thermal model to accurately predict the recession of the ablative material is accounted for by an RSS recession margin that is added to the thickness. Finally, depending on the program, design approach, and remaining uncertainty between the ground and flight models, a factor of safety is applied (which is set to 1.1 as a multiplication factor for MSL, for instance). With known tolerances for manufacturing the final sized thickness of the TPS is obtained. According to Wright et al. (2014), the discussed process is more rigorous and complex than that employed for previous robotic entry missions (see, for instance, Willcockson (1999a, b)). However, it was experienced that following this approach is more statistically defensible and quantifiable when off-nominal TPS performance effects have to be taken into account.

It is noted that to analyse the ablative process often a 1D analysis is acceptable, since for the majority of the aeroshell the vehicle configuration is uniform and surface heating varies gradually. This may not be true, however, for particular locations where surface heating and structure geometry/configuration vary more abruptly on a smaller scale (such as the heat-shield shoulder or components of the parachute closeout cone at the back shell). In those cases the one-dimensional analysis may be too conservative.

In literature, there are quite a few examples of reducing the modelling complexity by considering two-dimensional and one-dimensional cases. Chen and Milos (2001) propose a two-dimensional approach to the thermal transfer problem and suggest that simplifying the model from three-dimensional to two-dimensional settings could expedite the design process of entry vehicles, while still maintaining a certain level of solution quality.[37] One-dimensional approaches have been studied more extensively, particularly during the early stages of the Apollo program. Swann et al. (1965), Swann and Pittmann (1962), for example, developed a comprehensive one-dimensional finite difference model of a charring ablative thermal protection system. Pittman and Brinkley (1976) applied this model to analyse the transient ther-

[37] In all fairness, this model is part of a larger methodology and toolset, consisting of a high fidelity Navier-Stokes equation solver (GIANTS), a two-dimensional implicit thermal response, pyrolysis and ablation program (TITAN), and a commercial finite-element thermal and mechanical analysis code (MARC).

mal response of a multi-layered insulated TPS. Such an approach would provide quick insight in the TPS performance at the conceptual level.

A relevant example of this is discussed by Ridolfi et al. (2012), and deals with the robust multi-disciplinary optimisation of unmanned entry capsules (in Sect. 10.6 some of the results will be shown as a case study). In their integrated approach an entry vehicle's mission is optimised, not just focussing on the trajectory, but addressing the vehicle's (aerodynamic) shape at the same time. Instead of simply using a maximum stagnation heat flux as trajectory constraint, the transient thermal conditions are taken into account by modelling the TPS and performing a thermal-response analysis. To be able to evaluate many different capsule shapes with a variety of realistic entry trajectories, a relatively fast model is required to analyse the TPS. In Sect. 10.5.5 the mathematics and physics behind such a model, a one-dimensional finite-difference model that can be applied to a variety of TPS models with stacked layers, will be discussed, including several verification cases.

10.5.2 TPS Mass Estimation Relationships

During the conceptual design phase it is not always possible or even efficient to immediately use high-fidelity tools to predict the required TPS thickness of both front heat shield and the back shell. Also, when a one-dimensional finite difference tool is not readily available it may very well go beyond temporal and monetary budgets to develop just that. Having mass estimating relationships (MERs) to provide a reliable, first estimate, would thus be very convenient to have. Sepka and Samareh (2015) discuss the development of such MERs for the TPS of blunt-body, Earth-entry spacecraft, using NASA's high-fidelity FIAT analysis program to generate reference data for a range of capsule shapes, and heating and pressure environments. A total of 840 Earth entry trajectories were simulated and analysed (with V_E ranging from 10 to 16 km/s and γ_E varying between $-5°$ and $-25°$). The capsule under consideration is a 60° sphere-cone shape, a common shape for planetary entry capsules, with a nose radius of 0.352 m, a diameter varying between 0.75 m and 2.25 m, and a vehicle mass ranging from 16 to 159 kg. On top of that a payload mass in between 5 and 30 kg is considered. For the thermal analysis a maximum (cold-wall) heat flux, $q_{c,\max}$, ranging from 1.5 to 37.67 MW/m^2 for the front heat shield and 20–580 kW/m^2 for the aftbody was applied. Corresponding heat loads, Q, were 38.55–344.53 MJ/m^2 and 590–5310 kJ/m^2, respectively.

The materials applied to the forebody are the ablators Phenolic Impregnated Carbon Ablator (PICA) and Carbon Phenolic atop Advanced Carbon-Carbon. For the aftbody, the materials are Silicone Impregnated Reusable Ceramic Ablator (SIRCA), Acusil II, SLA-561V, and LI-900. More specific details about the methodology, the used materials properties, and the detailed results can be found in the cited paper. A summary of the results is listed in Tables 10.4 through 10.6, where it is stressed that all credit goes to Sepka and Samareh (2015) for the excellent work that they did.

In these tables one may find the maximum allowable cold-wall heat flux, $q_{c,\max}$, the recession of the ablator, Δr, the accuracy of the MER with respect to the FIAT values, where the relative 1σ bound is given, the largest under-prediction of the MER thickness relative to the FIAT value, the minimum thickness of the TPS, $t_{\text{MER,min}}$, and finally, the MER equation for the thickness, t_{MER}. It is noted that these MERs have been derived for values of Q given in J/cm^2 and V_E in km/s, giving a thickness in cm. For other units, the curve fits have to be redone. To give a single example of how the curve fits were obtained, Fig. 10.43 shows the FIAT thickness predictions as a function of heat load divided by the entry velocity, $\frac{Q}{V_E}$, as well as the goodness of fit versus the FIAT predicted thickness. From the latter plot, it is clear that for larger thicknesses there is more of an under-prediction. The MER in these plots is shown for the forebody PICA ablator.

10.5.3 Ablation Modelling

Due to the violent process that an ablative heat shield is facing—the high temperatures "consume" the material—it is important that the TPS is sufficiently thick such that the underlying structure is always protected, but also not too thick as this gives a mass penalty. In general terms, the thickness of an ablative heat shield depends on various factors such as the spacecraft's re-entry velocity, flight-path angle, the material used in the heat shield, and the desired level of protection. To estimate the thickness of an ablative heat shield, the following general steps are commonly followed:

1. Determine the maximum temperature, $T_{w,\max}$, that the heat shield will experience during re-entry. This can be done using CFD simulations, empirical data from similar re-entry scenarios, or a (less-accurate) scaling process using stagnation-point data only;

Table 10.4 Mass estimating relationships for forebody TPS of blunt-body, Earth-entry spacecraft (Sepka and Samareh 2015). Note that the unit of $\frac{Q}{V_E^2}$ is in J/cm^2 per km^2/s^2

	PICA	CP/ACC
$q_{c,\max}$ (MW/m^2)	12	300
Δr (cm)	0.6–1.15	0.0002–0.141
$\frac{t_{\text{MER}}}{t_{FIAT}}$ (1σ)	6.3%	7.3%
$\frac{t_{\text{MER,min}}}{t_{FIAT}}$	11.7%	16.6%
$t_{\text{MER,min}}$ (cm)	3.27	2.27
t_{MER} (cm)	$1.8696\left(\frac{Q}{V_E^2}\right)^{0.1879}$	$1.1956\left(\frac{Q}{V_E^2}\right)^{0.2102}$

Table 10.5 Mass estimating relationships for aftbody TPS of blunt-body, Earth-entry spacecraft (Sepka and Samareh 2015). Note that the unit of $\frac{Q}{V_E^2}$ is in J/cm^2 per km^2/s^2

	SIRCA	Acusil II	LI-900
$q_{c,\max}$ (kW/m^2)	100	100	75
Δr (cm)	None	None	None
$\frac{t_{\text{MER}}}{t_{FIAT}}$ (1σ)	8.5%	7.6%	14.0%
$\frac{t_{\text{MER,min}}}{t_{FIAT}}$	15.6%	15.1%	18.7%
$t_{\text{MER,min}}$ (cm)	0.518	0.614	0.686
t_{MER} (cm)	$0.5281\left(\frac{Q}{V_E^2}\right)^{0.5416}$	$0.623\left(\frac{Q}{V_E^2}\right)^{0.5697}$	$q_{c,\max} \leq 100$ kW/m^2 $0.6961\left(\frac{Q}{V_E^2}\right)^{0.656}$ $q_{c,\max} > 100$ kW/m^2 $-0.0306\left(\frac{Q}{V_E^2}\right)^2 + 0.5896\left(\frac{Q}{V_E^2}\right) + 0.6739$

Table 10.6 Mass estimating relationship for aftbody TPS of blunt-body, Earth-entry spacecraft—concluded (Sepka and Samareh 2015). Note that the unit of $\frac{Q}{V_E^2}$ is in J/cm^2 per km^2/s^2

	SLA-561V
$q_{c,\max}$ (kW/m^2)	100
Δr (cm)	None
$\frac{t_{\text{MER}}}{t_{FIAT}}$ (1σ)	8.5%
$\frac{t_{\text{MER,min}}}{t_{FIAT}}$	15.7%
$t_{\text{MER,min}}$ (cm)	0.454
t_{MER} (cm)	$0.0064\left(\frac{Q}{V_E^2}\right)^2 + 0.0961\left(\frac{Q}{V_E^2}\right) + 0.3322$

2. Select a suitable ablative material for the heat shield. Ablative materials are designed to slowly burn away and absorb heat during re-entry, thereby protecting the spacecraft. Common ablative materials include phenolic resins, carbon phenolics, and silica phenolics;
3. Determine the material's ablation rate, $\dot{m}_{char}$, which is the rate at which the material erodes or burns away. This can be done experimentally by subjecting samples of the material to high temperatures and measuring their erosion rate;
4. Calculate the required thickness of the heat shield, t, with $t = \frac{T_{w,\max}}{\dot{m}_{char}} SF$, with SF being a safety factor, which accounts for uncertainties in the estimated maximum temperature and ablation rate. Typical values are in the range 1.5 to 2.

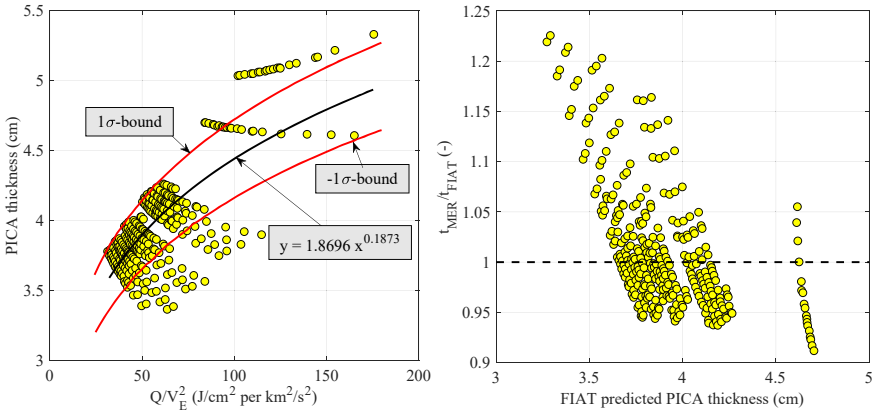


Fig. 10.43 The PICA mass-estimation relationship with 1σ error bounds (left) and goodness of fit (right). Data taken from Sepka and Samareh (2015)

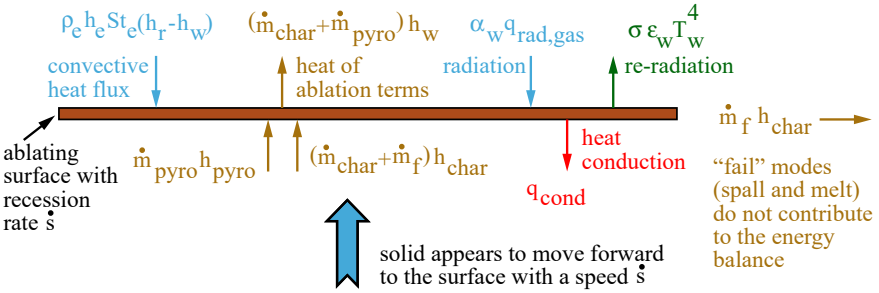


Fig. 10.44 Surface energy balance with ablative processes

5. Finally, validate the estimated thickness using further simulations and tests to ensure that it provides the desired level of protection during re-entry.

In addition, the actual thickness of an ablative heat shield may depend on specific design requirements and constraints of the vehicle.

The first step mentions to establish the maximum temperature from either CFD or empirical data. From the beginning of this chapter onwards it may have become clear that this is not a trivial exercise, when accurate data for the complete vehicle surface are required. This holds even more so for the analysis of an ablative heat shield, as there are complicated chemical processes involved. Figure 10.44 summarises the surface energy balance.

Given the overview of incoming and outgoing energy, the following formulations hold for the individual components:

1. Convective heat flux: following Eq. (3.287), the convective flux to the surface is formulated as $q_c = St_H \rho_e u_e (h_r - h_w)$, where St_H is the Stanton number for

heat transfer, ρ_e and u_e are the density and velocity at the edge of the boundary layer, h_w is the enthalpy of the wall, and h_r is the so-called recovery enthalpy.[38] This enthalpy refers to the transfer of heat energy between two fluid streams during a heat exchange process.[39] Note that in this case the Stanton number for mass transfer is assumed to be equal to that of heat transfer, so $St_M = St_H$.

2. Radiative heat flux: this flux, $q_{rad,gas}$, is an additional input and comes from the hot gasses surrounding the vehicle, see the general formulation by Tauber and Sutton, Eq. (3.331). Depending on the surface absorptance, α_w, the flux input is given by $\alpha_w q_{rad,gas}$.
3. Reradiated radiative heat flux: for a given wall temperature, T_w, the outgoing radiative heat flux is given by $q_{rad} = \sigma \varepsilon_w T_w^4$, with ε_w being the emittance of the wall.
4. Heat of ablation terms: the energy of the gases leaving the surface is due to the combined mass flow of the charring products, subscript $char$, and the gases from the pyrolysis process, subscript $pyro$, so $(\dot{m}_{char} + \dot{m}_{pyro})h_w$. A second energy component coming from the charring material enters the surface from below, and is given by $\dot{m}_{char} h_{char}$. Similarly, a third component comes from the pyrolysis gases and also enters the surface from below: $\dot{m}_{pyro} h_{pyro}$. Note that the unit of $\dot{m}$ is kg/(s m^2).
5. Conduction into the material: any hotter surface will conduct energy to areas with lower temperatures, unless those areas are perfectly insulated. Assuming there is no such insulator (or a vacuum), the conduction heat term is simply given by q_{cond}, and follows directly from Fourier's Law: $q_{cond} = -k \frac{\partial T}{\partial y}$, assuming that conduction takes place in one direction (and thus perpendicular to the layers of the TPS).
6. Note that the so-called fail products due to melting (the heat of fusion is not very significant) and spallation (the crumbling of material into smaller parts, which means a mechanical material loss with minimal energy accommodation) do not contribute (much) to the surface energy balance, and are therefore ignored.

With the above terms, the following surface energy balance applied to an ablative heat shield can be formulated (Milos et al. 2006):

$$St \rho_e u_e (h_r - h_w) + \dot{m}_{char} h_{char} + \dot{m}_{pyro} h_{pyro} + \alpha_w q_{rad,gas} = (\dot{m}_{char} + \dot{m}_{pyro}) h_w + \sigma \varepsilon_w T_w^4 - q_{cond} \tag{10.37}$$

[38] This is different from Eq. (3.287), where an adiabatic wall was considered. The ablation process is *not* adiabatic.

[39] The recovery enthalpy in ablation refers to the enthalpy gained by the gas as it passes over the ablating surface. This enthalpy gain occurs due to the energy released by the surface as it undergoes ablation. The ablating material absorbs heat from the gas, causing an increase in the gas enthalpy. This recovered enthalpy can have an impact on the heat-transfer characteristics and overall energy balance of the system.

The physics behind material ablation is very complex and difficult to model correctly. In the 1960s a lot of effort was put in the development of approximate methods to model this process. Rindal (1968), for instance, presents an approach for characterising the charring ablator response with in-depth coking reactions, part of a larger study on the coupled chemically reacting boundary layer and charring ablator. Coking, in this context comes from one of the three layers identified in the ablation process, starting from the uncharred, virgin material towards the surface: (i) the initial pyrolysis leads to the decomposition of the gas including the thermal decomposition of high-molecular-weight hydrocarbons and dissociation of, for instance, CO_2, H_2O, and H_2, (ii) the further decomposition of the hydrocarbons resulting in precipitation of carbon (the process called coking) on the adjacent char passages resulting in the buildup of the char density, and (iii) the chemical erosion of the char layer by the gases, including carbon vaporisation resulting in a char density reduction near the heated surface.

The description of ablation modelling is in anticipation of the one-dimensional finite difference model that will be presented later on. In light of the engineering spirit of that model, the ablation process will not be modelled in full detail. Nonetheless, the two main phenomena related to the ablation will be taken into account, i.e., the pyrolysis, at the interface between charred and uncharred material, and the charring at the external surface. Such a model has been clearly described by, for instance, Swann and Pittmann (1962) and Swann et al. (1965), stemming from the early days of the Apollo era. In more recent times, this model has been adopted by, for instance, Delgado Montes (1988) and Ridolfi et al. (2012), and, as such, it will be summarised below. It is noted that Swann et al. (1965) provide many references for the individual steps in the derivation; these references are not repeated here, but the interested reader is referred to the report by Swann et al. instead.

To have a global overview of the implementation, Fig. 10.45 gives a schematic layout of a charring ablator with the coordinate system used. The (simplified) representation of the charring ablator comprises of various layers: a char layer, a pyrolysis zone, an uncharred layer, and an insulation layer. The front surface of the char layer is subjected to heat and interacts with the external boundary layer. This leads to erosion of the char layer. The uncharred layer undergoes degradation at the pyrolysis zone, where pyrolysis products permeate through the char layer, absorbing energy and engaging in chemical reactions, both homogeneous and heterogeneous. If heterogeneous reactions are involved, deposition or internal erosion can occur, thereby affecting the density of the char layer. Consequently, an accurate description of the char layer, acting as a porous solid through which pyrolysis gases flow, necessitates the inclusion of conservation equations for energy in both the solid and gas phases, gas momentum, chemical species, total mass, and an equation accounting for variations in porosity.

In this analysis, the pyrolysis zone is considered as a plane with negligible thickness, which was found to be a satisfactory assumption (analysis done in the 1960s showed agreement between numerical solutions for char-layer temperatures obtained using a plane pyrolysis zone and those obtained through in-depth pyrolysis). The assumption of a plane pyrolysis zone was crucial in providing a second boundary condition for the gas-momentum conservation equation.

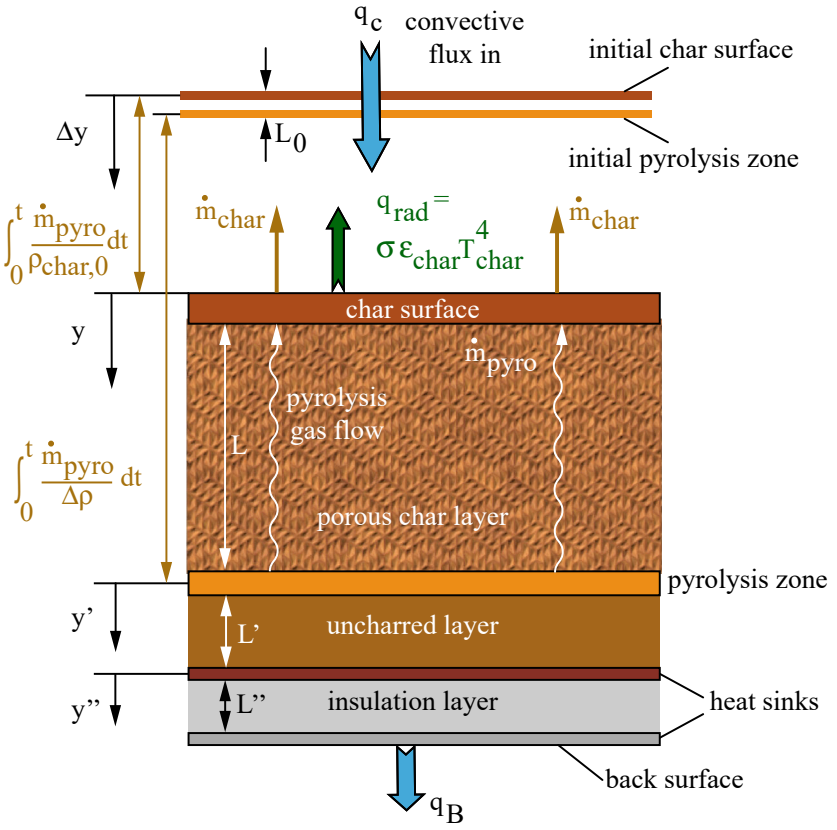


Fig. 10.45 Layout of the charring ablator

For the ablation process to begin, two cases can be considered. The first case considers the absence of ablation when the surface temperature is below the ablation temperature, i.e., $T_w < T_{char}$, such that the ablation rate is also zero, $\dot{m}_{char} = 0$. Then, after the ablation has started it continues to do so at the fixed temperature of $T_w = T_{char}$, as it is assumed that the ablation process consumes so much energy that a higher temperature will be lowered to T_{char}. The second case considers a char mass-loss rate that relates to T_w as follows:

$$\dot{m}_{char} = Ae^{-\frac{B}{T_w}} \tag{10.38}$$

which seems physically logical, as the ablation reactions happen more rapidly with increasing temperature. Here, A and B are constants, corresponding to a specific reaction rate and activation energy, respectively.

Properly selecting A and B yields results similar to those obtained by specifying T_{char}. The above expression needs to be adjusted slightly to take the oxidation of

the char process into account, as it was confirmed experimentally to be an important mechanism. So, Eq. (10.38) changes to

$$\dot{m}_{char} = Ae^{-\frac{B}{T_w}}\sqrt{C_{O_2,w}\,p_w} \tag{10.39}$$

where $C_{O_2,w}$ is the oxygen concentration by weight at the wall. p_w is the pressure at the wall, which, for hypersonic flow, can be related to the stagnation heat rate and enthalpy. The stagnation pressure in hypersonic flow can be approximated by

$$p_{w,s} = \frac{11}{12}\rho_\infty v_\infty^2 \tag{10.40}$$

with ρ_∞ and v_∞ being the free-stream density and speed. The proportionalities of heat rate and enthalpy are given by

$$q \propto \sqrt{\frac{\rho_\infty}{R_N}}v_\infty^3 \quad \text{and} \quad h_e \propto v_\infty^2 \tag{10.41}$$

with R_N being the effective nose radius, such that $p_{w,s}$ can be expressed with

$$p_{w,s} = GR_N\left(\frac{q_s}{h_{e,s}}\right)^2 \tag{10.42}$$

where, of course, a constant of proportionality, G, had to be introduced (with $G = 56.2\,\mathrm{m^3 s^2 atm/kg^2}$). Multiplying both sides with p_w and shifting $p_{w,s}$ to the right side yields:

$$p_w = G\frac{p_w}{p_{w,s}}R_N\left(\frac{q_s}{h_{e,s}}\right)^2 \tag{10.43}$$

Here, $\frac{p_w}{p_{w,s}}$ depends on the vehicle attitude and location at the body.

The rate at which char is removed by oxygen must be proportional to the net rate at which oxygen diffuses to the surface, which is represented by a parameter λ:

$$\dot{m}_{O_2} = \frac{Le^{0.6}q_{c,net}}{h_e - h_w}(C_{O_2,e} - C_{O_2,w}) \tag{10.44}$$

where $C_{O_2,e}$ is the oxygen concentration by weight external to the boundary layer and Le is the Lewis number.[40] The Lewis number can be computed with

$$Le = \frac{\alpha}{D_m} = \frac{k}{\rho D c_p} \tag{10.45}$$

[40] The Lewis number is defined as the ratio of thermal diffusivity to mass diffusivity, and relates to fluid flows with simultaneous heat and mass transfer. The Lewis number puts the thickness of the thermal boundary layer in relation to the concentration boundary layer.

Here, the following variables have been used: α is the thermal diffusivity, D_m is the mass diffusivity, k is the thermal conductivity, ρ is the material density, D is the mixture-averaged diffusion coefficient, and c_p is the specific heat capacity at constant pressure.

Furthermore, $q_{c,net}$ is the hot-wall convective heating rate corrected for transpiration:

$$q_{c,net} = q_c\left(1-\frac{h_w}{h_e}\right)\left\{1-(1-\beta)\left[0.724\frac{h_e}{q_c}\left(w_{char}\dot{m}_{char}+w_p\dot{m}_{pyro}\right)\right.\right.$$
$$\left.\left.-0.13\left(\frac{h_e}{q_c}\right)^2\left(w_{char}\dot{m}_{char}+w_p\dot{m}_{pyro}\right)^2\right]-\beta\eta\left(w_{char}\dot{m}_{char}+w_p\dot{m}_{pyro}\right)\left(\frac{h_e}{q_c}\right)\right\} \tag{10.46}$$

where w_{char} and w_p are the weighting factors for transpiration effectiveness of char mass loss and pyrolysis products, respectively; η is the transpiration coefficient.

Understanding what the terms in the above equation represent is important, so in $q_c\left(1-\frac{h_w}{h_e}\right)$ the cold-wall convective flux without mass flow, q_c, is multiplied by the hot-wall correction $\left(1-\frac{h_w}{h_e}\right)$. The terms within the curly brackets represent the so-called aerodynamic blocking due to the combined mass flows of charred and uncharred material: $\dot{m}_{blocking} = w_{char}\dot{m}_{char} + w_p\dot{m}_{pyro}$. Using transpiration theory, Swann et al. (1965) derived a second-order approximation, which implies that $\beta = 0$. This simplifies the expression for $q_{c,net}$:

$$q_{c,net} = q_c\left(1-\frac{h_w}{h_e}\right)\left\{1-\left[0.724\left(\frac{h_e}{q_c}\right)\dot{m}_{blocking}-0.13\left(\frac{h_e}{q_c}\right)^2\dot{m}_{blocking}^2\right]\right\} \tag{10.47}$$

Combining Eqs. (10.39) and (10.44) allows for the elimination of $C_{O_2,w}$, the concentration of oxygen at the wall, such that the rate of char removal by oxidation is found to be:

$$\dot{m}_{char} = \frac{1}{2}\left\{-\frac{(h_e-h_w)K^2p_w}{q_{c,net}\lambda Le^{0.6}}+\sqrt{\left[\frac{(h_e-h_w)K^2p_w}{q_{c,net}\lambda Le^{0.6}}\right]+4\,K^2p_wC_{O_2,e}}\right\} \tag{10.48}$$

where $K = Ae^{-\frac{B}{T_w}}$.[41] Note that Eq. (10.48) is easily found by rewriting Eq. (10.44) as

$$C_{O_2,w} = C_{O_2,e} - \frac{\dot{m}_{char}(h_e - h_w)}{\lambda Le^{0.6} q_{c,net}} \tag{10.50}$$

and substituting this into Eq. (10.39). Due to the square root involved, left and right side are squared to give a quadratic equation in $\dot{m}_{char}$ of the form $a\dot{m}_{char}^2 + b\dot{m}_{char} + c = 0$. Taking the positive root to force a positive $\dot{m}_{char}$ will result in Eq. (10.48).

When the charring process begins, as a consequence the char surface will recede (see Fig. 10.45 once more). The distance between the current and the initial surface location is given by

$$\Delta L = \int_0^t \frac{\dot{m}_{char}}{\rho_{char}} \mathrm{d}t \tag{10.51}$$

The thickness of the char layer at any point in time, L, is then equal to the initial char thickness, L_0, plus the thickness of char formed by the pyrolysis process, but minus the thickness of the char removed. Thus

$$L = L_0 + \int_0^t \frac{\dot{m}_{pyro}}{\rho_{unchar} - \rho_{char}} \mathrm{d}t \tag{10.52}$$

Similarly, it is easily understood that the instantaneous thickness of the uncharred material, L', is

$$L' = L'_0 - \int_0^t \frac{\dot{m}_{pyro}}{\rho_{unchar} - \rho_{char}} \mathrm{d}t \tag{10.53}$$

In Eqs. (10.52)–(10.53), $\dot{m}_{pyro}$ is the pyrolysis mass rate, which is the last variable to be determined. Clark (1973) proposes to use a similar expression as Eq. (10.38):

$$\dot{m}_{pyro} = A' e^{-\frac{B'}{T_{pyro}}} \tag{10.54}$$

---

[41] In case a first-order (= linear) oxidation reaction had been assumed, which would have implied $\beta = 1$, then the resulting removal of char by oxidation would be

$$\dot{m}_{char} = \frac{K p_w C_{O_2,e}}{1 + \frac{K p_w (h_e - h_w)}{q_{c,net} \lambda Le^{0.6}}} \tag{10.49}$$

Note that when $\beta = 1$, the aerodynamic blocking effect is represented by $\eta \dot{m}_{blocking} \left(\frac{h_e}{q_c}\right)$, rather than the term indicated in Eq. (10.47).

where A' and B' are constants similar to A and B, corresponding to a specific reaction rate for pyrolysis of the uncharred material (in kg/(m^2 s atm)), and the activation energy for pyrolysis of the uncharred layer in K. Values for A' and B' are not easy to find, as they follow typically from measurement campaigns. However, for comparison, Clark provides values for low-density phenolic-nylon ablation material: $A = 4.9 \cdot 10^{10}$ kg/(m^2 s atm) and $B = 4.25 \cdot 10^4$ K, and $A' = 7.74 \cdot 10^6$ kg/(m^2 s atm) and $B' = 1.289 \cdot 10^4$ K. This shows that because A' is about four orders of magnitude smaller than A, the mass flow from the pyrolysis is proportionally lower than the char mass flow.

The implementation of the introduced ablation model as a one-dimensional finite-difference model is further detailed in Sect. 10.5.5, where the surface-energy balance will be properly taken into account. A final note on this energy balance, though. In several older publications, the heating-up the uncharred material is a combustive process, adding a component to the incoming flux. Swann et al. (1965) define this term to be

$$q_{comb} = [1 - \mathcal{S}(T_w - T_{char})]\, \dot{m}_c \Delta H_{comb} \tag{10.55}$$

with ΔH_{comb} being the heat of combustion of the char. Similarly, on the outgoing side of the balance the heat of sublimation of the char layer is given by

$$q_{char} = \mathcal{S}(T_w - T_{char}) \dot{m}_{char} \Delta H_{char} \tag{10.56}$$

with ΔH_{char} being the heat absorbed by vaporisation of the char. Further, in these equations, $\mathcal{S}(T_w - T_{char})$ is a step function, which value is 0 when $T_w < T_{char}$ and 1 when $T_w \geq T_{char}$, such that only one term can be active. In Eq. (10.37), the combustion term is not included on the left-hand side, as it is assumed that $\dot{m}_{char}$ is zero for temperatures below the ablation temperature. The sublimation heat, q_{char}, is present on the right-hand side as a combined term $(\dot{m}_{char} + \dot{m}_{pyro})h_w$. For information, Clark (1973) provides separate values for ΔH_{comb} and ΔH_{char} (low-density phenolic nylon ablation material), being $1.20 \cdot 10^7$ and $5.00 \cdot 10^7$ J/kg, respectively.

In summary, and including the terms that were introduced in the ablator model, the surface-energy balance is given by

$$q_{c,net} + \dot{m}_{char} h_{char} + \dot{m}_{pyro} h_{pyro} + \alpha_w q_{rad,gas} = (\dot{m}_{char} + \dot{m}_{pyro}) h_w + \sigma \varepsilon_w T_w^4 - q_{cond} \tag{10.57}$$

with $q_{c,net}$, including the blockage effect, given by Eq. (10.47). The constants that appear in some of the equations have been gathered from literature for three ablative materials, i.e., phenolic nylon (Clark 1973), carbon phenolic, adapted from the work of Sutton (1970), and PICA (Tran et al. 1996; Parmenter et al. 2001), and listed in Table 10.7. Additional thermal properties can be found in the appendix to this chapter.

This section will conclude with a simple ablation model that is sometimes—but should not be—used for a coarse estimate of the required thickness (and thus mass) of an ablative heat shield.

Example 10.1 (*Simple Ablation Model and Why Not to Use It*) In conceptual design one often does not have access to elaborate software models, or the time to even develop simpler models such as the finite-difference model to be discussed later. A coarse analysis to establish the thickness and mass of the heat shield is often the only option available. In this example, such an analysis is introduced, and it is shown that such an analysis has to be used with caution, and perhaps even avoided. The approach of using mass-estimation relationships (MERs) (Sect. 10.5.2) is in that case a safer approach.

The first part of this example is based on the carbon-phenolic TPS design by Sutton (1970), which is also used later for validating the one-dimensional finite-difference model. Thus, it may give additional insight in the accuracy of the simplified approach. The original design concerns a TPS of approximately 3.81-cm thickness, and the finite-difference model assumes having a charred, uncharred, and insulator layer. The design is thermally loaded with a constant heat-flux of 1.44 MW/m^2.

The simplified approach is based on how the total amount of heat can be absorbed by a certain mass m_{TPS} of TPS, thereby ignoring any chemical reactions as well as internal conduction and convection processes:

$$m_{TPS} = \frac{QS_{TPS}}{c_p \Delta T} \tag{10.58}$$

Table 10.7 Ablative materials thermo-physical parameters

Parameter	Unit	Phenolic Nylon	Carbon Phenolic	PICA
ρ_{unchar}	kg/m^3	553	1,392	266
ρ_{char}	kg/m^3	143	1,184	210
ε	–	0.8	0.7	0.9
A	kg/(m^2 s bar)	$4.9{\cdot}10^{10}$	$4.9{\cdot}10^{10}$	$4.9{\cdot}10^{10}$
B	K	42,500	42,500	22,500
D	m^2/s	0.85	0.85	0.85
T_{char}	K	1,872	3,030	3,700
T_{pyro}	K	945	945	1,973
Δh_{subl}	MJ/kg	50	20.88	250
Δh_{pyro}	MJ/kg	1.28	0.465	0.5
$C_{O_2,e}$	–	0.23	0.23	0.23
λ	–	0.75	0.75	0.75
h_e	MJ/kg	3.1	3.1	3.1
w_{char}	–	0.2	0.2	0.2
w_{pyro}	–	0.1	0.1	0.1

where Q in J/m^2 is typically obtained by integrating the stagnation-point heat flux. Further, S_{TPS} is the TPS surface area exposed to Q, c_p is the specific heat capacity in J/(kg K), and ΔT is the total temperature increase. Effectively, this means that the TPS is considered to be a heat sink, and ignoring the fact that the ablation process takes away part of the incoming energy. Note that in Eq. (10.58), S_{TPS} could simply be omitted, since the flux is assumed to be uniformly distributed. In that case, the expression will provide the mass per square meter.

A few points need to be clarified. Heat-flux models used during conceptual design are typically the cold-wall Chapman model (see Eq. (10.1)), or, at best, the hot-wall radiation-equilibrium model, which effectively lowers the wall temperature by some 10–15%. The wall temperature can be calculated using Eq. (10.4):

$$q_c = q_{rad} = \varepsilon\sigma T_w^4 \quad \Rightarrow \quad T_w = \left(\frac{q_c}{\varepsilon\sigma}\right)^{0.25} \tag{10.59}$$

Assuming cold-wall equilibrium, for the given (constant) heat flux of $q_c = 1.44$ MW/m^2, $T_w = 2{,}454$ K. Here, $\varepsilon = 0.7$ for carbon phenolic.

Another point of attention is that the heat flux is not constant over the heat shield. For a ballistic flight, it is typically assumed that the stagnation point is in the middle of the shield. With non-zero angle of attack, this point shifts. In Fig. 10.46, it has been schematically indicated how the heat-flux variation could be distributed, based on a CFD analysis done by Marquardt et al. (2022). They analysed the entry of the 3.5 m diameter EURASTROS capsule at low and high hypersonic Mach numbers, with an angle of attack of 25°. Over the six regions indicated, the heat flux scaled with factors ranging from 1 (stagnation area 1) to 0.5 (area 6) for the low Mach number and 1 down to 0.7 at the high Mach number (peak heat-flux condition). With such scaling and the corresponding surface areas, a location-dependent equilibrium temperature can thus be calculated and for each area a corresponding TPS mass. Of course, this remains an approximation, but it will be more accurate.

However, there is more to it. Assuming that the stagnation point is the hottest point is not always correct. Figure 10.20, for instance, shows that the hottest point is, in fact, close to the shoulder of the heat shield of MSL. Two reasons can be identified: in the first place, the assumption of laminar flow did no longer hold, as the boundary layer transitioned to turbulent flow, yielding an immediate jump in heat flux. In the second place, the radius of curvature was not constant: close to the edge, the corner has a much smaller radius, and since the flux is inversely proportional to the reference radius, this resulted in another flux increase. Both effects combined made sure that the stagnation area is actually a region that is much less thermally loaded.

Let us now get the example started. The constant heat flux considered is $q_c = 1.44$ MW/m^2, with a test duration of 60 s. The total integrated heat load is thus $Q = 86.4$ MJ/m^2. During the test the heat flux was uniformly distributed, so there is no distinction between the stagnation area and other, cooler ones. The charring temperature of carbon phenolic lies between 750 and 925 K. For the sake of the example it assumed that charring will begin at $T_{char} = 854$ K. This means that $\Delta T = 1{,}600$ K (for the equilibrium wall temperature of $T_w = 2{,}454$ K), and for tem-

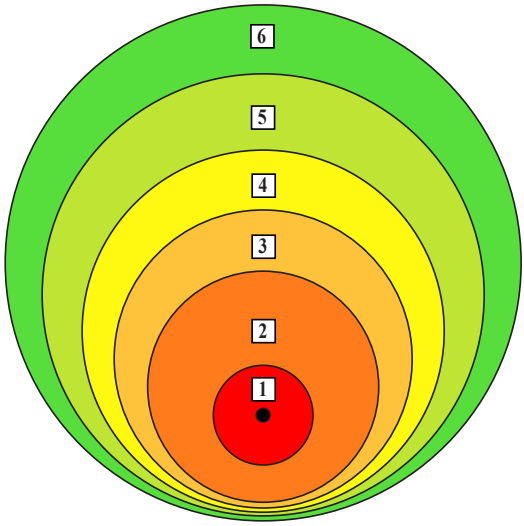


Fig. 10.46 Schematic areas of constant heat flux, based on CFD analysis by Marquardt et al. (2022)

peratures lower than 854 K the heat shield will simply reject the heat. The specific heat coefficient for the charring material at $T = 2,454\,\text{K}$ is $c_p = 2,168.5\,\text{J/(kg K)}$ (see the appendix for the material data).

Substituting the relevant numbers into Eq. (10.58) yields $m_{TPS} = 24.9\,\text{kg}$ for a normalised surface area of $1\,\text{m}^2$. With a density of uncharred carbon phenolic of $\rho = 1,392\,\text{kg/m}^3$, this gives a volume of $0.0179\,\text{m}^3$ ($= 17.9\,\text{L}$). Converting this to a thickness, where the curved heat shield is approximated by a flat disc, a thickness of 18 mm is found. Even though Sutton (1970) used a 31.8 mm sample for their experiments, the recession distance was less than 1 mm (see also the validation data, Fig. 10.58). This shows already a severe limitation of this approach. A final comparison was made with the MER listed in Table 10.4, i.e., $t_{MER} = 1.1956 \left(\frac{Q}{V_E^2}\right)^{0.2102}$, which resulted in $t_{MER} = 12.9\,\text{mm}$. Also in this case quite an overshoot is found, but this could also be because the input heat flux is on the very low end of the validity range.

A final evaluation will be done for a Stardust-type of entry, where both the standard version of Eq. (10.58) will be used, as well as the approach based on Fig. 10.46. The entry will be ballistic, with an entry velocity of $V_E = 12.6\,\text{km/s}$, and the capsule has an ablative heat shield made of PICA. Olynick et al. (1997) have done an extensive analysis of the Stardust heat shield, and in this example the same load case is considered. Figure 10.47 shows the thermal loads. Two different load cases are shown, i.e., one with ablation using some of the incoming energy, and one without ablation. The curves have been obtained by spline-fitting the original data points. For both cases, the equilibrium wall temperature is shown as well, with the remark that Olynick et al. (1997) report a maximum wall temperature of 3,380 K (with ablation) and 3,735 K (without ablation), respectively. To be consistent with their TPS thick-

ness calculation, the computed cold-wall temperatures have been scaled down by 11 and 12%, respectively. It is noted that the total integrated heat load for the ablative case is about 280 MJ/m^2, with a radiative component (by the surrounding gas) of about 7% of the total.

The approach followed to calculate the TPS thickness is as follows. The (discrete) heat-load and temperature profiles (as a function of time) are parsed,[42] and for each time interval the heat-load and temperature increments, ΔQ and ΔT are determined. Since the specific heat coefficient is a function of temperature, for the current temperature c_p is determined and a mass increment Δm_{TPS} is obtained (the TPS surface area is $S_{TPS} = 0.554\,\mathrm{m}^2$ for a diameter of $d = 0.84\,\mathrm{m}$). Doing this exercise, thereby assuming that the heat load is constant over the surface yields a total mass of $m_{TPS} = 51.86\,\mathrm{kg}$. With a density of the uncharred material of $\rho_{PICA,uncharred} = 266\,\mathrm{kg/m}^3$) and the given surface area this results in a thickness of $t_{TPS} = 35.2\,\mathrm{cm}$. The corresponding MER, $t_{MER} = 1.8696\left(\frac{Q}{V_E^2}\right)^{0.1879}$ gives $t_{MER} = 2.1\,\mathrm{cm}$ (which would lead to a TPS mass of 3.1 kg).

Olynick et al. (1997) state that their maximum calculated TPS thickness was about 4.8 cm; the maximum predicted surface recession rate was about 1.2 cm at the stagnation point. It may be clear that the heat-sink approach is not a good approach

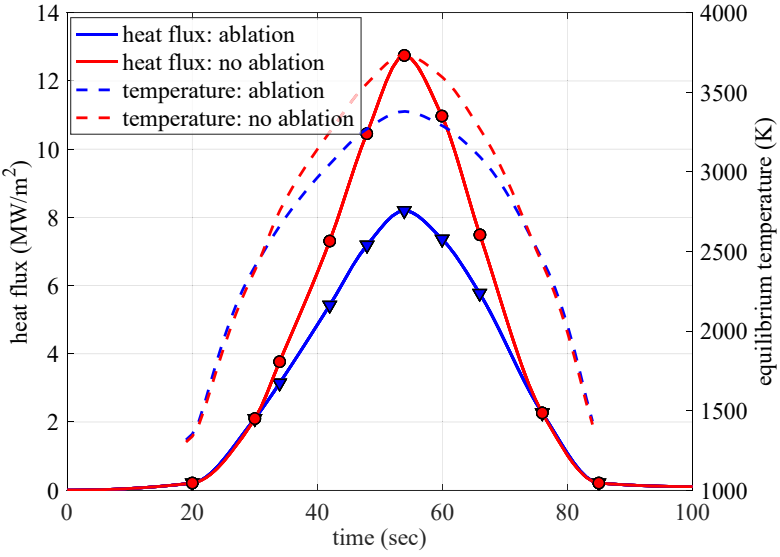


Fig. 10.47 The thermal load on the Stardust heat shield (heat flux and equilibrium wall temperature). Data taken from Olynick et al. (1997)

[42] Bessire and Minton (2017) have done testing on PICA samples and found that the material starts its two main competing decomposition processes when the temperature passes from roughly 300–500 °C. For the current example we assume that this is 400 °C. However, since this is always the case for the given temperature profiles, there will always be ablation.

to estimate the mass of an ablative heat shield and should therefore never be used, especially not when the heat load is large. The MER gives a reasonable approximation and is obviously preferred.

Note that when the scaling shown in Fig. 10.46 is used the total TPS mass is reduced, but not the thickness of the TPS in the stagnation region. With a heat-flux scaling of 1.00, 0.72, 0.61, 0.50, 0.39, and 0.28 for the six areas, and relative areas of <1, 3, 10, 36, 24 and 27% of the total surface areas, we find a total mass of 27.45 kg, and thicknesses of the six regions of 35.2, 27.6, 24.3, 20.9, 17.3, and 13.5 cm. Even though the mass has almost reduced by half, it is still way off.

This approach may serve a better purpose to determine the heat input to a non-ablative heat shield, or to evaluate a heat-sink concept.

■

10.5.4 Active-Cooling Modelling

The principle of active cooling involves both conduction and convection (Fig. 10.48). To understand the capabilities of convective cooling, a simplified analysis is done, inspired by Blosser (2000). Assume a TPS system that consists of a (metal) wall with thickness t, under which a coolant flows. The coolant is required to keep some inner structure at a temperature T_f. The outer wall is exposed to a heat flux q_0, which may be the combination of a convective and radiative heat flux; this is not relevant for the discussion. The net effect of the influx is that the outer wall has a temperature T_0. Through conduction, the inside of the outer wall has a temperature T_i that serves as a boundary condition (input) for the coolant. It is assumed that all heat conducted through the outer wall can be "transported" by convection of the coolant:

$$\frac{k}{t}\left(T_0 - T_i\right) = h\left(T_i - T_f\right) \tag{10.60}$$

with k and h being the thermal conductivity and convection coefficient, respectively. Note that any additional coolant passage or other substructure is neglected in this simple example.

For this cooling system to be effective, the heat that is conducted through the outer skin must be greater than or equal to the incoming heat flux, so

$$\frac{k}{t}\left(T_0 - T_i\right) \geq q_0 \tag{10.61}$$

With the two equations the inner temperature T_i can be eliminated:

$$\frac{k}{t}\left(T_0 - T_i\right) \geq q_0 \quad \text{and} \quad h\left(T_i - T_f\right) \geq q_0 \tag{10.62}$$

Rewriting yields:

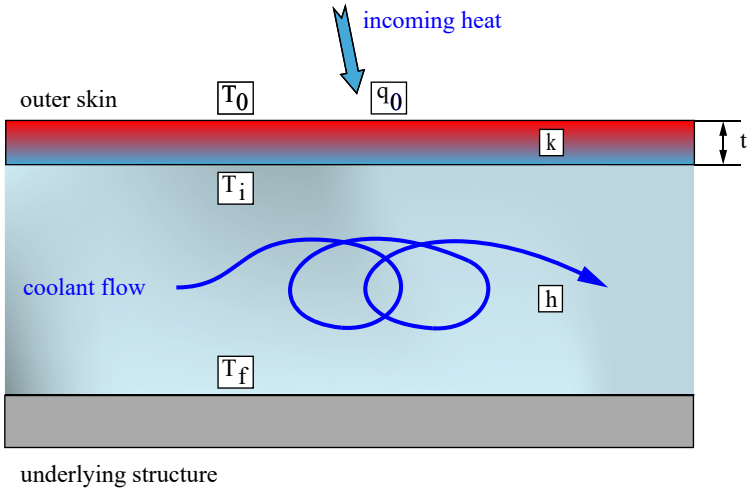


Fig. 10.48 Simplified representation of an actively cooled system

$$\frac{k}{t}\left(T_0 - T_f\right) + \frac{k}{t}\left(T_f - T_i\right) \geq q_0 \tag{10.63}$$

$$h\left(T_0 - T_f\right) + h\left(T_i - T_0\right) \geq q_0 \tag{10.64}$$

Equation (10.63) can be rewritten as

$$\frac{k}{t} + \frac{k}{t}\frac{T_f - T_i}{T_0 - T_f} = \frac{k}{t}\frac{T_0 - T_i}{T_0 - T_f} \geq \frac{q_0}{T_0 - T_f} \tag{10.65}$$

or

$$\frac{T_0 - T_i}{T_0 - T_f} \geq \frac{\left(\frac{q_0}{T_0 - T_f}\right)}{\frac{k}{t}} \tag{10.66}$$

Similarly:

$$\frac{T_i - T_f}{T_0 - T_f} \geq \frac{\left(\frac{q_0}{T_0 - T_f}\right)}{h} \tag{10.67}$$

Adding Eqs. (10.66) and (10.67) together yields

$$\frac{\left(\frac{q_0}{T_0 - T_f}\right)}{\frac{k}{t}} + \frac{\left(\frac{q_0}{T_0 - T_f}\right)}{h} \leq 1 \tag{10.68}$$

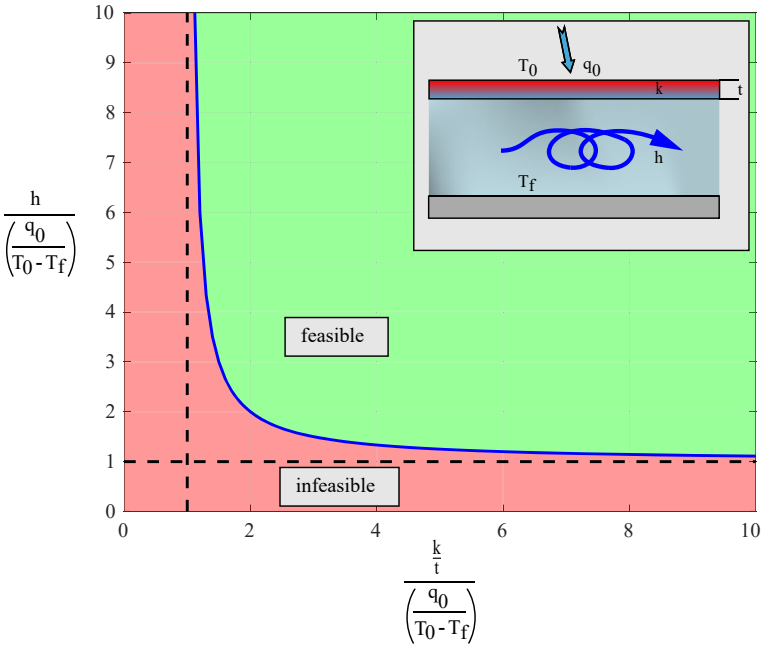


Fig. 10.49 Heat-transfer limits for convectively cooled structures: the incoming heat flux is conducted through the outer wall to the underlying coolant

as, after summation, the temperature terms on the left-hand side of the equations give equal numerator and denominator.

With $T_0 > T_f$, it means that the quantities on the left-hand side of Eq. (10.68) are positive, so it must be true that

$$\frac{k}{t} \geq \frac{q_0}{T_0 - T_f} \quad \text{and} \quad h \geq \frac{q_0}{T_0 - T_f} \tag{10.69}$$

to fulfil the inequality. Using these constraints, the heat-transfer limits for convectively cooled structures can now be easily visualised.

Figure 10.49 shows an inverted form of Eq. (10.68), using the "normalised" constraints of Eq. (10.69) as the two design parameters. The limit-case constraints when either one of the variables approaches 1 are indicated by dashed lines. In the figure, both feasible and infeasible regions have been indicated. The reddish area under the curve represents the combination of conductance and convection coefficient that is insufficient to maintain the surface temperature below T_0 for a given heat flux, q_0. Contrary, the green area shows the combination of h and k, in relation to q_0 and the limit temperature values, that is a feasible one. Of course, there may still be other considerations related to the structural implementation and coolant properties, but as a first rough estimate this analysis can be valuable.

For one of the implemented metallic TPS solutions, i.e., the enhanced-radiation cooling system, water as a cooling medium will be relied upon. The density of the water[43] is 1,000 kg/m^3, whereas its heat of evaporation is $h_{evaporation} = 2.26$ MJ/kg. The conductivity as a function of temperature is given by (Buursink 2005)

$$k_{H_2O} = -0.7676 + 7.536 \cdot 10^{-3} T - 9.825 \cdot 10^{-6} T^2 \tag{10.70}$$

in J/(m K), and the specific heat

$$c_{p,H_2O} = \frac{917.5 - 10.1016\,T + 0.0454134\,T^2 - 9.0751710^{-5} T^3 + 6.807010^{-8} T^4}{0.018} \tag{10.71}$$

in J/(kg K). Note that in both equations, T is expressed in Kelvin.

For the situation when direct-water cooling is applied, the moment the temperature of the water reaches its boiling limit, the heat from the skin will be absorbed by the water to enable its evaporation. This will keep the temperature of the skin at a constant value. Therefore, the water consumption in kg/(s m^2) can be expressed as follows:

$$\dot{m}_{H_2O} = \frac{\Delta q_{in}}{h_{evaporation}} \tag{10.72}$$

under the assumption that the water is cooling through nucleate boiling. The innermost skin is in contact with the water and will therefore have the same temperature. In case enhanced-radiation cooling is considered the cooling principle is still based on evaporation, but the engineering solution, which is shown in Fig. 10.41, is substantially different. A porous material, e.g., ZAL-15 (see Sect. 10.3.1.4), is inserted as an additional layer albeit detached from the outer skin, and has absorbed a large amount of water. At the interface the heat is transferred through radiation, which means that in the transient model equilibrium between the innermost layer of the skin and the ZAL-15 has to be taken into account. The water consumption and temperature of the porous material can be computed with the following relationships, where t_{porous} indicates the thickness of the porous material and $\dot{m}$ the evaporating water flux (Buursink 2005):

$$\Delta T_{H_2O} = \begin{cases} \frac{\Delta q_{in}}{\rho_{H_2O} t_{porous} c_{p,H_2O}} & \text{if } T_{H_2O} < T_{boil} \\ 0 & \text{if } T_{H_2O} = T_{boil} \end{cases} \tag{10.73}$$

[43] According to the U.S. Geological Survey on their website, https://www.usgs.gov/special-topics/water-science-school/science/water-density, the density shows a temperature dependency, e.g., at $T = 0\,^\circ$C it is 999.87 kg/m^3, at $T = 4\,^\circ$C it is 1,000.00 kg/m^3, and up to $T = 100\,^\circ$C it gradually decreases to 958.65 kg/m^3. These values hold for a standard pressure of 1 bar. As a first approximation, a constant density can be assumed, but a refined analysis may require a temperature (and pressure) dependency.

$$\dot{m} = \begin{cases} 0 & \text{if } T_{H_2O} < T_{boil} \\ \frac{\Delta q_{in}}{h_{evaporation}} & \text{if } T_{H_2O} = T_{boil} \end{cases} \tag{10.74}$$

10.5.5 Finite-Difference Model

Building on the work done in the Apollo era, more recently, general approximated one-dimensional formulations of the heat transfer problem have been proposed, substituting classical heat-flux limits with temperature limits. This directly links the vehicle's trajectory to the design choices of the TPS, and demonstrates that at a preliminary design stage these methods can be used for fast analysis while still providing satisfactory results (Martinelli and Braun 2011; Ferraiuolo and Manca 2011). For a faster analysis, the model is centred around a one-dimensional finite-difference formulation to simulate the temperature conditions at the stagnation point along the thickness of the TPS. Because the stagnation point experiences the highest heat-flux and heat-load during entry, solutions for this point will provide a conservative estimate of the TPS performance.

The materials' thermo-physical properties are of crucial importance for the determination of the correct temperatures of the TPS layers, and material consumption in the case of ablation or evaporation/ablation phenomena. These properties are determined by means of wind-tunnel experiments (in arc-jet or plasma tunnels), but most of the time they are kept classified. Only some data on thermo-physical properties of TPS materials can be found, see, for instance the report by Williams and Curry (1992). If the thermo-physical properties of the materials are not known, or known with large uncertainties, the analyses could have a limited validity, even in the case of a detailed three-dimensional model of the heat-transfer problem.

In the following, the mathematical formulation of a one-dimensional finite-difference model will be presented, separated in a model for solids and one for ablators. After the model has been completely described, the use of the model will be illustrated through an extensive verification campaign.

10.5.5.1 Mathematical Model Solids

The origin of the one-dimensional model to calculate the heat flow through a layered structure can be found in applying the principle of conservation of energy, and following the energy (= heat) flows through the different layers of the modelled system. From the heat flows the corresponding temperature distribution can then easily be computed. The fundamental equations can be found in many textbooks on thermal analysis of structures, for instance the books by Thornton (1996) and Rohsenow et al. (1998). An application of this fundamental theory to the modelling of a layered TPS

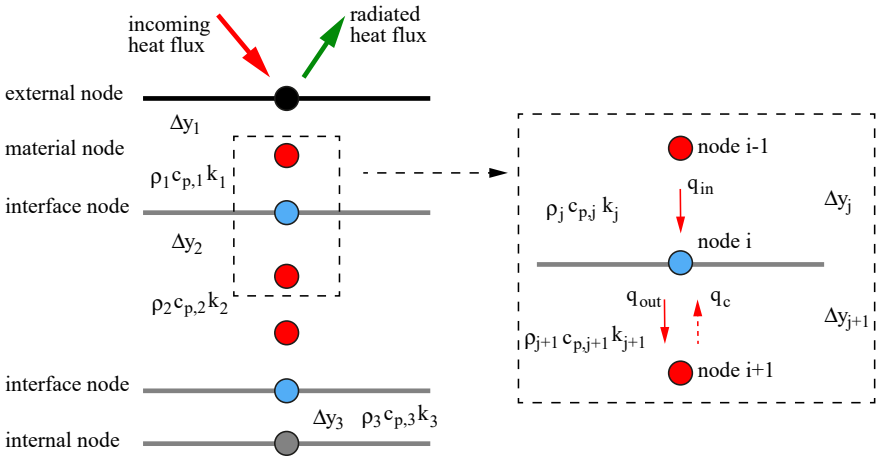


Fig. 10.50 Nodes description of a general finite difference model

can be found in, amongst others, the papers by Bolender and Doman (2006) and Liu and Zhang (2011), which has been adapted here to present it in its current form.

The one-dimensional finite-difference model introduced in this section is schematically shown in Fig. 10.50. The figure shows that the TPS is divided into different layers along its axial dimension (which, as physical layers go, could be a surface panel, honeycomb core, back panel, isolation material, bonding line, internal structure, etc.). Each layer has been discretised into isothermal nodes, with the distance in between layers that can vary from pair to pair. The thermo-physical properties of the layers are assumed to be isotropic.

The nodes are categorised as external, material, interface, and internal nodes. To address the thermal behaviour of the *external node*, we need to begin with a balance of incoming and outgoing fluxes. Starting from the conservation of energy, in the form of the time-rate-of-change of energy

$$\dot{\mathcal{E}}_{in} + \dot{\mathcal{E}}_{gen} = \dot{\mathcal{E}}_{stored} + \dot{\mathcal{E}}_{out} \tag{10.75}$$

where $\dot{\mathcal{E}}_{in}$ represents the rate of energy entering the control volume, $\dot{\mathcal{E}}_{gen}$ the rate of energy generated within the control volume, $\dot{\mathcal{E}}_{stored}$ the rate of energy stored within the control volume, and $\dot{\mathcal{E}}_{out}$ the rate of energy leaving the control volume. Each of these terms has the unit of J/s or W. Thornton (1996) shows that the energy balance for a control volume where no energy enters the system can be written as

$$\rho c_p \frac{\partial T}{\partial t} + \frac{\partial q_i}{\partial x_i} = Q \tag{10.76}$$

where the first term on the left-hand side is the rate of energy stored (per unit of volume), the second term is the rate of energy going out in direction i (typically x, y, and z), whereas the term on the right-hand side is the rate of energy generated internally in the control volume. Furthermore, in this equation ρ is the material's density in kg/m^3, $c_p = c_p(T)$ the temperature-dependent specific heat of the material in J/(kg K), and q_i the heat-flux component i in W/m^2.

Equation (10.76) can be rewritten in terms of heat fluxes, with the difference that in the case of a re-entry system there will be a heat flux entering the system.[44] To convert the rate of energy stored (per unit of volume, so W/m^3) as a flux (J per unit time per unit area, so W/m^2), the related term has to be integrated over the thickness of the layer. Also, to include the finite cross-sectional area of the layer, a multiplicative scaling with the area A is required. Thus

$$\rho c_p \frac{\partial T}{\partial t} \quad \Rightarrow \quad \rho c_p A \frac{\partial T}{\partial t} \mathrm{d}y \tag{10.77}$$

Note that the spatial direction y is used, to anticipate on further model development and following the notation in Fig. 10.50.

Internally, in addition to the heat stored in the material layer, it also conducts heat to the underlying layers. Fourier's Law gives the relation between a flux and temperature gradient (in this case in one dimension):

$$q_{cond} = k \frac{\partial T}{\partial y} \tag{10.78}$$

with k being the material's thermal-conductivity coefficient in W/(m K) for a conductive flux in W/m^2. Taking the finite layer-thickness into account, as well the cross-sectional area, this changes into

$$q_{cond} = \frac{\mathrm{d}}{\mathrm{d}y}\left(kA\frac{\partial T}{\partial y}\right) = kA \left.\frac{\partial T}{\partial y}\right|_{y+\Delta y} \tag{10.79}$$

The net input to the system is $\Delta q_{in} = q_c - q_{rad} - q_{abl}$, assuming that all is concentrated in a single direction; ingoing is the convective heat flux, and outgoing are, in general, the radiated heat plus the heat absorbed by transpiration gases in case of ablative material. Thus, the heat balance for the first material layer is in that case

$$q_{gen,1} + q_{cond,1} = \Delta q_{in} \quad \Rightarrow \quad \rho_1 c_{p,1} A \frac{\partial T}{\partial t} \mathrm{d}y + k_1 A \left.\frac{\partial T}{\partial y}\right|_{y+\Delta y} = \Delta q_{in} \tag{10.80}$$

[44] In general terms, this is the summation of the convective heat flux, q_c, and the flux radiated from the gas surrounding the vehicle, $q_{rad,gas}$. For low-speed re-entry, this second term is much smaler than q_c, but should be taken into account for interplanetary aerocapture, for instance.

To solve the above equation, forward finite-difference of the temperature in time will be used, so, with τ being the running time variable:

$$\frac{\partial T}{\partial t}\mathrm{d}y \approx \frac{T_1^{\tau+1} - T_1^{\tau}}{\Delta t}\Delta y_1 \quad \text{and} \quad \left.\frac{\partial T}{\partial y}\right|_{y+\Delta y} \approx \frac{T_2^{\tau} - T_1^{\tau}}{\Delta t}\Delta y_1 \tag{10.81}$$

Substituting Eq. (10.81) into Eq. (10.80), and solving for $T^{\tau+1}$ yields for the external node, after some manipulation:

$$T_1^{\tau+1} = \Delta q_{in}\frac{2\Delta t}{\rho_1 c_{p,1}\Delta y_1} + T_1^{\tau}\left(1 - \frac{2k_1\Delta t}{\rho_1 c_{p,1}\Delta y_1^2}\right) + T_2^{\tau}\frac{2k_1\Delta t}{\rho_1 c_{p,1}\Delta y_1^2} \tag{10.82}$$

with, in summary: τ indicates time, and the subscripts 1 and 2 the node numbers. The increments Δt and Δy are the temporal and spatial step-size, respectively.

In the *material nodes* only conduction and storage of heat are modelled:

$$\rho c_p A\frac{\partial T}{\partial t}\mathrm{d}y = \frac{\mathrm{d}}{\mathrm{d}y}\left(kA\frac{\partial T}{\partial y}\right) \tag{10.83}$$

Its finite-difference formulation follows in a similar manner as above, and is given by:

$$T_i^{\tau+1} = \left(1 - \frac{2k_j\Delta t}{\rho_j c_{p,j}\Delta y_j^2}\right)T_j^{\tau} + \frac{k_j\Delta t}{\rho_j c_{p,j}\Delta y_j}\left(T_{i+1}^{\tau} + T_{i-1}^{\tau}\right) \tag{10.84}$$

Here, j indicates the layer to which the nodes belong.

The model of the *interface node* may take an extra flux, q_{src}, into account, for instance in the presence of pyrolysis phenomena for ablative materials or heat absorption due to water evaporation. In that case:

$$-\frac{\mathrm{d}}{\mathrm{d}y}\left(k_j A\frac{\partial T}{\partial y_j}\right) = q_{src} - \frac{\mathrm{d}}{\mathrm{d}y}\left(k_{j+1}A\frac{\partial T}{\partial y_{j+1}}\right) \tag{10.85}$$

The assumption is that the pyrolysis manifests itself on a line rather than on a finite thickness region of the material. For interfaces, where no additional heat flux is present, the flux q_{src} is zero. Its finite-difference formulation can be expressed as follows:

$$T_i^{\tau+1} = \frac{2\Delta t}{\rho_j c_{p,j} \Delta y_j + \rho_{j+1} c_{p,j+1} \Delta y_{j+1}} \left(\frac{k_{j+1}}{\Delta y_{j+1}} T_{i+1}^{\tau} + \frac{k_j}{\Delta y_j} T_{i-1}^{\tau} \right) + \\ + \left[1 - \frac{2\Delta t}{\rho_j c_{p,j} \Delta y_j + \rho_{j+1} c_{p,j+1} \Delta y_{j+1}} \left(\frac{k_{j+1}}{\Delta y_{j+1}} + \frac{k_j}{\Delta y_j} \right) \right] T_i^{\tau} - \\ - \frac{q_{src}}{A} \frac{2\Delta t}{\rho_j c_{p,j} \Delta y_j + \rho_{j+1} c_{p,j+1} \Delta y_{j+1}} \tag{10.86}$$

10.5.5.2 Mathematical Model Ablator

Finite-difference models of ablative heat shields have been developed and extensively documented in the 1960 and 1970s, to support the development of the ablative heat shields of Mercury, Gemini, and Apollo (Swann et al. 1965; Clark 1973). Delgado Montes (1988) copied such model in the framework of an aerocapture study, and the modelling fits well in the scope of what Liu and Zhang (2011) has developed. Ridolfi et al. (2012) adopted the same approach, as discussed before, and combined the ablation model with the one-dimensional model described above. In this part of the model description, the finite-difference implementation will be summarised. As the ablation model is, in fact, very complex, the reader is referred to the reports by Swann et al. (1965) and Clark (1973), who give a detailed discussion on the mathematical and numerical formulation, including their verification results.

A typical implementation of an ablative heat shield is shown in Fig. 10.51. The model has several layers with thickness Δy from 1 to n_c being part of the char layer (total thickness L), then several layers $n_c + 1$ to $n_c + n_u$ with thickness $\Delta y'$ being the uncharred layer (total thickness L'), and finally several layers $n_c + n_u + 1$ to $n_c + n_u + n_i$ with thickness $\Delta y''$ being the insulation layer (total thickness L''). Interface layers are layer 1, which is the external char surface receiving the incoming heat flux, layer n_c, which represents the pyrolysis zone, layer $n_c + n_u$ being a heat sink in between the uncharred and insulation layer, and finally layer $n_c + n_u + n_i$ that is the back surface and links to the internal structure.

Clark (1973) discusses the boundary conditions in a clear manner. Even though in the proposed one-dimensional model mainly the temperature in between layers is considered as boundary condition, it is illustrative to go a little deeper with the discussion. For the char surface, i.e., the "outer wall" it is indeed assumed that the temperature is specified; alternatively, this could be the rate of material removal. The second condition is the energy balance, as discussed previously, Eq. (10.57). For the pyrolysis zone, which links the charred with the uncharred layer, also an energy balance should hold, i.e.,

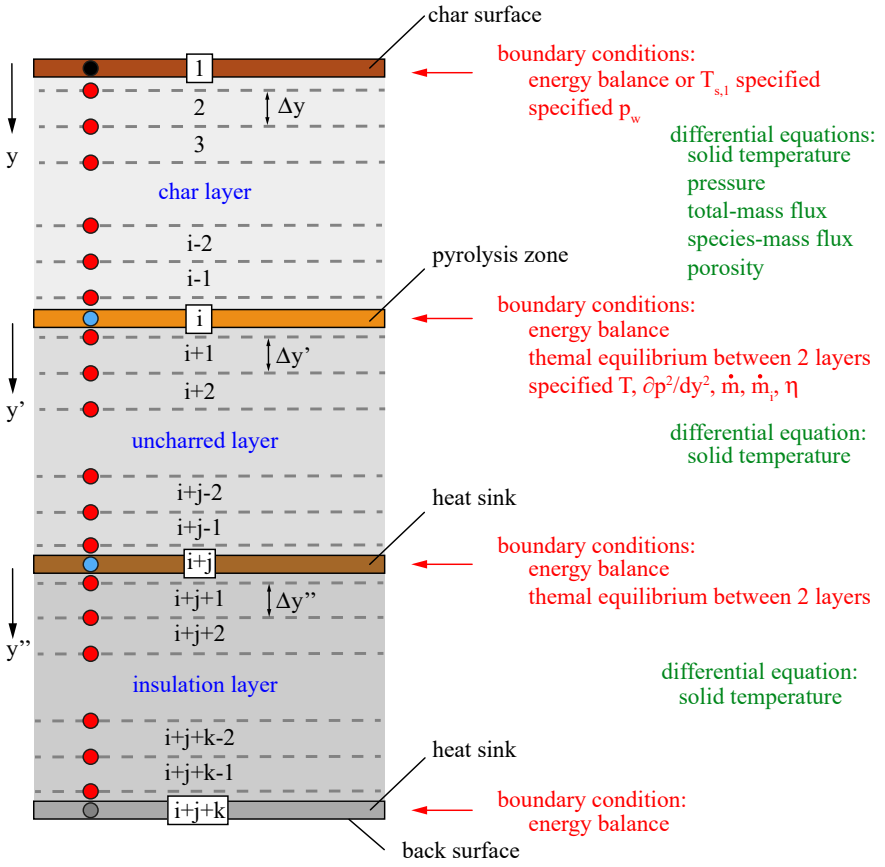


Fig. 10.51 Location of finite-difference stations for the general formulation of an ablation model; indicated are the boundary conditions and the equations underlying the main physical processes. Image inspired by Clark (1973)

$$-k_{char}\frac{\partial T_{char}}{\partial y} = \dot{m}_{pyro}\Delta H_{pyro} - k'\frac{\partial T'}{\partial y} \tag{10.87}$$

where the temperatures of the two layers are equal at their interface, i.e., $T_{char} = T'$; the prime indicates the uncharred layer. The rate of pyrolysis, $\dot{m}_{pyro}$, is given by Eq. (10.54). The boundary-layer condition at the interface between the uncharred layer and the insulation layer is following the same logic:

$$-k'\frac{\partial T'}{\partial y} = -k''\frac{\partial T''}{\partial y} + \rho_{hs}c_{p,hs}L_{hs}\frac{\partial T'}{\partial t} \tag{10.88}$$

where the subscript *hs* stands for *heat sink*, i.e., the interface layer in between with thickness L_{hs}, and the double prime relates to the insulation layer. In the upcoming

discrete model this constraint is ignored by assuming $L_{hs} = 0$. What does hold at the interface is $T' = T''$, though. Finally, the boundary condition between the insulation layer and the back surface has been covered before, and simply follows the energy balance.

To formulate the finite-difference model for the ablative TPS, the following approach is followed. At the outer surface of an ablative TPS several phenomena take place: conduction inwards and radiation outwards of the incoming heat, but also ablation of the charred material and blockage due to the mass transfer of material in the boundary layer. The *cooling flux*, q_{src}, in Eqs. (10.85) and (10.86) at the external surface of an ablative TPS is equal to $\dot{m}_{char}\Delta H_{char}$, with $\dot{m}_{char}$ being the mass removal rate of the charring material and ΔH_{char} the heat required for the ablation. Two situations can occur: in the lead up to full ablation the temperature of the adjoining layer is lower than the ablation temperature, and once it is on-going the temperature will be constrained by the ablation temperature. The recession of the combustion plane, as well as the location of the beginning of the uncharred layer, are given by Eqs. (10.51) through (10.53).

To derive the finite-difference equations the running y-coordinate can be transformed to coordinates in the char and uncharred layers such that the finite-difference stations remain fixed, and the coordinates themselves move to accommodate the changes in the locations of the surfaces of the different materials. These new coordinates become then a function of $\dot{m}_{char}$ and $\dot{m}_{pyro}$. This transformation is at the heart of the derivation of the finite-difference equations, starting from the set of governing (partial) differential equations and the corresponding boundary conditions. This derivation is elaborate and is not repeated here; it has been very well documented by Swann et al. (1965) and Clark (1973), who were part of the group of engineers that set up these models in the 1960 and 1970s. The final result will be provided, though, as this may aid the reader to set up their own model. It is strongly advised, though, to go through the original documentation for a complete understanding.

As shown in Fig. 10.51 and discussed above, the three layers of the ablative TPS consist of a number of stations for each of the layers, and a total of $n_c + n_u + n_i$ nodes is considered. The following notation will be used for the distance between the stations: $\ell = L/(n_c - 1)$ for the char layer, $\ell' = L'/n_u$ for the uncharred layer, and $\ell'' = L''/n_i$ for the insulation layer. As before, τ is used as running time variable, and i is used to indicate station (node) i. The full set of equations are presented in the intermezzo below.

*** Intermezzo 10.1 (Finite-Difference Equations Ablator)**
This intermezzo will list the finite-difference equations of the one-dimensional ablator model.

Char layer, combustion plane (first node, i = l)

$$-T_1^\tau = c_{0,1} + c_{1,1}T_1^{\tau-1} + c_{2,1}T_2^{\tau-1} + c_{3,1}T_3^{\tau-1} + c_{4,1}T_4^{\tau-1} \tag{10.89}$$

In case $T_1^{\tau-1} < T_{char}$ then:

$$c_{0,1} = \left\{ q_{aero}^\tau - \dot{m}_{char}^{\tau-1}\left[\mathcal{S}(T_1^{\tau-1} - T_{char})(h_{char} + \Delta h_{char}) - \Delta h_{char}\right] + \frac{\ell^{\tau-1}}{2}\Delta\mathcal{E}_{gen}\right\} \times \\ \times \frac{2\Delta t}{\rho_{\tau-1,1}c_{p,1}\ell_{\tau-1}} \tag{10.90}$$

with $\mathcal{S}(.)$ being the earlier defined step function: if the argument is positive, then its output is 1, and 0 otherwise. It is noted that when a variable time step is used, in the above equation $\Delta t = \Delta t_{\tau-1,\tau}$. However, in the formulation provided here a constant time step is assumed. Furthermore, $c_{p,1}$ is typically a function of temperature, i.e., $c_{p,1} = c_{p,1}(T_{\tau-1,1})$, which also holds for any further use of c_p, and also the other temperature-dependent properties ρ, k and ε. This means that the property values of the previous time step, $\tau - 1$, and thus $T^{\tau-1}$, need to be inserted in the above equation. This will not be explicitly mentioned for all following equations, but should be done nonetheless. Last, but not least: $\Delta\mathcal{E}_{gen}$ is the heat generated within the control volume during the time considered, so $\Delta\mathcal{E}_{gen} = \dot{\mathcal{E}}_{gen}\Delta t$.

The other coefficients:

$$c_{1,1} = \left\{ -\frac{k_{1,2}}{\ell^{\tau-1}} - \sigma\varepsilon_1(T_1^{\tau-1})^3 - \frac{11}{12}\left(\dot{m}_{char}^{\tau-1}c_{p,1} + \dot{m}_{pyro}^{\tau-1}c'_{p,1}\right) + \frac{\rho c_{p,1}\ell^{\tau-1}}{2\Delta t}\right\} \times \\ \times \frac{2\Delta t}{\rho c_{p,1}\ell^{\tau-1}} \tag{10.91}$$

$$c_{2,1} = \left\{ \frac{k_{1,2}}{\ell^{\tau-1}} + \frac{18}{12}\left(\dot{m}_{char}^{\tau-1}c_{p,1} + \dot{m}_{pyro}^{\tau-1}c'_{p,1}\right) + \frac{\rho c_{p,1}\ell^{\tau-1}}{2\Delta t}\right\} \times \frac{2\Delta t}{\rho c_{p,1}\ell^{\tau-1}} \tag{10.92}$$

$$c_{3,1} = \left\{ -\frac{9}{12}\left(\dot{m}_{char}^{\tau-1}c_{p,1} + \dot{m}_{pyro}^{\tau-1}c'_{p,1}\right) + \frac{\rho c_{p,1}\ell^{\tau-1}}{2\Delta t}\right\} \times \frac{2\Delta t}{\rho c_{p,1}\ell^{\tau-1}} \tag{10.93}$$

$$c_{4,1} = \left\{ 2\frac{2}{12}\left(\dot{m}_{char}^{\tau-1}c_{p,1} + \dot{m}_{pyro}^{\tau-1}c'_{p,1}\right) + \frac{\rho c_{p,1}\ell^{\tau-1}}{2\Delta t}\right\} \times \frac{2\Delta t}{\rho c_{p,1}\ell^{\tau-1}} \tag{10.94}$$

Otherwise, when $T_{\tau-1,1} \geq T_{char}$:

$$c_{1,0} = c_{2,1} = c_{3,1} = c_{4,1} = 0 \quad \text{and} \quad c_{1,1} = 1 \tag{10.95}$$

Char layer, subsequent nodes ($\mathbf{2 \leq i \leq n_c - 2}$)

$$-T_i^\tau = c_{i,0} + c_{i-1,i}T_{i-1}^{\tau-1} + c_{i,i}T_i^{\tau-1} + c_{i+1,i}T_{i+1}^{\tau-1} + c_{i+2,i}T_{i+2}^{\tau-1} \tag{10.96}$$

with

$$c_{0,i} = \ell^{\tau-1}\Delta\mathcal{E}_{gen}\frac{\Delta t}{\ell^{\tau-1}\rho_i c_{p,i}} \tag{10.97}$$

$$c_{i-1,i} = \left\{\frac{k_{i-1,i}}{\ell^{\tau-1}} - \frac{1}{3}\left[\dot{m}_{char}^{\tau-1}c_{p,1} + \dot{m}_{pyro}^{\tau-1}c'_{p,i} + \frac{i-1}{n_c-1}\left(\frac{\dot{m}_{pyro}^{\tau-1}\rho_i}{\Delta\rho_i} - \dot{m}_{char}^{\tau-1}\right)c_{p,i}\right]\right\} \times \times \frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1}} \tag{10.98}$$

$$c_{i,i} = \left\{-\frac{k_{i-1,i}+k_{i,i+1}}{\ell^{\tau-1}} + \frac{\ell^{\tau-1}\rho_i c_{p,i}}{\Delta t} - - \frac{1}{2}\left[\dot{m}_{char}^{\tau-1}c_{p,1} + \dot{m}_{pyro}^{\tau-1}c'_{p,i} + \frac{i-1}{n_c-1}\left(\frac{\dot{m}_{pyro}^{\tau-1}\rho_i}{\Delta\rho_i} - \dot{m}_{char}^{\tau-1}\right)c_{p,i}\right]\right\} \times \frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1}} \tag{10.99}$$

$$c_{i+1,i} = \left\{\frac{k_{i,i+1}}{\ell^{\tau-1}} + \left[\dot{m}_{char}^{\tau-1}c_{p,1} + \dot{m}_{pyro}^{\tau-1}c'_{p,i} + + \frac{i-1}{n_c-1}\left(\frac{\dot{m}_{pyro}^{\tau-1}\rho_i}{\Delta\rho_i} - \dot{m}_{char}^{\tau-1}\right)c_{p,i}\right]\right\} \times \frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1}} \tag{10.100}$$

$$c_{i+2,i} = -\frac{1}{6}\left[\dot{m}_{char}^{\tau-1}c_{p,1} + \dot{m}_{pyro}^{\tau-1}c'_{p,i} + + \frac{i-1}{n_c-1}\left(\frac{\dot{m}_{pyro}^{\tau-1}\rho_i}{\Delta\rho_i} - \dot{m}_{char}^{\tau-1}\right)c_{p,i}\right] \times \frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1}} \tag{10.101}$$

In the above (and following) equations, $\Delta\rho_i = \rho'_i - \rho_i$.

Char layer, second-last node ($\mathbf{i = n_c - 1}$)

$$-T_i^\tau = c_{i,0} + c_{i-2,i}T_{i-2}^{\tau-1} + c_{i-1,i}T_{i-1}^{\tau-1} + c_{i,i}T_i^{\tau-1} + c_{i+1,i}T_{i+1}^{\tau-1} \tag{10.102}$$

with

$$c_{i,0} = \ell^{\tau-1}\Delta\mathcal{E}_{gen}\frac{\Delta t}{\ell^{\tau-1}\rho_i c_{p,i}} \tag{10.103}$$

$$\begin{aligned} c_{i-2,i} = \frac{1}{6}\Bigg[& \dot{m}^{\tau-1}_{char}c_{p,1} + \dot{m}^{\tau-1}_{pyro}c'_{p,i} + \\ & + \frac{i-1}{n_c-1}\left(\frac{\dot{m}^{\tau-1}_{pyro}\rho_i}{\Delta\rho_i} - \dot{m}^{\tau-1}_{char}\right)c_{p,i}\Bigg] \times \frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1}} \end{aligned} \tag{10.104}$$

$$\begin{aligned} c_{i-1,i} = & \left\{\frac{k_{i-1,i}}{\ell^{\tau-1}} - \left[\dot{m}^{\tau-1}_{char}c_{p,1} + \dot{m}^{\tau-1}_{pyro}c'_{p,i} + \frac{i-1}{n_c-1}\left(\frac{\dot{m}^{\tau-1}_{pyro}\rho_i}{\Delta\rho_i} - \dot{m}^{\tau-1}_{char}\right)c_{p,i}\right]\right\} \times \\ & \times \frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1}} \end{aligned} \tag{10.105}$$

$$\begin{aligned} c_{i,i} = & \left\{-\frac{k_{i-1,i}+k_{i,i+1}}{\ell^{\tau-1}} + \frac{1}{2}\left[\dot{m}^{\tau-1}_{char}c_{p,1} + \dot{m}^{\tau-1}_{pyro}c'_{p,i} + \frac{i-1}{n_c-1}\left(\frac{\dot{m}^{\tau-1}_{pyro}\rho_i}{\Delta\rho_i} - \dot{m}^{\tau-1}_{char}\right)c_{p,i}\right] + \right. \\ & \left. + \frac{\ell^{\tau-1}\rho_i c_{p,i}}{\Delta t}\right]\Bigg\} \times \frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1}} \end{aligned} \tag{10.106}$$

$$\begin{aligned} c_{i+1,i} = & \left\{\frac{k_{i,i+1}}{\ell^{\tau-1}} + \frac{1}{3}\left[\dot{m}^{\tau-1}_{char}c_{p,1} + \dot{m}^{\tau-1}_{pyro}c'_{p,i} + \frac{i-1}{n_c-1}\left(\frac{\dot{m}^{\tau-1}_{pyro}\rho_i}{\Delta\rho_i} - \dot{m}^{\tau-1}_{char}\right)c_{p,i}\right]\right\} \times \\ & \times \frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1}} \end{aligned} \tag{10.107}$$

Char layer, decomposition plane (interface node, $\mathbf{i = n_c}$)

$$\begin{aligned} -T^{\tau}_i = & \; c_{0,i} + c_{i-3,i}T^{\tau-1}_{i-3} + c_{i-2,i}T^{\tau-1}_{i-2} + c_{i-1,i}T^{\tau-1}_{i-1} + c_{i,i}T^{\tau-1}_{i} + \\ & + c_{i+1,i}T^{\tau-1}_{i+1} + c_{i+2,i}T^{\tau-1}_{i+2} + c_{i+3,i}T^{\tau-1}_{3+1} \end{aligned} \tag{10.108}$$

with, for $T^{\tau-1}_i < T_{char}$:

$$c_{0,i} = \left(-2\Delta h_{pyro}\dot{m}^{\tau-1}_{pyro} + \ell^{\tau-1}\Delta\mathcal{E}_{gen}\right)\left(\frac{\Delta t}{\ell^{\tau-1}\rho_i c_{p,i} + \ell'^{\tau-1}\rho'_i c'_{p,i}}\right) \tag{10.109}$$

$$c_{i-3,i} = -\frac{1}{3}\left(c'_{p,i} + \frac{\rho_i c_{p,i}}{\Delta\rho_i}\right)\dot{m}^{\tau-1}_{pyro}\left(\frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1} + \rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.110}$$

$$c_{i-2,i} = -\frac{9}{6}\left(c'_{p,i} + \frac{\rho_i c_{p,i}}{\Delta\rho_i}\right)\dot{m}^{\tau-1}_{pyro}\left(\frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1} + \rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.111}$$

$$c_{i-1,i} = \left[\frac{2k_{i-1,i}}{\ell^{\tau-1}} - 3\left(c'_{p,i} + \frac{\rho_i c_{p,i}}{\Delta\rho_i}\right)\dot{m}^{\tau-1}_{pyro}\right]\left(\frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1} + \rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.112}$$

$$\begin{aligned} c_{i,i} = &\left\{\left[-\frac{2k_{i,i+1}}{\ell^{\tau-1}} - \frac{2k_{i-1,i}}{\ell^{\tau-1}} + \frac{11}{6}\left(c'_{p,i} + \frac{\rho_i c_{p,i}}{\Delta\rho_i}\right)\dot{m}^{\tau-1}_{pyro}\right] - \right. \\ &- \left.\frac{11}{6}\left(c'_{p,i} + \frac{\rho'_i c_{p,i}}{\Delta\rho_i}\right)\dot{m}^{\tau-1}_{pyro} + \frac{\rho_i c_{p,i}\ell^{\tau-1} + \rho'_i c'_{p,i}\ell'^{\tau-1}}{\Delta t}\right\} \times \\ &\times \left(\frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1} + \rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \end{aligned} \tag{10.113}$$

$$c_{i+1,i} = \left[-\frac{2k_{i-1,i}}{\ell^{\tau-1}} + 3\frac{c'_{p,i}\rho'_i}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right]\left(\frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1} + \rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.114}$$

$$c_{i+2,i} = -\frac{9}{6}\frac{c'_{p,i}\rho'_i}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\left(\frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1} + \rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.115}$$

$$c_{i+3,i} = \frac{2}{6}\frac{c'_{p,i}\rho'_i}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\left(\frac{\Delta t}{\rho_i c_{p,i}\ell^{\tau-1} + \rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.116}$$

Otherwise, when $T_{\tau-1,i} \geq T_{char}$, all coefficients are zero, except for $c_{i,i} = 1$.

Uncharred layer, nodes $\mathbf{n_c + 1 \leq i \leq n_c + n_u - 2}$

$$-T^{\tau}_i = c_{i-1,i}T^{\tau-1}_{i-1} + c_{i,i}T^{\tau-1}_i + c_{i+1,i}T^{\tau-1}_{i+1} + c_{i+2,i}T^{\tau-1}_{i+2} \tag{10.117}$$

with

$$c_{i-1,i} = \left[\frac{k'_{i-1,i}}{\ell'^{\tau-1}} - \frac{1}{3}\left(\frac{n_c + n_u - i}{n_u}\right)\left(\frac{\rho'_i c'_{p,i}}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right)\right]\left(\frac{\Delta t}{\rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.118}$$

$$c_{i,i} = \left[-\frac{k'_{i-1,i}}{\ell'^{\tau-1}} - \frac{k'_{i,i+1}}{\ell'^{\tau-1}} - \frac{1}{2}\left(\frac{n_c + n_u - i}{n_u}\right)\left(\frac{\rho'_i c'_{p,i}}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right) + \frac{\rho'_i c'_{p,i}\ell'^{\tau-1}}{\Delta t}\right] \times \times \left(\frac{\Delta t}{\rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.119}$$

$$c_{i+1,i} = \left[\frac{k'_{i,i+1}}{\ell'^{\tau-1}} + \left(\frac{n_c + n_u - i}{n_u}\right)\left(\frac{\rho'_i c'_{p,i}}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right)\right]\left(\frac{\Delta t}{\rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.120}$$

$$c_{i+2,i} = -\frac{1}{6}\left(\frac{n_c + n_u - i}{n_u}\right)\left(\frac{\rho'_i c'_{p,i}}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right)\left(\frac{\Delta t}{\rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.121}$$

Uncharred layer, second-last node ($\mathbf{i = n_c + n_u - 1}$)

$$-T_i^{\tau} = c_{i-2,i}T_{i-2}^{\tau-1} + c_{i-1,i}T_{i-1}^{\tau-1} + c_{i,i}T_i^{\tau-1} + c_{i+1,i}T_{i+1}^{\tau-1} \tag{10.122}$$

with

$$c_{i-2,i} = \frac{1}{6}\left(\frac{n_c + n_u - i}{n_u}\right)\left(\frac{\rho'_i c'_{p,i}}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right)\left(\frac{\Delta t}{\rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.123}$$

$$c_{i-1,i} = \left[\frac{k'_{i-1,i}}{\ell'^{\tau-1}} - \left(\frac{n_c + n_u - i}{n_u}\right)\left(\frac{\rho'_i c'_{p,i}}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right)\right]\left(\frac{\Delta t}{\rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.124}$$

$$c_{i,i} = \left[-\frac{k'_{i-1,i}}{\ell'^{\tau-1}} - \frac{k'_{i,i+1}}{\ell'^{\tau-1}} + \frac{1}{2}\left(\frac{n_c + n_u - i}{n_u}\right)\left(\frac{\rho'_i c'_{p,i}}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right) + \frac{\rho'_i c'_{p,i}\ell'^{\tau-1}}{\Delta t}\right] \times \times \left(\frac{\Delta t}{\rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.125}$$

$$c_{i+1,i} = \left[\frac{k'_{i,i+1}}{\ell'^{\tau-1}} + \frac{1}{3}\left(\frac{n_c + n_u - i}{n_u}\right)\left(\frac{\rho'_i c'_{p,i}}{\Delta\rho_i}\dot{m}^{\tau-1}_{pyro}\right)\right]\left(\frac{\Delta t}{\rho'_i c'_{p,i}\ell'^{\tau-1}}\right) \tag{10.126}$$

Uncharred layer, interface node with insulation layer ($\mathbf{i = n_c + n_u}$)

$$-T_i^{\tau} = c_{i-1,i}T_{i-1}^{\tau-1} + c_{i,i}T_i^{\tau-1} + c_{i+1,i}T_{i+1}^{\tau-1} \tag{10.127}$$

with

$$c_{i-1,i} = \left(\frac{k'_{i-1,i}}{\ell'^{\tau-1}}\right)\left(\frac{\Delta t}{\rho'_i c'_{p,i}\frac{\ell'^{\tau-1}}{2} + \rho''_i c''_{p,i}\frac{\ell''^{\tau-1}}{2} + C^{\tau-1}_{hs,n_c+n_u}}\right) \tag{10.128}$$

with $C_{hs,nc+nu}$ in J/(m^2 K) being the capacity of the heat-sink located in the uncharred layer-insulation interface.

$$c_{i,i} = \left[-\frac{k'_{i-1,i}}{\ell'^{\tau-1}} - \frac{k''_{i,i+1}}{\ell''^{\tau-1}} + \left(\frac{\rho'_i c'_{p,i}\frac{\ell'^{\tau-1}}{2} + \rho''_i c''_{p,i}\frac{\ell''^{\tau-1}}{2} + C^{\tau-1}_{hs,n_c+n_u}}{\Delta t}\right)\right]\times \\ \times \left(\frac{\Delta t}{\rho'_i c'_{p,i}\frac{\ell'^{\tau-1}}{2} + \rho''_i c''_{p,i}\frac{\ell''^{\tau-1}}{2} + C^{\tau-1}_{hs,n_c+n_u}}\right) \tag{10.129}$$

$$c_{i+1,i} = \left(\frac{k''_{i,i+1}}{\ell''^{\tau-1}}\right)\left(\frac{\Delta t}{\rho'_i c'_{p,i}\frac{\ell'^{\tau-1}}{2} + \rho''_i c''_{p,i}\frac{\ell''^{\tau-1}}{2} + C^{\tau-1}_{hs,n_c+n_u}}\right) \tag{10.130}$$

Insulation layer, all nodes except last ($\mathbf{n_c + n_u + 1 \le i \le n_c + n_u + n_i - 1}$)

$$-T_i^{\tau} = c_{i-1,i}T_{i-1}^{\tau-1} + c_{i,i}T_i^{\tau-1} + c_{i+1,i}T_{i+1}^{\tau-1} \tag{10.131}$$

with

$$c_{i-1,i} = \left(\frac{k''_{i-1,i}}{\ell''^{\tau-1}}\right)\left(\frac{\Delta t}{\rho''_i c''_{p,i}\ell''^{\tau-1}}\right) \tag{10.132}$$

$$c_{i,i} = \left[-\frac{k''_{i-1,i}}{\ell''^{\tau-1}} - \frac{k''_{i,i+1}}{\ell''^{\tau-1}} + \left(\frac{\rho''_i c''_{p,i}\ell'^{\tau-1}}{\Delta t}\right)\right]\left(\frac{\Delta t}{\rho''_i c''_{p,i}\ell''^{\tau-1}}\right) \tag{10.133}$$

$$c_{i+1,i} = \left(\frac{k''_{i,i+1}}{\ell''^{\tau-1}}\right)\left(\frac{\Delta t}{\rho''_i c''_{p,i}\ell''^{\tau-1}}\right) \tag{10.134}$$

Insulation layer, interface with back surface ($\mathbf{i = n_c + n_u + n_i}$)

$$-T_i^{\tau} = c_{0,i} + c_{i-1,i}T_{i-1}^{\tau-1} + c_{i,i}T_i^{\tau-1} \tag{10.135}$$

with

$$c_{0,i} = \left(\sigma\varepsilon_i T_B^4 - \mathcal{S}(T_i^{\tau-1} - T_{char})\Delta W_f \Delta h_f\right)\left(\frac{\Delta t}{\rho''_i c''_{p,i}\frac{\ell''^{\tau-1}}{2} + C^{\tau-1}_{hs,n_c+n_u+n_i}}\right) \tag{10.136}$$

where $C_{hs,n_c+n_u+n_i}$ in J/(m^2 K) is the capacity of the heat sink located in the back surface, $\varepsilon_{hs,n_c+n_u+n_i}$ is the emissivity of the inner surface of the TPS, T_B is the internal temperature (of the back surface), ΔW_f is the rate of coolant consumption (kg/s m^2), and Δh_f is the heat capacity of the coolant (J/kg). Of course, ΔW_f and Δh_f are only defined, when applicable.

$$c_{i-1,i} = \left(\frac{k''_{i-1,i}}{\ell''^{\tau-1}}\right)\left(\frac{\Delta t}{\rho''_i c''_{p,i}\frac{\ell''^{\tau-1}}{2} + C^{\tau-1}_{hs,n_c+n_u+n_i}}\right) \tag{10.137}$$

$$c_{i,i} = \left[-\frac{k''_{i-1,i}}{\ell''^{\tau-1}} - \sigma\varepsilon_i (T_i^3)^{\tau-1} + \left(\frac{\rho''_i c''_{p,i}\frac{\ell''^{\tau-1}}{2} + C^{\tau-1}_{hs,n_c+n_u+n_i}}{\Delta t}\right)\right]\times \times \left(\frac{\Delta t}{\rho''_i c''_{p,i}\frac{\ell''^{\tau-1}}{2} + C^{\tau-1}_{hs,n_c+n_u+n_i}}\right) \tag{10.138}$$

At the interface between charred and uncharred material, the pyrolysis cooling flux is modelled as $\dot{m}_{pyro}\Delta H_{pyro}$, where ΔH_{pyro} is the heat of pyrolysis, $\dot{m}_{pyro}$ is the mass flux of the pyrolysis material, which is, for $T_i^{\tau-1} \geq T_{pyro}$:

$$\dot{m}_{pyro} = \left\{T_i^\tau - \frac{2\Delta t}{\rho_i c_{p_i}\Delta y_i + \rho_{i+1}c_{p_{i+1}}\Delta y_{i+1}}\left[\frac{k_i}{\Delta y_i}T^\tau_{i-1} + \frac{k_{i+1}}{\Delta y_{i+1}}T^\tau_{i+1}\right] - \left[1 - \frac{2\Delta t}{\rho_i c_{p_i}\Delta y_i + \rho_{i+1}c_{p_{i+1}}\Delta y_{i+1}}\left(\frac{k_i}{\Delta y_i} + \frac{k_{i+1}}{\Delta y_{i+1}}\right)\right]T_i^{\tau-1}\right\} \cdot \cdot \frac{-\left(\rho_i c_{p_i}\Delta y_i + \rho_{i+1}c_{p_{i+1}}\Delta y_{i+1}\right)}{2\Delta t\,\Delta H_{pyro}} \tag{10.139}$$

In case $T_i^{\tau-1} < T_{pyro}$, the mass flux of the pyrolysis material is zero. The quantities needed to model the ablative TPS, for the three types of ablative materials considered, are shown in Table 10.7. For the sake of convenience, we repeat here: ρ, c_p, and k represent the density, specific heat and conductivity, respectively; ε is the emissivity, while A, B, and D represent the reaction rate, activation energy, and diffusion coefficient, respectively; p_w is the environmental pressure at the outer edge of the boundary layer, $C_{O_2,e}$ is the oxygen concentration there, and λ is the weight of char removed per unit weight of oxygen. The factor K is computed as Ae^{-B/T_w}.

For the mass flow of the char, when $T^{\tau-1} < T_{char}$ then

$$\dot{m}_{char} = \frac{1}{2}\left[-\frac{(h_e - h_w)\, K^2 p_w}{q_c \lambda Le^{0.6}} + \sqrt{\left(\frac{(h_e - h_w)\, K^2 p_w}{q_{eff} \lambda Le^{0.6}} \right)^2 + 4\, K^2 p_w C_{O_2,e}} \right] \tag{10.140}$$

which is the same as Eq. (10.48), introduced earlier. However, in case $T^{\tau-1} = T_{char}$ then $\dot{m}_{char}$ follows from the energy balance and can be written as:

$$\dot{m}_{char} = \frac{1}{\Delta H_{char}} \left[(T_i^{\tau} - T_i^{\tau-1}) \frac{2k_i \Delta t}{\rho_i c_{p,i} \Delta y^2} - T^{\tau} - \ell^{\tau-1} \frac{2k_{i+1} \Delta t}{\rho_i c_{p,i} \Delta y^2} \right] \frac{\rho c_{p,i} \Delta y}{2\Delta t} - \\ - q_c + \sigma \epsilon \left(T_i^{\tau-1} \right)^4 \tag{10.141}$$

The finite-difference formulation presented here allows for solving the problem with a number of nodes that may vary case-by-case, and with an explicit forward integration technique. To obtain a faster execution of the thermal transient analysis, the results presented in this chapter were obtained with an implicit formulation of the problem. It was decided to describe here the explicit formulation, because it is more direct and it allows for linking it directly to the involved physical phenomena. Having an implicit formulation means that the state at time t depends on previous states but also on the current one. In that way, even when the time step is large, the integration error can be kept relatively low. There are references in literature for solving this type of problems with an implicit formulation, see, for instance, the book by Press et al. (2007). For the verification analyses presented here, an explicit formulation was used.

■

10.5.5.3 Verification and Validation

The one-dimensional finite-difference model as presented before has been verified for a number of TPS designs. The verification cases have been documented by Ridolfi et al. (2012) and have been re-evaluated for the purpose of this chapter. For all cases, the material properties have been introduced earlier (Sect. 10.3).

Metallic uncooled TPS, hot-structure approach

The uncooled metallic TPS, which is considered to be a hot structure (the TPS is both thermal-load carrying and mechanical-load carrying), consists of an outer layer of 10 mm thick PM-1000, with underneath an insulator layer of Saffil with a thickness

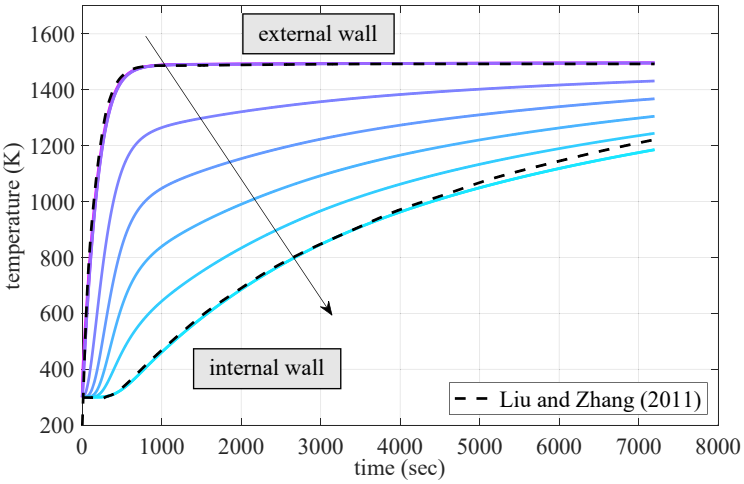


Fig. 10.52 Metallic uncooled TPS, computed temperature profiles versus the benchmark data provided by Liu and Zhang (2011). Incoming heat flux is 240 kW/m^2

of 50 mm.[45] The internal structure is a 5 mm thick titanium layer. An incoming heat flux of $q_{in} = 240$ kW/m^2 is applied as thermal load. For the modelling, four nodes are used for the outer skin (so $\Delta y_1 = 3.33$ mm), six for the insulation layer (with $\Delta y_2 = 10$ mm), and two for the internal structure ($\Delta y_3 = 5$ mm). The response of the TPS is compared with the results obtained by Liu and Zhang (2011) for the same TPS design and with a similar one-dimensional tool, and is shown in Fig. 10.52. The verification data are limited to the outer skin and the innermost layer of the TPS, but the related two temperature profiles provide a very good match. The temperature histories for the internal nodes show profiles that correspond well to intermediate profiles. A few observations are made. The four PM-1000 temperature profiles are all overlapping, which means that heat conduction through the metal layer is very quick. There is a significant delay in the insulator layer, and even after 7,200 s no steady state has been reached. If the maximum operational temperature of titanium is assumed to be 600 °C (873 K), then this temperature is only reached after more than 3,000 s. But, in all fairness, to have a constant heat flux of 240 kW/m^2 for more than 50 min, is an unrealistically long time. On a final note: initially the evaluation was run with time steps of $\Delta t = 1$ s, and the difference in temperature for the outer layer was significant. The time shift was 350 s at $T = 1,400$ K. Due to the rapid increase of the temperature in the metal layer, it was therefore required to reduce the time step to $\Delta t = 0.1$ s. This is a consideration for the analysis of any TPS where materials with high conductivity are involved.

[45] The Saffil used here is the one with a density of 92 kg/m^3, where the properties are the effective properties. This means that it is assumed that the porous insulation material is isotropic and homogeneous (Liu and Zhang 2011).

Metallic cooled, nucleate boiling and enhanced-radiation cooling approaches

For the verification of the TPS model with cooling mechanisms, the design of an entry capsule, designated DART and conceptualised at Delft University of Technology, is considered (Buursink 2005). DART is a bi-conic vehicle with a small, water-cooled nose (with radius $R_N = 0.25\,\text{m}$), a conic section (with a cone angle of $\theta_c = 15°$), and a short flare at a relative angle of 15° with the cone (so $\theta_f = 30°$). The total vehicle length is $L = 1.6\,\text{m}$, whereas the base diameter is $d = 1.3\,\text{m}$. The vehicle has a mass of 150 kg and enters the atmosphere into a ballistic flight with a velocity of $V_E = 7,000\,\text{m/s}$ at an angle of $\gamma_E = -5°$, at $h_E = 120\,\text{km}$. The aerodynamic properties have been calculated using the modified Newtonian method.

The applied cooling mechanisms are nucleate boiling at the nose and enhanced-radiation cooling at the conic section. For the outer skin of the nose, 1-mm thick PM-2000 is considered, whereas for the cone this is 0.5-mm thick PM-2000. The water-soaked insulator material is ZAL-15, with a layer of 5 mm. For the calculations of TPS temperatures and water consumption, the one-dimensional, finite-difference model was used and its output compared with benchmark data obtained with an engineering model by Buursink (2005). This model takes the heat flux as input and from that calculates the temperature of the skin, the wet porous layer and the evaporation of water. Only the specific-heat capacity of the material is considered, conduction is ignored. Details of the model can be found in the cited reference. One assumption is mentioned, though. The model estimates the amount of water needed for cooling the nose by simply integrating the applied heat load over time and over the nose surface, assuming that all this energy was used to evaporate water. The heat flux on a point of a hemispherical nose has been related to the heat flux at the stagnation point by:

$$\frac{q_{c,hemisphere}}{q_{c,stagnation}} = (\cos\phi)^{1.5} \tag{10.142}$$

where ϕ is the angle of the line connecting the stagnation point with a point on the hemisphere and the horizontal centre axis, with $\phi \leq 70°$. This formulation allows for calculating the integrated heat load entering the nose area, assuming that ϕ runs from 0° to a certain $\phi_{\max}$:

$$Q = \int_0^{\phi_{\max}} q_{c,hemisphere}(\phi)\mathrm{d}A = \int_0^{\phi_{\max}} q_{c,stagnation}\,(\cos\phi)^{1.5}\,2\pi R\sin\phi R\mathrm{d}\phi \tag{10.143}$$

yielding

$$Q = -\frac{4}{5}\pi R^2 q_{c,stagnation}\,(\cos\phi)^{2.5}\Big|_0^{\phi_{\max}} \tag{10.144}$$

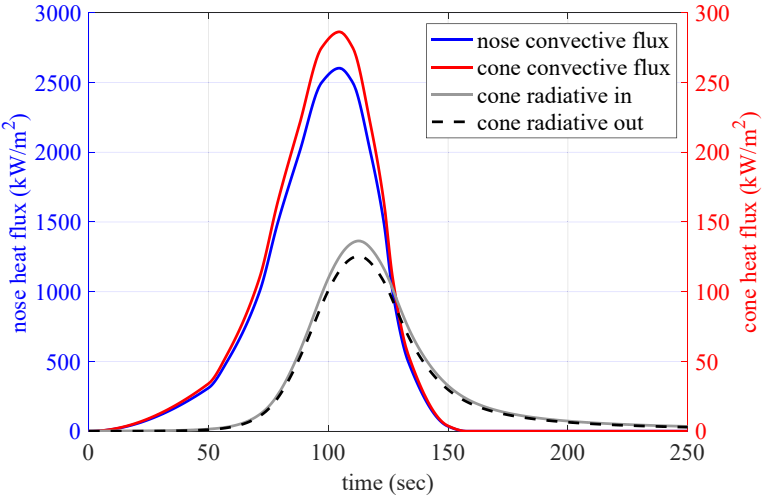


Fig. 10.53 Metallic cooled TPS: incoming heat fluxes

For the one-dimensional model, the following six nodes have been defined. For the nose, an outer-wall skin node, a material node, and an interface node with the ZAL-15 layer, to allow for calculating the cooling effect and to provide input to calculating the water consumption. For the cone, a similar setup: an outer-wall PM-2000 node, an interface node with the ZAL-15 layer, and in addition an interface node between the ZAL-15 and the underlying PM-2000 wall. The initial temperature of the water is assumed to be 0 °C, whereas the boiling temperature is taken as 150 °C (at a pressure of 5 bar).

The flux profile for the (ballistic) entry, a combination of convective and radiative flux, is shown in Fig. 10.53. Note that the flux has been evaluated at the stagnation point using a cold-wall Chapman model. The cone convective heat flux is assumed to be 11% of the stagnation flux. Also included in the figure are the radiative fluxes for the cone; it can be seen that the emitted (= outgoing) flux is almost as large as the radiative flux transmitted into the ZAL-15 layer of the cone.

Figure 10.54 shows that both the computed and the benchmark temperature profiles are very close to each other, although there is a small difference for the nose profile. The computed temperature rises earlier compared to the benchmark, which can be explained by the difference in the way the heat enters the nose. In the one-dimensional model the heat flux is considered to be a point load, which will force a more rapid increase of the temperature. Also the thermal conduction will contribute to this. The water will reach its boiling point more quickly, and the water cooling will be effective from an earlier point in time. Also, since for the benchmark the flux is divided over an area, the integrated water flow is slightly larger, which leads to a slightly larger water consumption. For the cone the match is almost perfect; the wall thickness of the cone is very small, and equilibrium will be reached very quickly. A final remark to be made is that the cone area is a lot larger than the nose area (1.86

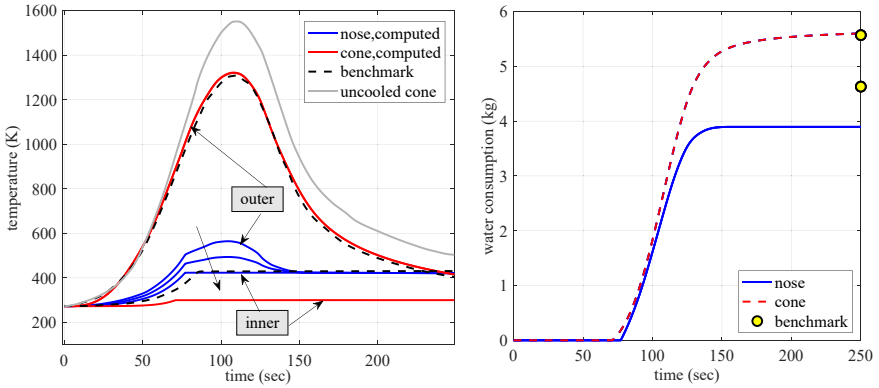


Fig. 10.54 Metallic cooled TPS: layer temperatures and comparison with benchmark (left) and cooling-water consumption (right)

versus 0.0784 m^2), such that any potential difference between the computations and benchmark is more likely to be averaged out for the cone.

Concerning the water consumption, the moment the cooling becomes active an initial waterflow of about 5 g/s is required for the cooling, which gradually increases to about 9 g/s. For the cone the initial value is close to 2.5 g/s and increases to 11 g/s. The total amount of water used is 3.9 kg for the nose and 5.6 kg for the cone. The latter value is larger, because of the large area to be cooled.

Ceramic TPS

The ceramic TPS is built up as follows. The outer layer is UHTC ZrB_2-SiC with a thickness of 10 cm. A Saffil insulation layer of 4 mm is directly underneath, separating the UHTC from the titanium internal structure with a thickness of 4 mm. The Saffil has a density of 96 kg/m^3. A total of ten nodes has been defined, i.e., four for the UHTC (the first one represents the wall temperature, the next two are material nodes, and the last one the interface node with the insulator), five for the insulator (the first four nodes being material nodes and the last one being the insulator-titanium interface node), and finally one for the inner wall. Both the environment temperature and the initial node temperatures are set to 300 K.

For the comparison, a benchmark obtained from Savino et al. (2005) who studied the aerothermodynamic of UHTC-based thermal protection systems for the Unmanned Space Vehicle (USV), is used. The performance of the thermal model with the UHTC ZrB_2-SiC TPS is presented in Fig. 10.55, with on the left the convective heat-flux profile going up to 5 MW/m^2 and on the right the thermal response, which has been compared with the temperature profile obtained for the same material in literature. The results show that discrepancies between the benchmark temperature-profile and the computed one are well within 5%, whereas the internal temperature is kept at a value close to the initial 300 K.

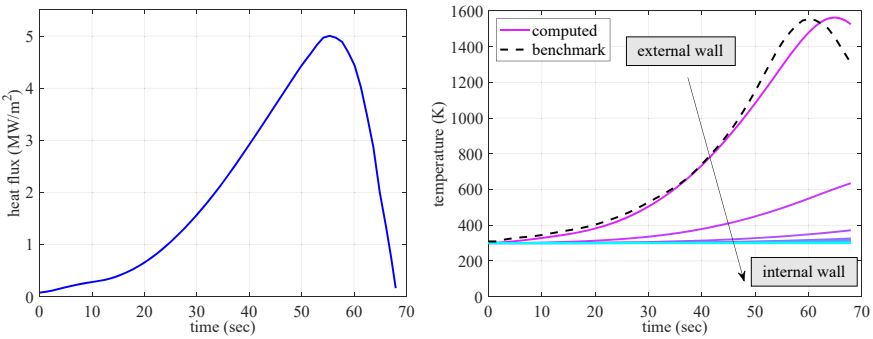


Fig. 10.55 UHTC ZrB_2-SiC TPS: incoming heat flux and temperature profiles

Phenolic-nylon ablative TPS

For verifying a proper implementation of the ablative model, as discussed in Sect. 10.5.3, a phenolic-nylon ablative material is applied. The layout of the TPS is as follows, using a comparable implementation published by Delgado Montes (1988). Also their published material data have been used. The TPS model consists of three layers:

1. Charred ablative layer. This layer with a thickness of 1 mm consists of three nodes, one representing the wall temperature, one a material node, and the third one being the interface node between the charred and uncharred layer. The used material properties for the charred phenolic-nylon is a density of $143\,\mathrm{kg/m^3}$, specific heat coefficient of $c_p = 500\,\mathrm{J/(kg\ K)}$ and a thermal conductivity of $k = 2\,\mathrm{W/(m\ K)}$;
2. Virgin (uncharred) ablative layer. This layer is 18 mm thick, has a density of $\rho = 553\,\mathrm{kg/m^3}$, a temperature-dependent specific heat coefficient of $c_p = 764.04 + 2.812T$, and a temperature-dependent thermal conduction of $k = 0.7 + 3.457 \cdot 10^{-5}T$. The layer has five nodes, the first four being material nodes and the last one being the interface node between the uncharred layet and the insulator;
3. Insulator. This last layer has three nodes, with the first two being the material nodes and the last one the interface with the inner structure, representing the inner temperature. The layer is 10 mm thick, with $\rho = 800\,\mathrm{kg/m^3}$, $c_p = 1,000\,\mathrm{J/(kg\ K)}$, and $k = 0.1\,\mathrm{W/(m\ K)}$.

The convective and radiative heat fluxes used to validate the performance of the thermal model with the phenolic-nylon TPS are provided in Fig. 10.56; these values are typical for a high-speed ballistic entry. Figure 10.57 shows the nodal temperatures and material-layer thickness, respectively. With an initial thickness of the uncharred layer of 18 mm, the output of the model is compared with the results obtained in the original report. The temperature and material recession trends are almost a per-

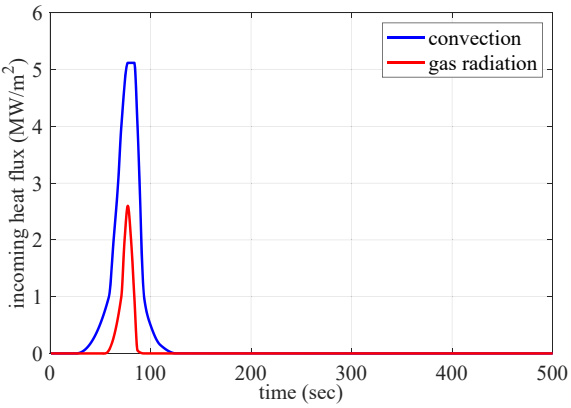


Fig. 10.56 Phenolic-Nylon TPS: benchmark heat fluxes on the external surface

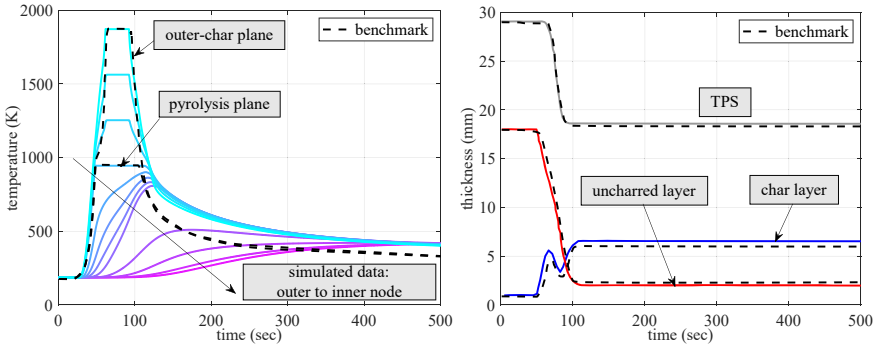


Fig. 10.57 Phenolic-Nylon TPS: nodal temperatures (left) and material-layer thickness (right)

fect match, with the only difference being a mild underestimation of the material recession.

Carbon-phenolic ablative TPS

The carbon-phenolic designed TPS is based on an original design by Sutton (1970). The performance of the thermal model is compared with test results obtained with ground-tests with experimental billets of approximately .81 cm thickness. The same layout of the TPS has been used as in the previous case, i.e., having a charred, uncharred, and insulator layer. However, in this case the uncharred layer is 38 mm thick. The material properties are now those presented in the related tables in the appendix, apart from the insulator that has been taken the same as in the previous case. The node division is three for the charred layer, and four for each of the uncharred and insulator layer. The model was initialised with an environment, wall, and back-part temperature of 800 K, to represent the test conditions.

The computed temperature profile in relation to the experimental temperature data is plotted in the left graph of Fig. 10.58. The temperature data in the high-enthalpy test

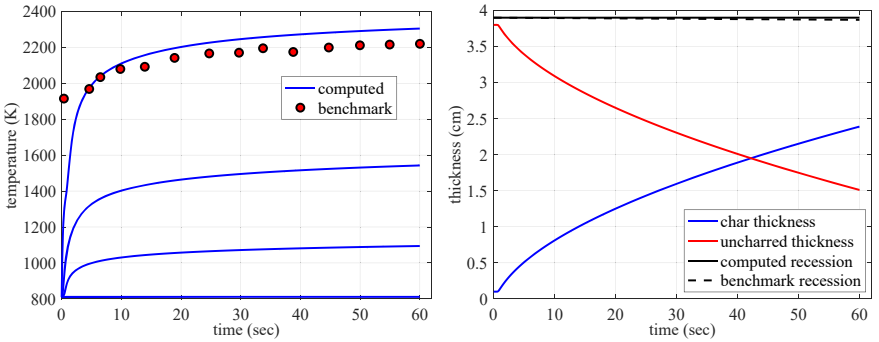


Fig. 10.58 Carbon-Phenolic TPS: comparison between benchmark and computed temperatures (left) and material recession (right)

facility was obtained with an equivalent constant heat-flux of 1.44 MW/m^2, using an atmosphere with an oxygen mass fraction equal to 0.23. From the results it is confirmed that the surface temperature shows a good match with the experimental data, whereas the temperatures of the inner nodes show a logical decrease. In terms of regression, shown in the right graph, the match is excellent.

PICA ablative TPS

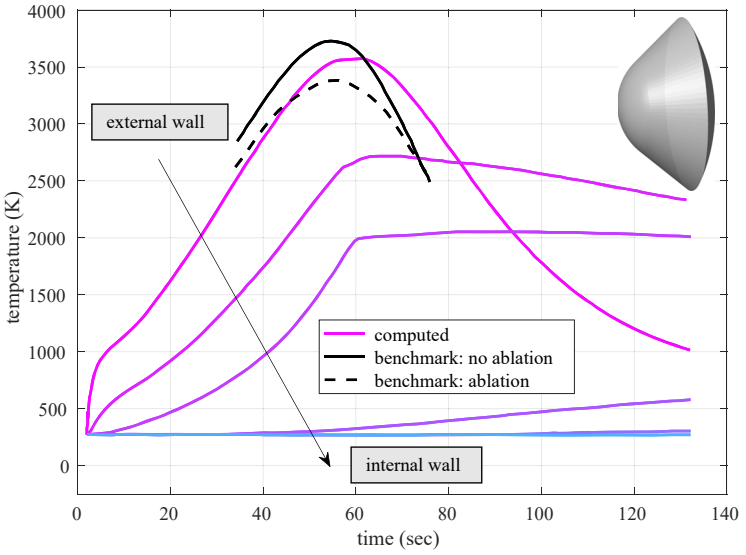


Fig. 10.59 PICA TPS. Comparison between benchmark and computed external layer temperature trends of the PICA TPS for the Stardust sample return capsule

The last test to be performed in this section is centred around a third ablative design, i.e., the PICA TPS of the Stardust entry capsule. As was discussed in the introduction to this chapter, the capsule entered Earth's atmosphere at very high speed, which necessitated the application of a PICA heat shield to withstand the large thermal flux. For verifying the developed model, data from the flow-field analysis of the Stardust sample return capsule is used (Olynick et al. 1997). The benchmark data from this study were obtained by using a two-temperature, non-equilibrium, axisymmetric flow solver, GIANTS2 (Gauss-Seidel Implicit Aerothermodynamic Navier-Stokes Code), coupled with a radiation process model, NOVAR2 (Nonequilibrium Optimized Vectorizable Radiation process model), adapted to incorporate the ablation phenomena. The entry trajectory was then discretised with seven points and laminar flow solutions with coupled radiation were generated with and without ablation. The GIANTS/NOVAR and FIAT codes were combined to perform the TPS sizing. As an interesting example of how to do a high-fidelity numerical analysis of an ablative TPS, the reader is referred to the above cited paper for more detailed information.

As before, the model has three layers with the familiar layout of a charred, uncharred, and insulator layer (where yet again the same insulator properties have been used). The thickness of each layer was 2, 57.2, and 10 mm, respectively, with three, four, and four nodes each. The model was initialised with an environment, wall, and back-part temperature of 253 K. As mentioned, the results obtained with the one-dimensional stagnation-point analysis are compared with the results obtained with the much more complete analysis performed in support of the Stardust-capsule design. The benchmark data were only available for a relatively short time interval, but nonetheless the match shown in Fig. 10.59 is acceptable. Whereas the benchmark data are provided as either an ablative state or a non-ablative state, the computed temperature fits relatively well. Another aspect that was different in the simulation was that Stardust's trajectory was recomputed due to the lack of heat-flux data as a function of time. Even though the altitude-velocity profile matched very well, the corresponding peak heat-flux was equal to 11 MW/m^2 rather than the documented 12 MW/m^2. All differences considered, it may be concluded that the verification of the PICA TPS was successful.

10.6 Case Study

Earlier, in Sect. 10.5.5 a one-dimensional finite-difference model was presented that can be integrated in an optimisation framework to find optimal re-entry trajectories, while simultaneously finding the best aerodynamic shape of the considered entry capsule and including realistic thermal constraints. In this case study, some of the results obtained by Ridolfi et al. (2012) while using that methodology will be shown. Their objective was to do a robust optimisation of unmanned entry capsules, while simultaneously considering continuous shape-variation models, aerothermodynamics, flight mechanics, and TPS designs considering a transient analysis, all at the

same level of fidelity, to provide a valuable methodology at the conceptual-design level.

The analysis that has been performed focussed on the application of six different TPS concepts, similar to those described before while verifying the one-dimensional tool. As a metallic TPS solution the application of PM-2000 is considered, which is a highly oxidation-resistant alloy. Due to the presence of a significant amount of aluminium and chromium, an Al_2O_3 oxidation layer can be formed at high temperatures in an oxidising atmosphere, which protects the TPS from the harmful effects of oxidation. Without any additional cooling, this concept is a heat sink. A variation of this metallic TPS is that it can include water as cooling mechanism, i.e., when the temperature of the water reaches the boiling limit, the heat from the skin is absorbed by the water for its evaporation, which leaves the temperature of the skin at a constant value. Two variations can be selected as an active TPS: the water is cooling through (1) nucleate boiling (the innermost skin is in direct contact with the water, such that its temperature is the same as the water itself), or (2) enhanced-radiation cooling, a concept originally developed at Delft University of Technology (Buursink 2005). When enhanced-radiation cooling is adopted, the cooling principle is still based on evaporation, but now the water is contained in a porous material, for instance, ZAL-15, detached from the external skin. At the interface the heat is transferred through radiation, therefore, the equilibrium between the innermost layer of the skin and the ZAL-15 shall be taken into account in the transient model. The third TPS concept considered is based on ceramic materials, most notably the Ultra High Temperature Ceramics (UHTC), using ZrB_2-SiC as a potential implementation. The last category with three TPS solutions focusses on the use of ablative materials, in particular phenolic nylon, carbon phenolic, and the phenolic-impregnated carbon ablator, PICA.

One of the criteria considered in the design optimisation is the extent of reusability of the TPS, i.e., (1) being fully re-usable (re-usability index equal to 3), when the nose and rear part of the capsule are made of re-usable materials, (2) being partially re-usable (re-usability index is equal to 2), when only the nose or the rear part of the capsule is made of re-usable materials and the other is made of ablative material, and (3) being non-reusable (re-usability index is equal to 1), when both the nose and the rear part present ablative solutions. The fully re-usable engineering solutions are metallic uncooled and ZrB_2-SiC, for either the nose or rear part, or both. Further, active cooling mechanisms can be selected by the optimiser for nose and cone. In particular, for the nose direct water-cooling is considered, while enhanced-radiation cooling can be applied to the rear part. The re-usability index is also related to the maximum temperature reached during the re-entry of the capsule. After a certain temperature threshold the capsule is considered to be less reusable if the external temperature gets closer to the limit temperature of the material. All TPS solutions are layered with an external skin, an intermediate layer of Saffil insulator with a thickness of 100 mm, and an internal structure of titanium with a thickness of 3 mm. The insulator and the internal structure were not included in the optimisation, and were kept equal for all the TPS solutions used in this case study. The assumptions and the specific challenges related to each of the TPS solutions are not considered any

further, as the focus will be on some of the results. They have been further detailed by Ridolfi et al. (2012), and the interested reader is referred to that publication.

The mission that is optimised is that of a cargo service executed by an unmanned capsule returning from the International Space Station. The baseline shape of the capsule is that of an Apollo capsule (see Fig. 10.60 for a representation of the common shape), with a large variation of the geometric shape parameters, i.e., a nose radius R_N (= heat-shield radius) varying between 1 and 5 m, the corner radius, R_s (0.001–0.4 m), the capsule radius, R_m (0.5 to 5 m), the inverted cone angle, θ_c(−60 to 30°), the capsule length (0.2–3 m), the z-location of the centre of mass, Δz_{com} (between 0 and 0.1 m), the type of TPS in the nose and rear, with ranges for the thickness, t_{nose} and t_{rear} between 0.0015 and 0.1 m. In case of water cooling, the initial water mass for nose and cone, m_{nose} and m_{cone}, varying between 5 and 1,000 kg. For the initial entry conditions, the flight-path angle, γ_E, was varied between −6 and −2°, whereas the initial bank angle, σ_0, ranged from 0 to 90°. One can imagine that with so many design degrees of freedom, the analysis of which design parameter influences a certain objective is extensive. In the spirit of this case study, the discussion is focussed on a limited number of results, with the remainder of the discussion covered in the cited paper.

The objective is to find trade-offs between minimum mass and maximum volume of the capsules, while maximising the re-usability at the same time. This is done by optimising the geometric parameters as discussed above, where geometric constraints are taken into account to guarantee feasible capsule geometries. For the nose TPS, a choice can be made out of (1) ZiB_2SiC, (2) metallic uncooled (heat sink), (3) phenolic-nylon ablator, (4) carbon phenolic ablator, (5) metallic direct water cooling, and (6) PICA ablator. For the conic rear part, the choice is mainly the same, except for number 5, which is a metallic enhanced radiation cooling concept. Concerning the thermal constraints, the external-skin temperature of nose and conic rear part shall remain under the limit temperature of the applied materials. The bond-line temperature, i.e., the interface temperature between the TPS and the internal structure, shall not exceed 600 K. Of course, when either an ablator or water cooling is used,

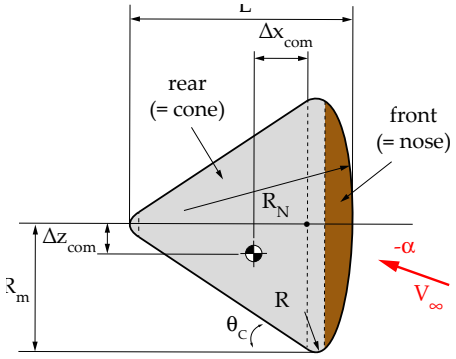


Fig. 10.60 The ballistic capsule geometry, with the baseline shape derived from the apollo command module

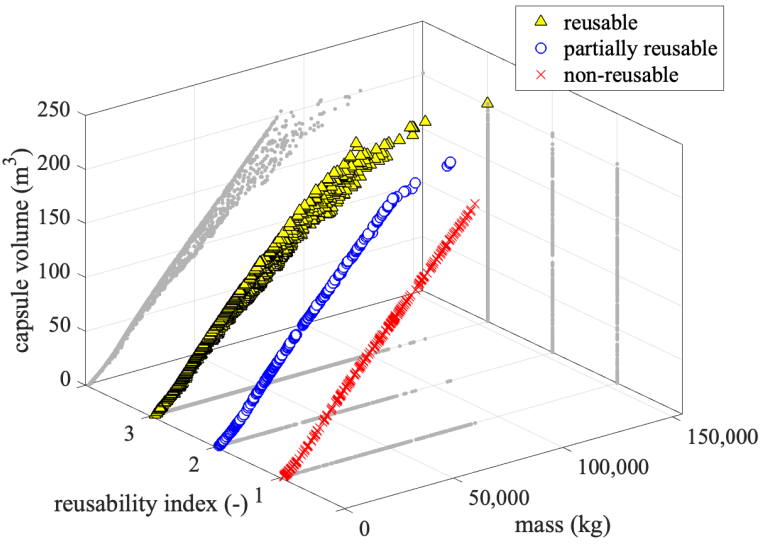


Fig. 10.61 Complete Pareto front, robotic capsules servicing the international space station

the consumed materials shall always remain below the available amount. Finally, the flight-path constraint *g-load* shall not exceed 8 g during the entire entry.

Figure 10.61 shows the Pareto fronts of the optimisation process, where the front has been divided into three parts, i.e., for the fully-reusable, partially reusable, and non-reusable capsules. The optimiser, the multi-objective evolutionary algorithm MOEA/D (Zhang and Li 2007), pushed the solutions toward a minimum mass, maximum re-usability and maximum volume, which is the preferred (and expected) solution. Looking at the front projections at the different planes shows that the performance of the various capsules with different materials is comparable. Amongst others, this means that from a mass-volume trade-off point of view there exists entry conditions and design-parameter settings for which, for instance, a capsule with the nose made of ZrB_2SiC and the rear part with metallic enhanced-cooling behaves like a capsule with a fully ablative nose and conic part. It is noted that the mass range for the Pareto-front capsules is very large, providing both very small and light, and heavy and large designs. On the high end, these capsules may simply not be launched into LEO orbit with the current launchers. However, the purpose of this case study is to study the evolution of the shapes and the engineering solutions while optimising potentially conflicting objectives. Setting wide limits on the geometric variables allowed us to do just that.

Zooming in on the results, Fig. 10.62 provides a detailed view on the Pareto front of the fully-reusable capsules shown before in Fig. 10.61. It is striking to see that almost half of the re-usable TPS is made of ZrB_2SiC, which is true for both the nose and rear part. However, looking carefully confirms that there are also capsules with a nose made of ZrB_2SiC and the rear part made of metallics. The latter include

both metallic uncooled and the metallic enhanced radiation-cooling concepts. For the nose TPS, not many metallic solutions with water cooling are found amongst the optimal solutions, which could be explained by the fact that the optimiser seems to prefer large values of nose radius to maximise the available capsule volume. This has a negative effect on the amount of water required to cool the corresponding larger nose area for a given heat-load. It is true that a large nose radius will lower the peak heat-flux on the front shield, but since the flux is given per unit time and per unit area, the total heat load on the nose will still increase such that there will be larger demand on the cooling mechanism of the TPS. It may be obvious that the increased amount of water required for the cooling results in a higher overall vehicle mass, and requires a larger volume for storage, both aspects that will degrade the performance of the concept. However, there may be some cases that the combination of peak heat-flux and integrated heat load is such that a thin metallic external surface cooled by water is more efficient than a thicker metallic skin protecting the capsule with the heat-sink principle and, thus, without water cooling from the inside.

It is mentioned once more that the re-usability index will be degraded once the maximum occurring temperature reached by the material is approaching the limit temperature. Amongst the solutions on the Pareto front of Fig. 10.62, an improvement in the re-usability index causes the capsules to be less efficient in terms of mass and volume, or, *vice versa*. This effect is more evident as the capsules get larger and heavier. Indeed, the combination of a large thickness of nose and rear-part skin, and a large amount of water when applicable, yields a lower maximum temperature at the surface, and, consequently, a larger design margin, and also a larger mass and smaller available volume.

In Fig. 10.62 five capsules following from the optimisation process have been indicated. It may be clear that these capsules get worse in terms of aerodynamics and flight-mechanics characteristics as their size increases. To support this, in Tables 10.8 and 10.9, both the design parameters and the relevant performance for these five reusable capsules are listed. The discussion of these results is further supported by the trajectory and performance plots of the five capsules given by Figs. 10.63 through 10.65.

In general, the maximum L/D decreases when the capsules are getting larger, as the drag area is the dominating effect. This means that the larger the vehicle, the more it will behave like a ballistic capsule. For the few capsules with a metallic nose, in general the optimiser pushes the *g-load* towards its limit of 8 g: the trajectories are steep, where also the peak heat-flux increases to larger values, but the heat-load reduces to lower values due to the shorter flight times. The latter is the main driver for the amount of water that is needed to cool the capsule's nose. Consequently, a low heat-load will also have an effect on reducing the amount of water to be stored on board, thereby reducing the total capsule mass. The L/D, in combination with γ_E that seems to be the dominating effect, also determines the flight range that the capsule covers from entry at 120 km altitude to the stop criterion at Mach 3. A shallower trajectory indeed results in a larger flight range, but, as was mentioned earlier, due

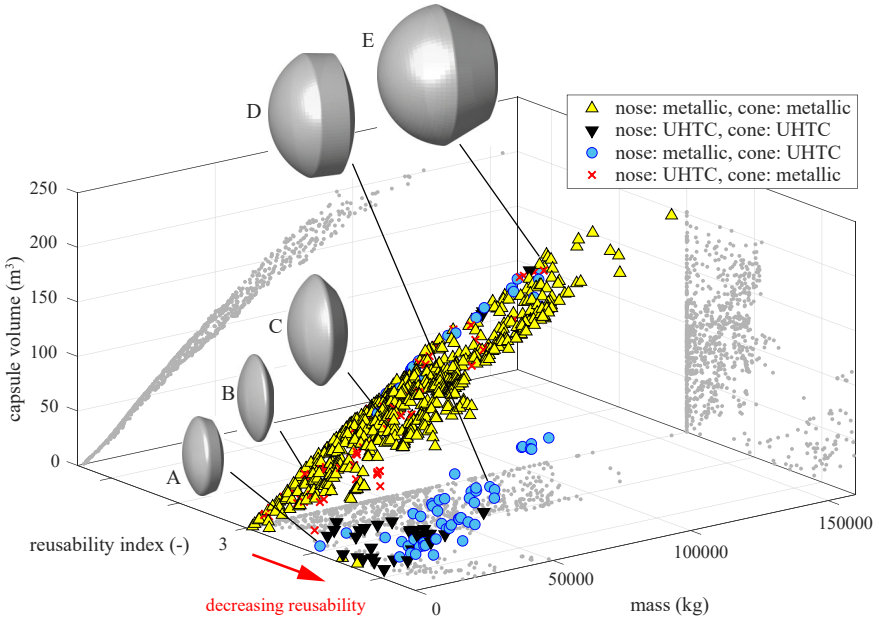


Fig. 10.62 Pareto front with fully re-usable, unmanned capsules, returning from the international space station

Table 10.8 Design-parameter settings for the reusable capsules shown in Fig. 10.62.

	A	B	C	D	E
R_N (m)	3.64	4.99	4.89	5	5
R_s (m)	0.34	0.25	0.33	0.4	0.3
θ_c (deg)	−19.5	−60.2	−45.3	−3.4	−26.9
R_m (m)	2.27	3.07	3.88	4.35	4.95
L (m)	0.6	0.2	0.97	1.95	3
Δz_{com} (m)	0.1	0.065	0.07	0.09	0.08
TPS type nose (–)	ZrB_2-SiC	water-cooled	ZrB_2SiC	ZrB_2SiC	ZrB_2SiC
t_{nose} (m)	0.0015	0.0045	0.0015	0.002	0.011
	Metallic	Enhanced	Enhanced	Enhanced	ZrB_2-SiC
TPS type rear part (–)	Uncooled	Radiation	Radiation	Radiation	
		Cooling	Cooling	Cooling	
t_{cone} (m)	0.0087	0.0015	0.0025	0.0015	0.0015
$m_{H_2O,nose,i}$ (kg)	–	730	–	–	–
$m_{H_2O,cone,i}$ (kg)	–	19.5	175	200	–
γ_E (deg)	−2.75	−3.95	−2.89	−2.00	−2.20
σ_0 (deg)	84.9	60.7	32.1	20.0	40.0

Table 10.9 Relevant performance for the reusable capsules shown in Fig. 10.62

	A	B	C	D	E
m (kg)	7,768	10,100	24,938	45,479	88,623
V (m^3)	10	13	52	100	182
d (km)	1910	1350	1940	3040	2580
L/D_{max} (–)	0.26	0.17	0.12	0.10	0.03
$g\text{-}load_{max}$ (–)	6.9	7.9	5.2	4.5	7.7
$q_{c,max}$ (MW/m^2)	0.772	0.654	0.672	0.633	1.000
$T_{nose,max}$ (K)	2,250	577	2,150	2,160	2,300
$T_{rear,max}$ (K)	700	897	900	919	1030
$T_{bond-line,max}$ (K)	550	423	547	600	460

to the longer time this takes the integrated heat load increases and, depending in the TPS design, the TPS temperature also increases (see Figs. 10.64 and 10.65).

Zooming in on capsule B, even if the heat-load is the lowest (the heat-load is determined as the total area under the heat-flux curve) the amount of water needed to keep the nose relatively cool is almost 8% of the entire capsule mass. Further, it is found that the nose turned out to be quite large (Table 10.8). Thus, what can be concluded for the given design-parameter ranges, is that to design a metallic, water-cooled nose that meets the temperature constraints, the amount of water coupled has to be substantial, but it has to be coupled with some form of heat-sink design, i.e., a rather bulky skin that will absorb and store part of the incoming heat. Irrespective of the amount of water required, water cooling is a very effective mechanism to keep the temperatures at bay. For Concept B this is confirmed by Fig. 10.65, where the temperature profile remains far below that of Concept C, despite a similar peak heat flux. Thus, since the temperature of the metallic TPS is far from its operating-limit temperature, this concept is more reusable, according to the definition that has been used here.

The heat-sink TPS, i.e., the uncooled metallic solution, is not as efficient as the others. There are not many capsules applying this TPS concept on the Pareto front, and each one of them only has a heat sink on the rear part of the vehicle, where the heat-flux is significantly lower (assumed to be 30% of the stagnation-point heat-flux). Capsule A is representative for this group of vehicles. Table 10.8 shows that the thickness of the rear part is relatively large, which, in general, makes this type of thermal protection heavy, especially when a large cone area is to be considered.

Unlike the other metallic solutions, the TPS based on enhanced-radiation cooling for the rear part of the capsules actually survives the optimisation process, even for large capsules with a large exposed surface area. This can be explained from the fact that for this rather unconventional approach to keep a metallic surface cool, only a very thin skin is required and yet only requires a limited amount of water when compared to the direct water-cooling mechanism. The Ultra High Temperature Ceramics, of which ZrB_2SiC is just one example, performs very well in terms of

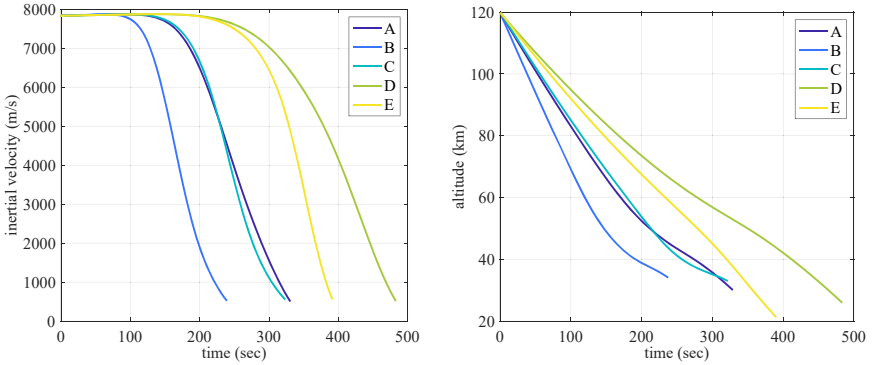


Fig. 10.63 Altitude and velocity profiles for the selected capsules *A* through *E* with reusable TPS

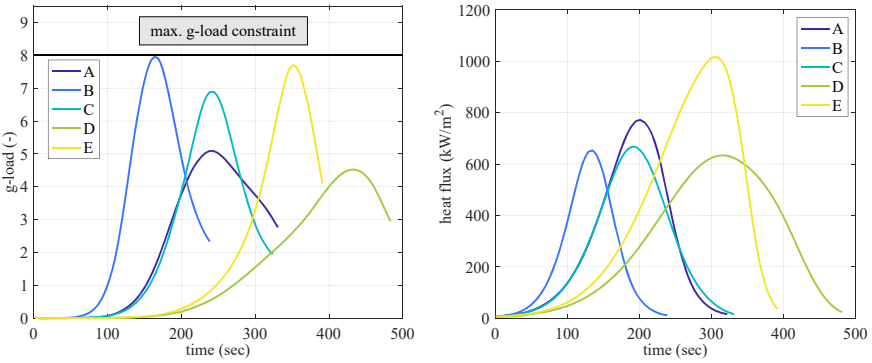


Fig. 10.64 Thermo-mechanical loads for the selected capsules *A* through *E* with reusable TPS

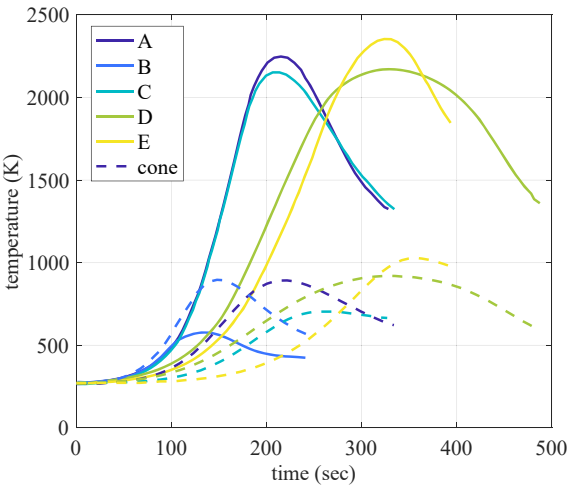


Fig. 10.65 Nose and cone (rear-part) temperature profiles for the outer skin of the capsules *A* to *E*

mass-volume efficiency. In general, a thin skin provides a good performance given the maximum temperature reached by the material, even with very large nose and conic surfaces.

Appendix—Thermal Properties

In this appendix, the thermal properties of many materials are listed, without any further discussion or interpretation (several discussions can be found in earlier sections, most notably Sect. 10.3). For many materials, different sources can be found, not all of them specifying the exact composition, any heat treatment, or test conditions. Therefore, the values provided are indicative, even though they are accurate enough to serve as input for conceptual design studies at various levels of detail. Several of the data tables have been used during the verification process of the one-dimensional finite difference model (Sect. 10.5.5) and the subsequent case study (Sect. 10.6).

The references and websites used to obtain the data are provided with each table. One reference is highlighted here, i.e., the NASA report by Williams and Curry (1992). This report lists the thermo-physical property data of many thermal protection materials, divided into two categories: ablative materials (nylon-phenolics, teflon-teflons, phenolic-nylons, phenolic-phenolics, carbon-phenolics, carbon materials with ablation data, silica-teflon, silica-phenolic, silica-silicones, polybenzimidazoles, inorganics-silicone binders, and organics-silicone binders, with a total of 84 materials) and reusable materials (carbon-based, silicon-based, polyiamides, adhesives, coating, and metals, with a total of 40 materials). Note that the properties in this report, but also in several others, are provided in US units; in this appendix they have been converted to SI units, without further listing the original values with their units.

It is noted that the temperature data in the tables are provided in degrees Kelvin, as this is the commonly applied unit in the (engineering) models. However, for ease of comparison and interpretation, in the corresponding overview figures, the data are presented as a function of temperature in degrees centigrade. This also makes these figures more consistent with those presented in earlier sections of this chapter.

Metals

Steel is not typically recognised as a thermal-protection material in re-entry technology, although it has its uses. It is listed here mainly to have some frame of reference, as it is a more commonly known material. The selected material is stainless steel AISI 310, with a density that varies between 7,840 and 8,040 kg/m^3 at room temperature. Note that the density decreases with temperature, as the material typically expands with increasing temperature. However, here only the nominal value at room temperature will be given. More detailed information can usually be found

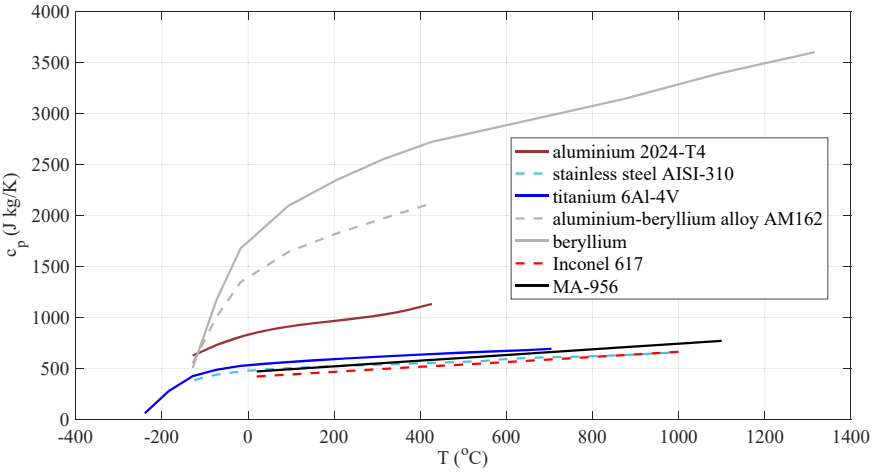


Fig. 10.66 The specific-heat coefficient as a function of temperature for several metals

in the cited literature. The thermal properties as a function of temperature are listed in Table 10.10. The next more well-known material is 2024-T4 aluminium, with a density of 2,768 kg/m^3. Its thermal properties are listed in Table 10.11. Two types of titanium, 6Al-4V titanium with a density of 4,429 kg/m^3 and grade-2 unalloyed titanium, standard oxygen (TA2), with a density 4,500 kg/m^3, have their thermal properties listed in Tables 10.12 and 10.13, respectively.

For beryllium, with a density of 1,854.6 kg/m^3, the data are shown in Table 10.14, whereas for the aluminium-beryllium alloy AM162 (density is 2,098 kg/m^3), the corresponding data have been listed in Table 10.15. The thermal properties of Inconel 617, with a density of 8,360 kg/m^3, are given in Table 10.16, whereas those of the ODS material MA956 (density of 7,250 kg/m^3) are shown in Table 10.17. The equivalent Plansee material PM-2000 is listed in Table 10.18. The density of PM-2000 is 7,180 kg/m^3, its melting point is 1,756 K, while the maximum service temperature is 1,623 K. A temperature of 1200 K is considered the limit for full re-usability of PM-2000. Note that the validation of the one-dimensional finite difference model has been performed with the material PM-1000. PM-1000 was used in the original publication, but since it is no longer in production, Ridolfi et al. (2012) decided to use PM-2000 for the actual robust-optimisation process described in the case study.

The density of PM-1000 is 8,240 kg/m^3. For a temperature range in between $T = 273.15$ and 1,023 K, its thermal conductivity in W/(m K) is given by $k_{PM-1000} = 11 + 0.025333(T - 273.15)$, whereas for temperatures in between 1,023 and 1,470 K, the expression is $k_{PM-1000} = 30 + 0.026667(T - 1023)$. Finally, the specific heat coefficient of PM-1000 in J/(kg K) is $c_{p,PM-1000} = 410 + 0.408333T$ for the full temperature range of $T = 273.15 - 1,470$ K.

For ease of interpretation, Figs. 10.66 and 10.67 have been added. These figures show the variation with temperature of c_p and k for a number of metals.

Table 10.10 Stainless steel AISI 310 thermal properties as a function of temperature (Engineering ToolBox, (2019). AISI 310 Stainless Steel—Properties. Available at: https://www.engineeringtoolbox.com/stainless-steel-310-properties-d_2167.html. Date accessed: May 21, 2023)

T K	c_p J/(kg K)	k W/(m K)
148	382.6	9.4
173	412.8	10.1
198	435.9	10.8
223	452.2	11.2
248	465.1	11.7
273	475.7	12.1
298	484.1	12.7
323	491.0	13.2
373	502.2	14.1
423	512.8	15.0
473	519.2	15.8
523	526.2	16.6
573	534.2	17.3
623	542.1	18.0
673	549.2	18.8
723	556.2	19.4
773	563.2	20.1
823	572.4	20.9
873	590.9	22.2
923	602.0	23.0
973	607.9	23.7
1023	612.5	24.0
1073	618.4	24.5
1123	625.4	25.2
1173	634.4	26.0
1223	646.3	26.9
1273	654.7	27.7

Ceramics

The ultra-high temperature ceramics (UHTC) material used in Sects. 10.5.5 and 10.6 is ZrB_2SiC. This heat-resistant material has undergone several experiments in the high-enthalpy test facility of CIRA, thus providing some information in literature (Francese 2007). The density of ZrB_2SiC is 5,610 kg/m^3 and its melting temperature is 3,020 K. The maximum temperature for single use of the material is 2,860 K, while it can be reused multiple times if the maximum temperature remains below 2,300 K.

Table 10.11 2024-T4 aluminium thermal properties as a function of temperature (Brown Jr. et al. 1994)

T (K)	c_p J/(kg K)	k W/(m K)
144	623.8	88.2
200	728.5	101.7
256	808.0	111.4
311	866.7	122.6
367	908.5	134.6
422	942.0	149.5
478	967.1	172.0
533	992.3	183.9
589	1025.8	186.9
644	1071.8	181.7
700	1130.4	172.7

Table 10.12 6Al-4V titanium thermal properties as a function of temperature (Brown Jr. et al. 1994)

T (K)	c_p J/(kg K)	k W/(m K)
33	58.6	2.3
89	276.3	3.6
144	422.9	4.9
200	485.7	6.0
256	523.3	7.1
311	544.3	8.1
367	561.0	8.8
422	577.8	9.9
478	590.3	10.7
533	602.9	11.7
589	615.5	12.9
644	628.0	13.9
700	640.6	15.0
756	648.9	16.1
811	661.5	17.2
867	669.9	18.2
922	678.3	19.4
978	690.8	20.3

Table 10.13 Grade-2 unalloyed titanium (TA2) thermal properties as a function of temperature (Liu and Zhang 2011)

T (K)	c_p J/(kg K)	k W/(m K)
293	544	16.33
373	544	16.33
473	628	16.33
573	670	16.75
673	712	17.17
773	754	18.00
873	837	18.00

Table 10.14 Beryllium thermal properties as a function of temperature (Williams and Curry 1992)

T (K)	c_p J/(kg K)	k W/(m K)
144	502.4	112.2
200	1172.3	138.3
256	1674.7	171.2
367	2093.4	169.7
478	2344.6	142.1
589	2553.9	130.1
700	2721.4	117.4
922	2930.7	102.4
1144	3140.1	88.2
1367	3391.3	84.5
1589	3600.6	62.3

Table 10.15 Aluminium-beryllium alloy AM162 thermal properties as a function of temperature (Grensing and Hashiguchi 1995)

T (K)	c_p J/(kg K)	k W/(m K)
144	548.5	343.9
200	1004.8	276.6
256	1343.9	224.3
367	1641.2	186.9
478	1821.2	164.5
589	1972.0	145.8
700	2118.5	134.6

Table 10.16 Inconel 617 thermal properties as a function of temperature (https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-617.pdf. Date accessed: May 21, 2023)

T (K)	c_p J/(kg K)	k W/(m K)
293	419.0	13.4
373	440.0	14.7
473	465.0	16.3
573	490.0	17.7
673	515.0	19.3
773	536.0	20.9
873	561.0	22.5
973	586.0	23.9
1073	611.0	25.5
1173	636.0	27.1
1273	662.0	28.7

Table 10.17 MA956 thermal properties as a function of temperature (https://www.specialmetals.com/. Date accessed: 2005. No longer available)

T (K)	c_p J/(kg K)	k W/(m K)
293	469.0	10.9
373	491.0	12.2
473	519.0	13.9
573	547.0	15.4
673	575.0	16.9
773	602.0	18.4
873	630.0	19.8
973	658.0	21.2
1073	686.0	22.6
1173	714.0	24.1
1273	741.0	25.5
1373	769.0	27.0

In Table 10.19 the thermo-physical properties of the material are listed. In this table, ε is the dimensionless emissivity coefficient. The next material is graphite/epoxy fabric, with a density of 1,577.8 kg/m^3. The corresponding properties are provided in Table 10.20, where the in-plane properties are given for the direction parallel to the plane of the fibres, whereas the perpendicular ones are through the thickness of the fabric, i.e., perpendicular to the main surface area.

The reinforced carbon-carbon (RCC) thermal properties as a function of temperature are shown in Table 10.21, whereas the density is 1,656.3 kg/m^3. A second

Table 10.18 PM-2000 thermal properties as a function of temperature (Plansee, Schwarzkopf, "PM-2000 data sheet", 2012. Retrieved from http://www.matweb.com. Date accessed: March 19, 2023)

T (K)	c_p J/(kg K)	k W/(m K)
373	500	–
473	480	16.0
773	610	21.0
1023	680	22.0
1273	740	25.5
1473	–	28.0

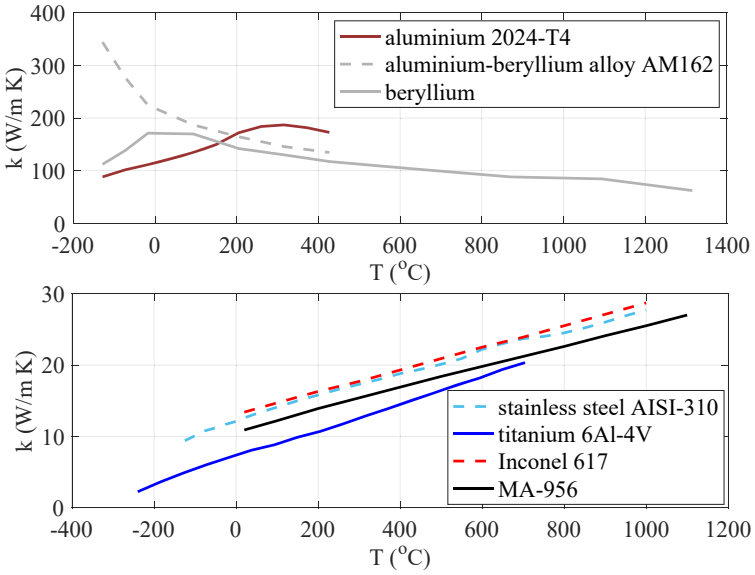


Fig. 10.67 The thermal conductivity as a function of temperature for several metals

material used for the Space Shuttle is LI-900 (LRSI/HRSI), for which the specific heat capacity as a function of temperature is listed in Table 10.22, the in-plane thermal conductivity in Table 10.23, and the perpendicular thermal conductivity in Table 10.24. Its density is 144.2 kg/m^3.

For the global overview, Fig. 10.68 shows the variation with temperature of c_p for graphite-epoxy fabric, the reinforced carbon-carbon, and LI-900. In Fig. 10.69 the corresponding curves for k have been plotted, where for RCC and LI-900 the distinction has been made for the parallel and perpendicular fibre orientation, and for LI-900 also for a low and a high ambient pressure. This pressure has a distinctive

effect on the thermal conduction, i.e., for the lower pressures a reduction of up to 60%. The fibre orientation has an even larger impact.

Insulators

As insulators, the thermal properties of Nomex felt (density is 86.5 kg/m^3), Q-felt (density is 56.1 kg/m^3), Cerrachrome (density is 96.1 kg/m^3), and three Saffil insulators with various densities (24, 48, and 92 kg/m^3) are considered. Subsequently, the properties can be found in:

1. Nomex felt thermal properties as a function of temperature and pressure, Table 10.25,
2. Q-felt thermal properties as a function of temperature and pressure, Tables 10.26 and 10.27,
3. Cerrachrome thermal properties as a function of temperature and pressure, Tables 10.28 and 10.29,

Table 10.19 ZrB_2SiC thermal properties as a function of temperature (Francese 2007)

T (K)	c_p J/(kg K)	T (K)	k W/(m K)	T (K)	ε (–)
298	459	300	103.8	1037	0.775
300	462	900	83.07	1169	0.712
400	554	1500	76.15	1263	0.72
500	603	2100	69.23	1334	0.734
600	633	2400	67.50	1485	0.748
700	656	–	–	1681	0.724
800	673	–	–	1842	0.663
900	687	–	–	–	–
1000	699	–	–	–	–
1100	711	–	–	–	–
1200	721	–	–	–	–
1300	731	–	–	–	–
1400	741	–	–	–	–
1500	750	–	–	–	–
1600	759	–	–	–	–
1700	768	–	–	–	–
1800	776	–	–	–	–
1900	785	–	–	–	–
2000	794	–	–	–	–
2100	802	–	–	–	–
2200	811	–	–	–	–

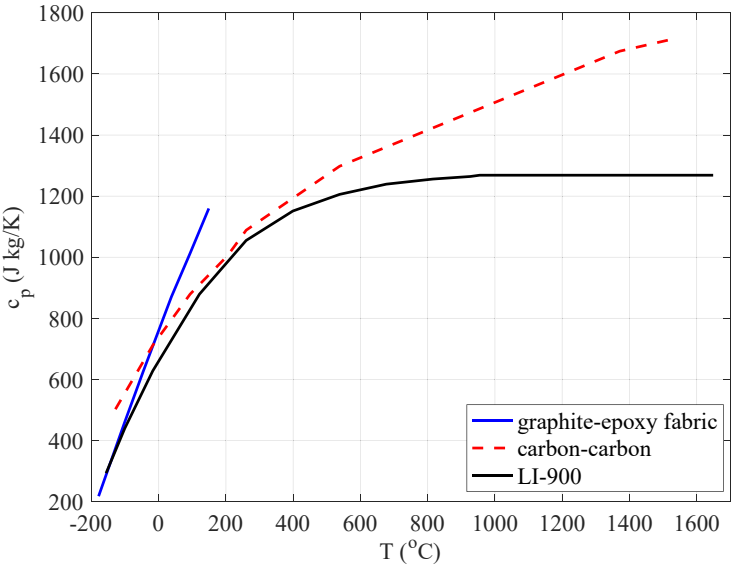


Fig. 10.68 The specific-heat coefficient as a function of temperature for several ceramics

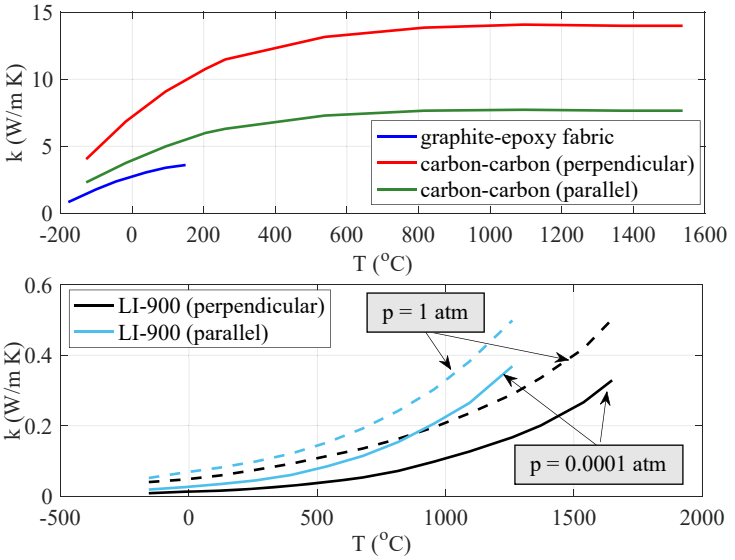


Fig. 10.69 The thermal conductivity as a function of temperature for several ceramics

Table 10.20 Graphite/epoxy fabric thermal properties as a function of temperature (Ehrlich Jr. et al. 1992)

T (K)	c_p J/(kg K)	k (in-plane) W/(m K)	k (perpendicular) W/(m K)
94	217.7	0.83	0.36
172	460.5	1.78	0.57
228	628.0	2.37	0.69
311	870.8	3.04	0.84
367	1013.2	3.39	0.92
422	1159.7	3.60	0.99

Table 10.21 Reinforced carbon-carbon (RCC) thermal properties as a function of temperature (Williams and Curry 1992)

T (K)	c_p J/(kg K)	k (in-plane) W/(m K)	k (perpendicular) W/(m K)
144	502.4	4.0	2.3
256	711.7	6.9	3.8
367	879.2	9.1	5.0
478	1004.8	10.8	6.0
533	1088.6	11.5	6.3
811	1297.9	13.2	7.3
1089	1423.5	13.8	7.6
1367	1549.1	14.1	7.7
1644	1674.7	14.0	7.6
1811	1716.6	14.0	7.6

4. Saffil (density 24 kg/m^3) thermal properties as a function of temperature and pressure, Tables 10.30 and 10.31,
5. Saffil (density 48 kg/m^3) thermal properties as a function of temperature and pressure, Tables 10.32 and 10.33, and
6. Saffil (density 92 kg/m^3) thermal properties as a function of temperature and pressure, Tables 10.34 and 10.35. The properties of saffil with density 92 kg/m^3 at $p = 1$ atm are also used by (Liu and Zhang 2011) for their model evaluation and used in Sect. 10.5.5 for model verification.

In Fig. 10.70, the specific heat coefficient is plotted versus temperature for the Nomex felt, Q-felt, Cerrachrome, and Saffil insulator, and is independent of pressure. Note that c_p is the same for each of the three Saffil densities. Figure 10.71 shows the thermal conduction as a function of temperature, and is presented for the two limit pressures listed in the tables. As with the ceramics, the conductivity is larger for the higher pressure. To conclude, and to visualise not just the pressure dependency, but also the difference between the three Saffil densities, Fig. 10.72 presents the Saffil thermal conductivity's dependence on temperature, pressure, and density. The more

Table 10.22 LI-900 (LRSI/HRSI) specific heat capacity as a function of temperature (Williams and Curry 1992)

T (K)	c_p J/(kg K)
117	293.1
172	439.6
256	628.0
394	879.2
533	1055.1
672	1151.4
811	1205.8
950	1239.3
1089	1256.0
1200	1264.4
1228	1268.6
1367	1268.6
1922	1268.6

Table 10.23 LI-900 (LRSI/HRSI) thermal conductivity (in-plane) in W/(m K) as a function of temperature and pressure (Williams and Curry 1992)

T (K)	$p = 0.0001$ (atm)	$p = 0.001$ (atm)	$p = 0.01$ (atm)	$p = 0.1$ (atm)	$p = 1.0$ (atm)
117	0.02	0.02	0.04	0.04	0.05
256	0.03	0.03	0.05	0.06	0.07
394	0.03	0.04	0.06	0.07	0.08
533	0.04	0.05	0.07	0.09	0.10
672	0.06	0.07	0.09	0.11	0.12
811	0.08	0.09	0.11	0.14	0.15
950	0.11	0.12	0.14	0.18	0.19
1089	0.15	0.17	0.19	0.23	0.24
1228	0.21	0.22	0.24	0.29	0.30
1367	0.26	0.28	0.31	0.37	0.38
1533	0.37	0.38	0.42	0.47	0.50

compact Saffil fibre (higher density) has the lowest conductivity, and even more so at the lowest pressure.

Table 10.24 LI-900 (LRSI/HRSI) thermal conductivity (perpendicular) in W/(m K) as a function of temperature and pressure (Williams and Curry 1992)

	$p = 0$ (atm)	$p = 0.0001$ (atm)	$p = 0.001$ (atm)	$p = 0.01$ (atm)	$p = 0.1$ (atm)	$p = 1.0$ (atm)
T (K)	k (perpendicular) W/(m K					
117	0.01	0.01	0.01	0.03	0.04	0.04
256	0.01	0.01	0.02	0.03	0.04	0.05
394	0.02	0.02	0.02	0.04	0.05	0.06
533	0.02	0.02	0.03	0.05	0.07	0.07
672	0.03	0.03	0.04	0.06	0.09	0.09
811	0.04	0.04	0.05	0.07	0.10	0.11
950	0.05	0.05	0.06	0.09	0.13	0.14
1089	0.07	0.07	0.08	0.11	0.15	0.16
1228	0.10	0.10	0.11	0.13	0.18	0.20
1367	0.13	0.13	0.14	0.16	0.22	0.24
1533	0.17	0.17	0.18	0.20	0.27	0.29
1644	0.20	0.20	0.21	0.24	0.31	0.34
1811	0.27	0.27	0.28	0.31	0.38	0.42
1922	0.33	0.33	0.34	0.38	0.45	0.50

Table 10.25 Nomex felt thermal properties as a function of temperature and pressure (Williams and Curry 1992)

		k W/(m K)						
T (K)	c_p J/(kg K)	$p = 0$ (atm)	$p = 0.00001$ (atm)	$p = 0.0001$ (atm)	$p = 0.001$ (atm)	$p = 0.01$ (atm)	$p = 0.1$ (atm)	$p = 1.0$ (atm)
117	1256.0	0.011	0.011	0.012	0.014	0.016	0.018	0.019
256	1306.3	0.014	0.014	0.018	0.024	0.030	0.034	0.036
311	1323.0	0.015	0.015	0.021	0.029	0.036	0.041	0.043
367	1339.8	0.016	0.016	0.024	0.034	0.042	0.048	0.050
422	1373.3	0.018	0.018	0.027	0.038	0.048	0.056	0.058
478	1402.6	0.019	0.019	0.029	0.043	0.055	0.064	0.066
589	1444.4	0.023	0.023	0.036	0.055	0.070	0.082	0.084
700	1507.2	0.026	0.026	0.043	0.066	0.087	0.105	0.108
1367	1591.0	0.030	0.030	0.052	0.080	0.106	0.134	0.138

Table 10.26 Q-felt thermal properties as a function of temperature and pressure (Gorton et al. 1993)

		k W/(m K)		
T (K)	c_p J/(kg K)	$p = 5.13 \cdot 10^{-6}$ (atm)	$p = 1.31 \cdot 10^{-5}$ (atm)	$p = 1.31 \cdot 10^{-4}$ (atm)
256	757.0	0.00	0.00	0.01
311	800.9	0.00	0.00	0.01
367	844.9	0.01	0.01	0.01
422	884.7	0.01	0.01	0.02
478	949.6	0.02	0.02	0.02
589	1032.5	0.03	0.03	0.03
700	1103.6	0.05	0.05	0.05
811	1141.3	0.06	0.06	0.06
922	1167.3	0.08	0.08	0.08
1033	1176.9	0.11	0.11	0.11
1089	1183.6	0.13	0.13	0.13
1144	1190.7	0.16	0.16	0.16
1372	1218.3	0.25	0.25	0.25

Table 10.27 Q-felt thermal conduction as a function of temperature and pressure—concluded (Gorton et al. 1993)

	k W/(m K)				
T (K)	$p = 1.31 \cdot 10^{-3}$ (atm)	$p = 1.31 \cdot 10^{-2}$ (atm)	$p = 6.58 \cdot 10^{-2}$ (atm)	$p = 0.13$ (atm)	$p = 1.00$ (atm)
256	0.01	0.03	0.03	0.03	0.03
311	0.01	0.03	0.03	0.03	0.03
367	0.02	0.03	0.04	0.04	0.04
422	0.03	0.04	0.05	0.05	0.05
478	0.03	0.05	0.05	0.06	0.06
589	0.04	0.07	0.07	0.07	0.08
700	0.06	0.08	0.09	0.10	0.10
811	0.07	0.11	0.12	0.12	0.12
922	0.09	0.10	0.14	0.14	0.14
1033	0.12	0.15	0.17	0.17	0.18
1089	0.14	0.18	0.20	0.20	0.20
1144	0.17	0.20	0.22	0.22	0.23
1372	0.26	0.30	0.32	0.33	0.33

Table 10.28 Cerrachrome thermal properties as a function of temperature and pressure (Gorton et al. 1993)

		k W/(m K)		
T (K)	c_p J/(kg K)	$p = 5.13 \cdot 10^{-6}$ (atm)	$p = 1.31 \cdot 10^{-5}$ (atm)	$p = 1.31 \cdot 10^{-4}$ (atm)
256	690.0	0.01	0.01	0.01
367	819.8	0.01	0.01	0.01
478	949.6	0.02	0.02	0.02
589	1039.2	0.03	0.03	0.03
700	1105.3	0.04	0.04	0.04
811	1141.3	0.06	0.06	0.06
922	1167.7	0.09	0.09	0.09
1033	1175.6	0.12	0.12	0.12
1144	1188.6	0.16	0.16	0.16
1256	1202.0	0.20	0.20	0.20
1367	1215.0	0.25	0.25	0.25
1478	1228.4	0.32	0.32	0.32
1589	1241.4	0.38	0.38	0.38

Table 10.29 Cerrachrome thermal conduction as a function of temperature and pressure—concluded (Gorton et al. 1993)

	k W/(m K)				
T (K)	$p = 1.31 \cdot 10^{-3}$ (atm)	$p = 1.31 \cdot 10^{-2}$ (atm)	$p = 6.58 \cdot 10^{-2}$ (atm)	$p = 0.13$ (atm)	$p = 1.00$ (atm)
256	0.02	0.03	0.04	0.04	0.04
367	0.02	0.03	0.04	0.04	0.04
478	0.03	0.05	0.05	0.05	0.06
589	0.04	0.06	0.07	0.07	0.07
700	0.05	0.07	0.09	0.09	0.09
811	0.07	0.10	0.11	0.12	0.12
922	0.10	0.12	0.14	0.15	0.15
1033	0.13	0.15	0.18	0.18	0.19
1144	0.17	0.19	0.22	0.22	0.23
1256	0.21	0.23	0.26	0.27	0.27
1367	0.26	0.29	0.32	0.32	0.33
1478	0.32	0.35	0.38	0.39	0.40
1589	0.38	0.41	0.44	0.45	0.46

Table 10.30 Saffil (density 24 kg/m^3) thermal properties as a function of temperature and pressure (Daryabeigi 1999)

		k W/(m K)			
T (K)	c_p J/(kg K)	$p = 1.31 \cdot 10^{-7}$ (atm)	$p = 1.31 \cdot 10^{-6}$ (atm)	$p = 1.31 \cdot 10^{-5}$ (atm)	$p = 1.31 \cdot 10^{-4}$ (atm)
253	686.6	0.004	0.004	0.005	0.008
273	724.3	0.005	0.006	0.006	0.009
373	950.4	0.013	0.013	0.014	0.017
473	1021.6	0.026	0.026	0.026	0.030
573	1092.7	0.044	0.044	0.045	0.048
673	1138.8	0.069	0.069	0.069	0.073
773	1172.3	0.100	0.100	0.101	0.104
873	1197.4	0.139	0.140	0.140	0.143
973	1222.5	0.186	0.186	0.187	0.190
1073	1239.3	0.242	0.242	0.242	0.245
1173	1251.8	0.305	0.305	0.306	0.309
1273	1260.2	0.378	0.378	0.378	0.381
1373	1268.6	0.459	0.460	0.460	0.463
1473	1272.8	0.550	0.550	0.551	0.554

Table 10.31 Saffil (density 24 kg/m^3) thermal properties as a function of temperature and pressure—concluded (Daryabeigi 1999)

	k W/(m K)			
T (K)	$p = 1.31 \cdot 10^{-3}$ (atm)	$p = 1.31 \cdot 10^{-2}$ (atm)	$p = 0.13$ (atm)	$p = 1.00$ (atm)
253	0.020	0.026	0.027	0.027
273	0.021	0.028	0.029	0.029
373	0.032	0.042	0.043	0.044
473	0.046	0.060	0.062	0.063
573	0.065	0.083	0.087	0.087
673	0.091	0.112	0.117	0.117
773	0.123	0.148	0.153	0.154
873	0.163	0.191	0.197	0.198
973	0.210	0.241	0.249	0.250
1073	0.265	0.298	0.308	0.309
1173	0.329	0.365	0.375	0.377
1273	0.402	0.440	0.452	0.453
1373	0.483	0.523	0.537	0.538
1473	0.574	0.616	0.631	0.633

Table 10.32 Saffil (density 48 kg/m^3) thermal properties as a function of temperature and pressure (Daryabeigi 1999)

		k W/(m K)			
T (K)	c_p J/(kg K)	$p = 1.31 \cdot 10^{-7}$ (atm)	$p = 1.31 \cdot 10^{-6}$ (atm)	$p = 1.31 \cdot 10^{-5}$ (atm)	$p = 1.31 \cdot 10^{-4}$ (atm)
253	686.6	0.002	0.002	0.003	0.005
273	724.3	0.003	0.003	0.003	0.005
373	950.4	0.007	0.007	0.007	0.009
473	1021.6	0.013	0.013	0.013	0.015
573	1092.7	0.022	0.022	0.022	0.024
673	1138.8	0.034	0.034	0.035	0.036
773	1172.3	0.050	0.050	0.050	0.052
873	1197.4	0.070	0.070	0.070	0.072
973	1222.5	0.093	0.093	0.093	0.095
1073	1239.3	0.121	0.121	0.121	0.123
1173	1251.8	0.153	0.153	0.153	0.154
1273	1260.2	0.189	0.189	0.189	0.191
1373	1268.6	0.230	0.230	0.230	0.232
1473	1272.8	0.275	0.275	0.275	0.277

Table 10.33 Saffil (density 48 kg/m^3) thermal properties as a function of temperature and pressure—concluded (Daryabeigi 1999)

	k W/(m K)			
T (K)	$p = 1.31 \cdot 10^{-3}$ (atm)	$p = 1.31 \cdot 10^{-2}$ (atm)	$p = 0.13$ (atm)	$p = 1.00$ (atm)
253	0.014	0.023	0.024	0.024
273	0.015	0.024	0.026	0.026
373	0.020	0.033	0.036	0.037
473	0.027	0.044	0.049	0.049
573	0.036	0.057	0.064	0.064
673	0.049	0.073	0.081	0.082
773	0.065	0.092	0.102	0.103
873	0.084	0.114	0.126	0.128
973	0.108	0.141	0.154	0.156
1073	0.135	0.170	0.185	0.188
1173	0.167	0.203	0.221	0.224
1273	0.203	0.242	0.261	0.264
1373	0.244	0.283	0.305	0.308
1473	0.289	0.330	0.353	0.357

Table 10.34 Saffil (density 92 kg/m^3) thermal properties as a function of temperature and pressure (Daryabeigi 1999)

		k W/(m K)			
T (K)	c_p J/(kg K)	$p = 1.31 \cdot 10^{-7}$ (atm)	$p = 1.31 \cdot 10^{-6}$ (atm)	$p = 1.31 \cdot 10^{-5}$ (atm)	$p = 1.31 \cdot 10^{-4}$ (atm)
253	686.6	0.002	0.002	0.002	0.003
273	724.3	0.002	0.002	0.002	0.003
373	950.4	0.004	0.004	0.004	0.005
473	1021.6	0.007	0.007	0.007	0.008
573	1092.7	0.011	0.012	0.012	0.012
673	1138.8	0.018	0.018	0.018	0.019
773	1172.3	0.025	0.025	0.025	0.026
873	1197.4	0.035	0.035	0.035	0.036
973	1222.5	0.047	0.047	0.047	0.048
1073	1239.3	0.061	0.061	0.061	0.061
1173	1251.8	0.076	0.076	0.076	0.077
1273	1260.2	0.094	0.094	0.095	0.096
1373	1268.6	0.115	0.115	0.115	0.116
1473	1272.8	0.138	0.138	0.138	0.138

Table 10.35 Saffil (density 92 kg/m^3) thermal properties as a function of temperature and pressure—concluded (Daryabeigi 1999)

	k W/(m K)			
T (K)	$p = 1.31 \cdot 10^{-3}$ (atm)	$p = 1.31 \cdot 10^{-2}$ (atm)	$p = 0.13$ (atm)	$p = 1.00$ (atm)
253	0.010	0.021	0.024	0.024
273	0.010	0.022	0.025	0.025
373	0.012	0.027	0.033	0.034
473	0.015	0.034	0.042	0.043
573	0.020	0.041	0.052	0.053
673	0.026	0.050	0.063	0.065
773	0.034	0.059	0.076	0.078
873	0.043	0.071	0.089	0.092
973	0.055	0.084	0.105	0.108
1073	0.068	0.099	0.123	0.126
1173	0.084	0.115	0.141	0.146
1273	0.102	0.135	0.163	0.167
1373	0.123	0.155	0.186	0.191
1473	0.145	0.178	0.211	0.217

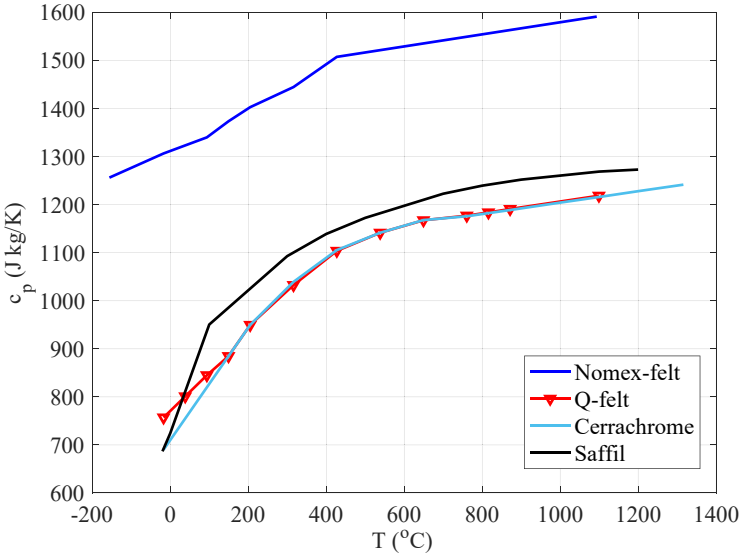


Fig. 10.70 The specific-heat coefficient as a function of temperature for several insulators

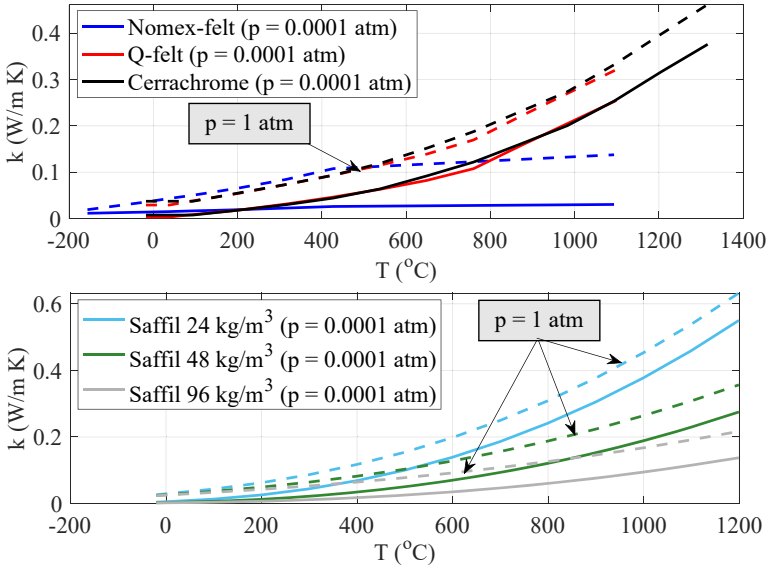


Fig. 10.71 The thermal conductivity as a function of temperature for several insulators

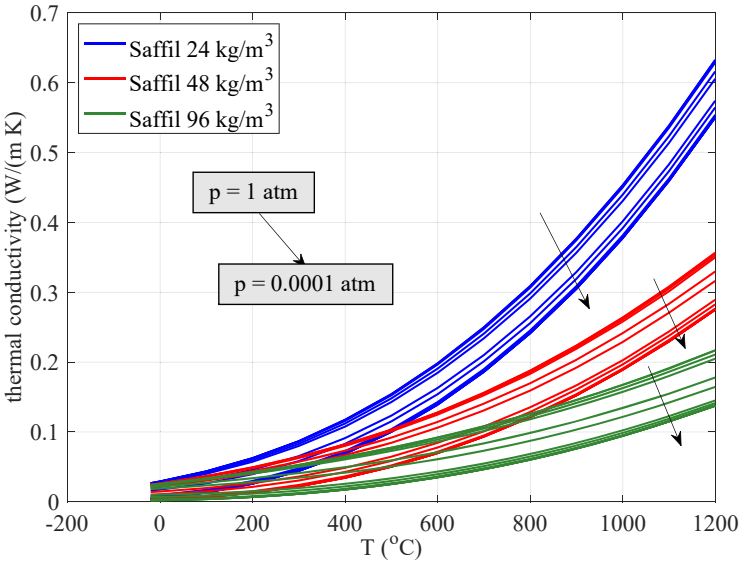


Fig. 10.72 The thermal conductivity as a function of temperature for several insulators

Ablative Materials

To conclude this appendix, the thermal properties of two commonly applied ablative materials are given, i.e., carbon phenolic (Table 10.36) and PICA (Table 10.37). Both materials can be in two states, an uncharred and a charred one. The phenolic-nylon ablative material has a specific heat that varies with temperature according to the following expression: $764.04 + 2.812\,T$ J/(kg K) in the uncharred form. For the charred phenolic nylon a constant specific heat equal to 800 J/(kg K) may be considered. Its conductivity is equal to $0.7 + 3.457 \cdot 10^{-5} T$ W/(m K) when uncharred, whereas for charred phenolic nylon the conductivity is equal to 2 W/(m K).

Problems

Judge whether the following statements are TRUE or FALSE (Problems 10.1 to 10.10).

10.1 Between an actively-cooled TPS and an ablative heat shield, the ablative heat shield can withstand higher thermal loads

10.2 Reusable thermal-protection systems eliminate the need for refurbishment or replacement after each mission.

10.3 The choice of thermal-protection system depends solely on the re-entry velocity of the vehicle.

Table 10.36 Carbon phenolic thermo-physical properties

Uncharred			Charred		
T (K)	c_p J/(kg K)	k W/(m K)	T (K)	c_p J/(kg K)	k W/(m K)
256	990	0.561	278	1,000	0.81
311	1,220	0.639	556	1,380	0.872
367	1,330	0.693	811	1,610	0.935
422	1,390	0.742	1,089	1,860	0.977
477	1,450	0.774	1,366	2,010	1.128
533	1,510	0.774	1,645	2,060	1.189
589	1,560	0.742	1,923	2,110	1.502
644	1,620	0.698	2,200	2,150	2.065
700	1,680	0.630	2,478	2,170	2.790
811	1,800	0.479	2,756	2,190	3.502
–	–	–	3,030	2,210	4.88
–	–	–	3,311	2,240	6.39
–	–	–	3,590	2,260	7.38
–	–	–	3,867	2,280	9.325

Table 10.37 PICA thermo-physical properties

Uncharred		Charred		
T (K)	c_p J/(kg K)	k W/(m K)	c_p J/(kg K)	k W/(m K)
256	879	0.0397	733	0.0397
294	984	0.0402	783	0.0402
444	1,298	0.0416	1,093	0.0416
556	1,465	0.0453	1,319	0.0453
644	1,570	0.0470	1,432	0.0470
833	1,716	0.0486	1,674	0.0486
1,111	1,863	0.0523	1,842	0.0523
1,389	1,934	0.0560	1,967	0.0560
1,667	1,980	0.0698	2,051	0.0605
1,944	1,988	0.0872	2,093	0.0729
2,222	2,001	0.111	2,110	0.0922
2,778	2,009	0.175	2,135	0.146
3,333	2,009	0.278	2,152	0.223

10.4 The ablative thermal-protection system works by dissipating heat through melting and vaporisation of the material.
10.5 Thermal-protection systems are primarily designed to prevent the heating of the vehicle during re-entry.
10.6 Thermal-protection systems for re-entry vehicles are only necessary for spacecraft returning from interplanetary missions.
10.7 Non-ablative thermal-protection systems rely on reflecting or refracting heat away from the vehicle's surface.
10.8 Thermal-protection systems are only necessary for the vehicle's exterior surfaces.
10.9 The design of the thermal-protection system is decoupled from the structural design of the vehicle.
10.10 Assuming a cold-wall condition for the design of a thermal-protection system will lead to higher equilibrium wall temperature.

TPS design and analysis (Problems 10.11 to 10.15).

During flight, a hypersonic aircraft has a constant heat load in the wing leading edge of $q_c = 200\,\text{kW/m}^2$. *This heat load has been calculated assuming a (constant) wall temperature of 300 K.* However, q_c is actually a function of the wall temperature that can be determined by:

$$q_c = \alpha\,(T_{aw} - T_w) \tag{10.145}$$

where q_c is the convective heat flux, α is the heat transfer coefficient, T_{aw} is the adiabatic wall temperature ($T_{aw} = 2500\,\text{K}$), and T_w is the wall temperature.

The wall thickness is relatively thin and so it can be assumed that the wall is in radiation equilibrium. The radiation heat flux can be determined by:

$$q_{rad} = \varepsilon\sigma T_w^4 \tag{10.146}$$

with ε is the emission coefficient and σ the Stefan-Boltzmann constant ($\sigma = 5.6704 \cdot 10^{-8}\frac{W}{m^2K^4}$). The maximum allowable wall temperature is 1300 K to assure structural integrity. To enhance the radiation cooling of the wall a coating with a high emission coefficient of $\varepsilon = 0.8$ is applied.

For the calculations you can make use of Table 10.38.

10.11 Explain how the temperature of the leading edge can be determined and determine the heat transfer coefficient, α.
10.12 Determine the leading edge wall temperature during flight. Is that temperature acceptable? Motivate your answer.
10.13 Determine the maximum allowable convective heat flux for the leading edge.
10.14 During flight the coating of the leading edge is damaged and is locally removed, exposing the bare structural material that has an emission coefficient of only 0.4. Calculate the resulting wall temperature. Is that threatening the structural integrity of the wing leading edge? Explain.

Table 10.38 Temperature and flux values

T_w (K)	q_c (kW/m^2)	q_{rad} (kW/m^2)
1200	118.18	94.07
1210	117.27	97.24
1220	116.36	100.49
1230	115.45	103.83
1240	114.55	107.25
1250	113.64	110.75
1260	112.73	114.34
1270	111.82	118.01
1280	110.91	121.77
1290	110.00	125.62
1300	109.09	129.56
1310	108.18	133.59
1320	107.27	137.72
1330	106.36	141.94
1340	105.45	146.26
1350	104.55	150.67
1360	103.64	155.19
1370	102.73	159.80
1380	101.82	164.52
1390	100.91	169.34
1400	100.00	174.27
1410	99.09	179.30
1420	98.18	184.44
1430	97.27	189.69
1440	96.36	195.05
1450	95.45	200.53
1460	94.55	206.12
1470	93.64	211.82
1480	92.73	217.65
1490	91.82	223.59
1500	90.91	229.65

10.15 Determine the maximum allowable convective heat flux for the leading edge with the removed surface coating.

TPS design with Mass Estimation Relationships (Problems 10.16 to 10.20).

In the conceptual design phase, several materials are considered for the TPS of an Earth entry capsule, with an entry velocity of 11 km/s. The design specification states that the front of the capsule will be exposed to a maximum flux of $q_c = 8$

MW/m^2 and total integrated heat load of $Q_{front} = 100$ MJ/m^2. The aftbody has to withstand $q_{c,\max} = 100$ kW/m^2 and $Q_{aft} = 3$ MJ/m^2. For this design phase, the capsule is approximated by a truncated cone with a base diameter of $d_{front} = 4.2$ m, an aft diameter of $d_{aft} = 1.2$ m, and a height of 1.4 m. To correct for the curved front-shield surface area, a margin of 40% is taken.

10.16 For the front heat shield, PICA ($\rho = 266$ kg/m^3) is considered. What is the required thickness of the heat shield, using the Mass Estimation Relationships from Sect. 10.5.2.

10.17 Assume that the front of the heat shield can be approximated by a flat, circular disk, with a diameter of 4 m: what is the estimated mass of the shield?

10.18 For the TPS of the aftbody, three options are available: SIRCA ($\rho = 300$ kg/m^3), Acusil II ($\rho = 250$ kg/m^3), and SLA-561V ($\rho = 224.3$ kg/m^3). What will be the required TPS thickness for each of these materials?

10.19 Calculate the mass for these three options, given that the curved surface area of the truncated cone is 32.4 m^2.

10.20 What is the total TPS mass? Is this realistic?

References

Agrawal P, Prabhu D, Milos FS, Stackpoole M (2016) Investigation of performance envelope for phenolic impregnated carbon ablator (PICA). AIAA-2016-1515, In: AIAA SciTech forum: 54th AIAA aerospace sciences meeting. San Diego, CA, 4–8 Jan 2016

Anderson Jr JD (2019) Hypersonic and high-temperature gas dynamics, 3rd edn. American Institute of Aeronautics and Astronautics

Ball AJ, Garry JRC, Lorenz RD, Kerzhanovich VV (2007) Planetary landers and entry probes. Cambridge University Press

Balat-Pichelin M, Robert JF, Sans JL (2006) Emissivity measurements on carbon-carbon composites at high temperature under high vacuum. Appl Surf Sci 253:778–783

Banks BA, Miller SKR, De Groh KK (2003) Atomic oxygen effects on spacecraft materials. In: Ninth International symposium on materials in a space environment. Noordwijk, The Netherlands, 16–20 June 2003. Also: NASA-TM-2003-212484, NASA

Beck RAS, Driver DM, Wright MJ, Laub B, Hwang HH, Slimko EM, Edquist KT, Sepka SA, Willcockson WH, Thames TD (2009) Development of the Mars Science Laboratory heatshield thermal protection system. AIAA-2009-4229, In: 41st AIAA thermophysics conference. San Antonio, TX, 22–25 June 2009

Beerman AF, Lewis MJ, Starkey RP, Cybyk BZ (2009) Significance of nonequilibrium surface interactions in Stardust return capsule ablation modeling. J Thermophys Heat Transfer 23(3):425–432

Bessire BK, Minton TK (2017) Decomposition of phenolic impregnated carbon ablator (PICA) as a function of temperature and heating rate. ACS Appl Mater Interfaces 9:21422–21437. American Chemical Society, 2017

Black WE (1977) Summary report of evaluation of coated columbium alloy heat shields for Space Shuttle thermal protection system application. NASA-CR-2824, NASA

Blosser ML (2000) Advanced metallic thermal protection systems for reusable launch vehicles. PhD thesis, Graduate school of engineering and applied science. University of Virginia, USA

Blosser ML, Chen RR, Schmidt IH, Dorsey JT, Poteet CC, Bird RK, Wurster KE (2004) Development of advanced metallic-thermal-protection system prototype hardware. J. Spacecr. Rocket. 41(2):183–194

Bolender MA, Doman DB (2006) Modeling unsteady heating effects on the structural dynamics of a hypersonic vehicle. AIAA-2006-2246, In: AIAA atmospheric flight mechanics conference and exhibit, Keystone, CO, 21–24 Aug 2006

Böhrk H, Piol O, Kuhn M (2010) Heat balance of a transpiration-cooled heat shield. J Thermophys Heat Transf 24(3):581–588

Bouilly J-M (2004) Thermal protection of the Huygens probe during titan entry: last questions. In: 2nd international planetary probe workshop. NASA Ames Research Center, Moffett Field, CA, 23–27 Aug 2004

Bouilly JM (2006) Thermal protection system of the Huygens probe during Titan entry: flight preparation and lessons learned. In: 5th European workshop on thermal protection systems and hot structures. In: Noordwijk, The Netherlands, 17–19 May 2006. Published as ESA-SP-631

Brewer RA (1975) Selection of a heat protection system for Venusian entry. AIAA-1975-0731, In: 10th thermophysics conference. Denver, CO, 27–29 May 1975

Brown WF Jr, Mindlin H, Ho CY (1994) Aerospace structural metals handbook. Purdue University, CINDAS/USAF CRDA handbooks operation

Buckner JL (2016) Investigation Of TI-6AL-4V alloy response to atmospheric re-entry exposure. PhD thesis, University of Texas at El Paso

Buursink J (2005) On the development of a cooled metallic thermal protection system for spacecraft. PhD thesis, Faculty of aerospace engineering, Delft University of Technology, The Netherlands

Caram JM, Bouslog SA, Cunnington Jr GR (1993) Emittance measurements of Space Shuttle Orbiter reinforced carbon-carbon. AIAA-1993-0841, In: 31st aerospace sciences meeting & exhibit, Reno, NV, 11–14 Jan 1993

Çengel YA, Ghajar AJ (2015) Heat and mass transfer fundamentals & applications, 5th edn. McGraw-Hill Education, New York, NY

Chen YK, Milos FS (2001) Thermal response modeling system for a Mars sample return vehicle. In: Thermal and fluids analysis workshop. NASA Ames Research Center

Clark RK (1973) An analysis of a charring ablator with thermal nonequilibrium, chemical kinetics, and mass transfer. NASA-TN-D-7180, In: NASA Langley Reserach Center

Cunnington GR, Funai AI, McNab TK (1983) Radiative properties of advanced spacecraft heat shield materials. NASA-CR-3740, NASA

Curry DM (1993) Space Shuttle Orbiter thermal protection system design and flight experience (also published as NASA-TM-104773). In: First ESA/ESTEC workshop on thermal protection systems. Noordwijk, The Netherlands, 5–7 May 1993

Daryabeigi K (1999) Analysis and testing of high temperature fibrous insulation for reusable launch vehicles. AIAA-1999-1044, In: 37th AIAA aerospace sciences meeting and exhibit. Reno, NV

Daryabeigi K (2000) Design of high temperature multi-layer insulation, PhD dissertation, University of Virginia

Daryabeigi K (1999) Effective thermal conductivity of high temperature insulations for reusable launch vehicles. NASA-TM-1999-208972, NASA

Daryabeigi K (2001) Thermal analysis and design of multi-layer insulation for re-entry aerodynamic heating. AIAA-2001-2834, In: 35th AIAA thermophysics conference, Anaheim, CA, 11–14 June 2001

Del Campo L, Pérez-Sáez RB, González-Fernández L, Esquisabel X, Fernández I, González-Martín P, Tello MJ (2010) Emissivity measurements on aeronautical alloys. J Alloy Compd 489:482–487

Delgado Montes I (1988) Study on aerocapture guidance and navigation, analysis of a charring ablator thermal protection system. GMVSA 2067/88, prepared under ESA Contract. Grupo de Mecanica del Vuelo, S.A

Devi GR, Rao KR (1993) Carbon-Carbon composites: an overview. Def Sci J 43(4):369–383

Dorsey JT, Poteet CC, Wurster KE, Chen RR (2004) Metallic thermal protection system requirements, environments, and integrated concepts. J Spacecr Rocket 41(2):162–172

Dotts RL, Battley HH, Hughes JT, Neuenschwander WE (1982) Space Shuttle Orbiter reusable surface insulation subsystem thermal performance. AIAA-1982-0005, In: AlAA 20th aerospace sciences meeting, Orlando, FL, 11–14 Jan 1982

Duffa G (2013) Ablative thermal protection systems modeling, AIAA education series

Dulak RE, Gecka AM (1970) Low cost ablative heat shields for Space Shuttles, NASA-CR-111814, NASA

Edquist KT, Dyakonov A, Wright MJ, Tang CY (2009) Aerothermodynamic design of the Mars Science Laboratory heatshield, AIAA-2009-4075. In: 41st AIAA thermophysics conference, San Antonio, TX, 22–25 June 2009

Ehrlich Jr CF et al (1992) Advanced manned launch system (AMLS) study—final report, NASA-CR-189673, NASA

Eidenoff HL, Rose L (1974) Thermal-structural evaluation of TD Ni-20Cr thermal protection system panels, NASA CR-132487, NASA

Erb RB, Greenshields DH, Chauvin LT, Pavlosky JE, Statham CL (1970) Apollo thermal-protection system development. J Spacecr Rocket 7(6):727–735

Ewing ME, Laker TS, Walker DT (2013) Numerical modeling of ablation heat transfer. J Thermophys Heat Transf 27(4):615–632

Ferraiuolo M, Manca OA (2011) New methodology to preliminary design structural components of re-entry and hypersonic vehicles. In: Lerner JC, Boldes U (eds) Wind tunnels and experimental fluid dynamics research, Chap. 19, pp 409–426, InTech

Finckenor MM, Dooling D (1999) Multilayer insulation material guidelines, NASA/TP-1999-209263, NASA Marshall Space Flight Center

Fischer WP, Gülhan A, Esser B, Faerber J, Reger N (2003) Plasma-windtunnel tests on a new metallic TPS material. In: 4th European workshop on TPS for space vehicles, Palermo (Nov 2002), ESA SP-521. ESA, Noordwijk

Fitzer E (1988) Composites for high temperatures. Pure Appl Chem 60(3):287–302

Forsberg CW, Burchell TD, Williams DF, Holcomb DE, Holdaway RF, Qualls AF (2005) Carbon-carbon-composite salt-cooled electric space reactors, Paper 1148. In: Space nuclear conference, San Diego, CA, 5–9 June 2005

Francese A (2007) Numerical and experimental study of UHTC materials for atmospheric re-entry. PhD thesis, Universita' degli studi di Napoli "Federico II"

Glass DE, Camarda CJ (1992) Preliminary thermal/structural analysis of a carbon-carbon/refractory-metal heat-pipe-cooled wing leading edge. In: Thornton EA (ed) Thermal structures and materials for high speed flight, vol 140. AIAA, Progress in Astronautics and Aeronautics, pp 301–322

Gnoffo PA, Weilmuenster KJ, Hamilton HH II, Olynick DR, Venkatapathy E (1999) Computational aerothermodynamic design issues for hypersonic vehicles. J Spacecr Rocket 36(1):21–43

Gonzales D, Eisman M, Shipbaugh C, Bonds T, Tuan Le A (1995) Proceedings of the RAND project AIR FORCE workshop on transatmospheric vehicles, 18–19 April 1995

Gorton MP, Shideler JL, Webb GL (1995) Static and aerothermal tests of a superalloy honeycomb prepackaged thermal protection system, NASA-TP-3257, NASA

Grensing FC, Hashiguchi D (1995) Mechanical and thermal properties of aluminum-beryllium alloy AM162. In: Advances in powder metallurgy & particulate materials—1995, proceedings of the 1995 international conference & exhibition on powder metallurgy & particulate materials

Groh KK, de Roig DM, Burke CA, Shah DR (1994) Performance and durability of high emittance heat receiver surfaces for solar dynamic power systems. In: American society of mechanical engineers (ASME) international solar energy conference. San Francisco, CA, 27–30 Mar 1994. Also: NASA-TM-106549, NASA

de Groh KK, Banks BA, McCarthy CE, Rucker RN, Roberts LM, Berger LA (2008) MISSE 2 PEACE polymers atomic oxygen erosion experiment on the international space station. High Perform Polym 20:388–409

Gogonin II (2023) Heat transfer in nucleate boiling. Cambridge Scholars Publishing

Groninger BV (1971) Titanium and superalloy heat shield design for Space Shuttle application. In: Proceedings of the 8th space congress—technology today and tomorrow, vol 1. Cocoa Beach, FL, pp 7–1 to 7–20, 19–23 April 1971
Gupta RN, Moss JN, Simmonds AL, Shim JL, Zoby EV (1984) Space Shuttle heating analysis with variation in angle of attack and catalycity. J Spacecr Rocket 21(2):217–219
Hargis Jr CB (1964) The X- 20 (Dyna- Soar) progress report, NASA-TM-X-861. In: Proceedings of the second manned space flight meeting, Dallas, TX, 22–24 April 1963. Published May 1964
Hays D (1982) An assessment of alternate thermal protection systems for the Space Shuttle Orbiter, vol I. Executive summary, NASA-CR-3548, NASA
Helenbrook RG, Anthony FM (1971) Design of a convective cooling system for a Mach 6 hypersonic transport airframe, NASA-CR-1918, NASA Langley Research Center
Helenbrook RG, McConarty WA, Anthony FM (1971) Evaluation of active cooling systems for a Mach 6 hypersonic transport airframe, NASA-CR-1917, NASA Langley Research Center
Heppenheimer TA (2007) Facing the heat barrier: a history of hypersonics, NASA SP-2007–4232, Government Printing Office, NASA
Hill AD, Anderson DM, Coughlin DJ, Chowdhry RS (1998) X-33 trajectory optimization and design, AIAA-1998-4408. In: 36th AIAA aerospace sciences meeting and exhibit, Reno, NV, 12–15 Jan 1998
Holland LR (1963) Physical properties of titanium. III. The specific heat. J Appl Phys 34(8):2350–2357
Hollis BR, Horvath TJ, Berry SA, Hamilton HH, Thompson RA, Alter SJ (2001) X-33 computational aeroheating predictions and comparisons with experimental data. J Spacecr Rocket 38(5):658–669
Howe JT (1985) Introductory aerothermodynamics of advanced space transportation systems. J Spacecr Rocket 22(1):19–26
Hurwicz H, Kratsch KM, Rogan JE (1972) Ablation, AGARDograph, no 161. North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development
Ifti HS, Hermann T, McGilvray M, Larrimbe L, Hedgecock R, Vandeperre L (2022) Flow characterization of porous ultra-high-temperature ceramics for transpiration cooling. AIAA J 60(5):3286–3297
Ifti HS, Hermann T, Rocher ME, Doherty L, Hambidge C, McGilvray M, Vandeperre L (2022) Laminar transpiration cooling experiments in hypersonic flow. Exp Fluids 63(102):2022
Inger GR, Baker RL (2005) Nonequilibrium viscous shock-layer heat transfer with arbitrary surface catalycity. J Spacecr Rocket 42(2):193–200
Jackson LR, Dixon SC (1980) A design assessment of multiwall, metallic standoff, and RSI reusable thermal protection systems including Space Shuttle application, NASA-TM-81780, NASA
Janna WS (2009) Engineering heat transfer, 3rd edn. CRC Press, Boca Raton, FL
Jenkins DR (2008) Space Shuttle: History Of The National Space Transportation System the first 100 missions, 3rd edn. Midland Publishing
Jenkins DR (2007) X-15 - extending the frontiers of flight, NASA
Johnson R Jr, Killpatrick DH (1976) Evaluation of dispersion strengthened nickel-base alloy heat shields for Space Shuttle application, NASA-CR-2614, NASA
Johnson CE, Tran HK, Smith M, Dill H (1997) Stardust backshell and back interface plate design verification tests in the NASA ames arc jet facilities, AIAA-1997-2483. In: 32nd thermophysics conference. Atlanta, GA, 23–25 June 1997
Johnson SM (2016) Coatings and surface treatments for reusable entry systems. In: Interagency coordinating committee on ceramics research and development (ICCCRD). Washington, D.C., 7 Mar 2016
Keldysh MV (1977) Venus exploration with the Venera 9 and Venera 10 spacecraft. Icarus 30:605–625
Keller BP, Nelson SE, Walton KL, Ghosh TK, Tompson RV, Loyalka SK (2015) Total hemispherical emissivity of Inconel 718. Nucl Eng Des 287:11–18

Keller K, Hoffmann M, Zörner W, Blumenberg J (1992) Application of high temperature multilayer insulations. Acta Astronaut 26(6):451–458
Kelly HN, Blosser ML (1994) Active cooling from the sixties to NASP, NASA-TM-109079, NASA Langley Research Center
Kim D, Jang C, Ryu WS (2009) Oxidation characteristics and oxide layer evolution of alloy 617 and Haynes 230 at 900 and 1100 °C. Oxid Met 71:271–293
Kontinos DA, Palmer G (1998) Numerical simulation of metallic TPS panel bowing, AIAA-1998-0866. In: 36th AIAA aerospace sciences meeting and exhibit, Reno, NV, 12–15 Jan 1998
Kumar S, Mahulikar SP (2016) Selection of materials and design of multilayer lightweight passive thermal protection system. J Therm Sci Eng Appl 8:021003-1–021003-9
Labaste V, Mignot Y (2003) EADS launch vehicles main achievements for planetary exploration missions. In: 4th european workshop on hot structures and thermal protection systems for space vehicles, Palermo, Italy, 26–29 Nov 2002. Published in ESA-SP-521
Linton RC, Vaughn JA, Finckenor MM, Kamenetzky RR, DeHaye RF, Whitaker AF (1993) Orbital atomic oxygen effects on materials: an overview of MSFC experiments on the STS-46 EOIM-3, AIAA-1993-4102. In: AIAA space programs and technologies conference and exhibit. Huntsville, AL, 21–23 Sept 1993
Liu S, Zhang B (2011) Effects of active cooling on the metal thermal protection system. Aerosp Sci Technol 15:526–533
Mahzari M, Beck RAS, Hwang H, Monk J, Morgan J, Williams J, Edquist KT (2022) Development and sizing of the Mars 2020 thermal protection system, AIAA-2022-3951. In: AIAA AVIATION forum. Chicago, IL & Virtual, June 27–July 1, 2022
Mahzari M, Braun RD, White TR, Bose D (2015) Inverse estimation of the Mars Science Laboratory entry aeroheating and heatshield response. J Spacecr Rocket 52(4):1203–1216
Marquardt P, Fertig M, Reimer T, Gülhan A (2022) EURASTROS capsule design and re-entry analysis, Paper 7249. In: 9th European conference for aeronautics and space sciences (EUCASS). Lille, France, June 27–July 1, 2022
Martinelli SK, Braun RD (2011) Centerline heating methodology for use in preliminary design studies, IEEEAC-paper-1322. In: IEEE aerospace conference. Big Sky, MT
McConarty WA, Anthony FM (1971) Design and evaluation of active cooling systems for Mach 6 cruise vehicle wings, NASA-CR-1916, NASA Langley Research Center
McDonnel D (1972) Study of structural active cooling and heat sink systems for Space Shuttle. NASA-CR-123912 (also Final Report MDE E0638), McDonnel Douglas, 30 June 1972
Meltzer M (2007) Mission to Jupiter: a history of the Galileo project, NASA-SP-2007-4231, NASA
Merlin PW (2009) Design and development of the Blackbird: challenges and lessons learned, AIAA-2009-1522. In: 47th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition. Orlando, FL, 5–8 Jan 2009
Miller J (2001) The X-planes—X-1 to X-45, 3rd edn. Midland Publishing
Milos FS (1997) Galileo probe heatshield ablation experiment. J Spacecr Rocket 34(6):705–713
Milos FS, Chen Y-K, Squire TH (2006) Updated ablation and thermal response program for spacecraft heatshield analysis, Paper TFAWS06-1008. In: 17th thermal and fluids analysis workshop. University of Maryland, 7–11 Aug 2006
Milos FS, Chen Y-K, Squire TH, Brewer RA (1999) Analysis of Galileo probe heatshield ablation and temperature data. J Spacecr Rocket 36(3):298–306
Mooij E, Vo QT, Offerman JWPI (2006) Technology status and developments of metallic hot structures and thermal protection systems. In: 5th European workshop on thermal protection systems and hot structures. ESA/ESTEC, Noordwijk, The Netherlands, 17–19 May 2006
Mooij E, Vo QT, Boers T (2006) Metallic hot structures and thermal protection systems: technology status and future developments. In: The 1st international ara days, atmospheric re-entry systems missions and vehicles. Arcachon, France, 3–5 July 2006
Muylaert J, Walpot LMGFM, Ottens H, Cipollini F, Tumino G, Marraffa L, Caporicci M, Stavrinidis C (2003) EXPERT: European experimental re-entry testbed. The ESA In-flight ATD research

programme. In: 3rd international symposium on atmospheric re-entry vehicles and systems. Arcachon, France
Myers DE, Martin CJ, Blosser ML (2000) Parametric weight comparison of advanced metallic, ceramic tile, and ceramic blanket thermal protection systems, NASA-TM-2000-210289, NASA Langley Research Center
Niblock GA, Miller CC, Holmgren JS (1972) Study of structural active cooling and heat sink systems for Space Shuttle, NASA-CR-123912, NASA
Olynick D (1998) Trajectory-based thermal protection system sizing for an X-33 winged vehicle concept. J Spacecr Rocket 35(3):249–257
Olynick D, Ghent YK, Tauber ME (1997) Forebody TPS sizing with radiation and ablation for the Stardust sample return capsule, AIAA-1997-2474. In: 32nd AIAA thermophysics conference. Atlanta, GA, June 23–25 1997
Palmer GE, Kontinos D, Sherman B (1999) Surface heating effects of X-33 vehicle thermal-protection-system panel bowing. J Spacecr Rocket 36(6):836–841
Parmenter KE, Shuman K, Milstein F, Szalai CE, Tran HK, Rasky DJ (2001) Compressive response of lightweight ceramic ablators: phenolic impregnated carbon ablator. J Spacecr Rocket 38(2):231
Paglia L, Genova V, Tirillò V, Bartuli C, Simone A, Pulci G, Marra F (2021) Design of new carbon-phenolic ablators: manufacturing, plasma wind tunnel tests and finite element model rebuilding. Appl Compos Mater 28:1675–1695
Pavlovsky JE, StLeger LG (1974) Apollo experience report - thermal protection subsystem. NASA-TN-D-7564. Lyndon B. Johnson Space Center, Houston, TX
Peterson DL, Nicolet WE (1974) Heat shielding for Venus entry probes. J Spacecr Rocket 11(6):382–387
Pittman CM, Brinkley KL (1976) One-dimensional numerical analysis of the transient thermal response of multilayer insulative systems, NASA-TM-X-3370, NASA Langley Research Center
Pitts WC, Wakefield RM (1980) Performance of entry heat shields on Pioneer Venus probes. J Geophys Res 85(AI3):8333–8337, 30 Dec 1980
Prabhu DK, Loomis MP, Venkatapathy E, Polsky S, Papadopoulos P, Davies CB, Henline WD (1998) X-33 aerothermal environment simulations and aerothermodynamic design, AIAA-1998-0868. In: 36th AIAA aerospace sciences meeting and exhibit. Reno, NV, 12–15 Jan 1998
Prabhu DK, Venkatapathy E, Kontinos D, Papadopoulos P (1998) X-33 catalytic heating, AIAA-1998-2844. In: 7th AIAA/ASME joint thermophysics and heat transfer conference. Albuquerque, NM, 15–18 June 1998
Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes: the art of scientific computing. 3rd edn. Cambridge University Press
Pulci G, Tirillò J, Marra F, Fossati F, Bartuli C, Valente T (2010) Carbon-phenolic ablative materials for re-entry space vehicles: manufacturing and properties. Composites: Part A 41:1483–1490
Reay D, Kew P (2006) Heat pipes–theory: design and applications, 5th edn. Butterworth-Heinemann, Oxford, UK
Reich G, Jahn G (1996) MIRKA heat shield experiment HEATIN-theoretical and experimental approach to the surface protected ablator. SAE Trans Section 1: J Aero 105:481–488
Ridolfi G, Mooij E, Dirkx D, Corpino S (2012) Robust multi-disciplinary optimization of unmanned entry capsules, AIAA-2012-6575. In: AIAA modelling and simulation technologies conference. Minneapolis, Minnesota, 13–16 Aug 2012
Rindal RA (1968) An analysis of the coupled chemically reacting boundary layer and charring ablator. In: Part VI—an approach for characterizing charring ablator response with in-depth coking reactions, NASA-CR-1065, NASA
Robinson JS, Wurster KE, Mills JC (2009) Entry trajectory and aeroheating environment definition for capsule-shaped vehicles. J Spacecr Rocket 46(1):74–86
Rodriguez AC, Snapp CG (2011) Orbiter thermal protection system lessons learned, AIAA-2011-7308. In: AIAA SPACE conference & exposition. Long Beach, CA, 27–29 Sept 2011
Rohsenow WR, Hartnett JP, Cho YI (1998) Handbook of heat transfer, 3rd edn. McGraw-Hill
Rotelli R (1965) X-20 design and development. Ann N Y Acad Sci 134(1):80–92

Sahoo N, Kulkarni V, Saravanan S, Jagadeesh G, Reddy K (2005) Film cooling effectiveness on a large angle blunt cone flying at hypersonic speed. Phys Fluids

Savino R, De Stefano Fumo M, Paterna D, Serpico M (2005) Aerothermodynamic study of UHTC-based thermal protection systems. Aerosp Sci Technol 9:151–160

Sawant SS, Harpale A, Jambunathan R, Chew HB, Levin DA (2017) High fidelity and multi-scale thermal response modelling of an avcoat-like thermal protection material, AIAA-2017-0438. In: 55th AIAA aerospace sciences meeting. Grapvine, TX, 9–13 Jan 2017

Schwanekamp T, Meyer F, Reimer T, Petkov I, Tröltzsch A, Siggel M (2014) System studies on active thermal protection of a hypersonic suborbital passenger transport vehicle, AIAA-2014-2372. In: 19th AIAA international space planes and hypersonic systems and technologies conference. Atlanta, GA, 16–20 June 2014

Scoville CL, Gorsuch PD (1970) Thermal protection systems for the Space Shuttle. In: Astronautical research 1970, proceedings of the 21st congress of the international astronautical federation. Konstanz, West Germany, pp 424–443, 4–10 Oct 1970

Seifert RW (1973) Ablative heat shield design for Space Shuttle, NASA-CR-132282, NASA

Sepka SA, Samareh JA (2015) Thermal protection system mass estimating relationships for blunt-body, Earth entry spacecraft, AIAA-2015-2507. In: AIAA AVIATION forum: 45th AIAA thermophysics conference. Dallas, TX, 22–26 June 2015

Sepka SA, Wright MJ (2009) FIAT uncertainties using a Monte-Carlo approach: improvements and results for MSL, AIAA-2009-4234. In: 41st AIAA thermophysics conference. San Antonio, TX, 22–25 June 2009

Shukla KN (2015) Heat pipe for aerospace applications–an overview. J Electron Cool Therm Control 5:1–14

Siegel R, Howell JR (1992) Thermal radiation heat transfer, 3rd edn. Taylor & Francis Ltd, London

Simmons DJ, Meritt RJ (2022) Hypersonic instrumentation testing for thermal protection systems, AIAA-2022-3582. In: AIAA AVIATION forum, Chicago, IL & Virtual, June 27–July 1 2022

Sloof WG (2007) Self healing in coatings at high temperatures. In: van der Zwaag S (ed) Self healing materials. Springer series in materials science, vol 100. Springer, Dordrecht, pp 309–321

Stackpoole M, Sepka S, Cozmuta I, Kontinos D (2008) Post-flight evaluation of Stardust sample return capsule forebody heatshield material, AIAA-2008-1202. In: 46th AIAA aerospace sciences meeting and exhibit. Reno, NV, 7–10 Jan 2008

Stewart M, Koenig WJ, Harris RF (2018) Thermal protection systems technology transfer from Apollo and Space Shuttle to the Orion program, AIAA-2018-5134. In: AIAA SPACE and astronautics forum and exposition. Orlando, FL, 17–19 Sept 2018

Strauss EL (1967) Superlight ablative systems for Mars lander thermal protection. J Spacecr Rocket 4(10):1304–1309

Sudmeijer KJ, Buursink J, Lopes C (2002) Enhanced radiation cooling system for metallic TPS. In: 4th European workshop on hot structures and thermal protection systems for space vehicles. Palermo, Italy, 26–29 Nov 2002. Published in ESA SP-521, pp 259–266

Sudmeijer KJ, Ambaum MHA, Thierfelder W (2001) Development of an ODS sandwich structure for the stiffened load carrying skin of the rudder of a reusable launcher. In: 2nd international symposium on atmospheric reentry vehicles and systems, Arcachon

Sutheesh PM, Chollackal A (2018) Thermal performance of multilayer insulation: a review, Paper 012061. In: IOP conference series: materials science and engineering, vol 396

Sutton K (1970) An experimental study of a carbon-phenolic ablation material, NASA-TN-D-5930, NASA Langley Research Center

Swann RT, Pittmann CM (1962) Numerical analysis of the transient response of advanced thermal protection systems, NASA-TN-D-1370, NASA Langley Research Center

Swann RT, Pittmann CM, Smith JC (1965) One-dimensional numerical analysis of the transient response of thermal protection systems, NASA-TN-D-2976, NASA Langley Research Center

Swenson Jr LS, Grimwood JM, Alexandder CC (1989) This new ocean: a history of project Mercury, NASA-SP-4201, NASA History Series

Szalai C, Slimko E, Hoffman P (2014) Mars science laboratory heatshield development, implementation, and lessons learned. J Spacecr Rocket 51(4):1167–1173

Thornton EA (1996) Thermal structures for aerospace applications, AIAA education series

Throckmorton DA (1982) Benchmark aerodynamic heat transfer data from the first flight of the Space Shuttle Orbiter, AIAA-1982-0003. In: AlAA 20th Aerospace sciences meeting. Orlando, FL, 11–14 Jan 1982

Tran HK, Johnson CE, Rasky DJ, Hui FCL, Hsu M, Chen T, Chen YK, Paragas D, Kobayashi L (1996) Phenolic impregnated carbon ablators (PICA) as thermal protection systems for discovery missions, AIAA-1996-1911. In: 31st thermophysics conference. New Orleans, LA, 17–20 June 1996

Trumble KA, Cozmuta I, Sepka S, Jenniskens P, Winter M (2010) Postflight aerothermal analysis of Stardust sample return capsule. J Spacecr Rocket 47(5):765–774

Van Baten T, Buursink J (2002) Spaceship with heat-isolating outer skin, European patent office, patent WO 02066326 (A1)

Varisco A, Bell P, Wolter W (1978) Design and fabrication of metallic thermal protection systems for aerospace vehicles, NASA CR-145313, NASA

Wijker JJ (2008) Spacecraft structures. Springer, Berlin Heidelberg

Willcockson WH (1999) Mars Pathfinder entry heatshield design and flight experience. J Spacecr Rocket 36(3):374–379

Willcockson WH (1999) Stardust sample return capsule design experience. J Spacecr Rocket 36(3):470–474

Williams SD, Curry DM (1992) Thermal protection materials: thermophysical property data, NASA-RP-1289, NASA

Wright MJ, Beck RAS, Edquist KT, Driver D, Sepka SA, Slimko EM, Willcockson WH (2014) Sizing and margins assessment of Mars Science Laboratory aeroshell thermal protection system. J Spacecr Rocket 51(4):1125–1138

Zhang Q, Li H (2007) MOEA/D: a multiobjective evolutionary algorithm based on decomposition. IEEE Trans Evol Comput 11(6):712–731

Chapter 11
Aero-assisted Manoeuvres



In Chap. 5 the different entry mechanisms were discussed, i.e., ballistic, gliding, and skipping flight. Each mission type had in common that they led to a landing on the target body, starting from a (closed) orbit around this body or reaching the atmospheric interface on an inbound trajectory, e.g., the Apollo returning from the Moon. Such a trajectory is a typical example of an *aerocapture* trajectory, or, to be more specific, a *direct-entry* aerocapture. The direct entry of the Apollo capsule was followed by a semi-ballistic flight and concluded with a sea landing in the Pacific Ocean. During this single-pass trajectory, all kinetic energy had to be converted to heat before the capsule's speed had reduced to an acceptable level for parachute deployment. One can imagine that this resulted in a very high thermal loading that could only be withstood by an ablative thermal-protection shield. An alternative to this rather violent entry is, of course, the skipping flight, executed by several of the Zond spacecraft, and as a backup also present in the Apollo guidance system. The skipping in and out of the atmosphere reduces the velocity in such a way that the overall thermo-mechanical loading on the vehicle reduces significantly.

The aerocapture does not necessarily have to be executed as a direct-entry flight. Coming from an interplanetary approach trajectory, one (or more) pass(es) through the atmosphere can reduce the orbital energy of the vehicle to such an extent that, when leaving the atmosphere for the last time, a closed, elliptical orbit is obtained. Using the atmosphere in this way will reduce the need for a propulsion system and significant mass savings can be achieved.

Aerocapture is not the only type of manoeuvre that can be used to improve mission performance. A more gradual approach to reduce the orbital energy is the so-called *aerobraking*, where the spacecraft will only pass through the upper layers of the atmosphere and thus many passes are required. On the other side of the spectrum is the *aerogravity assist*, where the vehicle will soar through the atmosphere with very high velocity to change its hyperbolic orbit. This manoeuvre is so critical, though, that it has not yet been executed in real life, whereas the other two have been at multiple occasions.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1_11

This chapter will focus on these three mentioned aero-assisted manoeuvres: aerobraking will be discussed in Sect. 11.1, aerocapture in Sect. 11.2, and the aerogravity assist in Sect. 11.3. Due to space limitations, aeromanoeuvring executed to change (arbitrary) orbital elements, such as the inclination will not be covered.

11.1 Aerobraking

11.1.1 Concept and Missions

It is a known fact that satellites orbiting a body with an atmosphere at low altitudes will feel the effect of atmospheric drag, and continuously lose energy. Even though the decelerations may be minute, this will invariably lead to orbital decay. For space debris this is an effect that will ultimately result in atmospheric entry and the removal of debris from orbit, but for satellites that have strict orbital requirements this means that corrective manoeuvres have to be performed to compensate for the energy loss.

It was with the launch of Explorer 51, also called AE-C (Atmospheric Explorer C), on 16 December, 1973, that flying through the upper layers of the atmosphere was used to actively collect measurements. In its first year of operations, the perigee of AE-C's initial elliptical orbit (149 $\times$ 4,294 km) was changed many times and lowered to 129 km.[1] It thus became the first satellite to use the concept of aerobraking to its advantage, i.e., to do in situ atmospheric studies not possible from a typical spacecraft orbit, and it was followed by several of its successors (Marcos et al. 1977).

This feat was repeated by the Pioneer Venus Orbiter at Venus (Strangeway 1993; Siddiqi 2018). Launched on 20 May 1978, the orbiter was designed to spend an extended period in orbit around Venus mapping the surface using a radar package. Towards the end of its mission, in May 1992, it lowered its periapsis and used its thrusters to keep it between 150 and 250 km until propellant depletion. It used its accelerometers for drag-measurement experiments to extend the knowledge about the upper layers of Venus' atmosphere.

Whereas technically the above two missions are considered to have performed aerobraking, actively using aerobraking as a mission-operations concept came with subsequent missions. Before those are discussed, let us consider for a moment why aerobraking would actually be a useful concept. When spacecraft are sent to other planets they will reach their target with a hyperbolic velocity that is in general too large to be captured in a closed orbit around the planet. Thus, commonly a propulsive "braking" or orbit-insertion manoeuvre is executed. However, depending on the incoming velocity, the mass of the spacecraft and the orbital parameters of the target orbit, this may cost a large amount of fuel. In case the target planet has an atmosphere, it could be used to safely absorb some of the spacecraft's kinetic energy. When doing this repeatedly, the need to change orbit with a propulsive ΔV reduces. In particular

[1] https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=1973-101A, last accessed February 15, 2020.

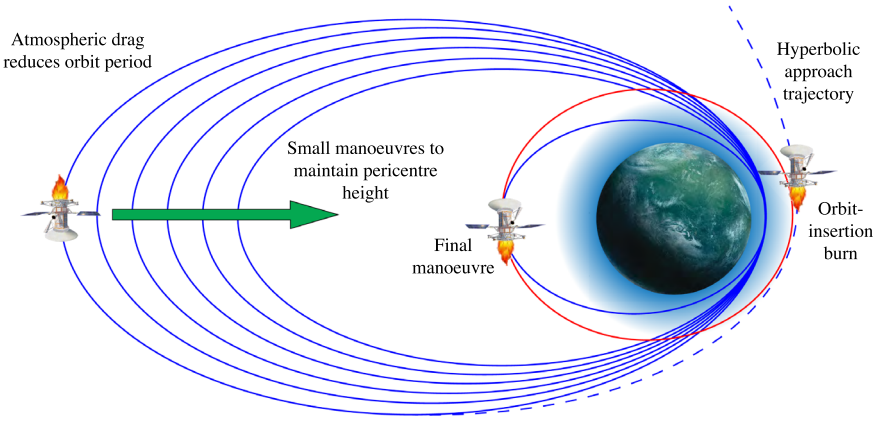


Fig. 11.1 The aerobraking manoeuvre

for large-mass vehicles this can be a large savings. Of course, this will come at the expense of a longer mission, larger support staff to plan all periapsis manoeuvres (where typically the aerobraking takes place), as well as more *Deep Space Network* coverage for intense communication during the planning and the manoeuvres.

Figure 11.1 shows the mission concept of aerobraking. Consider a spacecraft approaching a planet with a sufficiently thick atmosphere. Having reached the potential periapsis of the closed, elliptical orbit around the planet, an orbit-insertion burn is given. This results in a highly elliptical orbit, which requires a second burn at the apocentre to lower its periapsis to the target altitude inside the atmosphere. During each pass at the periapsis, atmospheric drag causes the apoapsis altitude to be lowered and trim the orbit into a more circular flight path. However, also the periapsis altitude will be lowered, albeit small, and it is thus necessary that at apoapsis a small, corrective burn will fire to avoid lowering the periapsis too deep into the atmosphere. A too low altitude could obviously jeopardise the spacecraft due to excess thermo-mechanical loads. Once the target apoapsis-altitude is reached, a final burn will raise the periapsis out of the atmosphere and the actual mission can begin.

In principle, three distinct phases can be discerned: walk-in, main phase, and walk-out, which is, in fact, the general approach to aerobraking. During the walk-in, the periapsis is (slowly) lowered into the atmosphere taking density uncertainties into account, the apoapsis lowering is achieved during the main phase, whereas during the walk-out the periapsis altitude is increased again to the altitude of the final orbit.

It may be clear that although aerobraking is very challenging, the effort is worth the risk, because it eliminates the need for the heavy load of propellant that would otherwise be needed to place the spacecraft in its desired orbit (Smith and Bell 2005). For example, the additional mass needed to deliver the Mars Odyssey spacecraft to its final orbit would have increased the spacecraft mass beyond the capability of the low-cost launch vehicle chosen for that mission. It would have required a larger, and potentially much more expensive, rocket. Besides atmospheric uncertainties, a major

reason that aerobraking is difficult is that towards the end it "speeds up", because the orbital period gets shorter and shorter.

With the main concept clear, some aerobraking missions will now be briefly discussed. The Hiten spacecraft was Japan's first lunar mission, and launched on 24 January, 1990 (Uesugi 1996). It entered a highly elliptical orbit around Earth, but its apogee of "only" 290,000 km was significantly short of the 476,000 km target. Using Weak Stability Boundary Theory, developed by Belbruno and Miller of JPL (Belbruno 2004), put the spacecraft on a low-energy ballistic transfer and into the required elliptical swing-by orbit. The small propulsive "nudge" required to initiate this manoeuvre was sufficiently small to be achievable by the spacecraft's propulsion system. After the eighth swing-by, on March 19, 1991, Hiten became the first spacecraft to successfully demonstrate aerobraking as a manoeuvre to lower the apoapsis of its orbit. It flew by the Earth at a perigee altitude of 125.5 km at 11.0 km/s, the encountered atmospheric drag lowered the velocity by 1.712 m/s, and the apoapsis altitude by 8,665 km (see also the summary Table 11.1). A second manoeuvre followed on March 30, 1991, after the ninth lunar flyby.

The next extra-terrestrial aerobraking manoeuvre was performed by the Magellan spacecraft on Venus (Croom and Tolson 1994; Lyons et al. 1995), see Fig. 11.2. Launched from the Space Shuttle payload bay by inertial upper stage on May 4, 1989, Magellan was tasked to map the surface and to measure the planetary gravitational field. The spacecraft was inserted in a highly-elliptical orbit around Venus, but to increase both coverage and resolution of the mapping the orbit needed to be circularised. Its thruster system was not designed for major thrust usage, so the only remaining option was to use aerobraking to lower its apoapsis by over 8,000 km in 700 passes. The manoeuvre was initiated in late May 1993 and lasted to early August, and the related data, taken from several references (Lyons 1999; Spencer and Tolson 2007; Tolson et al. 2013), are provided in Table 11.1. For aerobraking, the spacecraft's configuration was such that the maximum solar-panel area was exposed to the oncoming flow. A fortunate characteristic of this configuration was that the centre of pressure was located aft of the centre of mass, which enhanced its longitudinal stability.

Table 11.1 Analysis of several aerobreaking manoeuvres (references listed in the text)

Mission	Body	R_p [km]	Passes [#]	ΔV [m/s]	Δr_a [km]	Fuel saved [kg (%Dry)]
Hiten #1	Earth	125.5	1	1.712	8,665	–
Hiten #2	Earth	120	1	2.8	14,000	–
Magellan	Venus	170	700	1,200	8,000	490 (≈50%)
MGS	Mars	120	446	1,200	54,000	330 (≈50%)
MO	Mars	110	332	1,080	26,000	320 (≈85%)
MRO	Mars	110	420	1,200	43,500	580 (≈60%)
MCO	Mars	57	0.5	–	–	–

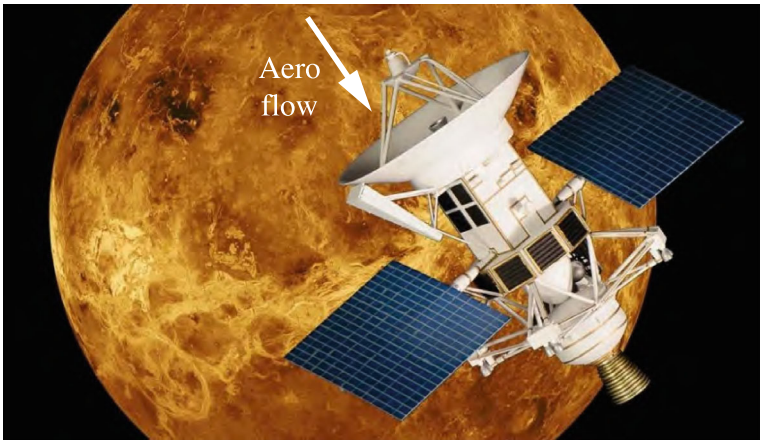


Fig. 11.2 Magellan spacecraft; during aerobraking, the solar array was oriented perpendicular to the airflow. Adapted from original image: NASA/JPL

Magellan's aerobraking manoeuvre was divided into the three standard phases. The walk-in phase was executed to slowly lower Magellan's periapsis from the initial 172 to 150 km, because accurate density models were not available. For the main phase, a periapsis dynamic-pressure corridor was established. The limit values for this corridor were set to 0.27 and 0.32 N/m^2, respectively, but due to the lower than expected density fluctuations and solar-array heating, it was later increased to 0.29 and 0.35 N/m^2. Primary drivers for setting the limits on maximum dynamic pressure were the maximum temperatures for solar-array joints, antenna delamination, and limits on the attitude control system (Spencer and Tolson 2007). The lower bound was limited by the constraint of 100 days for the total aerobraking time, as a too low dynamic pressure would not give sufficient deceleration and corresponding apoapsis decrease. The main goal of the eight-day walk-out was to raise the periapsis outside of the noticeable atmosphere by means of propulsive manoeuvres.

In general, the main challenges faced by the Magellan team were the uncertainties in atmospheric and gravitational conditions at Venus, and lack of automation in the orbit determination and guidance system. Still, the benefits of aerobraking were so obvious that the experience with Magellan paved the way for using this manoeuvre in future Mars missions as well. Magellan's end-of-life mission was the so-called windmill experiment. The termination sequence was initiated late August 1994, and consisted of a series of orbital trim manoeuvres. These lowered the spacecraft into the outermost layers of the Venusian atmosphere such that the experiment could begin on September 6. The induced rotations due to aerodynamic torques were countered by the spacecraft's attitude thrusters, and from this activity useful information about the interaction with the atmosphere could be derived. The mission officially ended on 13 October, 1994.

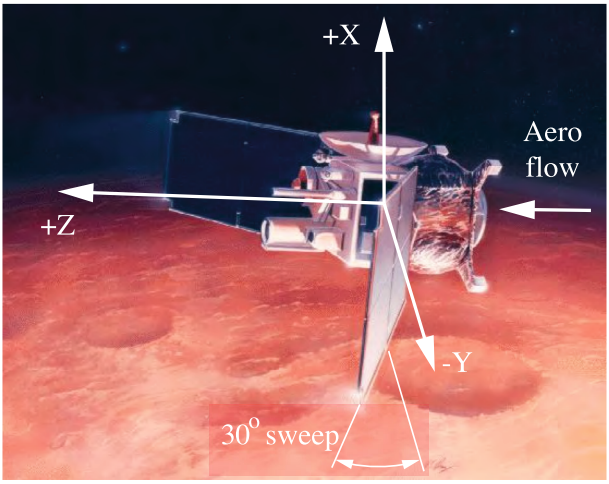


Fig. 11.3 Mars Global Surveyor Spacecraft in aerobraking configuration. Adapted from original image: NASA/JPL

The first Mars mission that was designed with aerobraking in mind before it was launched on November 7, 1996, was Mars Global Surveyor (MGS), which mission was to map the entire Martian surface. Due to the low-density atmosphere surrounding Mars,[2] MGS (and every subsequent aerobraking mission to Mars) used its solar panels to increase its surface area perpendicular to the velocity to increase its drag area. Figure 11.3 shows MGS in this "aerobraking configuration". The solar panels were swept back to reduce the mechanical loading during the aerobraking manoeuvre (Lyons 1999).

However, due to an issue with solar-array deployment right after launch, the solar array was tilted 20° with respect to the nominal position. It did not seem to affect the panels during the walk-in phase, as no major deflection of the solar panels was noticed. After an atmospheric sweep with a much higher than expected density, though, the solar panels experienced substantial damage. The MSG team halted the aerobraking, such that they could assess the damage and choose a new strategy. To avoid further damage, aerobraking was resumed after one month with the maximum allowed dynamic pressure reduced by two thirds. As a result, the manoeuvre took one year more than planned. Details of the manoeuvre are taken from Lyons (1999), and Spencer and Tolson (2007), and have been included in Table 11.1.

The first failure in an aerobraking manoeuvre occurred during the Mars Climate Orbiter (MCO) mission. This mission was launched in 1998 to investigate the Martian atmosphere. As before with MGS, an aerobraking manoeuvre was proposed to be executed after orbit insertion about Mars to circularise the orbit. This orbit insertion

[2] Even though the (upper) atmosphere of Mars is thin, one the challenges of aerobraking is to account for the density differences on either day or night side, which makes operations difficult and intense for the mission operators.

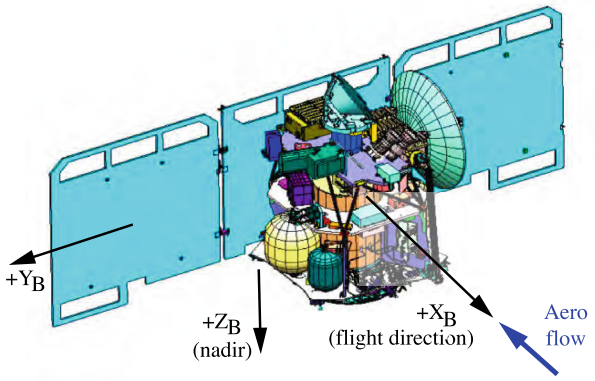


Fig. 11.4 Mars Odyssey Spacecraft in aerobraking configuration (incoming flow in negative x_B direction). Adapted from original image: NASA/JPL

is where the mission failed, due to a mix-up of imperial and metric units when issuing commands.[3] The insertion manoeuvre placed the vehicle in an orbit with a periapsis of 57 km, 170 km below the planned 226 km, and 23 km below the minimum altitude considered survivable by MCO of 80 km. At this altitude the atmospheric density was too high, and most likely resulted in the disintegration of the spacecraft in the Martian atmosphere, even before the actual aerobraking could begin.

The next Mars mission to successfully use aerobraking was Mars Odyssey (MO), launched on 7 April, 2001, to investigate the planet's geology and radiation environment, as well as to search for evidence of past or present water or ice. After the interplanetary transfer of well over six months, MO was inserted into Mars orbit on 24 October 2001, with an orbital period of 18.6 hours. The aerobraking manoeuvre had to reduce this period to two hours within a 90-day window to assure the proper local solar time in the final science orbit (Smith and Bell 2005). Figure 11.4 shows the aerobraking configuration of the spacecraft. To increase the reference area to 11 m^2, the solar array was placed perpendicular to the flow.

Spencer and Tolson (2007) and Tolson et al. (2013) provide details on the manoeuvre that are given in Table 11.1. For this mission, the aerobraking was split into four phases: walk-in, main, endgame, and walk-out (Smith and Bell 2005). During the walk-in the periapsis altitude was gradually lowered from 292 to 111 km. During this process the on-board models were calibrated and the spacecraft performance evaluated. During the main phase the heating of the solar array constrained the amount of braking that could be done. As the atmospheric conditions were unpredictable, the margins on the reference thermal constraints resulted in much lower actual solar-panel limits. Whereas the peak heat rate was used as a limit in the main phase, during the endgame phase this was changed to cumulative heating. The walk-out phase began the moment that the predicted time to reach an apoapsis altitude

[3] Mars Climate Orbiter Mishap Investigation Board, "Phase I Report", November 1999. Available at https://llis.nasa.gov/llis_lib/pdf/1009464main1_0641-mr.pdf.

of 300 km was estimated to be less than one day. During the walk-out phase, the periapsis altitude is gradually increased outside the atmosphere, because below 300 km it becomes almost inevitable that the spacecraft spirals inwards toward the planet.

The details of the manoeuvre were obtained from Tolson et al. (2013), and Spencer and Tolson (2007), and are presented in Table 11.1. Compared with the other Mars missions, it was the most successful one in terms of mass savings relative to the spacecraft's dry weight: a staggering 85%!

The Mars Reconnaissance Orbiter (MRO), launched in 2005, was sent to Mars to map its surface in a higher resolution than was done by MGS so that landing sites for future surface missions could be determined. During the MRO mission, aerobraking manoeuvres were used for three main reasons: increase or decrease atmospheric drag (i.e., increasing or decreasing periapsis altitude), collision avoidance with other Mars-orbiting spacecraft, and for lifetime maintenance. The latter means that the time needed to reach a 300-km apoapsis altitude had to be greater than a pre-specified number of days (two, in this case).

The aerobraking was divided into four phases: walk-in, main, walk-out and transition. In total, a propulsive ΔV of 44 m/s was used for aerobraking, spread over corridor control and collision avoidance. During the walk-in the periapsis altitude was gradually lowered into the Martian atmosphere, and the main phase was used to control the periapsis altitude while staying within the limits of the atmospheric corridor. During walk-out, aerobraking manoeuvres were mainly carried out to avoid collisions and maintain lifetime above the two-day limit. Finally, during the so-called transition phase that began at an apoapsis altitude of 485 km, MRO was placed in its target scientific orbit.

Figure 11.5 shows MRO in aerobraking configuration, i.e., MRO's attitude was such that its solar panels were aligned roughly perpendicular to the flow (with the backside facing it), with a total surface area of 37.5 m^2. For the whole aerobraking phase, the configuration was kept constant to avoid entering the atmosphere with the wrong orientation.

Before we move on to a more in-depth analysis of aerobraking, two more missions are worth mentioning. These are the ESA missions Venus Express and the ExoMars Trace Gas Orbiter (Fig. 11.6). To begin with the former, Venus Express, launched

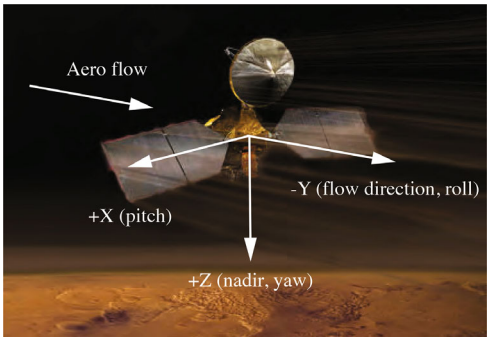
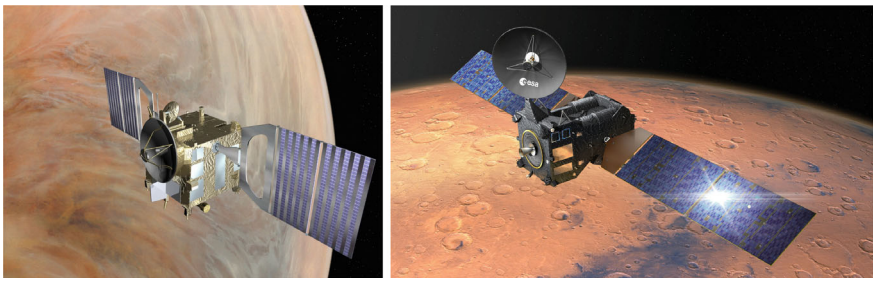


Fig. 11.5 Mars Reconnaissance Orbiter in aerobraking configuration. Image adapted from original: NASA/JPL



(a) Venus Expres (b) ExoMars Trace Gas Orbiter

Fig. 11.6 ESA missions to perform aerobraking manoeuvres. Images: ESA

in November 2005, was the first Venus exploration mission of ESA. The spacecraft arrived at Venus in April 2006, after having executed several propulsive manoeuvres. It began its science mission in a polar orbit, and continuously sent back its observational data. The nominal mission ended after two years, but due to its success the mission was extended five times, the last time on 20 June, 2013, until mission end in 2015. It was during this final extension that an end-of-life aerobraking experiment was planned. Initiated in May 2014, it successfully performed multiple passes at 131 to 135 km altitude until July 2014.

Val Serra et al. (2011) discuss in detail the planned aerobraking experiment. In the particular case of Venus Express, the manoeuvre was divided into five phases: initiation, walk-in, main, walk-out, and termination. During initiation, the altitude was slowly lowered, and during walk-in, the atmosphere variabilities and induced heating effects were examined, to compare the results with the theoretical predictions. Furthermore, the experienced aerodynamic torques would be used to calibrate the centre-of-pressure and centre-of-mass alignment, as to guarantee aerodynamic stability and the (relaxed) pointing accuracy. During walk-out, it was decided to use a minimum lifetime as driving constraint as well. Finally, with termination, the periapsis altitude was raised above the atmosphere. Orbit determination was performed with the ESA deep-space antennae, using two-way Doppler (Damiani et al. 2015). The atmospheric model used was an exponential model. After each atmospheric passage, the model was adjusted for the observed discrepancies, and a new scale height was used for the successive sweep. In particular, high variability of the scale height was noticed at high altitudes, due to the small effects (and thus sensitivity) of density. This analysis was also used to check that the dynamic pressure did not exceed the set threshold. Six passes through the atmosphere were performed, and results of the successful execution of the experiment are summarised by Damiani et al. (2015).

The ExoMars Trace Gas Orbiter (TGO or ExoMars Orbiter) is a joint mission between ESA and Roscosmos that, on 14 March 2016, launched an atmospheric research orbiter and the Schiaparelli demonstration lander to Mars, as part of the ExoMars programme. After separating the Schiaparelli lander, TGO performed a Mars-orbit injection and soon after began its aerobraking manoeuvre (March 2017),

the first operational execution in an ESA mission. Aerobraking was suspended in July and August due to a solar conjunction (Mars, as seen from Earth, is hidden by the Sun). Aerobraking resumed in September and continued until March 2018. During that time, close to a thousand (952) passes through the atmosphere took place (Denis et al. 2018). Initial key constraints for the aerobraking were an aerodynamic heating of 2.8 kW/m^2 and a corresponding integrated heating of 500 kJ/m^2—with the actual corridor constraints set to half their values to account for atmospheric uncertainties and general safety concerns, i.e., a three-hour eclipse time and two-day lifetime. The integrated-heating constraint became more apparent towards the end of the main phase, since more time was spent in the atmosphere, whereas the lifetime limit dominated the walk-out phase, when the spacecraft was close to spiralling toward the surface. Total savings in terms of propulsive ΔV was 1,017 m/s.

11.1.2 Theory

In the previous section, while giving an overview of previous missions that used aerobraking, it was already mentioned that the reason to do so is to save on fuel for major orbit corrections. Also, for several missions the different guidance phases were mentioned. Before we will move to potential guidance algorithms in the next section, we will briefly revisit the aerobraking objective, and look at some related definitions.

The main constraints on aerobraking used by MO were:

- **Thermal**: periapsis heat-flux corridor, where the upper limit was based on maximum temperature of solar arrays (175 °C) with a safety margin of 80–100%, and the lower limit based on total aerobraking time.
- **Manoeuvres**: aerobrake trim manoeuvres were used to keep the periapsis in the heat-flux corridor; manoeuvre magnitudes were chosen from a predefined list (updated weekly).
- **Lifetime**: the time needed to reach 300 km apoapsis altitude had to be greater than one day.

Apart from the need to define an entry corridor in relation to thermo-mechanical loads, both the periapsis and apoapsis play an important role in the operation of the trajectory definition and the guidance-system design. For an elliptical orbit, defined by its Kepler elements (see Sect. 4.2, Figs. 4.1 and 4.2), the apoapsis and periapsis distance to the centre of main attracting body are defined by

$$r_a = a(1+e) \qquad r_p = a(1-e) \tag{11.1}$$

with a being the semi-major axis and e the eccentricity. The corresponding altitudes above the surface, h_a and h_p, are simply obtained by subtracting the equatorial radius, R_e: $h_a = r_a - R_e$ and $h_p = r_p - R_e$.

The effect of passing through the atmosphere is noticeable as a thermo-mechanical load: the drag force will slow down the vehicle, experienced as a mechanical load, and at the same time will heat up the exposed surfaces. These two loads are taken into account while defining the lower bound of the entry corridor. However, integrating the retarding effect of the drag force will also provide the total velocity loss, and an indication of the fuel savings by not using a propulsion system to achieve this.

The mechanical load is defined as

$$\bar{a} = -\frac{D}{mg_0} \tag{11.2}$$

whereas the incremental velocity loss can be obtained with

$$\Delta V = \int_0^t \frac{D}{m} \mathrm{d}t \tag{11.3}$$

Several authors have used a simple heat-flux indicator (Chambre and Schaaf 2017; McRonald and Nock 1977; Dec and Amundsen 2003) that works well enough to serve as a constraint. This indicator, which has been derived in Chap. 3, Sect. 3.6.2.3, is given by

$$q_{ab} \approx \frac{1}{2}\rho_\infty u_\infty^3 \tag{11.4}$$

where q_{ab} keeps the unit of W/m^2. The integrated heat load in J/m^2 is then given by

$$Q_{ab} = \int_0^{t_f} q_{ab} \mathrm{d}t \tag{11.5}$$

whereas the total absorbed heat in J is obtained by integrating over the surface area. For a flow perpendicular to all parts of the exposed vehicle, a conservative estimate is simply

$$Q_{ab,surf} = \int_{S_w} Q_{ab} \mathrm{d}S \approx Q_{ab} S_{ref} \tag{11.6}$$

To get a feeling for the occurring heat loads during aerobraking, in Fig. 11.7 the maximum heat rate experienced by Mars Odyssey during each pass through the atmosphere is shown. One can notice some interesting features:

1. The upper corridor limit was exceeded several times during aerobraking, however, another two limits were posed. The "flight-allowable thermal limit" is the actual maximum limit for the solar panels (which is twice as much as the upper corridor limit, due to the 100% margin, corresponding to a safety factor

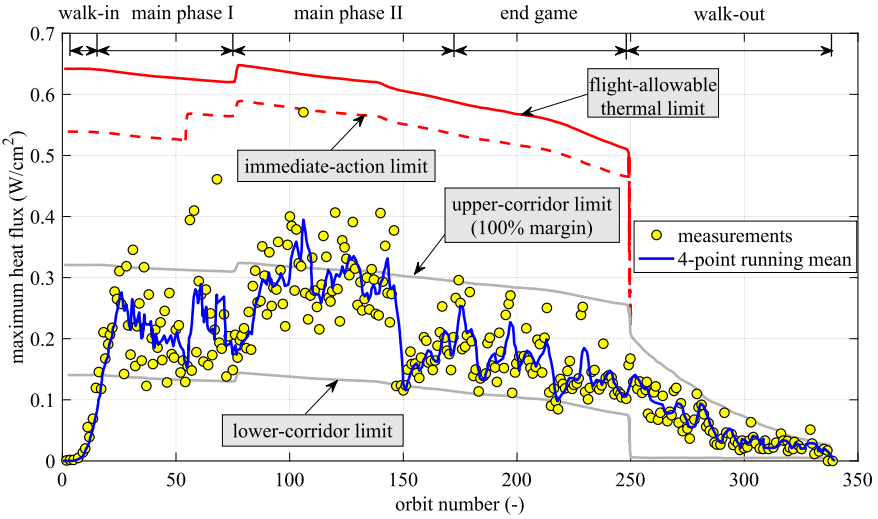


Fig. 11.7 Maximum heat flux during each atmosphere sweep of the aerobraking phase of Mars Odyssey. Data taken from (Smith and Bell, 2005)

of 2), and the "immediate action limit" corresponds to the threshold where the ground control-centre would have commanded an altitude raise for the next pass.

2. The heat rates during the endgame and walk-out phases became more strict, because of the different strategies applied, i.e., the cumulative heating and orbit lifetime limitations.

Mars Odyssey was also equipped with a so-called periapse time estimator (PTE) (Johnson and Willcockson 2003). The role of PTE was to be a technology demonstrator, to see if aerobraking could be partially automated. In fact, the PTE was used to correct the successive atmospheric sweep, based on the periapsis time error computed during the previous pass. The timing error was computed by analysing the acceleration data collected by the onboard inertial measurement unit (IMU). The correction for the successive sweep was simply to shift the predefined manoeuvres in time, by the error computed by the PTE. Despite its simplicity, this algorithm proved to be quite useful, especially in reducing the workload for the ground-station personnel.

While discussing aerobraking guidance techniques (Sect. 11.1.3), the PTE will be discussed in some more detail. However, before this comes up, the following example will support the concept of aerobraking.

Example 11.1 (*Open-loop aerobraking*) In this example the effect of passing through the upper layers of the Martian atmosphere is illustrated to provide some insight in the aerobraking mechanism. Let us consider a satellite with a mass of 1000 kg that has approached Mars, and has been propelled into a highly elliptical

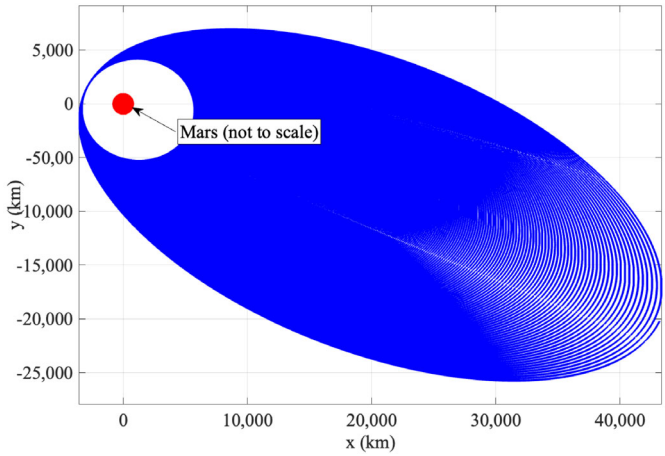


Fig. 11.8 Orbit bundle of an aerobraking mission through the Martian atmosphere

orbit with the following orbital parameters: a semi-major axis of $a = 27{,}228.5$ km, eccentricity $e = 0.871$, inclination $i = 93°$, longitude of ascending node $\Omega = 158.7°$, and argument of periapsis $\omega = 23.4°$. The initial position is set at the apoapsis, so the true anomaly is $\theta = 180°$. The satellite's solar panels have been oriented perpendicular to the oncoming flow, which yields a frontal area of 30 m^2. Finally, a (constant) drag coefficient of $C_D = 2.4$ is applied to represent an average effect of the free-molecular flow aerodynamics. This satellite configuration compares well with Mars Reconnaissance Orbiter, having similar mass and geometry properties.

The purpose of this example is to show how the orbit geometry will change without active guidance, i.e., when the satellite is flying through the upper layers of the Martian atmosphere and how this passage will affect both apoapsis and periapsis, as well as the induced thermal loads. The first simulation performed is one for 200 days. During this time, 947 passages through the atmosphere are achieved, which is shown in Fig. 11.8. Even though most orbits have visually merged, it is clear that the apoapsis decreases significantly, whereas the periapsis remains almost constant. The final orbit is still elliptical, but is moving towards a circular one.

Zooming in, the apoapsis altitude reduces from $h_a = 47{,}550$ to 2,927 km, whereas the periapsis altitude changes from $h_p = 116.3$ km down to $r_p = 113.8$ km (see Fig. 11.9). This corresponds to a change in eccentricity of $e = 0.8710$ to 0.2893. If the goal would have been to reach a final apoapsis altitude of 300 km, after which the orbit would be circularised, some more passages would be required, or the initial periapsis altitude should be lower. Below it will be shown that this is, of course, possible, but the resulting orbit will be sensitive to small changes in initial condition.

The change in h_a is more or less constant, shown as a decrease with an almost constant gradient. The change in h_p, on the other hand, is very slow for about 120 days, but starts to decline more rapidly after. Even though the total decrease is a mere

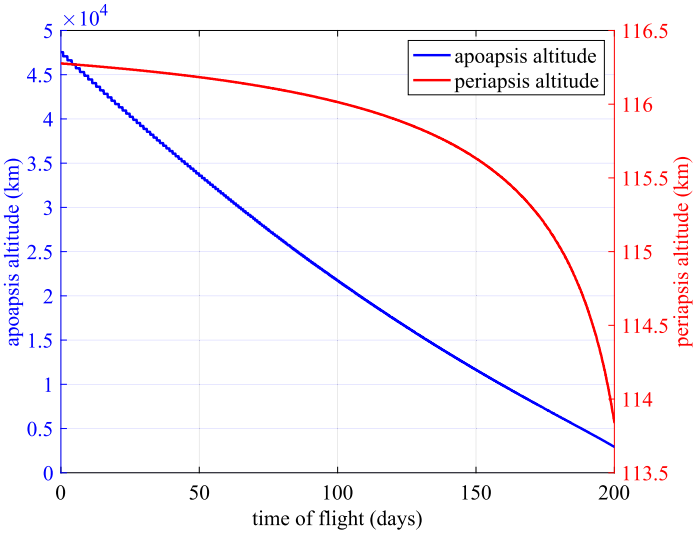


Fig. 11.9 Decrease in apoapsis and periapsis altitude for an open-loop aerobraking mission through the Martian atmosphere

2.5 km, it is not hard to imagine that this decrease will be increasingly getting larger with each successive pass through the atmosphere.

Figure 11.10 shows the effect of the atmosphere on the change in velocity, as well as the thermal loads experienced by the satellite. The total ΔV after 200 days is about 820 m/s. To get an idea about the fuel savings, let us use MRO as an example. This satellite used 70% of its total 1,187 kg of mono-propellant fuel to slow down its velocity with 1,400 m/s during Mars orbit insertion.[4] With 831 kg for 1,400 m/s, this would mean about 487 kg for 820 m/s, assuming a linear scaling. Even with this coarse approximation, it is clear that a significant fuel saving can be achieved with aerobraking.

Concerning the total integrated heat load (per surface area), after 200 days this amounts to 6.242 kJ/cm^2, or 62.42 MJ/m^2. With an exposed surface area of $S_{ref} = 30$ m^2 and assuming a homogeneous heating, this would mean a total energy absorption of 1873 MJ.[5] Of course, the heat is not all absorbed by the structure. One should realise that the atmospheric flight during which the structure heats up lasts only briefly, after which the major portion of the orbit is outside the atmosphere. During this extended period, the structure can cool down again, as most of the heat

[4] https://mars.nasa.gov/mro/mission/spacecraft/parts/propulsion/, date accessed: February 19, 2022.

[5] To get a feeling for what 1 MJ represents, two examples are provided. To heat up 10 liter of liquid water at constant pressure from 0 °C to 100 °C, about 4.2 MJ is required. Also, a mid-sized car with mass of 1600 kg (e.g., an Audi A4), travelling with a speed of 127 km/hr (35.4 m/s) has an equivalent of 1 MJ kinetic energy.

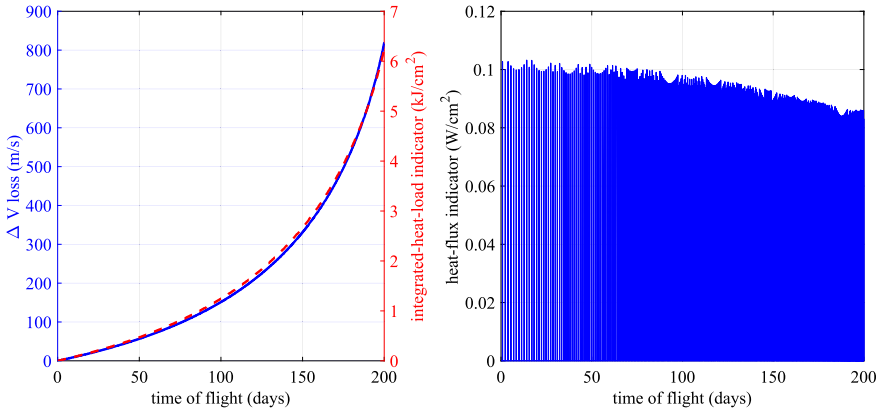


Fig. 11.10 Incremental ΔV loss and integrated heat load (left), and maximum heat flux (right) as a function of time of flight for an open-loop aerobraking mission

will be radiated into space again. Thus, only a very small fraction will be absorbed by the structure.

It is better to consider the maximum heat flux, as this parameter is a direct measure of the heating up of the structure, and gives an indication of the maximum allowable thermal load. It is used as constraints for the definition of the entry corridor, as shown in Fig. 11.7. The lower and upper constraint values for MO were defined to be $q_{ab,min} = 0.14\,\mathrm{W/cm^2}$ and $q_{ab,max} = 0.32\,\mathrm{W/cm^2}$, respectively. Note that it is common for aerobraking (and also aerogravity-assist) missions to express the heat flux in $\mathrm{W/cm^2}$, rather than in $\mathrm{W/m^2}$, as is common practise for re-entry missions. However, the values are easily transformed from one to the other; the two corridor constraints are equivalent to $q_{ab,min} = 1.4\,\mathrm{kW/m^2}$ and $q_{ab,max} = 3.2\,\mathrm{kW/m^2}$.

Looking at the heat flux of open-loop aerobraking mission (the right plot in Fig. 11.10) shows that the maximum occurring value is just above $q_{ab} = 0.1\,\mathrm{W/cm^2}$, which is well below the maximum allowable of MRO. Moreover, the figure also shows that with time (and each successive passage) the maximum value decreases, even though the atmospheric density is slightly increasing. This can easily be explained when one considers the definition of the heat flux, Eq. (11.4). Despite an increasing ρ, the speed is decreasing, which has a cubic effect on the heat flux compared to the linear effect of a density increase. A counter effect, though, is the ever decreasing time between successive periapsis passages (shown as the denser spikes from 100 days onwards). Due to the fact that the apoapsis altitude is lowering, the orbital period reduces proportionally. Thus, the time that the structure can cool down is also reduced. It may be obvious that a much more detailed coupled finite-element and computational fluid-dynamics analysis is required to get a more realistic prediction of the heat balance in the structure. An example of such a coupled analysis can be found in the paper by Dec and Amundsen (2003), whereas Liechty (2007) has made a comparison of MRO pre-flight aeroheating analysis with flight data.

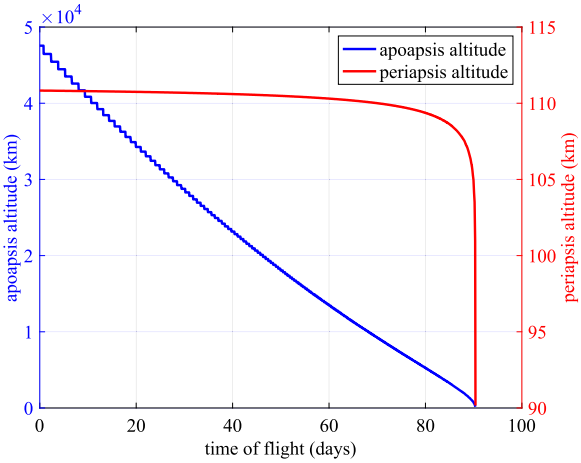


Fig. 11.11 Decrease in apoapsis and periapsis altitude for an open-loop aerobraking mission, with an initial eccentricity of $e = 0.8712$

To conclude this example, a small change in initial eccentricity is studied, with otherwise the same orbital and vehicle parameters as before. The eccentricity is increased from $e = 0.8710$ to $e = 0.8712$, which will have its effect on both apoapsis and periapsis altitude, see Eq. (11.1). The apoapsis altitude increases from 47,548.3 to 47,553.8 km, whereas the periapsis altitude reduces from 116.3 to 110.8 km. The simulation that we perform has to be much shorter to avoid the satellite from entering the atmosphere and crashing on the surface. Therefore, the simulation is stopped after 90.4 days, resulting in 499 orbits. During this time the eccentricity has reduced from $e = 0.8712$ to 0.0089, indicating a near-circular final orbit.

Figure 11.11 shows the decrease of both apoapsis and periapsis altitude. The former reduces from 47,548.3 to 154.6 km, whereas the latter decreases by just over 20 km, from 110.8 to 90.2 km. It is clear that when the atmosphere gets a "grip" on the satellite, the periapsis altitude starts decreasing rapidly, even though the linear decrease of the apoapsis altitude remains evident. Note that in an actual mission, the aerobraking would have stopped when h_a had reached 300 km, but this simulation serves the purpose to illustrate what happens when the satellite enters too deeply into the atmosphere. Indeed, that can lead to the destruction of the satellite.

Figure 11.12 shows both the ΔV_{loss} and the thermal loads. The velocity loss, and thus the fuel saving, reaches a large value and is asymptotically rising, because of the rapidly increasing atmospheric density. This seems like a positive result, but once the encountered heat flux is studied, one can see the danger of that statement. In just a few orbits the heat flux is rapidly rising, and jumps from $q_{ab} = 0.2$ W/cm^2 to about $q_{ab} = 0.45$ W/cm^2, a value well above the corridor limit of MO.

This example has shown the need for a robust guidance system, not only to keep the satellite within the entry corridor, but also to target the apoapsis altitude, and to finalise the manoeuvre by raising the perisapsis and circularising the final orbit. ■

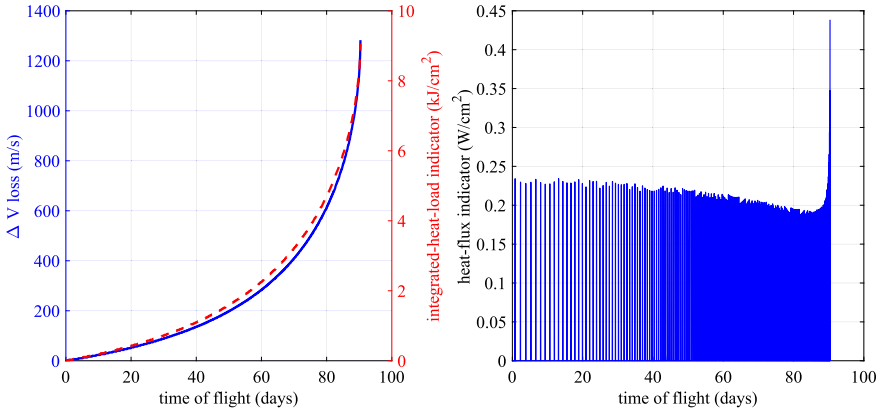


Fig. 11.12 Incremental ΔV loss and integrated heat load (left), and maximum heat flux (right) as a function of time of flight for an open-loop aerobraking mission, with an initial eccentricity of $e = 0.8712$

11.1.3 Guidance

In the previous section the aerobraking manoeuvre was introduced, as well as several missions that have executed such manoeuvres in the past. Distinct sub-phases were introduced as part of the overall mission. Each of these sub-phases has certain characteristics that would require specific attention when executed by a(n) (on-board) guidance, navigation, and control system. This was highlighted with an example of an open-loop aerobraking mission, which showed the necessity of periapsis-altitude control to avoid excessive heating. In this section we will look at the sub-phases of the aerobraking manoeuvre, and associated guidance issues.

The main task of the guidance system for an aerobraking manoeuvre is to determine the magnitude and direction of the corrective apoapsis manoeuvres, to achieve a gradual decrease of the apoapsis altitude, while respecting the heating and lifetime constraints dictated by the spacecraft material properties. These constraints define the periapsis corridor, in the same way the entry corridor is defined for entry vehicles (see Sect. 5.4.3). Taking data from some existing missions, a typical periapsis corridor is formed as follows. Because the satellites are not designed as entry vehicles and are thus lacking a proper heat shield, the constraint for the lower altitudes is determined by a maximum heat flux of 2.8 kW/m^2 (defined for TGO). Linked to this is an integrated heat-load constraint of 500 kJ/m^2 (Lyons 2012). To have sufficient atmospheric drag, the higher-altitude constraint is defined by a minimum dynamic pressure (and drag), which, for MRO, was 0.19 N/m^2 (Lyons 2002). In addition, one could think of a maximum bending moment (of, for instance, the solar panels). Obviously, this constraint is linked to a *maximum* dynamic pressure.

Finally, to avoid overshooting the lower corridor bound and consequently to terminate the manoeuvre, a so-called lifetime constraint is used, for MRO it was defined

to be 2 days (Long et al. 2008). The lifetime is defined to be the predicted time to reach the targeted apoapsis altitude. The smaller the lifetime is chosen, the longer the aerobraking manoeuvre will last, and the more exposure the satellite may have during the periapsis passage. Therefore, concerning thermo-mechanical loads, it is safer to avoid taking a too low value for the lifetime.

Before the satellite will start its main braking phase, the guidance system has some other tasks to take care of, and also after completion some actions have to be taken to formally terminate the aerobraking manoeuvre. In the previous section, several guidance phases were already mentioned. To summarise, these are:

1. **Walk-in**: the period in which the periapsis is slowly lowered in the highest layers of the atmosphere. During walk-in, the main objective is to initialise the atmosphere estimator, such that the risk of over- and (especially) underestimating the encountered density is minimised. To do so in a safe manner, during the first few orbits (Murri (2013) recommends seven orbits) the periapsis is slowly lowered to probe the atmosphere and update the on-board atmosphere model. This slow reduction is achieved by multiplying the estimated corridor boundaries—to be discussed below—with a simple constant:

$$\kappa_p = \begin{cases} 1.2 - \frac{0.2}{7}\mathcal{N} & \mathcal{N} < 7 \\ 1.0 & \text{otherwise} \end{cases} \tag{11.7}$$

   with $\mathcal{N}$ being the number of orbits performed. The constant factor 1.2 is an empirical value, chosen to introduce an offset between the actual and effective estimated corridor boundaries. Lowering the periapsis is achieved by applying a propulsive ΔV in the apoapsis. Even though not the main focus of the current discussion, in the intermezzo below, a simplified analysis is presented to determine the relation between a change in semi-major axis (and consequently, the periapsis altitude) and the required ΔV.
2. **Main**: the phase during which most of the braking occurs. This phase will be discussed in detail below.
3. **Walk-out**: the part during which the lifetime constraint takes effect. As mentioned earlier, towards the end of the aerobraking phase the lifetime constraint will come in effect, and will provide an extra check on the lower bound of the periapsis corridor. This phase becomes active whenever the value of eccentricity gets lower than a threshold value, e.g., 0.3 (this value is typically determined empirically, similar to the number of orbits for the walk-in phase). This value cannot be chosen too low, as is confirmed by Fig. 11.13. This figure shows evolution of eccentricity and the drop in apoapsis altitude per orbit (expressed as a percentage of the apoapsis altitude itself) during an uncontrolled aerobraking manoeuvre. Towards the end of the manoeuvre, the decline is very steep, and continuing the manoeuvre too long will lead to a potential violation of the lower bound of the periapsis corridor.
4. **Termination**: the orbit during which the periapsis is raised outside the atmosphere and the apoapsis is targeted to its final value. These manoeuvres will

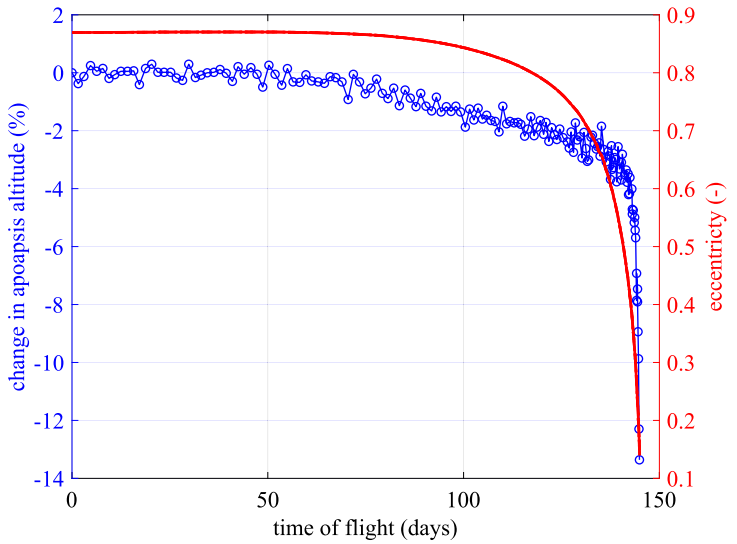


Fig. 11.13 Example evolution of eccentricity and apoapsis-altitude change, for an uncontrolled aerobraking manoeuvre

have to be propulsive, and will follow a similar procedure as for lowering the periapsis altitude discussed in Intermezzo 11.1.

5. **Aerobraking complete**: the final orbit is used to check the achieved orbit configuration; no special precautions or actions need to be taken.

The above phase separation is similar to those for Mars Reconnaissance Orbiter, of which more details and results of its flight performance are provided by Lyons (2001, 2002). Inspiration for the development of autonomous aerobraking guidance can also be found in the reports by Murri et al. (2012) and Murri (2013).

Intermezzo 11.1 (Periapsis-Lowering Manoeuvre)
Before beginning the aerobraking manoeuvre, the periapsis has to be lowered into the atmosphere to update the on-board atmospheric-density models with more accurate updates, and to achieve that the aerodynamic drag is sufficiently large to start lowering the apoapsis. In this intermezzo the effect of executing a thrusted manoeuvre on the semi-major axis, a, and eccentricity, e, is discussed. A direct application, as discussed in this section, is the apoapsis manoeuvre while aerobraking. This manoeuvre should lead to a change in periapsis altitude, r_p. To determine the required ΔV to achieve this, the change in r_p has to be related to a change in a. It is noted that this is a simplified analysis based on an impulsive manoeuvre.

For an elliptical orbit with given a and e, the apoapsis and periapsis altitude are defined by:

$$r_a = a(1+e) \qquad r_p = a(1-e) \tag{11.8}$$

It easily follows for the semi-major axis that

$$a = \frac{r_p + r_a}{2} \tag{11.9}$$

Thus, a change in r_p can be directly related to a change in a, when it is assumed that r_a remains constant. Using a superscript "+" to indicate the value of a parameter after the manoeuvre, it can be written that

$$a^+ - a = \frac{r_a + r_p^+}{2} - \frac{r_a + r_p}{2} = \frac{r_p^+ - r_p}{2} \tag{11.10}$$

or

$$\Delta a = \frac{\Delta r_p}{2} \tag{11.11}$$

The next step is determine the required ΔV to achieve this Δa. Starting point is given by Lagrange's planetary equations that allow for calculating the rates of change of the orbital elements if the form of the perturbing function is known. These equations can be found in many astrodynamics textbooks, see for instance those by Schaub and Junkins (2014) and Vallado (2013). In the current discussion, the Gauss form of the Lagrange's planetary equations for semi-major axis and (to be used at a later stage) eccentricity is used.

Let the non-gravitational acceleration in the local orbital frame $\mathcal{F}_L$—see Sect. 4.1.2 for the proper definition—be given by

$$\mathbf{a_L} = \begin{pmatrix} a_r & a_\theta & a_n \end{pmatrix}^{\mathrm{T}} \tag{11.12}$$

with the three components in the radial and azimuthal direction (in the orbital plane), and the third one normal to the orbital plane. The equations for rate of change of a and e are then given by:

$$\frac{\mathrm{d}a}{\mathrm{d}t} = 2\frac{a^2}{\sqrt{\mu p}}\left[a_r e \sin\theta + a_\theta\left(1 + e\cos\theta\right)\right] \tag{11.13}$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} = \sqrt{\frac{p}{\mu}}\left[a_r \sin\theta + a_\theta\left(\cos E + \cos\theta\right)\right] \tag{11.14}$$

with μ being the gravitational parameter of the target planet, $p = a(1-e^2)$ being the semi-latus rectum, and E the eccentric anomaly (see Eq. (4.5)).

To relate the continuous derivative $\frac{\mathrm{d}a}{\mathrm{d}t}$ to an impulsive shot, we need to evaluate the limit of both the left-hand side of Eq. (11.13) to obtain the discrete change in a, as well as the acceleration components, a_r and a_θ, on the right-hand side to obtain the corresponding ΔV:

$$\Delta a = \lim_{t \to 0} \int_0^t \frac{\mathrm{d}a}{\mathrm{d}t} \mathrm{d}\tau \qquad \Delta V_{(r \text{ or } \theta)} = \lim_{t \to 0} \int_0^t a_{(r \text{ or } \theta)} \mathrm{d}\tau \tag{11.15}$$

Substituting these equations into Eq. (11.13)—and assuming the motion is unperturbed—yields the relation for a discrete change in semi-major axis due to impulsive shots:

$$\Delta a = 2 \frac{a^2}{\sqrt{\mu p}} \left[\Delta V_r e \sin\theta + \Delta V_\theta \left(1 + e \cos\theta\right) \right] \tag{11.16}$$

Because the manoeuvre is executed at the apoapsis, where $E = \theta = \pi$ rad, the above equation can be simplified. Substituting these values for E and θ, and realising that the orbital mean motion, n, is defined by

$$n = \sqrt{\frac{\mu}{a^3}} \tag{11.17}$$

Eq. (11.16) reduces to

$$\Delta a = 2 \frac{a^2}{\sqrt{\mu p}} \left(1 - e\right) \Delta V_\theta = \frac{2}{n} \sqrt{\frac{1 - e}{1 + e}} \Delta V_\theta \tag{11.18}$$

Finally, with Eq. (11.11) the magnitude of the required ΔV_θ can be related to the change in periapsis altitude:

$$\Delta V_\theta = n \frac{\Delta r_p}{4} \sqrt{\frac{1 + e}{1 - e}} \tag{11.19}$$

■

After having covered the initial and final phases of aerobraking, the focus is shifted to the main phase. Also within this main phase, several sub-phases can be discerned, which have been depicted in Fig. 11.14. The sub-phases are executed/encountered from the current-orbit to the next-orbit apoapsis, where the states or conditions triggering an event will commonly have to be estimated onboard.

Pre-manoeuvring Phase/Periapsis Corridor Estimation

In the pre-manoeuvring phase several functions are carried out to prepare for executing the actual aerobraking manoeuvre when reaching the apoapsis. One of the

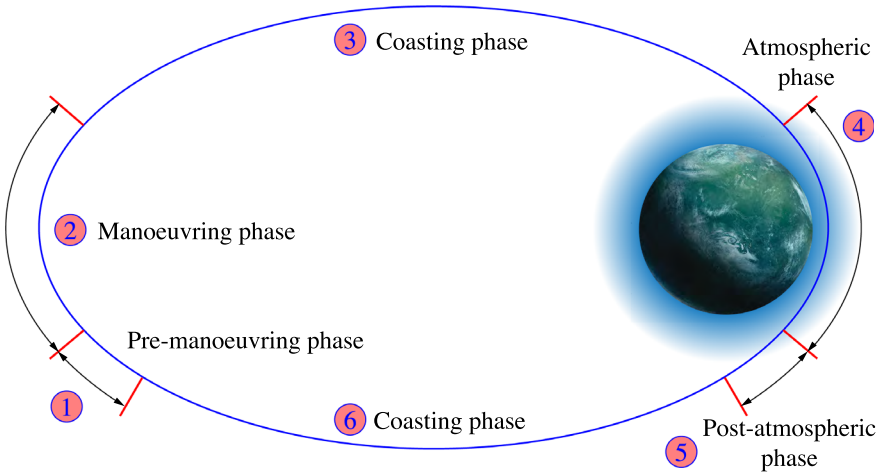


Fig. 11.14 In-orbit scheduling sequence for a spacecraft performing an aerobraking manoeuvre

tasks to be carried out is to estimate the periapsis-corridor boundaries. This is done by (on-board) propagation of a single orbit,[6] to estimate the thermo-mechanical loads on the structure. This estimation can be done by using a bisection root finder, as discussed in Sect. C.6 of the related appendix. From the orbit propagation, a guess of the periapsis altitude, $h_{p,guess}$, is made (with R_e being the equatorial radius of the target planet), which provide also an update of the semi-major axis and eccentricity:

$$r_{p,guess} = h_{p,guess} + R_e \tag{11.20a}$$

$$a_0 = \frac{r_a + r_{p,guess}}{2} \tag{11.20b}$$

$$e_0 = \frac{r_a - r_{p,guess}}{r_a + r_{p,guess}} \tag{11.20c}$$

and the other orbit elements simply keep their latest estimated value. For these new orbit parameters, the on-board propagator will be reinitialised. Then, the state is propagated for two-thirds of the orbital period, sufficiently long to include the atmospheric phase, and the resulting aerothermodynamics conditions are analysed. The upper altitude bound, the maximum dynamic pressure computed during propagation is compared with the minimum required value. This gives the root-finder function:

$$f_{upper}(h_{p,guess}) = \bar{q}_{max} - \bar{q}_{req,min} \tag{11.21}$$

[6] The on-board propagator could be any combination of propagator and integrator, as long as it is fast enough to execute on the on-board computer. This aspect will, of course, also limit the number and type of perturbations that can be taken into account.

For the lower altitude, in contrast, the heating conditions are used. Thus, the root-finder formula becomes:

$$f_{lower}(h_{p,guess}) = \max\left(q_{c,max} - q_{c,req,max},\, Q_{max} - Q_{req,max}\right) \tag{11.22}$$

In case the lifetime constraint is also active, i.e., the current orbit aerobraking phase is set to *walk-out*, the orbit is propagated nominally (thus with the current estimated state as initial conditions) for three days (i.e., 1.5 times the minimum lifetime). If a decay is detected, the lower-bound estimation is re-run, but this time the propagation is also carried out for three days, and the only parameter that is analysed is the lifetime:

$$f_{lower}(h_{p,guess}) = L - L_{req,min} \tag{11.23}$$

This extra margin is taken into account to be absolutely certain that the orbit decay will not violate the lower bound of the periapsis corridor, even though the on-board propagation may have neglected some of the perturbations, is using estimated values of the state and/or other model mismatches.

For the root-finding procedures the maximum number of iterations will have to be specified, in relation to, of course, the required accuracy. It may be clear that for an actual on-board implementation more advanced techniques for determining the periapsis corridor may be implemented. However, the current discussion serves as a simplified example to provide the reader with some insight about how it may work.

With all safety margins and possible corrections taken into account, in the end the result of this estimation process should lead to the upper and lower periapsis altitudes, $h_{p,upper}$ and $h_{p,lower}$. Once the corridor is known, the target periapsis altitude can be found and provided to the next sub-phase of the guidance logic.

If the predicted periapsis altitude already falls between the bounds, no manoeuvre is carried out. But, if this condition is not met, the target periapsis value is set to the midpoint of the corridor:

$$h_{p,\mathcal{N}} = \frac{h_{p,upper} + h_{p,lower}}{2} \tag{11.24}$$

with $\mathcal{N}$ being the current orbit number.

Manoeuvring Phase

The manoeuvring phase is the part of the orbit where propagation is halted, such that the estimated manoeuvre can be added to the simulated state of the spacecraft. Of course, the manoeuvre, and thus the propagation break, is only performed in case it is required by the corridor estimator, i.e., in case the predicted periapsis does not fall in-between the periapsis corridor bounds. See also the Intermezzo above for more information about the periapsis-lowering manoeuvre.

Coasting Phase

During the coasting phase, no actions related to aerobraking are taken by the guidance system. Of course, the on-board computer will keep monitoring the current state, and if anything is out the ordinary appropriate actions will be taken. The design of a complete flight software is far beyond the scope of this discussion, though.

Atmospheric Phase

The main element of aerobraking is the atmospheric phase. This is when the spacecraft exchanges its orbital energy with heat through friction to achieve a reduction in apoapsis altitude. During this phase, both the semi-major axis and the eccentricity are affected. To study the effect of aerodynamic drag on either a or e, a similar approach as taken in Intermezzo 11.1 can be followed, by calculating the discrete change in orbital elements and assuming that these changes can be achieved with impulsive shots. Starting from Gauss's planetary equations, Eqs. (11.13) and (11.14), the differential change in a was earlier found to be:

$$\Delta a = \frac{2}{n}\sqrt{\frac{1-e}{1+e}}\Delta V_\theta \tag{11.25}$$

Similarly, the change in eccentricity achieved by impulsive shots is:

$$\Delta e = \sqrt{\frac{p}{\mu}}\left[\Delta V_r \sin\theta + \Delta V_\theta\left(\cos E + \cos\theta\right)\right] \tag{11.26}$$

Because it is assumed that the change in a and e is discretely acquired at periapsis, consequently the true and eccentric anomaly are zero: $\theta = E = 0$. These conditions substituted in Eq. (11.26) yields

$$\Delta e = 2\sqrt{\frac{p}{\mu}}\Delta V_\theta = 2\sqrt{\frac{a(1-e^2)}{\mu}}\Delta V_\theta \tag{11.27}$$

Note that Eq. (11.25) remains unaffected, as Δa is not a function of either θ or E.

From a physical point of view, the change in velocity is of course acquired by the integrated effect of the aerodynamic deceleration,

$$\mathrm{d}V_\theta = c_{\Delta V}\int_{t_{in}}^{t_{out}} \|\mathbf{a}_{\mathbf{aero}}\|\mathrm{d}t \tag{11.28}$$

with t_{in} and t_{out} the time of entry into and exit out of the atmosphere, typically indicated by the accelerometers when a certain (small) deceleration level is exceeded. The constant $c_{\Delta V}$ has been included to account for the finite duration of the deceleration phase: in a realistic scenario, this phase can last several minutes, centred at the periapsis. This constant has been determined empirically by several authors, and established at 0.955 by Chabot et al. (2013) and 0.96 by Lyons (2001). The effect of $c_{\Delta V}$ becomes particularly useful for low-period orbits, where the time spent in the atmosphere is a considerable fraction of the orbital period.

To conclude the discussion of the atmospheric phase, MRO's capability of calculating autonomously when it had to re-orient itself for the next drag pass will be presented. This logic is known as the *periapsis time estimator* (PTE), and has been discussed in detail by Lyons (2001). Its main purpose is to prevent that the aerodynamic moments during the atmospheric pass become too large. Therefore, the spacecraft needs to be reoriented to an attitude that minimises these moments *before* the spacecraft enters the atmosphere. Since the drag pass is centred about the periapsis, the prediction of when this occurs is a convenient reference time for each orbit.

The approach that can be followed is centred around the notion that since at periapsis the deceleration will be at its maximum, this point divides the deceleration curve into two equal parts. Thus, the task at hand is to determine the condition at which time since periapsis passage, t_p, is zero.[7] One can make a prediction of this point by running a (simplified) dynamics model, determining the aerodynamic deceleration and using a recursive root finder, such as the bisection method:

$$f(\tilde{t}_p) = \int_{t_{in}}^{\tilde{t}_p} \|\mathbf{a}_{\mathbf{aero}}\| \mathrm{d}t - \int_{\tilde{t}_p+\Delta t}^{t_{out}} \|\mathbf{a}_{\mathbf{aero}}\| \mathrm{d}t \tag{11.29}$$

[7] The time since periapsis passage, obviously only defined for non-circular obits, is often used to define the mean anomaly, M:

$$M = \sqrt{\frac{\mu}{a^3}}(t - t_p) = n(t - t_p)$$

with n being the mean orbital motion. The relation between the mean and eccentric anomaly is provided by Kepler's equation

$$M = E - e \sin E$$

to be solved iteratively. The true anomaly, on the other hand, is related to E through

$$\theta = 2 \arctan\left(\sqrt{\frac{1+e}{1-e}}\right) \tan\frac{E}{2}$$

where $\tilde{t}_p$ is the estimate for t_p, which can be any time between t_{in} and t_{out}.[8] The corrected $\tilde{t}_p$ can then be used to correct the orbital elements and predict the next periapsis passage.

To aid the navigation function, the estimated values of t_p, Δa, and Δe can be used to correct the current estimates of the Keplerian elements. To do so, the estimated periapsis time is converted to an estimate of the true anomaly, $\tilde{\theta}_p$ (see Footnote 7), and treated as a correction factor of the true anomaly at exit of the atmosphere, θ_{out}. The Keplerian state can thus be corrected as follows. Inclination, i, argument of periapsis, ω, and longitude of the ascending node, Ω, are assumed to remain unaffected, and are taken as the estimates at the exit point. Semi-major axis and eccentricity are computed as the sum of the values at apoapsis, corrected by the change due to the atmospheric passage. Thus

$$\begin{pmatrix} a_{out} \\ e_{out} \\ i_{out} \\ \omega_{out} \\ \Omega_{out} \\ \theta_{out} \end{pmatrix} = \begin{pmatrix} a_a \\ e_a \\ i_{out} \\ \omega_{out} \\ \Omega_{out} \\ \theta_{out} \end{pmatrix} + \begin{pmatrix} \Delta a \\ \Delta e \\ 0 \\ 0 \\ 0 \\ \tilde{\theta}_p \end{pmatrix} \tag{11.30}$$

Obviously, the above discussion on guidance aspects is not complete to allow for a realistic guidance-system design. The purpose was merely to give a high-level overview of issues related to aerobraking manoeuvring.

11.2 Aerocapture

11.2.1 Concepts and Missions

Whereas with aerobraking the decrease in velocity is a slow and very gradual process, the objective of aerocapture is to change a hyperbolic interplanetary arrival trajectory into an elliptical orbit, by passing just once through the atmosphere. During this pass, kinetic energy is transformed into heat by aerodynamic friction, which has to be sustained by the vehicle's thermal protection system (TPS). The similarities between aerobraking and aerocapture stem from the fact that in both missions the atmosphere is used to decrease the energy of the vehicle passing through the atmosphere. The main difference between the two manoeuvres is the type of orbit with which the manoeuvre is performed, and the penetration depth into the atmosphere. Aerobraking is performed to lower the apoapsis of a closed orbit, when the vehicle has already

[8] Of course, with an IMU the actual measured accelerations can be used, although it should be noted that these measurements appeared to be noisy at best, in part due to density variations. Therefore, Lyons (2001), in his discussion, shows besides the use of the time-stamped measurements a 40-s running mean of the same data.

been captured about a celestial body, whereas an aerocapture manoeuvre is performed from an open (hyperbolic) orbit with the target to obtain a closed orbit. Aerobraking would be a logical follow-up to aerocapture to lower the apoapsis further if the current apoapsis were to be greater than the desired one. Additionally, aerobraking commonly happens in the upper parts of the atmosphere to limit the aerodynamic forces and aerothermal load, thus requiring many passes through the atmosphere; whereas aerocapture penetrates the atmosphere deeper to make use of the increased density at lower altitudes to increase the aerodynamic forces, and decrease the orbital energy faster.

Two different aerocapture strategies can be followed to achieve aerocapture, i.e., the direct entry and the periapsis approach. With direct entry the vehicle dives into the atmosphere, experience a severe thermo-mechanical load, depending on entry angle and velocity, and will subsequently follow its path to the planetary surface for either a touchdown or a splashdown. In most cases, the hyperbolic direct entry is supplemented by additional force inducing elements, such as parachutes or retrorockets. With the periapsis approach, the vehicle is targeted from its incoming approach orbit to an altitude that provides sufficient atmospheric drag such that the vehicle is decelerated into an intermediate orbit, with an apoapsis as shown in Fig. 11.15. After emerging from the atmosphere and reaching this (target) apoapsis, the periapsis must then be raised above the atmosphere to ensure that a second atmospheric passage does not occur. Such a passage could result in loss of the spacecraft, since its trajectory may have been sufficiently retarded to impact the surface of the planet. After this propulsive manoeuvre, the spacecraft will now be in its final elliptical (or circular) orbit.

On Earth, hyperbolic direct entry has been applied by several sample return missions returning from comets and asteroids. These include Stardust and Hayabusa. Stardust was a NASA sample return mission launched in 1999. The objective of the mission was to approach comet Wild-2 and collect cometary and interstellar particles (Atkins et al. 1997). These samples would then be returned to Earth. At Earth, the sample return capsule (Figs. 5.5 and 9.120) performed a ballistic entry. The inertial velocity of the entry was 12.9 km/s with a flight-path angle of $-8.2°$. During entry the vehicle encountered a peak heat flux of 12.0 MW/m^2 and a peak mechanical load of 34 g (Revelle and Edwards 2007). Its entry mission has been discussed previously in Sect. 9.7.4.

Hayabusa was a JAXA sample return mission launched in 2003. The objective of the mission was to rendezvous with the asteroid Itokawa, and collect samples from the asteroid's surface, and return them to Earth (Kawaguchi et al. 2006). At Earth, the sample return capsule (SRC) performed a ballistic entry with an initial inertial velocity of 12.035 km/s and a flight-path angle of $-12.35°$ (Cassel et al. 2011). During entry it encountered a peak heat flux of 11.12 MW/m^2 and a peak mechanical load of 49 g. The capsule, a 45° sphere-cone with a diameter of 40 cm and a mass of 16.3 kg, parachute-landed at the Woomera Test Range in South Australia and was safely recovered on June 13, 2010 (Fig. 11.16).

A follow-up to this mission, Hayabusa2, was launched in 2014 and has rendezvoused with asteroid Ryugy in 2018. Despite being a completely different mis-

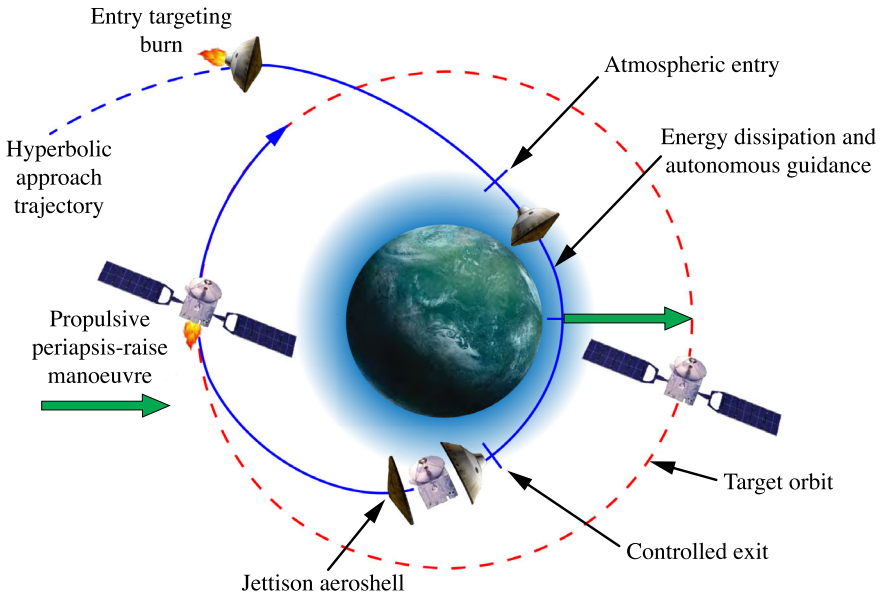


Fig. 11.15 The aerocapture manoeuvre



Fig. 11.16 The Hayabusa sample return capsule after landing, just before recovery. Image: JAXA

sion from its predecessor, the commonality is that of a sample return mission. Having successfully collected surface samples from Ryugy, they were stored in the SRC of the same design as the one for Hayabusa. The spacecraft left the asteroid in November 2019 and during its flyby along Earth, it released the SRC on 5 December 2020, which entered the Earth atmosphere with a velocity of about 12 km/s. As before, the entry was monitored in a dedicated airborne observation program (Scott Jr. et al. 2022). One of the many results was that the predicted peak radiative heating of 1.914 MW/m^2 at 69 s after entry interface could quite well be confirmed, which led to a successful validation of NASA CFD codes' ability to accurately predict shock-layer radiation.

In recent history, Mars missions have been performed using direct entry as well. Some examples of these missions are NASA's Pathfinder (1996) landing the Sojourner rover, NASA's Mars Exploration Rover (2003) landing the Spirit and Opportunity rovers, NASA's Mars Science Laboratory (2011) landing the Curiosity rover, ESA's ExoMars, in particular the Schiaparelli EDM Lander (2016), and NASA's Mars 2020 (2021). The NASA missions Mars Exploration Rover and Mars Science Laboratory have been discussed in detail in Sects. 9.7.5 and 9.7.6, so here only a few details on Mars 2020 will be provided. Mars 2020 followed the same mission profile as MSL, in the sense it was delivered to Mars with an orbital manoeuvring stage that was jettisoned prior to entry. Mars 2020 was heavily instrumented, such that an extensive post-flight analysis could be performed. Karlgaard et al. (2022) discuss the Mars Entry, Descent, and Landing Instrumentation 2 (MEDLI2), the instrument suite that provided the data for the trajectory, aerodynamics, and atmosphere reconstruction. More details on the entry, descent, and landing simulation is also provided by Way et al. (2022), whereas Dutta et al. (2022) zoom in on the atmospheric properties. Finally, Alpert et al. (2022) do an inverse estimation of the entry aeroheating environment using the MEDLI2 flight data.

A few relevant key results of these analyses are the following. The entry system entered the atmosphere with a flight-path angle of $\gamma_E = -15.48°$, with a velocity of 5.3247 km/s, and a cross-range error of 3.3 km. These values were all within the 1-σ error bound compared to the pre-flight predictions. Peak heating occurred 1 minute and 8 s after entry, giving a maximum heat flux of at most 890 kW/m^2 yielding a surface temperature of 1600 °C. The recession of the ablative heat shield was 1.9 mm. Again, these values were all within the 1-σ error bound. At 1 min and 23 s (after entry), the deceleration peaked at 10.7 g (Earth-referenced), which was outside the 2-σ error bound, but being lower than predicted. This was explained to be due to higher than expected atmospheric density in the upper atmosphere, such that the integrated drag effect led to a gradually lower velocity and deceleration. Three bank reversals were executed as planned, and it was concluded that the guidance-system performance was excellent.

For Venus, the only missions to perform hyperbolic direct entry into the atmosphere occurred during the space race in the 1960s and 1970s. According to Colin (1980) these were the Soviet Verena-3 through −12 from 1966 through 1978, and NASA's Pioneer-Venus 2 in 1978. Pioneer-Venus 2 consisted of four probes and the spacecraft bus. The relative velocity and relative flight-path angle at entry for

these four probes were 11.5377 km/s at −32.37°, 11.5372 km/s at −68.74°, 11.5371 km/s at −25.44°, and 11.5375 km/s at −41.50°, respectively. The bus entered with 11.5289 km/s at −9.38°.

The flight-path angles of these trajectories were found to be a lot steeper than those found characteristic for aerocapture, as will be discussed later, and are other examples of the direct entry aerocapture. To skip out of the atmosphere requires a much shallower entry, which will be discussed in more detail in Sect. 11.2.5, where allowable entry corridors are discussed. To anticipate on this, in the next section some characteristic trajectories will be introduced, to show the effect of a lift-up and a lift-down orientation in relation to the entry angle and corresponding thermo-mechanical loads.

11.2.2 Characteristic Trajectories

To anticipate on later discussions of optimal aerocapture trajectories, which fly with a combination of full lift-up and full lift-down, here some trajectory aspects are introduced that link strongly with those of skipping flight (Sect. 5.5). Full lift-up is typically required when the entry trajectory is relatively steep, and the vehicle loses quite some energy; the lift is required to make the vehicle leave the atmosphere again. Full lift-down, on the other hand, is used when the entry conditions are more shallow, and the vehicle would leave the atmosphere sooner rather than later, without shedding a significant portion of its energy. The lift-down orientation pulls the vehicle deeper into the atmosphere, such that with the increased drag more energy is dissipated.

Thus, in this section two trajectories will be introduced, i.e., the full lift-up trajectory and the full lift-down trajectory. The former flies the complete trajectory with a commanded bank angle of $\sigma_c = 0°$, whereas the latter is guided with $\sigma_c = 180°$. The simulations will be carried out for an Earth aerocapture, with an (inertial) entry velocity of $V_E = 13$ km/s at the entry-interface height of $h_E = 120$ km. Target apoapsis altitude for the simulation to stop is $h_a^* = 200$ km. The simulation model used is one for a non-rotating Earth, with central gravity field and a US76 atmosphere. This atmosphere model allows for a better density prediction for higher altitudes, as well as an altitude-dependent speed-of-sound that guarantees a more accurate Mach-number calculation and thus aerodynamic drag coefficient. To facilitate the simulation, a planar entry is assumed, using the base model given by Eqs. (5.8) through (5.10). The choice of this model is justified if one realises that adding a planetary rotation will obscure the aerocapture dynamics due to the relatively large perturbing effect of the Coriolis force, which is also heading dependent.

The simulations to be carried out are, in principle, an extension of the one introduced in Chap. 3, Example 3.7, where an aerocapture manoeuvre of the Crew Exploration Vehicle (CEV) was studied to assess the magnitude of the convective and radiative heat flux. In that example, the CEV was modelled after Apollo, albeit with a reduced ballistic coefficient. For the current simulations we will use the same aerodynamics, but then in combination with the original ballistic coefficient:

Table 11.2 Trim angle of attack and corresponding aerodynamic coefficients for the Crew Exploration Vehicle (Robinson et al. 2009)

Mach (–)	Angle of attack (°)	C_L (–)	C_D (–)	K (kg/m^2)
0.4	−12.86	0.24465	0.85300	537.39
0.7	−15.62	0.26325	0.98542	465.16
0.9	−18.30	0.32074	1.10652	414.26
1.1	−25.13	0.49373	1.16970	391.86
1.2	−24.87	0.47853	1.15600	396.50
1.35	−25.99	0.56282	1.27880	358.44
1.65	−26.78	0.55002	1.26570	362.14
2.0	−26.86	0.53247	1.27210	360.34
2.4	−26.38	0.50740	1.24120	369.32
3.0	−25.86	0.47883	1.21670	376.74
4.0	−23.88	0.44147	1.21480	377.32
10.0	−23.21	0.42856	1.22460	374.30
≤29.5	−19.94	0.38773	1.28910	355.61

$K = 355.61$ kg/m^2 ($m = 5{,}000$ kg, $S_{ref} = 12.017$ m^2 and $C_D = 1.28910$ at $M = 30$). Robinson et al. (2009) provide the aerodynamic data of Apollo in trim conditions in their paper, which they obtained by interpolation of the data of NAA (1965). These data can also be found in Appendix B, Sect. B.1, but since no interpolation of the trim angle of attack was done, the table provided by Robinson et al. (2009) is repeated here, see Table 11.2.

The one remaining initial condition is the entry angle, γ_E, which is to be found by trial and error. For the full lift-down trajectory, the entry trajectory cannot be too steep, as the lift is also pulling down the vehicle. The centrifugal force due to the curvature of the trajectory should be slightly larger than the lift force, to prevent the vehicle from crashing. Further, at the lowest point in the trajectory the energy of the system should be sufficiently large to allow for reaching the energy state at h_a^*, which should ideally be with a flight-path angle of $\gamma_F = 0°$. In that way, for circularisation only a horizontal ΔV would have to applied, in case an actual mission design would be considered. For the lift-up trajectory the entry trajectory has to be steeper, such that more energy is dissipated. The vehicle will dive deeper into the atmosphere, and the limit here is that the sum of lift and centrifugal force will be sufficient to pull the vehicle upwards again. The found entry angles are $\gamma_E = -5.528291°$ for the lift-down trajectory, and $\gamma_E = -8.4975°$ for the lift-up trajectory. It is noted that the lift-down trajectory is a lot more sensitive to (very) small initial flight-path angle variations compared to the lift-up trajectory, hence the more significant digits.

Figure 11.17 shows the altitude as a function of time. The full lift-down trajectory lasts about four times longer than the full lift-up trajectory. Both trajectories fly a major part of the ascending leg at an approximately constant altitude rate. The assumption of constant altitude rate is an important aspect of earlier guidance algo-

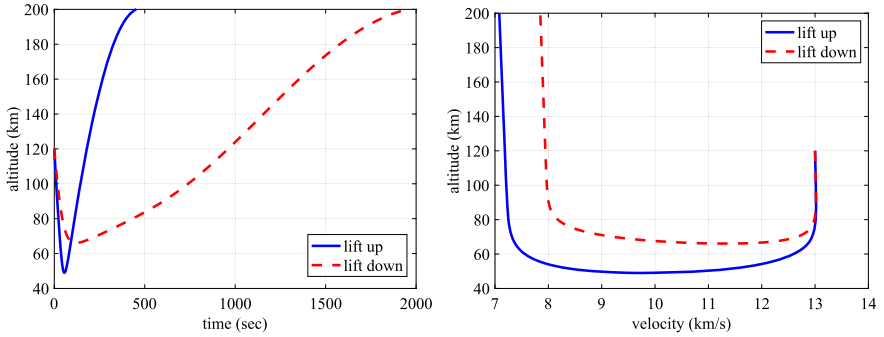


Fig. 11.17 Full lift-up and full lift-down trajectories to reach a target altitude of $h_a^* = 200$ km: altitude versus time (left) and altitude versus velocity (right)

rithms, i.e., the analytic predictor-corrector method (Cerimele and Gamble 1985; Rousseau et al. 2002), which will be briefly discussed in Sect. 11.2.3. In both cases, the ascending leg is the one that lasts the longest. The lift-down trajectory reaches a minimum altitude of 66 km, while the lift-up trajectory dives deeper into the atmosphere, to altitudes of 49 km. Looking at the altitude velocity profiles, it can be observed that the minimum altitude in the lift-up trajectory is reached at a velocity as low as 9.8 km/s, whereas it is reached with a velocity of about 11.3 km/s for the lift-down trajectory. Thus, it may be concluded that most of the energy for the lift-up trajectory is dissipated in about 55 s. This could be an important design driver for the design of the attitude controller: in case a bank reversal from full lift-down to full lift up has to be executed (which may take between 15 and 25 s), the rapid change in dynamics could pose a performance issue. To conclude, most of the energy is dissipated during the descending leg for a lift-up trajectory, and the opposite is true for the lift-down trajectory. It also shows that the final velocity of the lift-down trajectory is much larger (by about 715 m/s) than the final velocity of the lift-up trajectory, even though both aimed for the same apoapsis altitude.

Figure 11.18 shows both trajectories in the energy-dynamic-pressure plane. This visualisation is interesting, because the dynamic pressure is directly related to the load factor. In addition, it provides an indication of the aerocapture corridor: any trajectory starting with the same initial velocity as that of the boundary trajectories, and aiming for the same h_a^*, cannot have a single point in this plane outside of the area enclosed by the lift-down and the lift-up trajectories. These entry corridors will be discussed in a lot more detail in Sect. 11.2.5. It will then also be shown how for each point inside this corridor, a range of allowable flight-path angles exists. This, then, will describe the enclosure of all possible states for which an aerocapture can occur, for fixed V_E and h_a^*.

Going back to the figure at hand, the maximum dynamic pressure for a lift-down trajectory is about 6.4 times smaller than that for a lift-up trajectory. In addition, it is seen that the final energy of the lift-down trajectory is almost 6 MJ/kg larger than its lift-up counterpart. In the adjoining figure the velocity-flight-path angle relation is

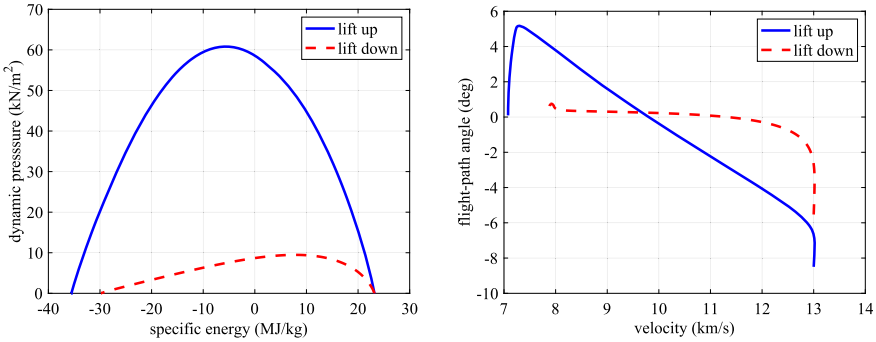


Fig. 11.18 Full lift-up and full-lift down trajectories to reach a target altitude of $h_a^* = 200$ km: dynamic pressure versus specific energy (left) and flight-path angle versus velocity (right)

depicted. It confirms what could already be derived from earlier results. The lift-up trajectory shows an almost linear relation between V and γ. Only towards the end of the flight gravity starts pulling the vehicle down to a more horizontal flight path. This could be optimised further, of course, such that any follow-up circularisation manoeuvre would be more fuel-optimal. The profile for the lift-down trajectory confirms that the initial flight-path angle is indeed very close to the value where skipping flight is no longer possible: a significant portion of the velocity decrease is flown in horizontal flight. The small flight-path angle in the ascending leg is consistent with the long time it takes to reach h_a^*.

It is stressed once more, that both trajectories forming the current entry corridor have been simulated without any active guidance system and have ignored any load constraints. To conclude this preliminary discussion, we will briefly discuss what that means for the occurring thermo-mechanical loads. Because of the definition of the *g-load*, for more or less constant aerodynamic properties the time history will follow the same profile as that of the dynamic pressure. Of interest are, of course, the peak values: for the full lift-down trajectory this is 3.13 g, whereas for the full lift-up trajectory this is significantly larger, i.e., 20.06 g. This should definitely be a point of attention while planning the aerocapture mission.

Figures 11.19 and 11.20 show part of the occurring heat fluxes, separated in a (cold-wall) convective and a radiative component (see Sect. 3.6 for the background on the used models), and the equilibrium wall-temperature for the full lift-down and full lift-up trajectories, respectively. In either case, it is evident that the radiative heat flux is significantly larger than the convective heat flux: for the lift-down case the peak value of the former is 6.26 MW/m^2 versus 1.79 MW/m^2 for the latter (a ratio of 3.5), whereas for the lift-up case the two peak values are 31.15 MW/m^2 and 3.91 MW/m^2, giving a ratio of 8.0. It is clear that the flux for the lift-up trajectory is extremely large, and that a proper modelling of the radiative heat flux is important. However, it is not only the peak values that matter, also the transient effect of the heating is a point of attention. The heat-flux gradient in the full lift-up trajectory is

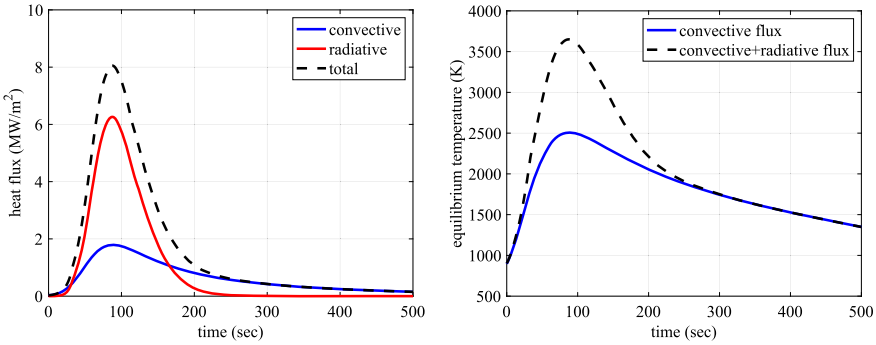


Fig. 11.19 Full lift-down trajectories to reach a target altitude of $h_a^* = 200$ km: heat flux (left) and corresponding equilibrium temperature (right)

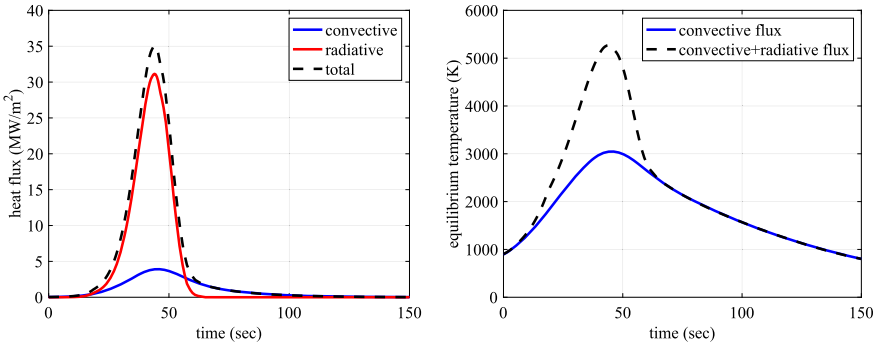


Fig. 11.20 Full lift-up trajectories to reach a target altitude of $h_a^* = 200$ km: heat flux (left) and corresponding equilibrium temperature (right)

very large as well (the peak value is reached in a very short time, but also the peak is relatively narrow, lasting only about 30 s, compared to about 160 s in the full lift-down case). These transient effects should always be considered when designing the thermal-protection system.

It is clear, though, that these fluxes could very well be too much for the heat shield to withstand. Looking at the equilibrium wall temperatures, for the full lift-down trajectory the maximum temperature is $T_{w,eq} \approx 3{,}650$ K ($\approx 2{,}500$ K when only convection is considered). For the full lift-up trajectory, $T_{w,eq} \approx 5{,}270$ K ($\approx 3{,}050$ K for the convective flux only). Without drawing very strict conclusions about this, two things are clear: a guidance system with robust constraint tracking is required, and the modelling of the (radiative) heat flux requires particular attention.

In more general terms, this preliminary analysis showed two benefits of the lift-down trajectory, namely the much smaller dynamic pressure (and thus *g-load* peak and the larger final velocity. Intuitively, one may assume that a larger final velocity

will lead to a smaller final ΔV needed to raise the periapsis. However, in the next section it will be discussed why, in reality, it is not advised to fly a full lift-down trajectory.

11.2.3 Analytic Predictor-Corrector Guidance

Over the years several aerocapture guidance schemes have been developed and tested in simulation environments; Rousseau et al. (2002) provide a nice overview of several algorithms (Analytical Predictor Corrector (APC), Numerical Predictor Corrector (NPC), Terminal Point Controller (TPC), and the Energy Controller (EC)) with several references, and compare their performance in terms of accuracy, robustness, thermo-mechanical loads control, and complexity.[9] Among those, the Analytic Predictor-Corrector (APC) guidance is, as its name implies, mostly based on analytical methods, and describing the method here would serve the spirit of this book. However, not only is it a relatively simple guidance system to implement, it is probably also one of the most widely used guidance algorithms in literature. It actually has quite a good performance and could very well serve as a benchmark for further studies. Therefore, the APC will briefly be discussed; for more detail and applications, the reader is referred to publications by, for instance, Gamble et al. (1988), Masciarelli et al. (2000), Perot et al. (2001).

The trajectory of an APC-driven trajectory consists of two different phases: an equilibrium glide and the exit phase. Each phase will now be detailed. Starting point for either is the set of equations of motion that we have seen many times before, i.e., the ones describing planar motion in the atmosphere of a non-rotating, spherical planet (as before, see Sect. 5.2 for the derivation of these equations). State variables considered are the speed V, flight-path angle, γ, and distance, R (or, equivalently, altitude h). Since R is also the radius of the instantaneous circular orbit, it can be treated as the radius of curvature.

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{D}{m} - g\sin\gamma \tag{11.31}$$

$$V\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \frac{L}{m} - g\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.32}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma \tag{11.33}$$

[9] The conclusion of Rousseau et al. (2002) is that the TPC ranked first with an excellent performance. The APC and EC are good alternatives to the TPC. The NPC, in the form it was tested, failed to compete. In all fairness, though, the authors do mention that its performance could be improved with further work, but that this may be reviewed in light of the complexity of the algorithm.

with $V_c = \sqrt{gR} = \sqrt{\frac{\mu}{R}}$ being the local circular velocity (which we also used extensively in Chap. 5). During the aerocapture manoeuvre, controlling the lift force through bank-angle modulation is standard practise, which means that the lift force in Eq. (11.32) is replaced by the vertical component of the lift force, i.e., $L_{vert} = L \cos\sigma$.

The aerocapture manoeuvre typically takes place at relatively small flight-path angle, so we can use this as an assumption to derive an expression for the derivative of the rate of change of the radius of curvature, R. We can easily do this by starting with the kinematic equation, Eq. (11.33). Taking the time derivative of both left and right side results in:

$$\dot{h} = V \sin\gamma \quad \Rightarrow \quad \ddot{h} = \dot{V} \sin\gamma + V \cos\gamma\dot{\gamma} \approx V\dot{\gamma} \tag{11.34}$$

Inserting this result into Eq. (11.32) yields:

$$\ddot{h} = \frac{L}{m}\cos\sigma - g\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.35}$$

This formulation will be used later on.

11.2.3.1 Equilibrium-Glide Entry Phase

The commanded bank angle during the entry phase comes from the simplified equations of planar motion, as discussed above. The first phase is approximated to be an equilibrium glide, derived in Sect. 5.4.1, Eq. (5.103), with the assumption that the flight-path angle is small. Setting Eq. (11.32) to zero yields:

$$L_{vert} = \bar{q} C_L S_{ref} \cos\sigma = mg\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.36}$$

Here, the $\cos\sigma$ term corresponds to the amount of banking involved, and $\bar{q}$ is the dynamic pressure. To damp potential oscillations, an altitude damper is required, such that equilibrium is achieved in short time, with only a small overshoot. This altitude damper is a corrective guidance law, derived from linearised error dynamics, and consists of both an altitude-rate and dynamic-pressure error component. For this, a reference dynamic pressure $\bar{q}_{ref}$ is needed to specify the glide condition:

$$\bar{q}_{ref} = -K_p \frac{mg}{C_L S_{ref}}\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.37}$$

which followed directly from Eq. (11.36) with $\sigma = 180°$; $\bar{q}_{ref}$ is computed to be K_p times the dynamic pressure that maintains equilibrium during a full lift-down flight. If K_p is equal to 1 the vehicle is in equilibrium glide, for values smaller than 1, the vehicle will dive under this condition and actual entry will occur, and when it

is larger, the vehicle skips out. According to aerocapture literature, a typical value for K_p is $K_p = 1.33$ (although tuning may be required). This gain value represents 75% of max lift-down capacity ($= 1/K_p$).

The guidance command for the bank angle to assure equilibrium glide is then:

$$\cos\sigma_{eg} = \frac{1}{\bar{q}}\left[\frac{mg}{C_L S_{ref}}\left(1 - \frac{V^2}{V_c^2}\right)\right] \tag{11.38}$$

with the dynamic-pressure tracking correction being:

$$\Delta\sigma = -K_{\dot{h}}\dot{h} + K_{\bar{q}}(\bar{q} - \bar{q}_{ref}) \tag{11.39}$$

Here, K_h and $K_{\bar{q}}$ are the gains for altitude rate and dynamic pressure, respectively.[10] Thus, including the corrective tracking guidance, Eq. (11.38) becomes

$$\cos\sigma_c = \frac{1}{\bar{q}}\left[\frac{mg}{C_L S_{ref}}\left(1 - \frac{V^2}{V_c^2}\right) + \Delta\sigma\right] \tag{11.40}$$

To establish values for $K_{\dot{h}}$ and $K_{\bar{q}}$, we need to take one step back and assume that the trajectory flown is "mildly" curved. With the equivalent expression derived for $V\dot{\gamma}$, we can now use Eq. (11.35), with σ_c as input,

$$\ddot{h} = \frac{L}{m}\cos\sigma_c - g\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.41}$$

and substitute Eqs. (11.39) and (11.40), then

$$\ddot{h} = \frac{L}{m}\left\{\frac{1}{\bar{q}}\left[\frac{mg}{C_L S_{ref}}\left(1 - \frac{V^2}{V_c^2}\right) - K_{\dot{h}}\dot{h} + K_{\bar{q}}(\bar{q} - \bar{q}_{ref})\right]\right\} - g\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.42}$$

With the term $g\left(1 - \frac{V^2}{V_c^2}\right)$ cancelling out, the above equation reduces to

$$\ddot{h} = \frac{L}{m}\left\{\frac{1}{\bar{q}}\left[-K_{\dot{h}}\dot{h} + K_{\bar{q}}(\bar{q} - \bar{q}_{ref})\right]\right\} \tag{11.43}$$

or

$$\ddot{h} + \frac{C_L S_{ref}}{m}K_{\dot{h}}\dot{h} - \frac{C_L S_{ref}}{m}K_{\bar{q}}\bar{q} = -\frac{C_L S_{ref}}{m}K_{\bar{q}}\bar{q}_{ref} \tag{11.44}$$

[10] Note that the minus sign in front of $K_{\dot{h}}$ is cosmetic, and only put there for the benefit of a later step. Obviously, when putting a plus sign, the minus sign would be included in $K_{\dot{h}}$.

The dynamic pressure can be linearised with respect to h over a small range, i.e.,

$$\bar{q} = Ah + B \tag{11.45}$$

such that Eq. (11.44) becomes a linear second-order differential equation, resembling the standard form of a damped mass-spring system $s^2 + 2\zeta\omega_n s + \omega_n^2 = F$, with forced input F. Comparing the two, the natural dynamic response characteristics become

$$\omega_n^2 = -\frac{C_L S_{ref}}{m} K_{\bar{q}} A \qquad 2\zeta\omega_n = \frac{C_L S_{ref}}{m} K_{\dot{h}} \tag{11.46}$$

with natural frequency, ω_n, and damping, ζ. For a given linearisation A, depending on the atmospheric properties and velocity at the considered range of flight conditions, one can now reverse the calculation in Eq. (11.46): by specifying target values for ω_n and ζ, the gains $K_{\dot{h}}$ and $K_{\bar{q}}$ can be determined. Or, choose the gains and verify that the response characteristics are acceptable. Cerimele and Gamble (1985) did just that: for their chosen vehicle and dynamic-pressure profile, values of $\omega_n = 0.049$ rad/s and $\zeta = 1.49$ were found to be acceptable for the gain values of $K_{\dot{h}} = 0.2$ and $K_{\bar{q}} = 5.0$.

Note that during this phase no particular prediction capabilities are required; the only goal of this phase is to keep the vehicle within a corridor that avoids skip-out, (crash)landing or other undesired situations, while decelerating to a speed V_s at which moment the exit phase can be initiated.

11.2.3.2 Exit Phase

The exit phase is based on a constant altitude rate scheme (see the earlier discussion on Fig. 11.17), and is, in nature, some form of predictor-corrector scheme, because in this phase the apoapsis can be accurately targeted. The principle behind this phase's guidance logic is that the bank angle is initialised, the trajectory simulated while keeping σ_c constant, the apoapsis will be calculated, and based on this prediction, σ_c will be adjusted such that the apoapsis is actually reached.

To begin the derivation, drag is considered to be the sole source of the deceleration, so Eq. (11.31) can be the starting point with the assumption of small flight-path angle:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma = -D = -C_D\frac{1}{2}\rho V^2 S_{ref} \tag{11.47}$$

or

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{C_D S_{ref}}{m}\frac{1}{2}\rho V^2 \tag{11.48}$$

Multiplying both sides with $\dot{h}$, and rearranging yields

$$\dot{h}\frac{\mathrm{d}V}{V^2} = -\frac{C_D S_{ref}}{m}\frac{1}{2}\rho \mathrm{d}h \tag{11.49}$$

Using the well-known assumption of an exponential atmosphere, the above equation can be rewritten as

$$\dot{h}\frac{\mathrm{d}V}{V^2} = -\frac{1}{2}\frac{C_D S_{ref}}{m}\rho_0 e^{-(h-h_0)/H_s}\mathrm{d}h \tag{11.50}$$

where the starting layer h_0 has been added to indicate the validity of the exponential approximation. Obviously, if there is only one layer then $h_0 = 0$, giving the familiar form that we have commonly used.

Integrating both sides from the vehicle's current state to the exit point, and recalling that $\dot{h} =$ constant, an expression for the predicted exit speed, V_F is found:

$$\frac{1}{V_F} = \frac{1}{V} - \frac{C_D S_{ref}\rho_0 H_s}{2m\dot{h}}\left(e^{-(h_F-h_0)/H_s} - e^{-(h-h_0)/H_s}\right) \tag{11.51}$$

However, by assuming a small flight-path angle the effect of gravity is ignored, so a correction is required to compensate for the velocity loss due to gravity. This can easily be derived by considering the energy balance, assuming that the only energy-transfer mechanism is the conversion of kinetic to potential energy. Assuming a flat Earth, the balance between the current and final state is:

$$gh + \frac{1}{2}V^2 = g_F h_F + \frac{1}{2}V_F^2 \tag{11.52}$$

or

$$V_F^2 = V^2 + 2\left(gh - g_F h_F\right) \tag{11.53}$$

The velocity change due to this energy exchange is

$$\Delta V_{F_{p/k}} = V_F - V = \sqrt{V^2 + 2\left(gh - g_F h_F\right)} - V \tag{11.54}$$

such that the actual exit velocity will be

$$V_F = V_F' + \Delta V_{F_{p/k}} \tag{11.55}$$

For an accurate calculation of the apoapsis altitude, one should realise that the considered velocities are relative ones, so a transformation to inertial velocity is required:

$$\Delta V_{F_{r/i}} = V_{F_i} - V_{F_r} \tag{11.56}$$

The predicted inertial exit velocity will become more accurate the closer the apoapsis is approached. The predicted inertial exit velocity is thus

$$V_{F_i} = V'_F + \Delta V_{F_{p/k}} + \Delta V_{F_{r/i}} \tag{11.57}$$

In reality, the constant $\dot{h}$ no longer holds. Because the altitude acceleration is more or less constant, this means that $\dot{h}$ becomes almost linear. Then, the altitude acceleration at exit is:

$$\ddot{h}_F = \frac{V_F^2}{R_F} - g_F - \frac{\rho_F V_F^2 S_{ref} C_L}{2m} \tag{11.58}$$

The altitude rate at exit can be computed using the former equation for $\ddot{h}_{ex}$:

$$\dot{h}_{ex} = \sqrt{\dot{h}_{const}^2 - 2(h - h_{ex})\ddot{h}_{ex}} \tag{11.59}$$

where $\dot{h}_{const}$ is the value that was assumed while obtaining Eq. (11.51). Since one of the main problems in predictor schemes is the error between the density in the model ρ_{mod}, and the real one, measured from drag ρ_d; thus, a filtered density multiplier K_ρ can be used:

$$\rho_d = \frac{2\,\dot{V}\,m}{C_D\,S\,V^2} \tag{11.60}$$

$$K_\rho = (1 - K)\,K_\rho + K\,\frac{\rho_d}{\rho_{mod}} \tag{11.61}$$

where K is a filter gain. Eventually, the commanded angle bank can be computed:

$$\cos\sigma_{cmd} = \frac{1}{\bar{q}}\left[\frac{m\,g}{C_L\,S}\left(1 - \frac{V^2}{V_c^2}\right) - K_{\dot{h}}(\dot{h} - \dot{h}_{ref}\right] \tag{11.62}$$

More details about this guidance algorithm can be found in the papers by Cerimele and Gamble (1985), and Gamble et al. (1988).

11.2.4 Optimal Aerocapture

Whereas in the previous section the fundamental principle of aerocapture guidance has been established, the current section will focus on optimal aerocapture manoeuvres. The optimality of an aerocapture trajectory depends on the mission objectives, which, in the current context, means that it is optimal for an apoapsis-targeting aerocapture. This translates in minimising ΔV_{total}, which is the summation of the

periapsis and apoapsis accommodating ΔV manoeuvres, as well as the path constraints of *g-load* and heat load. In this section we will address the optimal execution of an aerocapture for different performance objectives. Subsequently, the minimum ΔV, minimum peak load factor and peak heat flux, and the minimum radiative heat-load aerocapture will be discussed. Common to each is that the performance index as well as the constraints on the terminal conditions depend solely on a subset of the terminal vehicle state variables, notably the ones describing the planar motion of the vehicle, i.e., $\mathbf{x} = \begin{pmatrix} V & \gamma & R \end{pmatrix}^{\mathrm{T}}$. Furthermore, we assume this planar motion to take place relative to a non-rotating, spherical planet (or moon) with a central gravity field, $g = \frac{\mu}{R^2}$. This means that the dynamics model is identical to the one discussed in Chap. 5.

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \quad \Rightarrow \quad \frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{D}{m} - \frac{\mu}{R^2}\sin\gamma \tag{11.63}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \quad \Rightarrow \quad \frac{\mathrm{d}\gamma}{\mathrm{d}t} = \frac{L}{mV} - \frac{1}{V}\frac{\mu}{R^2}\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.64}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = V\sin\gamma \tag{11.65}$$

with the initial conditions given as $V(t_0) = V_0$, $\gamma(t_0) = \gamma_0$, and $R(t_0) = R_0$. The same change as in the previous section will be made: because during the aerocapture the planar lift needs to be controlled, the lift L in Eq. (11.64) will be replaced by $L\cos\sigma$, which is the actual vertical lift force in a three-dimensional model:

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \frac{L}{mV}\cos\sigma - \frac{1}{V}\frac{\mu}{R^2}\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.66}$$

where σ is the bank angle and is bounded by $0 \leq \sigma_{min} \leq \sigma \leq \sigma_{max} \leq \pi$). Any induced lateral motion due to the consequential horizontal lift force $L\sin\sigma$ will be ignored, such that the corresponding state variables defined by the heading angle, χ, and the combined angular position longitude and latitude, τ and δ, are not considered.

It is stressed that the bank angle is the only trajectory control variable. The angle of attack is therefore prescribed, and can be kept constant or defined to be a function of, for instance, flight time, Mach number (a combination of velocity and altitude), or any other variable that is convenient. Further, as is common practise for the planar-flight assumption, the angle of sideslip is set to zero such that no side force is induced.

Last, but not least, the above introduced problem describes a class of aerocapture problems with only in-plane targeting conditions. Any problem, which requires out-of-plane manoeuvres, such as plane changes or landing footprints involving a certain cross range when considering direct entry, are not optimised using the problem

solution to be presented below. In those cases the performance index also depends on the lateral motion of the vehicle and its position for position targeting and should be adjusted accordingly.

11.2.4.1 Minimum ΔV

The minimum-ΔV aerocapture manoeuvre will be derived using optimal-control theory in combination with Pontryagin's Maximum Principle (or, equivalently, the Minimum Principle). A summary of the theoretical background is provided in Chap. 7, or to be more specific in Sect. 7.5.4.1. It is noted that the original proof for a minimum ΔV trajectory was developed by Lu et al. (2015), and the presented solution will follow their line of thought.

With de definition of $\mathbf{x} = \begin{pmatrix} V & \gamma & R \end{pmatrix}^{\mathrm{T}}$, the performance index of the optimal trajectory is described as $\mathcal{J} = \mathbf{x}(t_f)$, and the terminal constraints must satisfy $\mathbf{s}\left[\mathbf{x}(t_f)\right] = \mathbf{0}$, where the dimension of $\mathbf{s}$ is $k \leq 3$. The Hamiltonian for this optimisation problem can be written as $\mathcal{H} = \boldsymbol{\lambda}^{\mathrm{T}}\dot{\mathbf{x}}$, so:

$$\mathcal{H} = \lambda_V \left(-\frac{D}{m} - \frac{\mu}{R^2}\sin\gamma\right) + \lambda_\gamma \left[\frac{L}{mV}\cos\sigma - \frac{1}{V}\frac{\mu}{R^2}\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right)\right] + \lambda_R V \sin\gamma \tag{11.67}$$

where, according to Pontryagin's Maximum Principle, the derivatives of the co-state variables, λ_V, λ_γ, and λ_R, are given by

$$\dot{\lambda}_V = -\frac{\partial \mathcal{H}}{\partial V} = \lambda_V \frac{\partial D}{\partial V} - \lambda_\gamma \frac{\partial \dot{\gamma}}{\partial V} - \lambda_R \sin\gamma \tag{11.68}$$

$$\dot{\lambda}_\gamma = -\frac{\partial \mathcal{H}}{\partial \gamma} = \lambda_V \frac{\mu}{R^2}\cos\gamma - \lambda_\gamma \frac{\partial \dot{\gamma}}{\partial \gamma} - \lambda_R V \cos\gamma \tag{11.69}$$

$$\dot{\lambda}_R = -\frac{\partial \mathcal{H}}{\partial R} = \lambda_V \left(\frac{\partial D}{\partial R} - \frac{2\mu\sin\gamma}{R^3}\right) - \lambda_\gamma \frac{\partial \dot{\gamma}}{\partial R} \tag{11.70}$$

The optimal bank-angle σ^* has to be such that, at any moment, $\mathcal{H}$ is maximised:

$$\sigma^* = \arg_{\sigma} \left[\max \mathcal{H}(\sigma)\right] \tag{11.71}$$

for a given range $\sigma \in [\sigma_{min}, \sigma_{max}]$, or

$$\sigma^* = \arg\max_{\sigma} \left\{ \lambda_V \left(-\frac{D}{m} - \frac{\mu}{R^2} \sin\gamma \right) + \lambda_\gamma \left[\frac{L}{mV} \cos\sigma - \frac{1}{V} \frac{\mu}{R^2} \cos\gamma \left(1 - \frac{V^2}{V_c^2} \right) \right] + \lambda_R V \sin\gamma \right\} \tag{11.72}$$

i.e., the maximum-principle equivalent of Eqs. (7.167) and (7.168). By inspecting Eq. (11.72), it can be seen that only the second component has a dependency on σ. Since $\cos\sigma$ is monotonic in $\sigma \in [\sigma_{min}, \sigma_{max}] \subseteq [0, \pi]$, it means that the expression for σ^* is linear in $\cos\sigma$, and with $L/mV > 0$, this means—without formal derivation—that the optimal bank-angle is given by

$$\begin{cases} \sigma^* = \sigma_{\min} & \text{if } \lambda_\gamma > 0; \\ \sigma^* = \sigma_{\max} & \text{if } \lambda_\gamma < 0; \\ \sigma^* \in [\sigma_{\min}, \sigma_{\max}] & \text{if } \lambda_\gamma = 0 \end{cases} \tag{11.73}$$

The third case implies that σ^* can have any value in the given interval for σ, and may even be time varying. Such a solution, where σ^* is different from its extrema, is called the singular optimal-control solution. However, this case does not exist for a finite interval of time, which Lu et al. (2015) proved as follows. If λ_γ is to be constant and equal to zero for a finite time, it is required that $\lambda_\gamma = \dot{\lambda}_\gamma = 0$. This implies, substituting $\lambda_\gamma = 0$ into Eq. (11.69):

$$\lambda_V \frac{\mu}{R^2} \cos\gamma - \lambda_R V \cos\gamma = \lambda_V \frac{\mu}{R^2} - V\lambda_R = 0 \tag{11.74}$$

provided that $\cos\gamma \neq 0$. Since the flight-path angle varies between $-\gamma_0$ and $+\gamma_0$, which in any practical scenario is never $\pm 90°$, this holds for all points in a skipping trajectory between entry and exit. Furthermore, since we are dealing with a free-time problem, i.e., t_f is not specified, the transversality condition $\mathcal{H} \equiv 0$ holds along the entire trajectory $\forall t \in [t_0, t_f]$. When λ_γ is put equal to zero, transversality condition Eq. (11.68) requires that

$$\mathcal{H} = -\lambda_V \left(\frac{D}{m} + \frac{\mu}{R^2} \sin\gamma \right) + \lambda_R V \sin\gamma = 0 \tag{11.75}$$

Equation (11.74) gives us $\lambda_V \frac{\mu}{R^2} = \lambda_R V$, and after substitution into Eq. (11.75) this leads to $\lambda_V = 0$, since $D > 0$. But, after looking at the relation between λ_V and λ_R, this automatically also leads to $\lambda_R = 0$. At this point, all the co-state variables are equal to zero, and this is a contradiction to the Maximum Principle that states that not all co-state variables can be zero. Therefore, it has been successfully shown that singular control cannot happen for any trajectory with free final time, and only terminal constraints and objectives. Therefore, the optimal trajectory-control will have to be a bang-bang style, i.e., starting with an initial value σ_0 and switching to its final value σ_f.

Of course, the above solution only holds within the assumptions that were made, and was possible, because neither the final conditions nor the objective included any term related to lateral motion. Generalising this problem to three-dimensional motion in a rotating atmosphere is relatively complicated, and typically not done mathematically, but numerically. However, two considerations can help. The first one is that the accelerations, neglected when using planar motion for a non-rotating primary body, are rather small. The largest one is the Coriolis acceleration, which is relatively small compared to the drag and lift acceleration. However, this is not true when the vehicle is in the upper, thin layers of the atmosphere. Although, in that case, the trajectory would be better approximated by Keplerian motion. During Keplerian motion, the parameters that define the orbit do not change, and, for this reason, nor does the objective function. The transition phase, where the Coriolis force is comparable to drag and lift, but Keplerian motion can no longer be assumed, is relatively short and could thus be ignored. This observation has been confirmed by Miele et al. (1990), who optimised the aerocapture trajectory for a rotating planet and found that the trajectory minimising the final ΔV is very close to a bang-bang trajectory. Even though not a mathematical proof, it gives confidence that the proof obtained for planar motion in a non-rotating atmosphere is at least a good approximation to consider, when optimising a three-dimensional trajectory in a rotating atmosphere.

11.2.4.2 Minimum Peak Load Factor and Heat Flux

For a typical mission design, the constraints imposed on a trajectory do not have to be minimised. As long as the constraints are respected, their actual value is of no importance. However, by finding the optimal trajectories for which the constraints are minimised, the limit values for the initial conditions can be found. No trajectory that starts with initial conditions beyond this boundary can be flown without violating these minimum requirements. This means that for these trajectories a pre-entry ΔV manoeuvre needs to be executed to change the initial conditions to avoid constraint violation.

Preliminary research has shown that there exist aerocapture trajectories that minimises both the peak load factor and the peak convective heat rate. Here, we will elaborate on that. To begin with, for the minimisation of the peak load factor, a corresponding control history $u(t)$ has to be found. However, for a vehicle with approximately constant aerodynamic properties, the load factor is proportional to the dynamic pressure according to the well-known expression

$$n = \frac{L^2 + D^2}{mg_0} = \frac{C_L^2 + C_D^2}{mg_0} \bar{q}_\infty S_{ref} \tag{11.76}$$

Therefore, we may write the equivalent expression to find $u(t)$ with

$$\min_{u(t)} \left[\max_t (\bar{q}_\infty) \right] \equiv \min_{u(t)} \left[\max_t (\rho_\infty V^2) \right] \tag{11.77}$$

Since it is known that both convective and radiative heat flux are, amongst others, a function of density and velocity—see the extensive discussions in Chaps. 3 and 5 (Sects. 3.6 and 5.3.2)—the proportionality to the density and velocity is used to obtain

$$\min_{u(t)} \left[\max_{t} \left(\rho_{\infty}^{n} V^{m} \right) \right] \tag{11.78}$$

In three successive steps we will now discuss the aerocapture-trajectory control profile, which minimises Eqs. (11.77) and (11.78). It is stressed, though, that this will not be a formal proof and does therefore not pose a generic truth[11]

Step 1: The ascending leg can be neglected when regarding peak values of the constraints

The first step can be proved easily by considering the altitude-density profile of the atmosphere (density decreases monotonously with altitude) and the relation between total, kinetic, and potential energy. During the ascending leg the altitude increases, so density decreases. Furthermore, both due to aerodynamic drag and energy exchange between kinetic and potential energy, the velocity of the vehicle decreases as well. Consequently, both the dynamic pressure and the proportionality of the heat flux decrease. Therefore, the peak value for the load factor and heat flux must occur somewhere during the downward leg. This has also been shown during the discussion of skipping flight.

Step 2: The lift-up lift-down bang-bang trajectory leads to the least negative flight-path angle (descending leg).

This second step will be discussed using the results of (numerical) simulations, starting with the bounds provided by the (saturated) limit trajectories shown in Figs. 11.17 through 11.20, i.e., the full lift-up and full lift-down trajectories. In addition, for a range of initial flight-path angles, varied between the two limit values, a switch in strategy is simulated. Using a bisection algorithm, the switch time from full lift-up to full lift-down and vice versa is determined, such that the vehicle will reach the apoapsis altitude (with $h_a^* = 200$ km, in this discussion) and a final flight-path angle of $\gamma_F = 0°$. The latter condition is enforced to have comparable trajectories that require only a tangential circularisation ΔV. The entry velocity for all simulations is set to $V_E = 13$ km/s. As before, the planar model for a non-rotating spherical Earth is used for the simulations, to avoid complications due to the Earth rotation as a function of vertical-plane orientation (represented by the flight heading). The

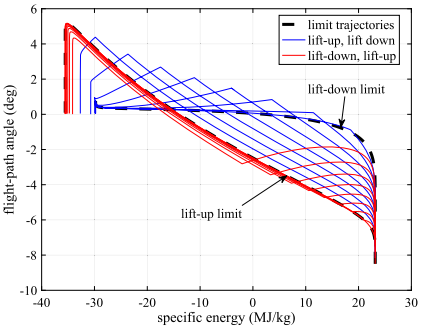
[11] A more formal analysis has been done by Lu (2014), applying the earlier introduced optimal-control theory. Such an analysis is quite involved, and requires the propagation of not just the states, but also the adjoint variables or co-states. Of course, once properly formulated this is a straightforward process, but since it requires the evaluation several derivatives, showing an analytical formulation is beyond the scope of the current discussion.

atmosphere is based on the US76 model, gravity is modelled with a central field, and the vehicle is the CEV (see Table 11.2, vehicle mass is $m = 5000$ kg).

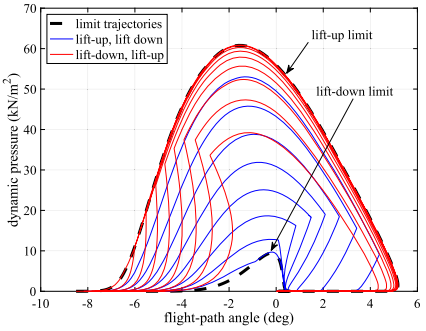
Combining the results from all simulations, an aerocapture corridor is obtained, including all possible boundary trajectories that can lead to the target apoapsis. Because of the applied model of planar motion, the corridor is represented by a three-dimensional space with dimensions of velocity, altitude, and flight-path angle. Because of their relations, velocity and altitude may be replaced by dynamic pressure and energy. Given the optimisation criteria Eqs. (11.77) and (11.78), it may be clear that dynamic pressure can also be substituted with the convective heat flux. The boundary surface consists of the encapsulation of all lift-up/lift-down and lift-down/lift-up trajectories, which are obtained from the range of allowable initial flight-path angles. The limit values of this range are $\gamma_E = -5.528291^\circ$ (full lift-down) and $\gamma_E = -8.4975^\circ$ (full lift-up). It is noted once more that these trajectories are unconstrained, i.e., the maximum thermo-mechanical loads are not observed, so no bank-angle modulation other than full lift up ($\sigma_c = 0^\circ$) or full lift down ($\sigma_c = 180^\circ$) is commanded. It is easily understood why the lift-up/lift-down and lift-down/lift-up trajectories are the corridor boundaries. Before the switch, the bank angle is saturated, and after the switch the only way to achieve the desired apoapsis is to have the command saturated in the opposite direction. A hypothetical vehicle inside such a corridor could not get out of it before crossing the surface defined by the collection of switch times, and could not go back inside if it were outside of it after that surface is crossed. Of course, this holds for the current planar motion over a non-rotating planet.

Figure 11.21 shows the resulting corridor in the form of two-dimensional planes: flight-path angle versus specific energy (Fig. 11.21a), dynamic pressure versus flight-path angle (Fig. 11.21b), and dynamic pressure versus specific energy (Fig. 11.21c). For any point on the $E - \bar{q}$ planes of Fig. 11.21c, if this is encountered by flying an up-down trajectory instead of a down-up trajectory, the corresponding flight-path angle will be the shallowest. This could be derived from Figs. 11.21a and 11.21b, if only they could be combined in a three-dimensional time-varying moving plot, unfortunately not possible on paper. However, intuitively one can reason this out as follows.

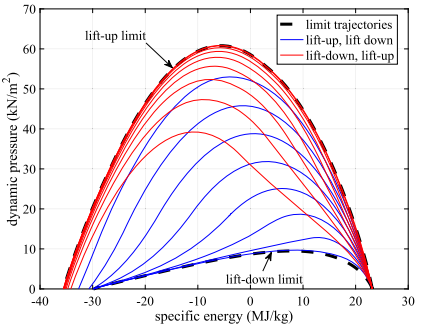
Following a line $\bar{q} =$ constant or $E =$ constant in one of those plots, one can find that most trajectories that are encountered first are down-up, which implies that those trajectories are always steeper. This holds for the descending leg only, corresponding to the point in which the flight-path angle is negative. For earlier discussed reasons, there is no need to consider the situation in which the flight-path angle is positive, i.e., the ascending leg. If Fig. 11.21c were three-dimensional, and the Z-axis were the flight-path angle (positive coming out of the paper), all the down-up trajectories would not be visible, because, for equal E and $\bar{q}$, they would have a steeper flight path (i.e., more negative angle). This observation is an important one, because a shallow flight-path angle implies smaller derivatives of both dynamic pressure and convective heat flux, which is confirmed by Lu (2014). This suggests that for a given moment in time, the lift-up lift-down trajectory is the one that minimises future peak load. Because this is valid at any moment in time, the peaks are minimised for the



(a) Flight-path angle versus specific energy



(b) Dynamic pressure versus flight-path angle



(c) Dynamic pressure versus specific energy

Fig. 11.21 Lift-up/lift-down and lift-down/lift-up trajectories for a range of initial flight-path angles, with entry velocity $V_E = 13$ km/s

complete trajectory. Moreover, the lift-up control is the control that maximises the derivative of the flight-path angle. The same reasoning applies to the convective heat flux.

Note that because the flight-path angle is always the least negative for lift-up/lift-down bang-bang or saturated lift-up trajectories, they reach the lowest point in the trajectory where $\gamma = 0$ at least as quick as, but generally quicker than, their lift-down/lift-up bang-bang or saturated lift-down trajectory counterpart. This means that the downward leg for lift-up trajectories has the shortest flight time possible.

It was also observed that for any initial condition, the trajectory where the load factor and the heat flux are the lowest will be one where the lift is kept full lift-up until the lowest point in the trajectory ($\gamma = 0°$). From this point onwards the bank angle can be set to any value that respects $\frac{\mathrm{d}\gamma}{\mathrm{d}t} \geq 0$ without the requirements being violated. This can even be lift-down, which is required to keep $\gamma = 0°$ in case the velocity is larger than the local circular velocity. The situation, where the bank angle is set such that $\frac{\mathrm{d}\gamma}{\mathrm{d}t} = 0$ is also referred to as equilibrium glide. If a horizontal equilibrium glide is flown, the velocity will eventually be low enough such that $\frac{\mathrm{d}\gamma}{\mathrm{d}t} < 0$, even with lift-up. When this occurs, it is expected that sufficient energy has been dissipated such that the

Table 11.3 Total flight time, switch time, and time of lowest point, for a series of lift-up/lift-down and lift-down/lift-up trajectories ($V_E = 13$ km/s). Limit trajectories are indicated in grey

	Lift-up/lift-down			Lift-down/lift-up		
γ (deg)	t_{final} (s)	t_{switch} (s)	$t_{h_{min}}$ (s)	t_{final} (s)	t_{switch} (s)	$t_{h_{min}}$ (s)
−8.4975	**473**	**No switch**	**55.7**	n.a.	n.a.	n.a.
−8.1676	613	83.1664	57.5	476	33.6097	57.1
−7.8377	934	78.1136	60.0	477	39.3661	59.4
−7.5078	1,430	76.8994	62.0	481	44.4769	62.5
−7.1779	1,801	76.9337	64.5	488	49.8055	65.8
−6.8479	1,926	76.8207	67.0	498	55.8791	70.9
−6.5180	1,990	75.6806	69.5	510	63.3699	76.9
−6.1881	2,026	72.2142	71.5	532	73.6102	86.6
−5.8582	2,049	63.6223	109.6	576	90.6982	102.7
−5.5283	n.a.	n.a.	n.a.	**2,061**	**No switch**	**120.5**

peak values of the new descending leg never surpass those of the initial descending leg, as long as the lift is kept upwards. However, this should be investigated on a case-by-case basis and is not a generalisation.

To conclude this brief analysis, Table 11.3 lists the total flight time, switch time, and time to reach the lowest altitude as a function of the initial flight-path angle, γ_E. Both γ_E and t_{switch} have been given with greater accuracy, as it turned out that the trajectories are very sensitive to just very small changes in either parameter. Especially the switch time turned out to be an issue, and the convergence criterion of the bisection algorithm had to be reduced to $t_{switch,i} - t_{switch,i-1} = 0.000025$ s.[12]

Beginning with the lift-up/lift-down trajectories, the total flight time drastically increases with decreasing entry angle (shallower flight). It is required, though, that the lift is switched to a down orientation to prevent skipping flight at an earlier stage. For the steeper entries this can take place after the lowest point in the trajectory has been passed: because of the centrifugal acceleration the vehicle experiences enough upwards "push" to be able to pull up even beyond this lowest point. However, at a certain point in time the upwards directed velocity needs to be compensated for. The shallower the flight-path is, the sooner this has to be initiated, even before the lowest point has been reached for the shallowest trajectories. For most trajectories the vehicle continues in a marginally ascending flight, resulting in very long flight times until the apoapsis altitude is reached. For the shallowest entry this flight time approaches, logically, the limit case of the full lift-down trajectory. Looking at the variation in switch time, which is a relatively small range, considering the large variation in entry angle and corresponding flight times, this demands most likely a very robust guidance and control system to execute such a manoeuvre.

[12] This meant that around the switch time the integration step size was reduced to that value as well, compared to $dt = 0.5$ s for the intervals outside $t_{switch} \pm 1$ s. The fourth-order Runge-Kutta method worked accurately and fast with this larger step size.

The lift-down/lift-up trajectories show a different pattern. To begin with, the hybrid trajectories all last a lot shorter, and the switch time always takes place before the lowest point. Not doing so would prevent the vehicle from pulling up, and would thus lead to a crash on the surface. The shallower the entry conditions, the more back in time this switch is pushed, as there is also more time until the lowest point is reached. Also in this case we see a sensitivity in the trajectory, albeit to the initial flight-path angle. The difference between the last two trajectories is only about 0.3°, yet without a switch the total flight time for the last trajectory is more than 3.5 times larger. It confirms the dependence on a robust guidance system, and a careful planning of the entry conditions.

Step 3: Shallower entry minimises peak values of dynamic pressure (load factor) and heat flux

The premises is that for the least negative flight-path angles (shallow flight) the future dynamic pressure (load factor) and heat flux will be smaller, which, consequently, minimises the corresponding peak values. This point has been proven by Lu (2014); he shows that a shallower flight path implies smaller derivatives of both dynamic pressure and heat flux. By flying either a bang-bang lift-up/lift-down trajectory, or a limit-case lift-up trajectory, the flight path is always the shallowest, and the derivatives of the dynamic pressure and the heat flux are kept minimal. This, combined with the fact that the flight time for the descending leg, where the peak value must occur, is shorter for lift-up trajectories, results in conclusion that indeed the peak value for the dynamic pressure, and thus the heat flux is lower for lift-up trajectories than it is for lift-down trajectories.

11.2.4.3 Minimum Radiative Heat-Load Aerocapture

Up to now, we have considered the optimal minimum ΔV aerocapture trajectory, which requires a flight with full lift-up, followed by full lift-down (Lu et al. 2015). Alternatively, when optimised for minimum integrated convective heat flux, the trajectory should be flown with a bank-angle history of full lift-down, full lift-up (Sigal and Guelman 2001). As is clear, these two objectives demand opposite bank-angle controls. However, there is one additional aspect that should be considered in this context. With increasing velocity during the atmospheric flight, the radiation emitted by the gases in the shock-layer around the vehicle becomes higher and higher. For lunar entry conditions, for example, the total radiative heat load is comparable to the total convective heat load. For even higher velocities, the ratio changes in favour of the radiative heat load: an entry velocity of 16 km/s results in a radiative heat load that is 6 to 12 times larger than the convective heat load. So, when an aerocapture is executed with high velocity, it means that minimisation of radiative heat load becomes much more important than minimisation of the convective heat load.

Zucchelli and Mooij (2018) proved that the integral of (almost) any monomial function of density and velocity is also minimised by a bang-bang trajectory. For the empirical formulations of the radiative heat flux, formulated by Tauber and Sutton

(1991), and Martin (1966), and discussed at length in Sect. 3.6.3, such a trajectory happens to be full lift-up, full lift-down. This corresponds to minimising ΔV, the heat-flux peak, and the load-factor peak! As the solution to this minimisation problem is not (yet) widely published, in the following the derivation steps will be summarised.

Given the monomial function $\mathcal{F} = f(\rho, V)$, where, for the sake of the derivation, we assume that the density follows an exponential profile: $\rho = \rho_0 e^{\beta h}$, with, as before, β is the inverse of the scale height. Starting point is again the formulation of planar motion with respect to a spherical, non-rotating planet, with the equations of motion given by Eqs. (11.63), (11.65), and (11.66). The cost function to be minimised is:

$$\mathcal{J} = \int_{t_0}^{t_f} f(\rho, V)\, \mathrm{d}t \tag{11.79}$$

The corresponding Hamiltonian is:

$$\begin{aligned} \mathcal{H} = f(\rho, V) + \lambda_V \left(-\frac{D}{m} - \frac{\mu}{R^2} \sin\gamma \right) + \\ + \lambda_\gamma \frac{1}{V} \left[\frac{L \cos\sigma}{m} - \frac{\mu}{R^2} \cos\gamma \left(1 - \frac{V^2}{V_c^2} \right) \right] + \lambda_R V \sin\gamma \end{aligned} \tag{11.80}$$

which is the same as Eq. (11.68), apart from the first term that is added on the right-hand side. Using Pontryagin's Maximum Principle gives a similar formulation as before:

$$\dot{\lambda}_V = -\frac{\partial \mathcal{H}}{\partial V} = -\frac{\partial f(\rho, V)}{\partial V} + \lambda_V \frac{\partial D}{\partial V} - \lambda_\gamma \frac{\partial \dot{\gamma}}{\partial V} - \lambda_R \sin\gamma \tag{11.81}$$

$$\dot{\lambda}_\gamma = -\frac{\partial \mathcal{H}}{\partial \gamma} = \lambda_V \frac{\mu \cos\gamma}{R^2} - \lambda_\gamma \frac{\partial \dot{\gamma}}{\partial \gamma} - \lambda_R V \cos\gamma \tag{11.82}$$

$$\dot{\lambda}_R = -\frac{\partial \mathcal{H}}{\partial R} = \rho\beta \frac{\partial f(\rho, V)}{\partial \rho} + \lambda_V \left(\frac{\partial D}{\partial R} - \frac{2\mu \sin\gamma}{R^3} \right) - \lambda_\gamma \frac{\partial \dot{\gamma}}{\partial R} \tag{11.83}$$

As before, the optimal bank angle σ^* that maximises H is:

$$\begin{aligned} \sigma^* = \arg\max_\sigma \left\{ f(\rho, V) + \lambda_V \left(-\frac{D}{m} - \frac{\mu}{R^2} \sin\gamma \right) + \right. \\ \left. \lambda_\gamma \left[\frac{L}{mV} \cos\sigma - \frac{1}{V} \frac{\mu}{R^2} \cos\gamma \left(1 - \frac{V^2}{V_c^2} \right) \right] + \lambda_R V \sin\gamma \right\} \end{aligned} \tag{11.84}$$

and following the same reasoning as before, the solution is provided by Eq. (11.73). Again, the optimal bank angle can only be different from its extrema if a singular arc exists ($\lambda_\gamma \equiv 0$), and it will be shown that also for this new formulation this cannot

happen. For λ_γ to be constant and equal to zero for a finite time, it is required to have $\lambda_\gamma = \dot{\lambda}_\gamma = 0$; all higher derivatives should also be equal to zero. With $\lambda_\gamma = 0$ we had already found that $\lambda_V \frac{\mu}{R^2} = \lambda_R V$, see Eq. (11.74). For $\mathcal{H} \equiv 0$, the existence of a singular arc would imply:

$$f(\rho, V) - \lambda_V \frac{D}{m} - \lambda_V \frac{\mu}{R^2} \sin\gamma + \lambda_R V \sin\gamma = 0 \tag{11.85}$$

or

$$f(\rho, V) - \lambda_V \frac{D}{m} = 0 \tag{11.86}$$

It thus follows for the two non-zero co-states that

$$\lambda_V = \frac{f(\rho, V)\, m}{D} \tag{11.87}$$

$$\lambda_R = m \frac{\mu}{R^2} \frac{f(\rho, V)}{DV} \tag{11.88}$$

A singular arc requires that also all higher derivatives of the co-state variable are zero. Thus, also $\ddot{\lambda}_\gamma = 0$. Taking the time derivative of Eq. (11.82) yields:

$$\ddot{\lambda}_\gamma = \left[\dot{\lambda}_V \frac{\mu}{R^2} - \lambda_V \frac{2\mu}{R^3} \dot{R} - \dot{\lambda}_R V - \lambda_R \dot{V} \right] \cos\gamma - \dot{\gamma} \sin\gamma \frac{\dot{\lambda}_\gamma}{\cos\gamma} = 0 \tag{11.89}$$

where it is known that $\dot{\lambda}_\gamma = 0$, as well as $\cos\gamma \neq 0$. This further simplifies the above equation:

$$\ddot{\lambda}_\gamma = \left[\dot{\lambda}_V \frac{\mu}{R^2} - \lambda_V \frac{2\mu}{R^3} \dot{R} - \dot{\lambda}_R V - \lambda_R \dot{V} \right] = 0 \tag{11.90}$$

Substituting $\dot{\lambda}_V$ and $\dot{\lambda}_R$, Eqs. (11.81) amd (11.83), together with λ_V and λ_R, Eqs. (11.87) and (11.88), and $\dot{V}$ and $\dot{R}$, Eqs. (11.63) and (11.65), results in:

$$\begin{aligned} &-\frac{\partial f(\rho, V)}{\partial V} \frac{\mu}{R^2} - m \frac{\mu^2 f(\rho, V)}{R^4 DV} \sin\gamma + \frac{2 f(\rho, V)}{V} \frac{\mu}{R^2} - m \frac{f(\rho, V)}{D} 2 \frac{\mu}{R^3} V \sin\gamma + \\ &-\frac{\partial f(\rho, V)}{\partial \rho} \beta \rho V + \beta f(\rho, V) V + 2m \frac{f(\rho, V)}{D} \frac{\mu}{R^3} V \sin\gamma + \frac{\mu}{R^2} \frac{f(\rho, V)}{V} + \\ &+ m \frac{\mu^2}{R^4} \frac{f(\rho, V)}{DV} \sin\gamma = 0 \end{aligned} \tag{11.91}$$

After some manipulation, this relation reduces to:

$$\frac{\mu}{R^2}\left(\frac{\partial f(\rho, V)}{\partial V} - 3\frac{f(\rho, V)}{V}\right) + \beta\rho V\left(-\frac{\partial f(\rho, V)}{\partial \rho} + \frac{f(\rho, V)}{\rho}\right) = 0 \quad (11.92)$$

From the above expression, we can derive that for a singular arc to be possible, the following two conditions must hold:

$$\frac{\partial f(\rho, V)}{\partial \rho} = \frac{f(\rho, V)}{\rho} \quad \text{and} \quad \frac{\partial f(\rho, V)}{\partial V} = 3\frac{f(\rho, V)}{V} \quad (11.93)$$

This implies $f(\rho, V) = \rho V^3$, which is not the correct expression for either radiative or convective heat flux.[13]

However, when the general monomial form for the radiative heat flux is assumed (Sect. 3.6.3), i.e.,

$$f(\rho, V) = c\rho^n V^m \quad (11.94)$$

Equation (11.92) changes to:

$$c\left[\frac{\mu}{R^2}\left(3\rho^n V^{m-1} - m\rho^n V^{m-1}\right) + \beta\rho V\left(\rho^{n-1} V^m - n\rho^{n-1} V^m\right)\right] = 0 \quad (11.95)$$

Dividing both sides by $c\rho^n V^{m-1}$ simplifies Eq. (11.95) to:

$$(3 - m)\frac{\mu}{R^2} = (n - 1)\beta V^2 \quad (11.96)$$

Because $\frac{\mu}{R^2} \approx g_0$—and is thus by approximation constant—Eq. (11.96) implies that in that case the velocity will be constant. Obviously, for any unpowered aerocapture flight this cannot be true, which counters the existence of a singular arc. Apart from the monomial function with $n = 1$ and $m = 3$—see above—Eq. (11.96) holds for any combination of n and m. Convective heat flux for a laminar flow is typically represented by $n = 0.5$ and $m = 3$, which has been proven by Sigal and Guelman (2001), and is a specific case of the more general proof shown above.

An additional observation is that, because for aerocapture it holds that $\frac{\mu}{R^2} \ll \beta V^2$, the direct consequence is that changes in m are almost negligible when put to comparison with changes in n. Thus, the sign of Eq. (11.96), and with it the order of commands, only depends on n. If $n > 1$, $\mathcal{J}$ will be minimised by a full lift-up, full lift-down trajectory, and when $n < 1$ the opposite will occur. For instance, when the radiative heat flux is estimated by Martin's approximation (Martin 1966),

$$\dot{q}_{rad} \propto R_n \rho^{1.6} V^{8.5} \quad (11.97)$$

[13] In Sect. 3.6.2, it has been derived that the average local heat flow for an arbitrary point at the wall is proportional to ρV^3.

then it is confirmed that the radiative heat load along the entire trajectory is also minimised by bang-bang control. And, since $n > 1$, the order of commands is full lift-up, full lift-down. When the Tauber-Sutton relation is used, one will reach a similar conclusion.

In case the convective and a radiative fluxes are combined into a single *total* flux, the proof becomes slightly more complex. Instead of dealing with a monomial function, now a polynomial with N terms should be considered. Then, the condition for the existence of a singular arc is:

$$\sum_{i=1}^{N} c_i \rho^{n_i} V^{m_i - 1} \left[(3 - m_i) \frac{\mu}{R^2} + (1 - n_i) \beta V^2 \right] = 0 \tag{11.98}$$

At this point we can only assume that some parametric trajectories $\rho = f(V)$ may exist such that Eq. (11.98) is satisfied, even if it is for just a (short) part of the trajectory. It is hard to believe, though, that this condition can hold for the entire aerocapture trajectory, since it has a descending and an ascending leg. As a consequence, there should be two solutions of velocity for each value of density. However, given a certain function it is always an option to numerically check whether the possible solutions are physical or not. If it is, a singular arc would be allowed, and if not, then the optimal solution will still be given by a bang-bang control.

To summarise, even though nothing can be said about the total heat load, it has been shown that the radiative heat load is minimised by a bang-bang trajectory with full lift-up and full lift-down commands. This is an important conclusion, because radiation is already comparable to convective heat flux at low-speed aerocapture, and becomes the main heat source for aerocapture with large entry velocities, as well as for vehicles with large nose radii: for entry velocities of 16 km/s, the radiative heat load can be as large as 15 times the convective heat load.

11.2.5 Entry Corridor

11.2.5.1 Methodology

In Chap. 5 the concept of the entry corridor was introduced, to show the maximum allowable load limits that an entry system can sustain. When plotting this with the vehicle's actual trajectory in the altitude-velocity plane it helps the mission designer to understand how much manoeuvring freedom remains; if the corridor is very narrow it means that any guidance system should be robust enough to avoid violating the constraints. Alternatively, when the corridor is plotted as a function of critical initial conditions, e.g., flight-path angle and velocity, it gives insight in the targeting conditions of the atmospheric interface to execute, for instance, aerocapture manoeuvres.

In this section the optimal aerocapture manoeuvre will be studied in the form of an entry corridor. Main objective is to establish the range of initial conditions for which

optimal aerocapture can be achieved without violating the imposed constraints, i.e., the combination of mechanical and thermal loads. For arbitrary initial conditions an optimal aerocapture is possible when the trajectory that minimises the loads does not violate the constraints, the trajectory that maximises the skip-out altitude does not undershoot the target altitude, and the trajectory that minimises the skip-out altitude does not overshoot the target.

Three considerations form the basis of studying aerocapture:

1. The peak *g-load* and heat flux occur during the descending leg of the trajectory, which was already established in the discussion of skipping flight (Sect. 5.5),
2. A lift-up configuration minimises the peak *g-load* and heat flux, and maximises the skip-out altitude, and
3. A lift-down orientation maximises the peak *g-load* and heat flux and minimises the skip-out altitude.

While considering this, it may not be hard to understand that for arbitrary initial conditions, optimal aerocapture is only possible if the full lift-up trajectory both respects the constraints and matches or overshoots the target, and the full lift-down trajectory matches or undershoots the target. Failing to meet with any of these conditions means that optimal aerocapture will not be not possible.

To establish the boundaries of the entry corridor, it suffices to determine the initial conditions for which the full lift-up or full lift-down trajectory reaches the target, or for which the loads experienced during the full lift-up trajectory matches the constraints. To provide additional information, besides these four boundaries two other boundaries can be observed to provide more insight. These are the no-skip-out boundary for the lift-up case, i.e., the vehicle remains in the atmosphere for the given initial conditions, and the parabolic skip-out boundary for the lift-down case, where the vehicle is not successfully captured.

It is noted that even by including the lift-down trajectory boundaries it is not guaranteed that when starting from the set of initial conditions contained within the boundaries an optimal trajectory is possible. However, what is certain, though, is that when the mission starts outside of these boundaries it *does* mean that an optimal trajectory is impossible without a ΔV manoeuvre. Such a manoeuvre can either be executed prior to entry or after the vehicle has left the atmosphere again. As an example, consider the intersection of two boundary trajectories, i.e., one flown full lift-down (for instance, an "overshoot target" trajectory) and one flown full lift-up (such as a "heat-flux constraint" trajectory). At this intersection the trajectory has to be flown in two conflicting orientations: full lift-up to prevent violating the constraints and full lift-down to avoid guaranteed overshoot. This obvious conflict demands an impossible trajectory requirement.

With a slight change to the definition of the lift-down boundaries by using the fact that the peak loads occur on the descending leg, this can be remedied, though. The alternative to performing a full lift-down flight, it can be separated into a lift-up flight until the lowest point in the trajectory is reached, after which a switch to a lift-down orientation is made. By flying the entire descending leg lift-up, the thermo-mechanical loads will be minimised, and, consequently, the contradiction

has been removed. In the previous section it was already concluded that the radiative heat load is minimised by a bang-bang trajectory with full lift-up and full lift-down commands, so this adjusted corridor constraint is fully compatible. This version of the entry corridor is hereinafter referred to as the "direct-entry entry corridor", to distinguish it from the altitude-targeting entry corridor. This corridor has applications for direct-entry missions as well, which will be discussed later.

11.2.5.2 Altitude Targeting Entry Corridor

To illustrate the introduced concept, constructing the aerocapture entry corridor will be shown for two different vehicles, i.e., a heavier Crew Return Module (CRM), which is suitable for manned flight, and a much lighter Sample Return Module (SRM). The shape, aerodynamic properties, and mass of both vehicles are based on the Apollo Command Module (CM), see Table 11.4, and for more detail on the aerodynamics also Appendix B, Sect. B.1. Besides studying two different vehicles, also the planetary environment will be varied: the corridors for aerocapture at Earth, Mars, and Venus will be compared. It is noted that the bulk of the results has been published by Engelsma and Mooij (2020), and is reproduced here with permission.

Both vehicles are subject to load-factor and heat-flux constraints. For the trajectories of the CRM a maximum load factor of 12 g is imposed, and for the SRM a maximum load factor of 20 g. For either vehicle a maximum heat flux of 7.95 MW/m^2 is considered.

Figure 11.22 shows the entry corridor for the Earth-CRM case. The earlier introduced constraints have been included, as well as the parabolic skip-out trajectory. Whereas for altitude targeting a full lift-down trajectory will determine the constraint, the parabolic skip-out trajectory is determined by flying the vehicle lift-up until the lowest point in the trajectory, and then switching to lift-down. The maximum reachable altitude is pre-specified for this constraint, and should be above the one for altitude targeting. Encapsulated by the constraints is the upper bound for the entry corridor width, indicated as a blue area in the figure. As was stated before, but repeated here to stress its importance, when the initial conditions are within this area it does *not* automatically mean that the trajectory is optimal, but when the initial conditions are outside this area it prohibits optimality.

To see the effect of a change in planetary environment, Fig. 11.23 shows the entry corridors for Mars and Venus. Besides the entry corridors, the bottom right graph shows the width of the individual entry corridors as a function of initial velocity. This

Table 11.4 Nominal values for the physical properties of the CRM and SRM. The CRM is based directly on the Apollo CM (Robinson et al. 2009)

Mission	Diameter (m)	Aerodynamic reference area (m^2)	Nose radius (m)	Mass (kg)
CRM	3.9116	12.017	4.694	5500
SRM	0.75	0.441	0.90	38.77

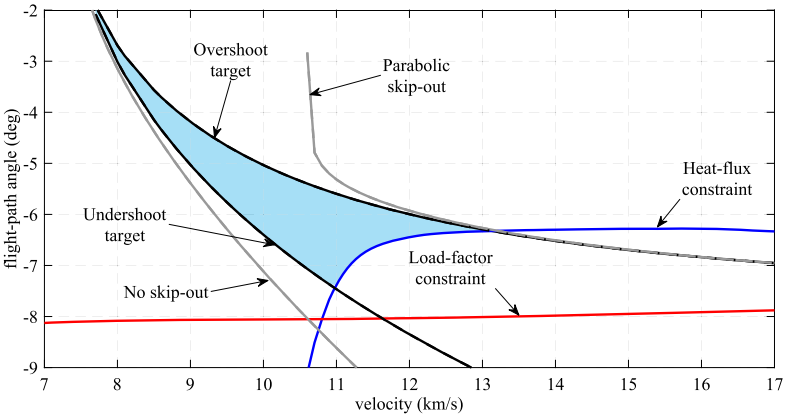


Fig. 11.22 Altitude-targeting entry corridor for the Earth-CRM case, flown with nominal vehicle configuration

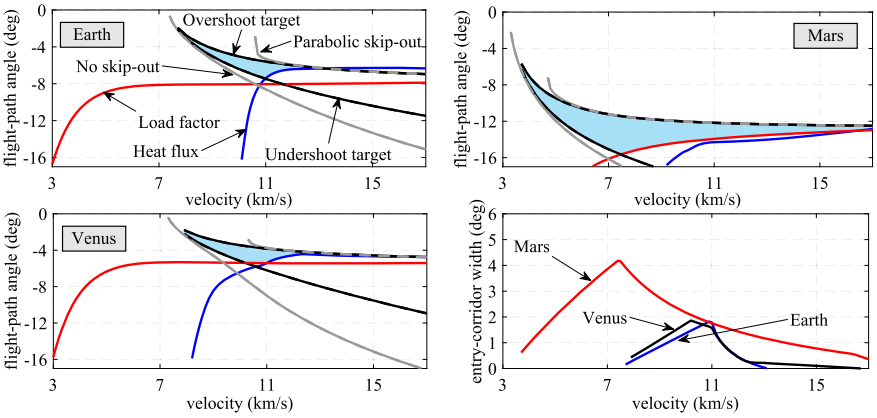


Fig. 11.23 Comparison between the altitude-targeting entry corridors for all three planets, flown with nominal CRM configuration

makes it easier to compare the position and size of the entry corridors. Earth and Venus show comparable corridors when considering velocity and width, but due to the higher density at Venus a shallower entry should be flown when compared to Earth and Mars entry to ensure aerocapture. For Mars the opposite holds due to the thin atmosphere: even though the corridor is significantly broader, a significantly steeper entry is required to prevent overshooting the target or skipping out without being captured. Another noticeable difference is that for Earth and Venus the maximum heat-flux is the major constraint, whereas for Mars it is the maximum *g-load*.

As the spacecraft enters the atmosphere with a low velocity (depicted on the left side of the plots), its flight-path angle boundaries are initially determined by the full

lift-up undershoot and the full lift-down overshoot boundaries. During this phase, the trajectory constraints primarily depend on the spacecraft's flight characteristics. However, as the velocity increases, these constraints gradually transition into being driven by other factors, leading to higher thermo-mechanical loads on the spacecraft.

When the spacecraft enters the Earth's atmosphere, the load factor is never a primary constraint as the heat-flux constraint takes precedence, becoming more demanding with increasing velocity. In contrast, when entering the atmosphere of Mars, the situation is reversed. The peak heat flux does not impose a driving constraint, but the peak load factor becomes the decisive factor in determining the maximum steepness of the entry trajectory. Venus exhibits a combination of both scenarios. At low initial velocities, the undershoot boundary is the primary driving constraint. However, as the spacecraft's velocity increases, it is surpassed by the load-factor boundary, and as the velocity continues to rise, the heat load takes over as the dominant driving constraint.

In summary, the behaviour of spacecraft entering different planetary atmospheres is influenced by various factors such as velocity, load factor, and heat flux, leading to distinct trajectories and thermal-mechanical considerations.

In Fig. 11.24, the entry corridor for the Earth-SRM case is depicted. A direct comparison with Fig. 11.22 immediately reveals that the entry corridor for the SRM is larger than the corridor for the CRM. This difference arises due to relaxed constraints concerning the load factor in the SRM compared to the CRM, and also because the SRM has a smaller nose radius. The reduced nose radius of the SRM leads to an increase in convective heat flux, but a decrease in radiative heat flux. Notably, the earlier discussions have shown that during aerocapture, the radiative heat flux generally surpasses the convective heat flux. As a result, the smaller nose radius in the SRM case leads to an overall lower total heat flux experienced during entry.

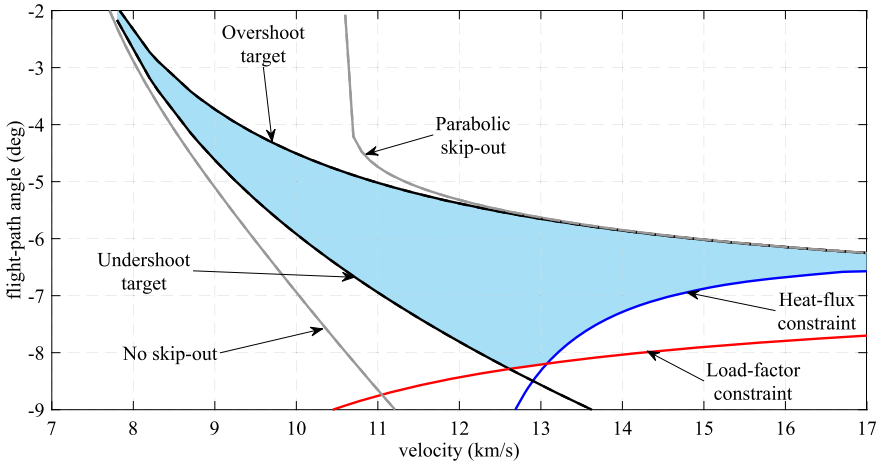


Fig. 11.24 Nominal entry corridor for Earth-SRM case with nominal vehicle configuration

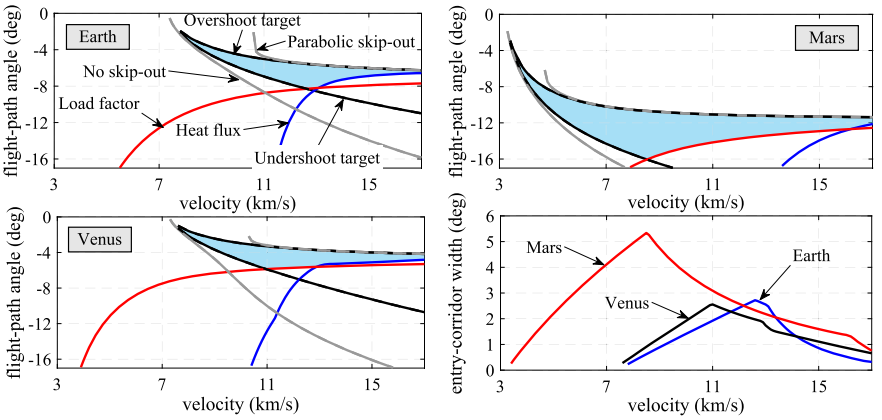


Fig. 11.25 Comparison between the altitude-targeting entry corridors for all three planets, flown with nominal SRM configuration

This characteristic is not only observed when comparing the CRM and SRM cases for Earth, but also holds true for entry scenarios involving Mars and Venus. The influence of the nose radius on heat flux is a consistent trend across these planetary atmospheres as well.

In Fig. 11.25, similar to the CRM case, we compare the entry corridors for the Mars and Venus SRMs with the one for Earth. Once more, we observe that the entry corridors for Earth and Venus exhibit comparable velocity ranges and corridor widths. However, the entry trajectory for Venus requires a shallower entry angle, while the Mars entry corridor is notably larger and demands a steeper entry flight. It is evident from the corridor width that the entry corridors for the SRMs are significantly larger than those for the CRM.

In contrast to the Mars-CRM case, in the Mars-SRM scenario, the peak heat flux becomes a decisive constraint at the highest simulated velocities. This may not be immediately apparent from the top right graph in Fig. 11.25, which illustrates the Mars entry corridor, but the discontinuity observed on the Mars entry corridor width in the bottom right graph clearly reveals this effect.

As for the Venus-SRM case, it follows a similar pattern to the Venus-CRM case, wherein the undershoot boundary initially dominates, followed by the load-factor constraint, and finally, the heat-flux constraint. The primary distinction is that in the SRM case, the load-factor boundary takes precedence over a broader range of velocities.

11.2.5.3 Direct-Entry Entry Corridor

The direct-entry corridor differs slightly from the altitude-targeting entry corridor. Figures 11.26 and 11.27 illustrate the direct-entry corridors for the Earth CRM and SRM cases, respectively. Three notable observations can be made:

1. The undershoot target boundary is absent, as there is no target to undershoot,
2. The no-skip-out boundary is not considered driving (and is shown dashed) since skipping out is not a concern in direct-entry scenarios, and
3. The "Overshoot Target" boundary is renamed to "Overshoot Max Altitude" to indicate its applicability for a skipping entry with a maximum skip-out altitude. In this case, the maximum allowed skip-out altitude is set at the same altitude as the target orbit (500 km).

The direct-entry corridor, as defined with these upper-limit boundaries, guarantees that direct-entry is possible without excessive skip-out and ensures the possibility of optimal altitude-targeting aerocapture. Initial conditions within these boundaries guarantee optimal aerocapture. However, this direct-entry corridor can also be considered conservative, providing guaranteed-aerocapture capabilities, as it excludes a region where aerocapture is potentially possible although not guaranteed. On the other hand, the altitude-targeting entry corridor represents the entire possible region, but without the guarantee that aerocapture is actually possible. Figures 11.28 and 11.29 present the direct-entry corridors for the other two planets (CRM and SRM).

For conditions within these regions, guaranteed aerocapture is possible, as described below. At the maximal skip-out and parabolic skip-out boundaries, the

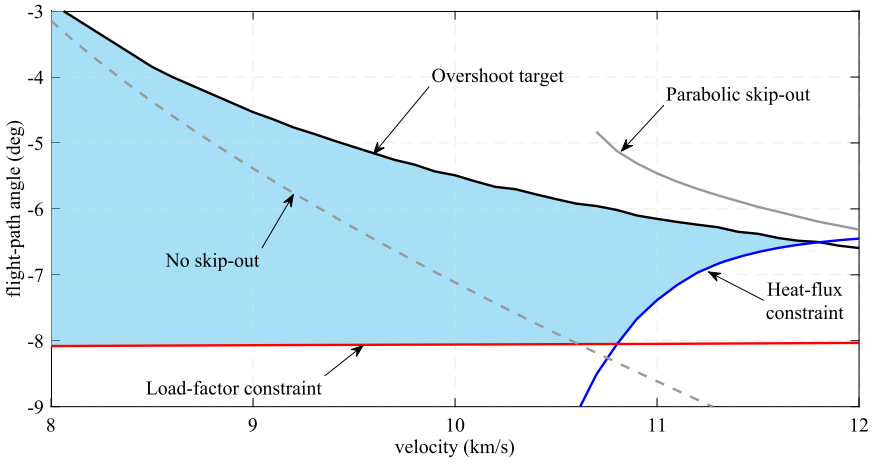


Fig. 11.26 Range-targeting and guaranteed-safety altitude-targeting entry corridor for the Earth-CRM case, flown with nominal vehicle configuration

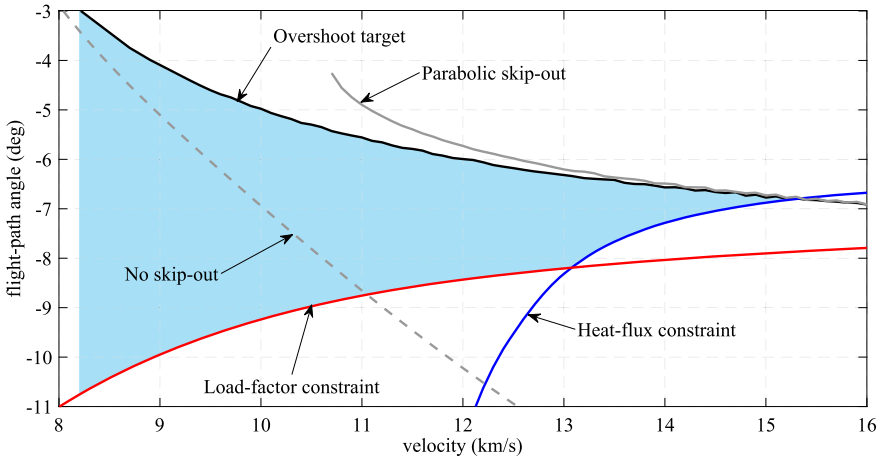


Fig. 11.27 Range-targeting and guaranteed-safety altitude-targeting entry corridor for the Earth-SRM case, flown with nominal vehicle configuration

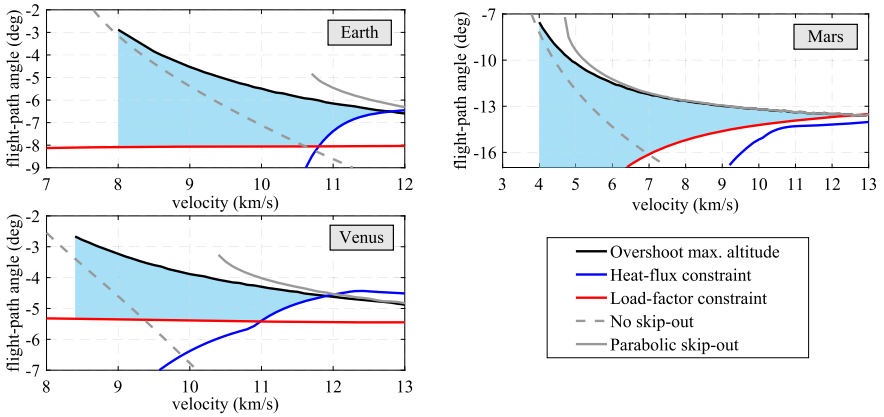


Fig. 11.28 Range-targeting and guaranteed-safety altitude-targeting entry corridor for the CRM for all planets, flown with nominal vehicle configuration

vehicle follows both a descending and ascending leg, ensuring it leaves the atmosphere. During the descending leg, the vehicle is flown lift-up, minimising peak load factor and heat load. Thus, there is no contradiction at the intersection between the upper parabolic and overshoot boundaries, and the lower peak load-factor and heat-flux boundaries.

By combining information from the altitude-targeting and direct-entry corridors, aerocapture feasibility can be determined for any initial state. It can be classified as impossible (outside altitude-targeting corridor), potentially possible (inside altitude-

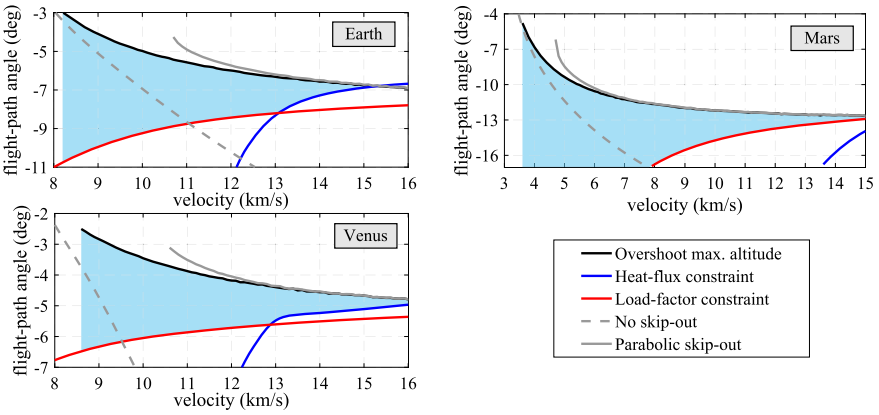


Fig. 11.29 Range-targeting and guaranteed-safety altitude-targeting entry corridor for the SRM for all planets, flown with nominal vehicle configuration

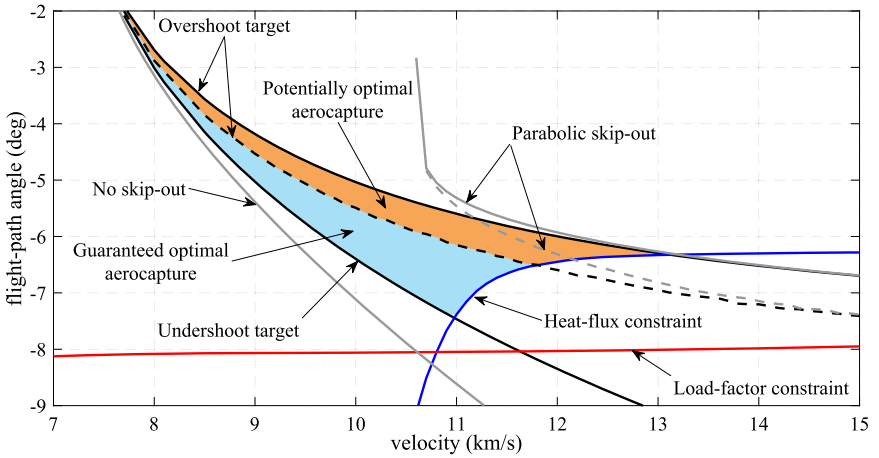


Fig. 11.30 Combined entry corridor showing the region for optimal aerocapture, and the region for which aerocapture is potentially possible

targeting corridor but outside direct-entry entry corridor, requiring optimisation), or guaranteed possible (inside direct-entry entry corridor). Figure 11.30 visualises this, with altitude-targeting and direct-entry entry corridors overlapped, and guaranteed and potentially possible aerocapture regions shaded light blue and orange, respectively.

11.2.6 Entry-Corridor Variations

This section focuses on investigating the impact of different vehicle configurations on the entry corridor. Two specific aspects of the vehicle configuration are considered: the aerodynamic model and the vehicle mass. Since the effects of varying the vehicle configuration are qualitatively similar for all six cases, including both altitude-targeting and direct entry, we will present a detailed analysis of the altitude-targeting Earth-CRM case. However, for the other cases, we will only illustrate the influence on the area bounded by the entry corridor.

To determine the area bounded by the entry corridor, the integration of the region between the upper and lower bounds is performed. By comparing the rates of change of these areas, valuable insights can be derived regarding the sensitivity of the entry corridor to the specific vehicle configuration.

Aerodynamics

The aerodynamic model is varied in two ways: scaling the lift coefficient up and down by 20% and varying the lift-to-drag ratio (L/D) from 0 to 3. Note that because the aerodynamic database, shown in the tables in Sect. B.1, has a variable L/D, scaling the lift coefficient retains the variable L/D nature of this model. The effect of increasing and decreasing the lift coefficient on the entry corridor is visualised in Fig. 11.31.

Increasing the lift coefficient leads to an overall widening of the entry corridor. This is because, with a larger lift coefficient, the vehicle maintains a higher altitude in the atmosphere during lift-up flight. Consequently, this lowers the load factor and heat flux compared to a nominal flight with the same entry angle, enabling a steeper entry trajectory.

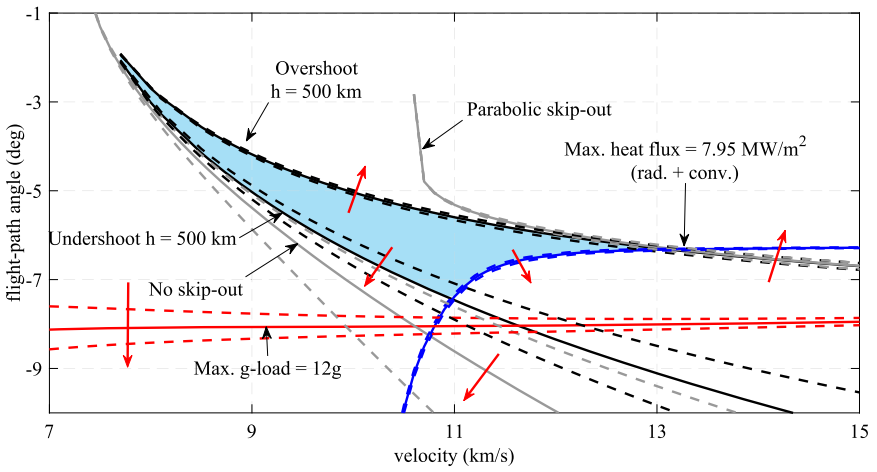


Fig. 11.31 Altitude-targeting entry corridor for the Earth-CRM case, flown with lift variations of vehicle configuration. The red arrows indicate the direction of increasing lift

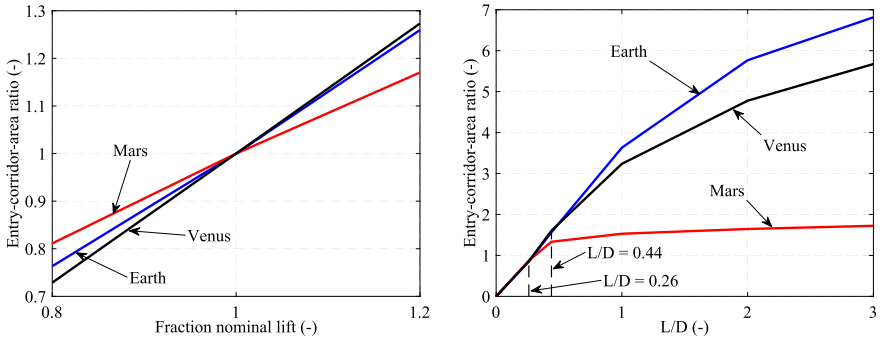


Fig. 11.32 Effect on the altitude-targeting entry corridor area due to variation of lift (left) and L/D (right) for the CRM

As the vehicle operates at higher atmospheric levels, the aerodynamic forces, primarily drag, are reduced compared to the standard scenario, resulting in less velocity loss. As a result, the vehicle can achieve a steeper entry while still having the capability to skip out and reach its intended target. During lift-down flight, with increased lift generated, the vehicle is "pulled" downward, allowing for a shallower entry that still ensures successful capture. Naturally, lowering the lift coefficient produces the opposite effects, narrowing the entry corridor and requiring more carefully-planned entry angles.

The effect on this lift variation on the size of the entry corridor is shown in the left plot of Fig. 11.32 for the CRM, and in the left plot of Fig. 11.33 for the SRM. The value for the entry corridor area in these figures has been normalised with respect to the nominal area.

In addition to the effect scaling C_L, the effect of varying the L/D ratio is also studied. By keeping the drag-coefficient dependency on the velocity and making the corresponding lift-coefficient a multiplication of these values, a fixed L/D profile is obtained. The L/D of the vehicle is varied from $L/D = 0$, a ballistic entry, to $L/D = 3$. In addition, the L/D ratios of 0.26 and 0.44 have also been selected, since they represent the minimum and maximum L/D ratio of the variable L/D profile. The right plots of Figs. 11.32 (CRM) and 11.33 (SRM) show the effect on the entry corridor area.

The presented figures indicate that as the lift-to-drag ratio increases, the entry corridor appears to become less sensitive to further increases in L/D, particularly evident in the Mars cases. However, this observation can be misleading for Mars due to the range of initial conditions imposed when generating the entry corridor. These artificial boundaries constrain the maximum size of the entry-corridor area.

Since the entry corridor for Mars lies close to two of these boundaries (steepest flight-path angle and slowest velocity), its potential for expansion is limited compared to Venus and Earth. For the latter planets, these limits are not reached even at $L/D = 3$, leading to the conclusion that sensitivity to L/D decreases with larger L/D values.

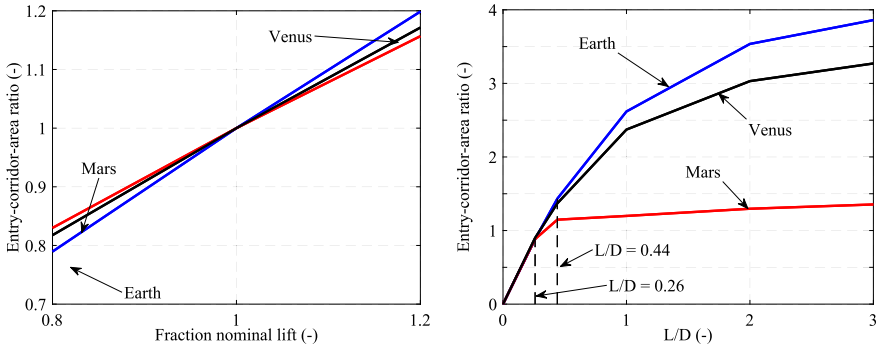


Fig. 11.33 Effect on the altitude-targeting entry corridor area due to variation of lift (left) and L/D (right) for the SRM

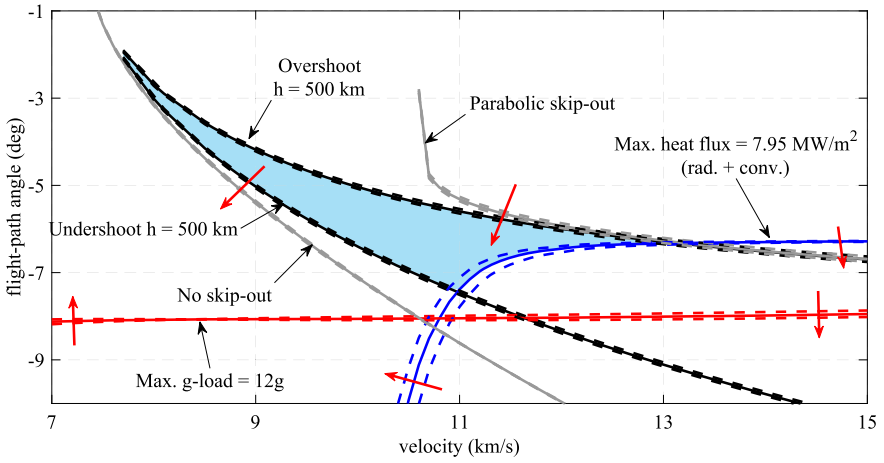


Fig. 11.34 Altitude-targeting entry corridor for the Earth-CRM case, flown with mass variations of vehicle configuration. The red arrows indicate the direction of increasing mass

Consequently, increasing the L/D of a low L/D vehicle has an overall positive effect on the entry corridor's size, but the effectiveness of such increases diminishes for vehicles with high L/D ratios.

Mass

The second investigation focuses on examining the impact of the vehicle's mass on the entry corridor. Similar to the variation in lift, the mass is altered by $\pm 15\%$. Figure 11.34 illustrates the effect of these mass variations on the entry corridor for the Earth-CRM case.

The effect of vehicle mass on the entry corridor differs from that of lift variations. Instead of widening the corridor, mass changes mainly cause a shift in its position. The no-skip, undershoot, overshoot, and parabolic skip boundaries all shift in the

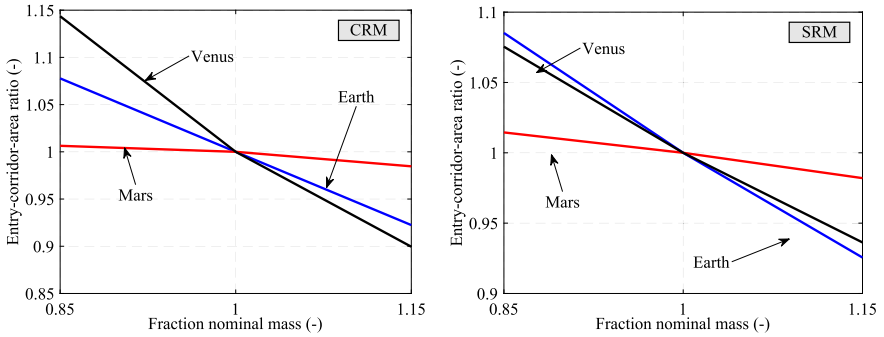


Fig. 11.35 Effect on the altitude-targeting entry corridor area due to variations in the mass of both the CRM (left) and the SRM (right)

same direction, resulting in either the possibility of a steeper entry (for the former two) or a requirement for a steeper entry (for the latter two) with increased mass.

The most significant impact on the entry-corridor area arises from changes in peak heat flux. For velocities relevant to the entry corridor, the peak heat flux increases compared to the nominal configuration. Due to reduced aerodynamic accelerations, the vehicle reaches a lower point in the atmosphere with higher velocity than in the nominal situation.

As the velocity increases, the effect of mass variations on the heat-flux boundary diminishes. At these high velocities, the vehicle's penetration is limited due to the high rate of change of the flight-path angle indicated by the heat-flux boundary, resulting in negligible aerodynamic acceleration.

Figure 11.35 demonstrates the effect of this mass variation on the size of the entry corridor for both CRM and SRM, with the entry-corridor area values normalised with respect to the nominal one. Comparing the entry-corridor ratios in these figures with those in Figs. 11.32 and 11.33, it becomes evident that the entry corridor is less sensitive to changes in mass compared to changes in aerodynamics. However, the sensitivities for individual planets differ significantly more than they did for lift. Nevertheless, it can be concluded that increasing the entry corridor area is more effectively achieved by increasing lift with no change in mass. Even if the vehicle's mass must increase, the overall effect on the entry corridor might still be positive, although the mass increase is counterproductive to the aeromanoeuvre's intended purpose.

11.3 Aerogravity Assist

11.3.1 Introduction

A gravitational slingshot, also known as gravity-assist manoeuvre or swing-by, is the use of a planet's (or any other astronomical object's) relative movement (e.g., orbit around the Sun) and gravity to alter the velocity and direction of a spacecraft. In general, this is done to save propellant and reduce cost. But it can also be the only option, i.e., the spacecraft cannot carry so much fuel to rely on a propulsive ΔV alone, so it has to rely on one or more gravity assists. The change in velocity can be either positive or negative, depending on which side of the planet the spacecraft will pass. The "assist" is provided by the motion of the gravitating body as it pulls on the spacecraft. NASA, on their website,[14] explains the concept using the analogy of a batter hitting a baseball, or a boy throwing a tennis ball at a train moving towards him. Suppose the train is moving with a speed of 80 km per hour, and the boy throwing the ball with a speed of 30 km per hour. The train driver sees the ball moving towards him with a speed of 110 km per hour. Once hitting the train with this speed, it rebounds with this same velocity, but at the same time the train transfers its momentum to the ball. This means that the train's speed has to be added to the ball's speed, which thus rebounds with a speed of up to 190 km per hour. Obviously, this is a mechanical interaction, but the gravity assist is its gravitational equivalent, with the train being replaced by the swing-by planet.

Figure 11.36 explains the concept of the gravity assist. When the velocity of the spacecraft is expressed with respect to the planet itself, it is easily understood that the planet's gravity will only deflect the spacecraft, but not affect its velocity magnitude. However, when an inertial frame is taken as reference, e.g., the heliocentric frame, then the change in velocity is clearly seen. The vectorial addition of the spacecraft's velocity before and after, and the planet's velocity yields a larger velocity after the assist. It could intuitively be explained by throwing a tennis ball at an on-coming train. Suppose we throw the tennis ball with velocity V_{ball}, after bouncing off the train it has the (large) velocity of the train added to its own, and the tennis ball will have a very large velocity!

The gravity assist was first used in 1959, when the Soviet probe Luna-3 was in a slingshot trajectory around the Moon, to take pictures of its far side. It took quite a long time before the next spacecraft executed a gravity assist. It was Pioneer-10 that used Jupiter's gravity in December 1973 to reach escape velocity and be the first artificial object to leave the Solar System. Pioneer-11 followed its "older sibling" and did flybys of Jupiter and Saturn in 1974 and 1979, respectively, to eventually also leave the Solar System. With its flybys of two planets, Pioneer-11 was not the first to do that. Mariner-10, launched at the end of 1973, to study both Mercury and Venus, did four flybys of both planets: Venus once (1974), and Mercury three times (two in 1974 and one in March 1975). However, Pioneer-11 was "pioneering" in another

[14] https://solarsystem.nasa.gov/basics/primer/, date accessed January 4, 2023.

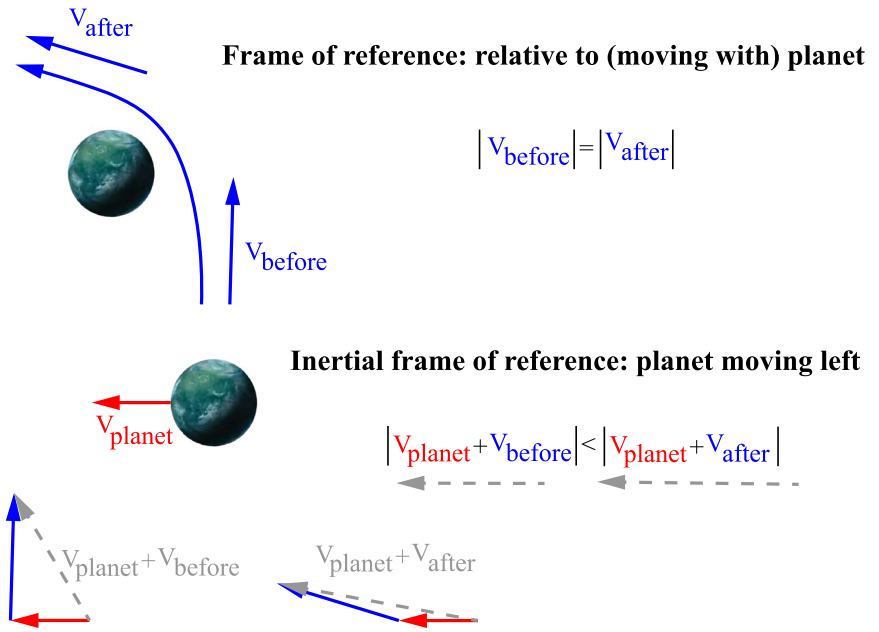


Fig. 11.36 Concept of the gravity assist

way: many elements of the mission proved to be critical in the planning of the Voyager program. The two spacecraft, Voyager-1 and Voyager-2, were both launched in 1977 and made use of a favourable alignment of Jupiter and Saturn to closely fly by them. Voyager-1 made also a flyby of Titan to do atmospheric measurements. On August 25, 2012, it crossed the heliopause and entered interstellar space and was the first spacecraft to do so. As of today, after more than 43 years, Voyager-1 is still active and transmitting data to Earth.

Besides Jupiter and Saturn, Voyager-2 also made flybys of Uranus and Neptune, although the latter two were only decided after launch. In November 2019, it was confirmed that Voyager-2 had entered interstellar space one year earlier. Galileo, the spacecraft sent to Jupiter in 1989, used flybys of Venus and Earth to reach its target in 1995, and became the first spacecraft to orbit the gas giant. An interesting mission to mention is Ulysses, a mission to study the poles of the Sun. Counterintuitively, it was launched in 1990 towards Jupiter to make use of its bulk to change its inclination with respect to the ecliptic of more than 80°, allowing the spacecraft to study all latitudes (including the poles) of the Sun. The MESSENGER mission (launched in August 2004) made extensive use of gravity assists to *slow* its velocity to allow for orbit insertion with available fuel. The mission included one flyby of Earth, two flybys of Venus, and three flybys of Mercury before finally arriving at Mercury in March 2011. The last mission that we will mention for its use of gravity assists is Cassini/Huygens. Launched in 1997, after its Venus-Venus-Earth-Jupiter flybys

it reached the Saturnian system in 2004, and, after releasing the Huygens probe, began its 13-year science mission. The mission ended on September 15, 2017, when Cassini's trajectory intersected with Saturn's upper atmosphere and it burned up.

The *aerogravity assist* (AGA) is a technique to change the hyperbolic orbit of a spacecraft by using a combination of the gravitational attraction and the atmosphere of a planet. The purpose of the AGA is to transform the hyperbolic interplanetary arrival trajectory[15] into a different hyperbolic trajectory by passing once through the atmosphere of the planet. One can use the induced aerodynamic force to compensate for the centripetal acceleration, by directing the lift vector of the spacecraft towards the planet (Casoliva et al. 2008). The effect of atmospheric drag has a negative impact on the trajectory: a percentage (typically in the order of 30%) of the energy of the vehicle is dissipated into heat by aerodynamic friction on the heat shield of the vehicle. However, the advantages of the AGA can outweigh the resulting velocity loss. Velocity increments in the order of 10-15 km/s can be achieved with an AGA at the terrestrial planets (McRonald and Randolph 1992). Through an effective guidance system, the AGA offers significant flexibility in turn angle (Casoliva et al. 2008). Drag losses can be alleviated by the use of a high L/D vehicle, such as the hypersonic waverider. Another advantage of the AGA is that the spacecraft, which produces lift, can also be used for an aerocapture manoeuvre at the destination planet (Bonfiglio et al. 2000).

Unlike the gravity assist, the aerogravity assist has not been used in an actual mission, as it requires substantial investment in technology to realise its potential. The large uncertainty in the atmospheric density remains one of the main risk factors. Currently, the gap between physical conditions (acceleration level, heating rate) and technical capabilities (thermal protection systems) prevents the usage of the AGA. However, the difference between the current and the required situation is such that one can envision that the technology will be developed within a realistic timeframe.

Intermezzo 11.2 (Hyperbolic Trajectory)
In Sect. 4.2 the orbital or Kepler elements were defined. These elements are widely used in the field of astrodynamics, and can define elliptical, parabolic and hyperbolic trajectories. Especially the latter are of interest for the current discussion, hence this very short refresher course. A hyperbolic trajectory, or in short a *hyperbola* (Fig. 11.37), being a conic section, is characterised by an eccentricity, e, that is larger than 1.

Because the radial distance, R, and the semi-latus rectum, p, are always positive, one can write for the true anomaly, θ:

$$\cos\theta > -\frac{1}{e} \tag{11.99}$$

[15] A parabolic orbit is not a usable arrival trajectory, since the atmospheric drag would inherently transform the orbit into an elliptical one, which means that the planet captures the satellite.

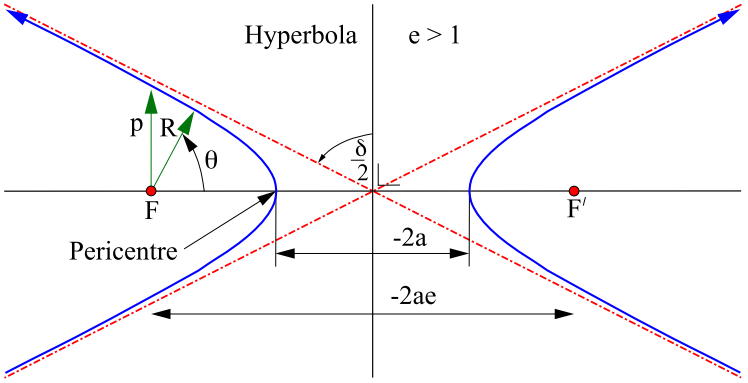


Fig. 11.37 Characterisation of a hyperbolic trajectory

Consider the case that a spacecraft is infinitely far away from a planet (located in the focal point, F), hence $R \to \infty$. In that case, the following equations are valid:

$$\cos\left(\frac{\pi}{2}+\frac{\delta}{2}\right) = -\frac{1}{e} \qquad R_p = a(1-e) \qquad V_\infty^2 = -\frac{\mu}{a} \tag{11.100}$$

with δ being the asymptotic deflection angle, R_p the periapsis distance, a the semi-major axis, V_∞ the spacecraft's hyperbolic velocity, and μ the gravitational parameter of the planet. In case the spacecraft would be in the periapsis of the hyperbola ($R = R_p$), then the corresponding speed, V_p, is:

$$V_p^2 = -\frac{\mu}{a}\left(\frac{e+1}{e-1}\right) \tag{11.101}$$

Using trigonometric relation $\cos\left(\frac{\pi}{2}+\frac{\delta}{2}\right) = -\sin\left(\frac{\delta}{2}\right)$ and Eqs. (11.100) one can find for δ:

$$\sin\left(\frac{\delta}{2}\right) = \frac{1}{e} = \frac{1}{1+\frac{R_p V_\infty^2}{\mu}} \tag{11.102}$$

which will be of particular interest for the discussion of the gravity assist. ■

11.3.2 Gravity Assist

To appreciate the benefits of performing an aerogravity assist, it is instructive to understand the concept of the gravity assist. A gravity assist (GA) is also known as a "swing-by". During a gravity assist, the (relative) velocity *with respect to the planet*

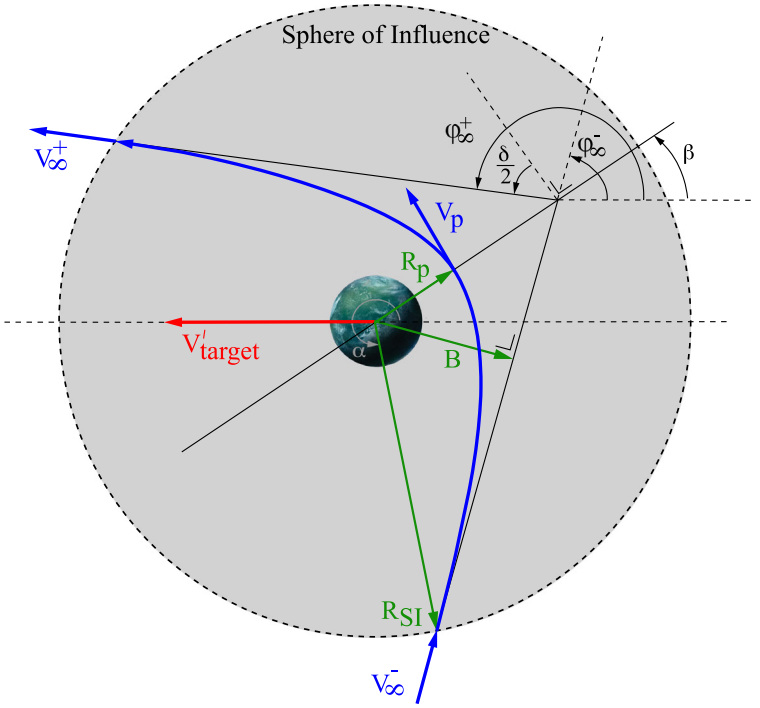


Fig. 11.38 The two-dimensional geometry of a gravity assist

does not change. The inertial (or, more accurately: heliocentric) velocity, however, *does* increase. Consider the geometry of the gravity assist shown in Fig. 11.38. A spacecraft on a heliocentric trajectory is entering the sphere of influence of the planet, i.e., its gravitational pull is dominating the one of the Sun, with a velocity $\mathbf{V}_\infty^-$. Let us define a vector **B**, from the primary body and perpendicular to the incoming asymptote, and use this one and $\mathbf{V}_\infty^-$ to define the so-called trajectory plane, such that we can project the planet's inertial velocity onto this plane, yielding $\mathbf{V}'_{\mathbf{target}}$.

When the spacecraft enters the sphere of influence of the target planet, its position is uniquely described by the radius of the sphere of influence, R_{SI} and the angle of entry, δ. The velocity at that moment is uniquely defined by the magnitude V_∞^- and the angle ξ, see also Fig. 11.39.

With m_s being the mass of the satellite and $\mathbf{V_s}$ its inertial velocity, m_{target} being the mass of the target planet and $\mathbf{V_{target}}$ its inertial velocity we can apply the principle of conservation of momentum to the gravity assist:

$$m_s\mathbf{V}_\mathbf{s}^- + m_{target}\mathbf{V}_\mathbf{target}^- = m_s\mathbf{V}_\mathbf{s}^+ + m_{target}\mathbf{V}_\mathbf{target}^+ \tag{11.103}$$

In the above equation, the subscripts $-$ and $+$ represent initial and final conditions, respectively, to be consistent with the notation used for the aerogravity assist later on

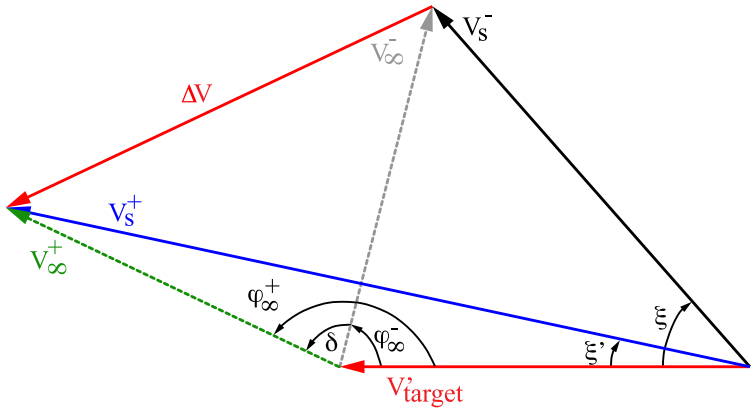


Fig. 11.39 Vector diagram of a gravity assist

in this chapter. As mentioned, we can project the velocities on the trajectory plane, such that scalar expressions can be derived; the velocity will simply have one scalar speed component. Below, we will therefore keep using the term "velocity", even though scalar equations are provided.

Equation (11.103) can be rewritten as the difference of the velocity of the target planet:

$$V^+_{target} - V^-_{target} = \frac{m_s}{m_{target}}(V^-_s - V^+_s) \tag{11.104}$$

When one realises that the mass of the satellite is small (on the order of 10^3 kg) compared to the mass of the target planet (on the order of $10^{24} - 10^{27}$ kg for the Venus/Jupiter range), this means that with $m_s \ll m_{target}$ it is easily understood we get $V^+_{target} - V^-_{target} \approx 0$. Thus, for practical purposes the velocity of the planet remains unchanged.

To analyse the situation of Fig. 11.38 in more detail, a velocity-vector diagram (hodograph) is constructed in Fig. 11.39. Using the same subscripts, V^-_∞ and V^+_∞ (dotted arrows) represent the hyperbolic excess velocities with respect to the planet before and after the gravity assist manoeuvre. The inertial velocity V^-_s of the satellite *before* passing the planet is given by $V^-_\infty + V_{target}$. Although the magnitude of the velocity of the satellite with respect to the planet does not change ($V^-_\infty = V^+_\infty = V_\infty$), its direction *does*. The velocity deflection is represented by the angle δ. Similar to V^-_s, the inertial velocity of the satellite *after* passing the planet, V^+_s, is given by $V^+_\infty + V_{target}$. Therefore, the *inertial* velocity of the satellite has changed in magnitude: $|V^+_s| > |V^-_s|$. We need to consider two cases, though, one where the satellite passes in front of the planet (decrease in inertial velocity) and one where the satellite passes behind the planet (increase in inertial velocity, which is the situation shown Fig. 11.39). Upon closer inspection four special cases can be identified:

1. The inertial speed change $\Delta V = |V_s^+| - |V_s^-|$ is maximal when V_∞^+ exactly aligns with V_{target}, i.e., $\varphi_\infty^+ = 180°$.
2. Every velocity change other than maximal can be obtained with two different deflection angles δ (with V_s^+ mirrored in $V_{target}^{'}$).
3. $\Delta V_{max} = 2|V_{target}^{'}|$ occurs when $\delta = 180°$. In physical terms, this would mean a head-on encounter of the satellite with the target planet.
4. A large enough deflection decreases the inertial speed of the satellite: $|V_s^+| < |V_s^-|$.

The limit case for the gravity assist is defined by its closest distance, R_p, at which the satellite passes the planet, with corresponding velocity, V_p. Recall that the orbital momentum is constant for the entire trajectory about the planet (Fig. 11.38):

$$B V_\infty = R_p V_p \tag{11.105}$$

The periapsis velocity in the hyperbolic trajectory, V_p, is the (vectorial) summation of the planetary escape velocity in the periapsis ($V_{esc,p} = \sqrt{2\frac{\mu_{target}}{R_p}}$) and the satellite's incoming velocity,[16] such that:

$$V_p^2 = V_{esc,p}^2 + V_\infty^2 = 2\frac{\mu_{target}}{R_p} + V_\infty^2 \tag{11.107}$$

where μ_{target} is the gravitational parameter of the target planet. Squaring Eq. (11.105) and substituting the above equation yields:

$$B^2 V_\infty^2 = R_p^2 \left(2\frac{\mu_{target}}{R_p} + V_\infty^2\right) \tag{11.108}$$

from which the periapsis distance, R_p, can easily be solved:

$$R_p = -\frac{\mu_{target}}{V_\infty^2} + \sqrt{\frac{\mu_{target}^2}{V_\infty^4} + B^2} \tag{11.109}$$

Substitution of Eq. (11.109) into Eq. (11.105) yields for the corresponding velocity, V_p:

$$V_p = \frac{\mu_{target}}{B V_\infty}\left(\sqrt{1 + \frac{B^2 V_\infty^4}{\mu_{target}^2}} + 1\right) \tag{11.110}$$

[16] Evaluating the energy balance at infinity (where potential energy is zero) and the periapsis, the *vis-viva* equation provides

$$\mathcal{E}_p = \frac{V_\infty^2}{2} = \frac{V_p^2}{2} - \frac{\mu_{target}}{R_p} \quad \Rightarrow \quad V_p^2 = V_\infty^2 + 2\frac{\mu_{target}}{R_p} \tag{11.106}$$

To avoid impact on the planet, for obvious reasons R_p has to be larger than the radius of the planet, R_t:

$$R_p > R_{target} \quad \Rightarrow \quad \sqrt{\frac{\mu_{target}^2}{V_\infty^4} + B^2} > \frac{V_\infty^2 R_t}{\mu_{target}} + 1 \tag{11.111}$$

If the escape velocity at the surface of the planet is defined to be $V_{esc,s} = \sqrt{2\frac{\mu_{target}}{R_{target}}}$, then the impact condition can be reformulated to get a minimum value for B:

$$B > R_{target}\sqrt{1 + \frac{V_{esc,s}^2}{(V_\infty^-)^2}} = R_{capture} \tag{11.112}$$

or, in words: if the incoming asymptote (distance given by B) crosses a sphere with radius equal to the right-hand side of Eq. (11.111), also known as the *capture radius*, $R_{capture}$, the spacecraft will crash somewhere on the surface.

An expression for the deflection angle δ can be readily obtained by substituting Eq. (11.109) into Eq. (11.102), thereby obtaining:

$$\sin\left(\frac{\delta}{2}\right) = \frac{1}{\sqrt{1 + \frac{B^2 V_\infty^4}{\mu_{target}^2}}} \tag{11.113}$$

Note that the maximum deflection angle easily follows from Eq. (11.102) by substituting $R_p = R_{target}$:

$$\sin\left(\frac{\delta_{max}}{2}\right) = \frac{1}{1 + \frac{R_{target} V_\infty^2}{\mu_{target}}} = \frac{1}{1 + \frac{V_\infty^2}{V_c^2}} \tag{11.114}$$

The angle, β, which defines the orientation of the hyperbola through its periapsis with the direction of the planet's projected motion, $\mathbf{V_t'}$, can be obtained from examining Fig. 11.38:

$$\beta = \varphi_\infty^- + \frac{\delta}{2} - \frac{\pi}{2} \tag{11.115}$$

and, after substitution of Eq. (11.113):

$$\beta = \varphi_\infty^- + \sin^{-1}\left(\frac{1}{\sqrt{1 + \frac{B^2 (V_\infty^-)^4}{\mu^2}}}\right) - \frac{\pi}{2} \tag{11.116}$$

The interplanetary approach trajectory determines the variables V_∞^-, B and β. These quantities, together with the characteristics of the planet $V_{target}^{'}$ and μ_{target}, uniquely determine all other parameters involved in a gravity-assist manoeuvre.

As mentioned at the beginning of this section, a gravity assist alters the heliocentric velocity of the spacecraft. In other words, the *total energy* of the spacecraft changes. This change of total energy per unit mass is equal to:

$$\Delta\mathcal{E} = \frac{1}{2}\left[(V_s^+)^2 - (V_s^-)^2\right] \tag{11.117}$$

Using Fig. 11.39, the velocities in the above equation can be expressed as:

$$(V_s^+)^2 = V_{target}^{'2} + (V_\infty^+)^2 - 2V_{target}^{'} V_\infty^+ \cos\varphi_\infty^+ \tag{11.118}$$

$$(V_s^-)^2 = V_{target}^{'2} + (V_\infty^-)^2 - 2V_{target}^{'} V_\infty^- \cos\varphi_\infty^- \tag{11.119}$$

$$\Delta V^2 = 2V_\infty^2 - 2V_\infty \cos\delta \Rightarrow \Delta V = 2V_\infty \sin\frac{\delta}{2} \tag{11.120}$$

where the angles follow from Fig. 11.38:

$$\varphi_\infty^+ = \frac{\pi}{2} - \frac{\delta}{2} + \beta \quad \text{and} \quad \varphi_\infty^- = \frac{\pi}{2} + \frac{\delta}{2} + \beta \tag{11.121}$$

With these angles and realising that $V_\infty^+ = V_\infty^- = V_\infty$, Eq. (11.117) can be rewritten as:

$$\Delta\mathcal{E} = 2V_{target}^{'} V_\infty \sin\frac{\delta}{2}\cos\beta \tag{11.122}$$

Using Eq. (11.113), the above equation can be reformulated as:

$$\Delta\mathcal{E} = 2\frac{V_{target}^{'} V_\infty \cos\beta}{\sqrt{1 + \frac{B^2 V_\infty^4}{\mu_{target}^2}}} \tag{11.123}$$

Inspecting this result reveals a number of physical implications:

1. The value of $\Delta\mathcal{E}$ is maximal if $V_{target}^{'}$ is maximal (for otherwise constant B, V_∞, β and μ_{target}). This occurs when the gravity assist occurs in the orbital plane of the planet.
2. For a fixed angle between $V_{target}^{'}$ and V_{target}, the planet characteristics V_{target} and μ_{target} determine the value of $\Delta\mathcal{E}$. However, an increase in V_{target} (planet closer to the Sun) is generally accompanied by a decrease in μ_{target} (smaller planet).
3. The angle β determines whether $\Delta\mathcal{E}$ is positive or negative. A passage in front of the planet ($90° < \beta < 270°$) decreases the total energy of the spacecraft, whereas a passage behind the planet ($\beta < 90°$ or $\beta > 270°$) increases the

total energy. The increase is maximal when $\beta = 0°$, which means a head-on encounter.

To obtain a (theoretical) maximum increase in total energy, not only should $\beta = 0°$, but also B should be minimised. The minimum allowable value of B would be the limit value of the impact distance, Eq. (11.111):

$$B = R_{target}\sqrt{1 + \frac{V_{esc,s}^2}{V_\infty^2}} \tag{11.124}$$

Physically, this means that the spacecraft's periapsis would be infinitely close to the surface of the planet. Inserting this limit distance into Eq. (11.123), one obtains:

$$\Delta\mathcal{E}_{max} = 2\frac{V_{target}^{'} V_\infty}{1 + 2\frac{V_\infty^2}{V_{esc,s}^2}} \tag{11.125}$$

The above maximum has been derived for constant (and given) hyperbolic approach velocity, V_∞. It is, of course, justified to ask what happens if V_∞ is optimised. That would, when possible, lead to the "absolute maximum increase" in total energy. Optimising V_∞ can be done by taking the derivative of Eq. (11.125) and equating it to zero, with, as a result: $V_{\infty,opt} = V_{c,s}$, i.e., the local circular velocity at the surface of the planet.[17] Substituting this result back into Eq. (11.125) yields, with $V_{target}^{'} = V_{target}$:

$$\Delta\mathcal{E}_{max,abs} = V_{target} V_{c,s} \tag{11.126}$$

[17] Taking the derivative of $\Delta\mathcal{E}_{max}$ with respect to V_∞ gives

$$2\frac{V_{target}^{'}}{1 + 2\frac{V_\infty^2}{V_{esc,s}^2}} - 8\frac{V_\infty^2}{V_{esc,s}^2}\frac{V_{target}^{'}}{\left(1 + 2\frac{V_\infty^2}{V_{esc,s}^2}\right)^2} = 0$$

This reduces to

$$4\frac{V_\infty^2}{V_{esc,s}^2}\frac{1}{1 + 2\frac{V_\infty^2}{V_{esc,s}^2}} = 1$$

and finally

$$2\frac{V_\infty^2}{V_{esc,s}^2} = 1 \quad \text{or} \quad V_\infty^2 = \frac{1}{2}V_{esc,s}^2$$

With $V_{esc,s} = \sqrt{2\frac{\mu_{target}}{R_{target}}}$ it follows that

$$V_\infty^2 = \frac{\mu_{target}}{R_{target}} = V_{c,s}^2$$

The condition $R_p = R_{target}$ (i.e., the periapsis is at the surface of the planet, which is, of course, a physical impossibility) is called a *grazing orbit*. In the example below, the maximum attainable ΔV and δ for several bodies in the Solar system will be calculated.

Example 11.2 (*Gravity Assists in the Solar System*) In this example several gravity-assist manoeuvres will pass review. To begin with, given the relevant planetary parameters, it is possible to calculate the gain in terms of bending angle and ΔV when a certain body of the Solar System is passed by at a certain distance. From earlier discussions it became clear that the maximum gain can be achieved in a so-called "grazing" orbit, i.e., passing the planet (or moon) at surface level (assuming the absence of an atmosphere). Even though this is not possible in reality, as this would lead to a crash of the satellite, it is instructive to see what will be the gain. Table 11.5 shows an overview of these values. In addition to the "grazing" flyby, results using flyby radii of 2, 5 and 10 times the radius of the planet are included. Note that the approach velocity, V_∞, is taken equal to the approach velocity resulting from a Hohmann transfer orbit. This kind of interplanetary transfer is discussed at length in Intermezzo 11.3.

As can be seen from the results, Jupiter is the most interesting candidate, closely followed by Saturn: the effect of the gravitational parameter μ is clearly seen. Even though the Moon is not a heavy body, relatively speaking, it can still be used as swing-by body, albeit with less efficiency compared to the larger planets. Of all celestial bodies listed, Pluto would be the worst choice as a swing-by target. Both δ and ΔV show a gradual decrease, once the satellite is passing by at a larger distance. This is, of course, a trivial observation, given the physical nature of the gravity assist.

Focussing on Jupiter, Fig. 11.40 shows the distance at closest approach and asymptotic deflection angle as a function of the normalised distance parameter, B. With

Table 11.5 Representative turn angle $\delta = \varphi_\infty^+ - \varphi_\infty^-$ and velocity change ΔV for several gravity-assist manoeuvres. The assumed hyperbolic velocity V_∞ is the approach velocity for a Hohmann transfer orbit (see Table 11.7); R_p denotes the distance of closest approach

Planet	Assumed	$R_p = R_{target}$		$R_p = 2R_{target}$		$R_p = 5R_{target}$		$R_p = 10R_{target}$	
	V_∞ (km/s)	δ (deg)	$\lvert\Delta V\rvert$ (km/s)	δ (deg)	$\lvert\Delta V\rvert$ (km/s)	δ (deg)	$\lvert\Delta V\rvert$ (km/s)	δ (deg)	$\lvert\Delta V\rvert$ (km/s)
Mercury	9.61	10.2	1.71	5.3	0.90	2.2	0.37	1.1	0.19
Venus	2.71	123.3	4.76	103.6	4.25	72.9	3.22	50.0	2.29
Moon	0.83	106.8	1.34	84.2	1.12	53.3	0.75	33.6	0.48
Mars	2.65	79.9	3.40	56.5	2.51	30.7	1.40	17.5	0.81
Jupiter	5.64	158.4	11.09	149.7	10.90	133.1	10.36	115.9	9.57
Saturn	5.44	145.5	10.39	132.1	9.95	108.1	8.81	85.7	7.40
Uranus	4.66	131.7	8.51	114.1	7.82	85.1	6.30	61.4	4.76
Neptune	4.05	141.5	7.64	126.7	7.24	100.9	6.24	77.7	5.08
Pluto	3.70	6.5	0.42	3.3	0.22	1.4	0.09	0.7	0.04

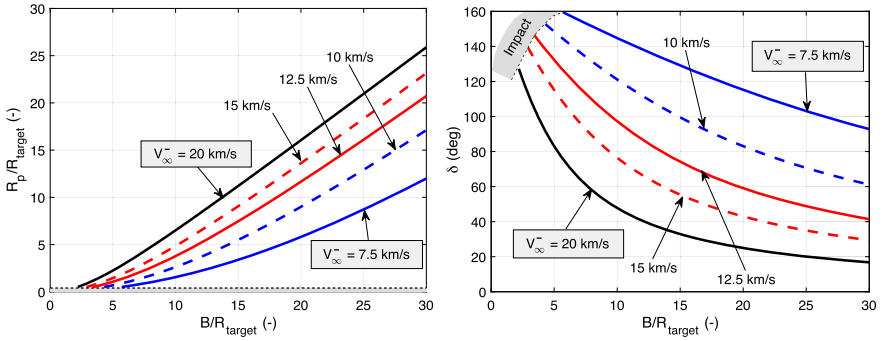


Fig. 11.40 Jupiter gravity assist: distance at closest approach (left) and asymptotic deflection angle (right) for several values of the approach velocity

increasing values of B, the distance at closest approach (Eq. (11.109)) increases as well, as these parameters are obviously linked. Below a certain value, the closest approach will lead to a crash on the planet, indicated by the grey constraint area (Eq. (11.111)). A second dependency is shown on the hyperbolic approach velocity, V_∞^-. At lower velocities the vehicle stays closer to the planet, but will also crash sooner, as it has less energy to conquer the gravity well. The figure on the right shows the asymptotic deflection angle, δ, as defined by Eq. (11.113). Increasing values of B and V_∞^- will both decrease δ. For small values of B, the deflection can be very large, with values well over 100°.

The maximum deflection angle that can be achieved, Eq. (11.114), has been plotted in Fig. 11.41, for several planets in the Solar System. It confirms the notion that with increasing μ the bending angle gets larger, but would decrease when the incoming velocity increases. The plot on the right shows the maximum achievable change in (kinetic) energy. Again, it is Jupiter that can provide the most change, even more than double the one achieved in a fly by of Saturn at high velocity. For the smaller planets there is a distinct maximum in energy gain at a certain value for V_∞^-; this maximum will be reached at very high velocities for Jupiter. For Venus, for instance, the absolute maximum value of $\Delta\mathcal{E}$ is about 257 km^2/s^2, for $V_\infty^- = 7.33$ km/s. Doing the same analysis for Earth, yields $\Delta\mathcal{E} = 235$ km^2/s^2, for $V_\infty^- = 7.91$ km/s, showing that Earth is an almost as good swing-by planet as Venus. There is, however, more to gain by using the atmosphere of the planet as well. This will be discussed in the next section. ■

Example 11.3 (*Simplified Analysis Gravity Assist*) Consider a gravity assist along Mars. Assume a relative velocity when entering the sphere of influence of Mars of $V_\infty = 4$ km/s, and a pericentre distance of 1.1 R_{Mars}. Other data that are relevant for this example are $\mu_{Sun} = 1.3271{\cdot}10^{11}$ km^3/s^2, $\mu_{Mars} = 42{,}832$ km^3/s^2, $R_{Mars} = 3397$ km, and the distance Mars-Sun = 1.52 au, with 1 au = $149.6{\cdot}10^6$ km.

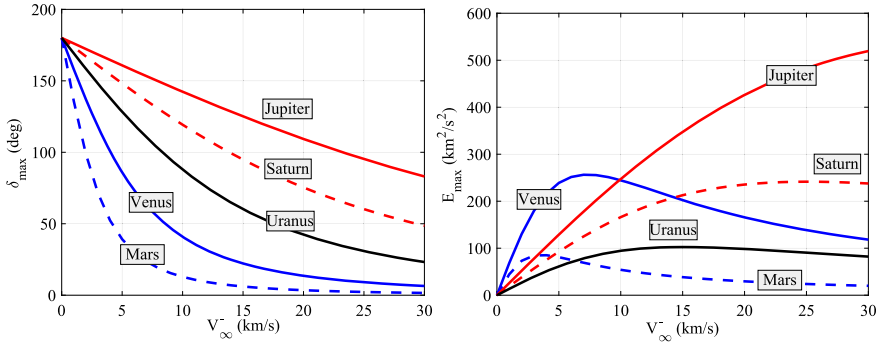


Fig. 11.41 Maximum asymptotic deflection angle (left) and maximum heliocentric orbital energy increase (right) for swing-by missions

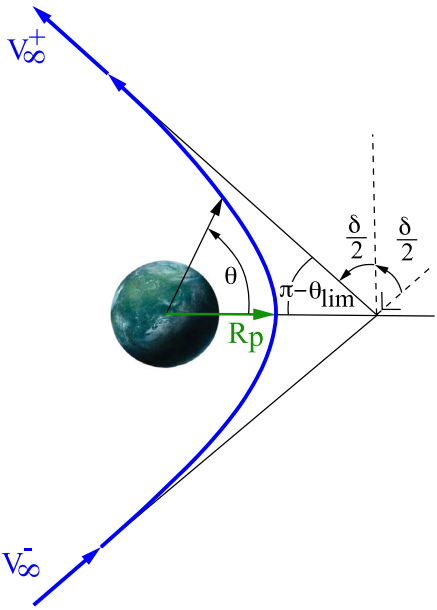


Fig. 11.42 Simplified representation of a gravity assist

To compute the orbital parameters of the hyperbolic orbit, we proceed as follows (see also Fig. 11.42). The semi-major axis follows from the energy equation,

$$\frac{V^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a} \tag{11.127}$$

At infinite distance $r \to \infty$, $V \to V_\infty$, such that $\frac{V_\infty^2}{2} = -\frac{\mu}{2a}$ and therefore

$$a = -\frac{\mu_{Mars}}{V_\infty^2} \tag{11.128}$$

With the given values it is found that $a = -2,677.0$ km. From the definition of pericentre radius,

$$R_p = a(1 - e) \tag{11.129}$$

it then follows that $e = 2.396$. It is noted that for any hyperbolic trajectory, $a < 0$ and $e > 1$. The true-anomaly limit, θ_{lim}, can be derived from the general expression of the conic section:

$$r = \frac{p}{1 + e\cos\theta} \tag{11.130}$$

The maximum radius is $r \to \infty$, which occurs when the denominator of the above equation is zero, or, when

$$\cos\theta_{lim} = -\frac{1}{e} \tag{11.131}$$

With $e = 2.396$ it follows that $\theta_{lim} = 114.67°$. From the figure the relation between the deflection angle and θ_{lim} is easily found:

$$\frac{\delta}{2} = \frac{\pi}{2} - (\pi - \theta_{lim}) \quad \Rightarrow \quad \delta = 2\theta_{lim} - \pi \tag{11.132}$$

Thus, $\delta = 49.34°$.

The final step is to calculate the vehicle's velocity before and after the gravity assist, i.e., V_s^- and V_s^+ (see the vector diagram of Fig. 11.39). For this, we need Mars's heliocentric velocity. Assuming a circular orbit around the Sun, with orbital radius $a = 1.52$ au $= 227.4{\cdot}10^6$ km, the corresponding velocity, $V_{Mars} = V_{target} = \sqrt{\frac{\mu_{Sun}}{a}}$ $= 24.158$ km/s. Thus, $V_s^- = 22.780$ km/s and $V_s^+ = 26.082$ km/s ($\Delta V = 3.302$ km/s). Last, but not least, the energy change is

$$\Delta\mathcal{E} = \frac{(V_s^+)^2}{2} - \frac{(V_s^-)^2}{2} \tag{11.133}$$

yielding $\Delta\mathcal{E} = 80.667$ km^2/s^2. ■

Intermezzo 11.3 (Hohmann Transfer)
To change from one orbit configuration to another, an orbital transfer is required. The most energy (or fuel) efficient way to do this is the so-called *Hohmann transfer*.[18]

[18] Walter Hohmann (18 March 1880–11 March 1945) was a German civil engineer and architect, who, in his free time, studied astronomy and became very interested in the problem of interplanetary spaceflight. With his orbital transfer method he made an important contribution to the field of orbital dynamics.

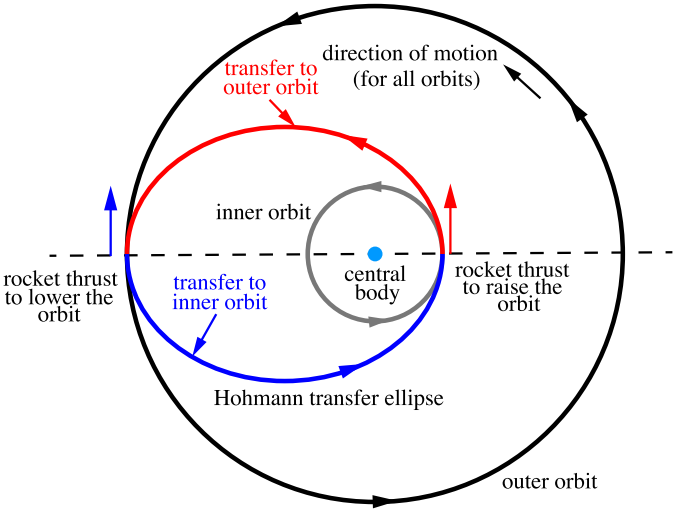


Fig. 11.43 The concept of a Hohmann transfer, i.e., the transfer from a lower to a higher orbit, or vice versa

Even though not the main focus of this chapter, understanding the Hohmann transfer can give us some insight in the approach conditions for gravity assists. Therefore, this intermezzo will briefly introduce the background and a computational scheme. For more details on orbital transfers in general and the Hohmann transfer in particular, the reader is referred to the books of, for instance, Wertz (2009) and Vallado (2013).

Let us consider an orbit transfer between two orbits around an arbitrary central body (Fig. 11.43) that, for the sake of argument, could be Earth. A Hohmann transfer is half of an elliptical orbit between the initial and the final orbit. Such an orbit can go either way, i.e., from the higher to the lower orbit or vice versa. In the former case, the rocket thrust will appear in the apoapsis of the transfer orbit, in the latter case in the periapsis; in either case the transfer orbit tangentially touches the two circular orbits. To execute the transfer two rocket thrusts are required, one to initiate the manoeuvre and to go from the initial circular to the elliptical transfer orbit, and one to circularise the final orbit. We will restrict the discussion to co-planar circular orbits (two-dimensional), but the theory could be generalised to three dimensions as well. The general concept of orbit transfer would also hold for transfers between inclined, elliptical orbits.

With the orbital radius of the initial and final orbit given by r_1 and r_2 (= semi-major axis for a circular orbit), the semi-major axis of the transfer orbit is given by

$$a_{transfer} = \frac{1}{2}(r_1 + r_2) \tag{11.134}$$

In the periapsis of the transfer orbit, the velocity follows from the *vis-viva* equation and is equal to

$$V_{p,transfer} = \sqrt{\mu\left(\frac{2}{r_1} - \frac{1}{a_{transfer}}\right)} \tag{11.135}$$

whereas the local circular velocity is

$$V_{c,1} = \sqrt{\frac{\mu}{r_1}} \tag{11.136}$$

Thus, the required ΔV to initiate the transfer is

$$\Delta V_1 = V_{p,transfer} - V_{c,1} = \sqrt{\mu\left(\frac{2}{r_1} - \frac{1}{a_{transfer}}\right)} - \sqrt{\frac{\mu}{r_1}} \tag{11.137}$$

with μ being the gravitational parameter of the central body.

In a similar fashion one can derive for the impulsive shot in the apoapsis of the transfer orbit:

$$\Delta V_2 = V_{c,2} - V_{a,transfer} = \sqrt{\frac{\mu}{r_2}} - \sqrt{\mu\left(\frac{2}{r_2} - \frac{1}{a_{transfer}}\right)} \tag{11.138}$$

The transfer time is equal to half the orbital period of the elliptical orbit, so

$$T_{transfer} = \pi\sqrt{\frac{a_{transfer}^3}{\mu}} \tag{11.139}$$

As an example we will consider the transfer of a satellite in a parking orbit around Earth to an encounter with the Moon, to be inserted into a low Lunar orbit (LLO). The algorithm as described above can be used, since the Moon is orbiting Earth as well. Assumed is that both the parking orbit and the orbit of the Moon are co-planar and circular. To calculate ΔV_1 and ΔV_2, first the parameters of the problem need to be specified. Assuming a parking-orbit altitude of $h_p = 200$ km (with the "periapsis" at the far side of Earth), the distance Earth-Moon being $2a =$ 384,400 km, and the Lunar-orbit altitude $h_a = 100$ km (with the "apoapsis" at the far side of the Moon), the following values for r_1 and r_2 are found: $r_1 = R_{Earth} + h_p$ $= 6578.1$ km (note that $R_{Earth} = 6378.1$ km is the equatorial radius of Earth), and $r_2 = a + R_{Moon} + h_a = 386{,}238.1$ km ($R_{Moon} = 1738.1$ km is the equatorial radius of the Moon). Doing the calculations, $a_{transfer} = 196{,}400.6$ km, $V_{p,transfer}$ $= 10{,}928.8$ m/s, and $V_{a,transfer} = 185.7$ m/s. However, since the Moon is moving in its orbit with $V_{Moon} = 1018.3$ m/s, the inertial approach velocity to the Moon is $V_{\infty,2} = |V_{a,transfer} - V_{Moon}| = 832.6$ m/s. Total transfer time is $T_{transfer} = 433{,}107$ s ≈ 5 days. Note that to enter the Hohmann transfer, using Eq. (11.137), $\Delta V_1 = 3.132$

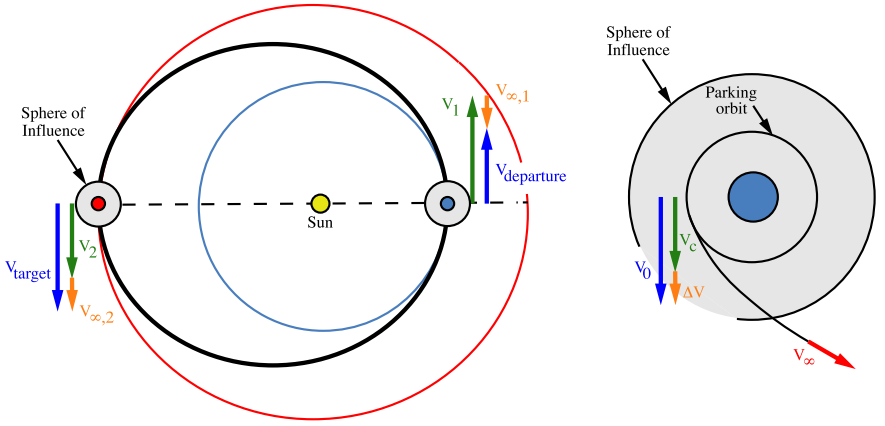


Fig. 11.44 Planetary sphere of influence and related velocities for an interplanetary transfer (explanation in the text)

km/s. To circularise the orbit at the location of the Moon, a second rocket boost yielding $\Delta V_2 = V_{\infty,2} = 0.833$ km/s is required. However, that would mean that the spacecraft is simply "chasing" the Moon. To enter into a lunar orbit, a relative motion around the Moon with $V_c = \sqrt{\frac{\mu_{Moon}}{R_{Moon}+h_{LLO}}} = 1.633$ km/s is required. This velocity has to be added to the earlier found ΔV, so $\Delta V_2 = 2.466$ km/s. Total ΔV to enter the LLO is thus $\Delta V_{total} = 5.597$ km/s.

The Hohmann transfer as described above also works for interplanetary transfers, with one fundamental difference, though. Instead of considering the orbits around a certain central (= planetary) body, the trajectories considered are all heliocentric trajectories, i.e., Sun-centred. Starting point is typically a parking orbit around Earth, but this could be any other planet or moon, of course. Important to realise is, though, that the spacecraft is inside the *sphere of influence* of the central body where the motion is primarily driven by its gravity field, as shown in Fig. 11.44 on the right. On the receiving end, the spacecraft will enter the sphere of influence of the target body (where the motion is determined by the target's gravity field); this means that there are essentially three different phases to consider while studying the interplanetary transfer, with the middle phase driven by the gravity of the Sun. It also means that we have to be aware of how the velocities are defined, i.e., relative with respect to the planet or with respect to the Sun. In the latter case, the velocity of either departure or target body should be taken into account.

Going back to Fig. 11.44, left, one has to deal with three velocity components on either side: the velocity of the departure and target body ($V_{departure}$ and V_{target}), the velocity of the satellite (V_1 at the departure body and V_2 at the target body), and finally the relative velocity change to the so-called excess velocity V_∞. The latter cannot be decoupled from the sphere of influence: zooming in, one can see that in the parking orbit the satellite's velocity will be the local circular velocity $V_{c,1}$,

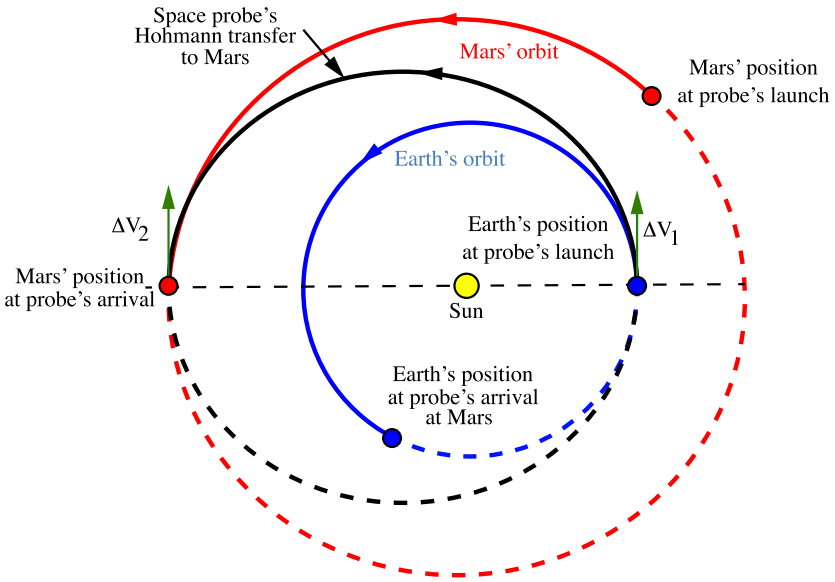


Fig. 11.45 A Hohmann transfer applied in interplanetary flight: example transfer from Earth to Mars

and, similarly when reaching the orbit, this will be indicated by $V_{c,2}$. A change in velocity, ΔV, will bring the spacecraft on a hyperbolic trajectory with velocity V_0, which should be larger than the body's escape velocity, moving towards the edge of the sphere of influence. Some gravity loss will incur, so once the sphere of influence is left, the (hyperbolic) velocity will be V_∞, which is the earlier mentioned hyperbolic excess velocity. Since these velocities are all expressed on a planetocentric scale, the velocity of the central body should be added to express this velocity on a heliocentric scale.

To discuss the Hohmann transfer in between planets, several steps are important to consider: (i) compute the semi-major axis of the transfer orbit (similar to Eq. (11.134), (ii) After executing the manoeuvre, compute V_∞ at the departure body, and calculate the one at the target body (such that the final manoeuvre can be planned), (iii) compute the periapsis velocity of planetocentric hyperbolae, and finally (iv) compute the ΔV on either side of the transfer. The computational scheme will be shown for the classical example of an Earth-Mars transfer, as shown in Fig. 11.45.

To start the calculations, first some velocities are computed, i.e., the velocity of the departure and target planets, Earth and Mars, as well as the local circular velocities. The latter are calculated for a parking-orbit altitude of 200 km, whereas the altitude of the target orbit is 400 km. The gravitational parameter of the Sun is equal to μ_{Sun} $= 1.32712440018 \cdot 10^{11}$ km^3/s^2, whereas other relevant parameters can be found in Table 11.6.[19]

[19] The Hohmann transfer as discussed here, is set up under the assumption of co-planar circular orbits. As one can see from the data in the table (eccentricity and inclination), this is not always

Table 11.6 Solar-System body characteristics[a]

Planet	μ (km^3/s^2)	R_e (km)	V_{esc} (km/s)	a (10^6 km)	e (–)	i (deg)
Mercury	22,031.868551	2,440.5	4.3	57.909	0.2056	7.004
Venus	324,858.592	6,051.8	10.36	108.210	0.0068	3.395
Earth	398,600.435507	6,378.1	11.186	149.598	0.0167	0.0
Moon	4,902.800118	1,738.1	2.38	0.3844	0.0549	18.28–28.58
Mars	42,828.375816	3,396.2	5.03	227.956	0.0935	1.848
Jupiter	126,712,764.1	71,492	59.5	778.479	0.0487	1.304
Saturn	37,940,584.8418	60,268	35.5	1,432.041	0.0520	2.486
Uranus	5,794,556.4	25,559	21.3	2,867.043	0.0469	0.770
Neptune	6,836,527.10058	24,764	23.5	4,514.953	0.0097	1.770
Pluto	975.5	1,188	1.21	5,869.656	0.2444	17.16

[a] Data taken from https://nssdc.gsfc.nasa.gov/planetary/factsheet/, date accessed: December 30, 2022

The heliocentric velocity of the departure and target planet is given by

$$V_{departure} = \sqrt{\frac{\mu_{Sun}}{r_{departure}}} \qquad V_{target} = \sqrt{\frac{\mu_{Sun}}{r_{target}}} \tag{11.140}$$

With $r_{departure} = r_{Earth} = 149.598 \cdot 10^6$ km and $r_{target} = r_{Mars} = 227.956 \cdot 10^6$ km, this yields $V_{departure} = 29.785$ km/s and $V_{target} = 24.130$ km/s. The local circular velocity is defined by

$$V_{c,departure} = \sqrt{\frac{\mu_{departure}}{r_{parking}}} \qquad V_{c,target} = \sqrt{\frac{\mu_{target}}{r_{final}}} \tag{11.141}$$

In this example, $r_{parking} = R_{e,Earth} + h_{parking} = 6378.1 + 200 = 6578.1$ km and $r_{final} = R_{e,Mars} + h_{final} = 3396.2 + 400 = 3796.2$ km. This substituted results in $V_{c,departure} = 7.784$ km/s and $V_{c,target} = 3.359$ km/s.

The characteristics of the transfer orbit, i.e., semi-major axis, $a_{transfer}$, and eccentricity, $e_{transfer}$, are calculated as follows

$$a_{transfer} = \frac{1}{2}(r_{departure} + r_{target}) \qquad e_{transfer} = \frac{|r_{target} - r_{departure}|}{r_{target} - r_{departure}} \tag{11.142}$$

true. However, for the sake of the examples covered in this section, it is justified to do so. It is noted, though, that for a proper analysis one should review these assumptions.

which are found to be $a_{transfer} = 188.777 \cdot 10^6$ km and $e_{transfer} = 0.2075$. The heliocentric velocity at the departure and target positions are, respectively:

$$V_1 = \sqrt{\mu_{Sun}\left(\frac{2}{r_{departure}} - \frac{1}{a_{transfer}}\right)} \qquad V_2 = \sqrt{\mu_{Sun}\left(\frac{2}{r_{target}} - \frac{1}{a_{transfer}}\right)} \tag{11.143}$$

which are, for this example, $V_1 = 32.730$ km/s and $V_2 = 21.479$ km/s. The excess velocity at departure and target planet follows from

$$V_{\infty,1} = |V_1 - V_{departure}| \qquad V_{\infty,2} = |V_2 - V_{target}| \tag{11.144}$$

yielding $V_{\infty,1} = 2.945$ km/s and $V_{\infty,2} = 2.649$ km/s. To calculate the manoeuvre ΔV in the periapsis around the departure and target planet, i.e.,

$$\Delta V_{departure} = |V_{0,departure} - V_{c,departure}| \qquad \Delta V_{target} = |V_{0,target} - V_{c,target}| \tag{11.145}$$

the velocities in the periapsis of the hyperbola around the departure and target planet, $V_{0,departure}$ and $V_{0,target}$, are required:

$$V_{0,departure} = \sqrt{\mu_{departure}\frac{2}{r_{parking}} + V_{\infty,1}^2} \qquad V_{0,target} = \sqrt{\mu_{target}\frac{2}{r_{final}} + V_{\infty,2}^2} \tag{11.146}$$

With the provided and calculated data it follows that $V_{0,departure} = 11.396$ kms and $V_{0,target} = 5.439$ km/s, and thus $\Delta V_{departure} = 3.612$ km/s and $\Delta V_{target} = 2.080$ km/s, yielding a total ΔV of $\Delta V_{total} = 5.692$ km/s. Finally, the total transfer time,

$$T_{transfer} = \pi\sqrt{\frac{a_{transfer}^3}{\mu_{Sun}}} \tag{11.147}$$

results in $T_{transfer} = 258.9$ days.

To conclude this intermezzo, in Table 11.7 some key characteristics of a Hohmann transfer to several bodies in the Solar System are provided. In this table the excess velocity at the target, $V_{\infty,2}$, can be found (to be used in the gravity-assist Example 11.2), as well as the manoeuvre ΔV, and the transfer time. The selected final orbits are circular with an altitude of 400 km for the terrestrial planets and 4,000 km for all the others. For the Moon, the low Lunar orbit of before has been selected ($h_p = 100$ km). ■

Table 11.7 Hohmann transfer from Earth to various bodies in the Solar System

Planet	$V_{\infty,1}$ (km/s)	$V_{\infty,2}$ (km/s)	$\Delta V_{departure}$ (km/s)	ΔV_{target} (km/s)	$T_{transfer}$ (days)
Mercury	7.533	9.612	5.555	7.602	105.5
Venus	2.495	2.707	3.504	3.298	146.1
Moon	–	0.832	3.132	2.466	5.0
Mars	2.945	2.649	3.612	2.080	258.9
Jupiter	8.793	5.643	6.305	17.244	997.7
Saturn	10.296	5.440	7.289	10.492	2,219.7
Uranus	11.280	4.661	7.977	6.341	5,846.8
Neptune	11.656	4.049	8.249	6.759	11,242.0
Pluto	11.810	3.695	8.361	3.312	16,480.0

11.3.3 Atmospheric Passage at Constant Altitude

Now that we have seen the benefits of the gravity assist, and looked at the orbital mechanics behind it, the aerogravity assist will be introduced. To start the discussion, let us look at the fundamental difference between the two manoeuvres. To this end, Fig. 11.46 compares the gravity assist with the aerogravity assist. The GA is now well understood: the gravitational pull of the planet deflects the trajectory and adds velocity to the inertial velocity of the spacecraft. On the right side of the figure, the same layout as for the GA, but now for the AGA is shown. Due to the flight in the atmosphere and the spacecraft following the curvature of the planet, the turn angle of the velocity vector is much larger. Despite the fact that the spacecraft loses velocity due to drag, i.e., the *relative* velocity decreases, the total sum is that the inertial velocity *increases*. This is clearly shown when vectorially adding the velocity vectors.

Figure 11.47 shows two vector diagrams that illustrate the increase or decrease of the heliocentric velocity of the spacecraft, $\mathbf{V_s}$. In terms of notation, the superscripts "$-$" and "$+$" indicate the situation *before* and *after* the AGA. The "∞" subscript on a velocity means that is related to the velocity measured very far from the planet. The drag losses during the AGA manoeuvre is indicated by ΔV_{aero}, which is of course zero for a GA. AGA always increases or decreases the heliocentric velocity, more than GA, because AGA can deflect further the incoming $\mathbf{V}_\infty$. In the figure, this is indicated by the angles δ_{GA} and δ_{AGA}.

The complete picture of the aerogravity assist is now shown in Fig. 11.48, summarising the earlier discussion. As before, the satellite enters the sphere of influence with velocity V_∞^-, and will leave the sphere again with V_∞^+. For reference, the osculating incoming and outgoing velocities have been indicated as well, where the latter would also have been the velocity of a pure gravity assist (dashed lines), as if there had been no contact with the atmosphere of the planet. For both GA trajectories the fictitious deflection angles, $\delta_{GA}(t_i)$ and $\delta_{GA}(t_f)$, are shown. The osculating periapses

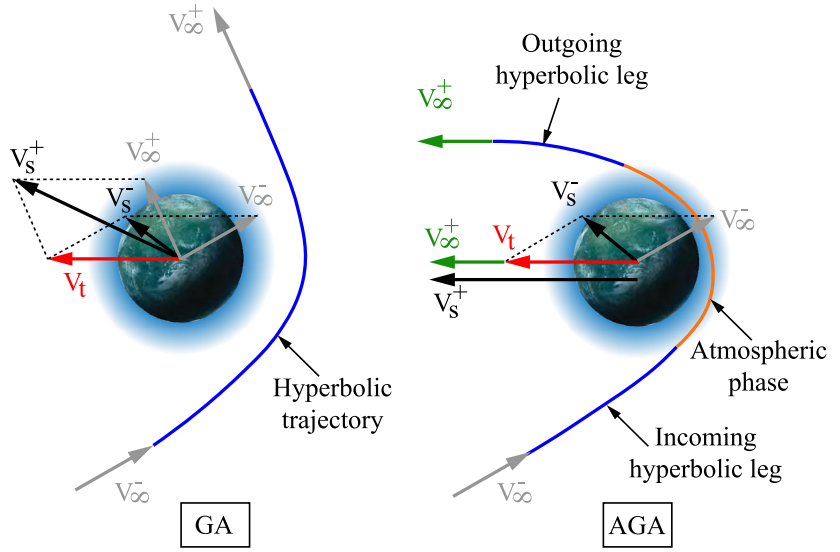


Fig. 11.46 Sketch of AGA manoeuvre compared to gravity assist manoeuvre

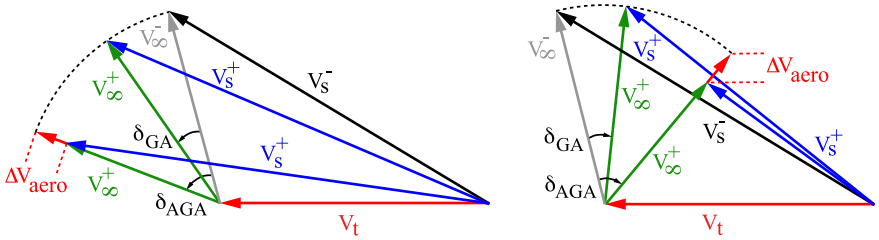


Fig. 11.47 Vector diagrams showing the effect of a gravity assist and AGA manoeuvre on the heliocentric velocity vector of the spacecraft: increase (left) or decrease (right)

$R_p(t_i)$ and $R_p(t_f)$ need not be the same, but they must be inside the atmosphere. The corresponding turn angles, defined as the angle between the position vector of entering/leaving the atmosphere, $R(t_i)$ and $R(t_f)$, and their corresponding periapsis radii is given by $\theta_p(t_i)$ and $\theta_p(t_f)$, respectively. Finally, at the point of entering/leaving the atmosphere the flight-path angles of the AGA trajectory (as well as the osculating trajectory) are given by $\gamma(t_i)$ and $\gamma(t_f)$. Finally, relevant for the discussion of the AGA, are the total (or differential) atmospheric turn angle, $\theta(t_f)$, and the deflection angle, δ_{AGA}. The latter has a practical limit of 180°.

The mathematical discussion of the aerogravity assist begins, as so many times before, with re-introducing the model for planar motion (Chap. 5):

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{11.148}$$

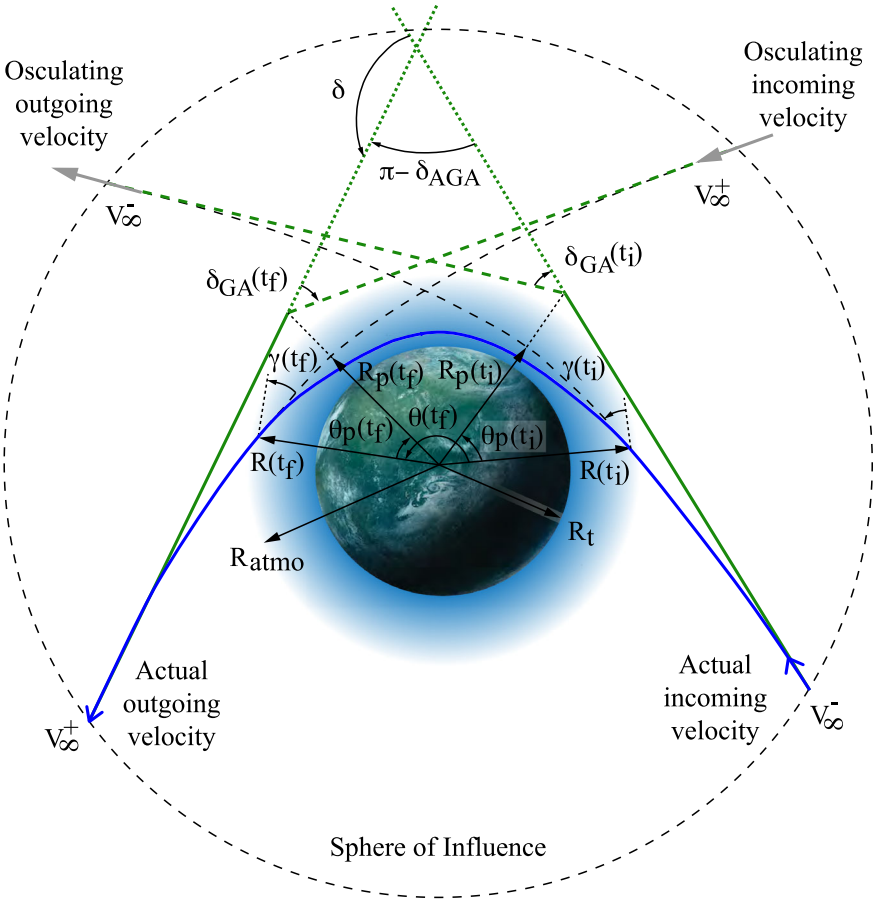


Fig. 11.48 Geometry of the aerogravity assist

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.149}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} = V\sin\gamma \tag{11.150}$$

To calculate the distance covered when an equatorial entry is considered, the time derivative of the longitude is added to this set:

$$\dot{\tau} = \frac{V\cos\gamma}{R} \tag{11.151}$$

The AGA will be analysed in two steps, i.e., a flight at constant altitude in the current section, and the more optimal flight when altitude is not constant in the final one. In their paper, Lewis and McRonald (1992) studied the design of hypersonic waveriders for aeroassisted interplanetary trajectories. As an assumption they looked at flights at constant altitude, stemming from the notion that minimum energy loss occurs then, where the required lift corresponds to the maximum $\frac{L}{D}$-ratio of the vehicle. As a consequence of the constant-altitude assumption ($\gamma = 0$), the atmospheric density, ρ, and gravitational acceleration, g, are therefore also constant.

11.3.3.1 Atmospheric Bending Angle

The dynamics model given by Eqs. (11.148) and (11.149) can be simplified (see also Fig. 11.49), but starting with arbitrary lift and drag:

$$\dot{V} = -\frac{D}{m} = -\frac{1}{2}\rho\frac{SC_D}{m}V^2 \tag{11.152}$$

$$\dot{\gamma} = 0 \quad \Rightarrow \quad \frac{V^2}{R} = g - \frac{L}{m} = g - \frac{1}{2}\rho\frac{SC_L}{m}V^2 \tag{11.153}$$

where, as before, $L = \frac{1}{2}\rho SC_L V^2$, $D = \frac{1}{2}\rho SC_D V^2$ (with C_L and C_D being the lift and drag coefficient of the spacecraft, depending on its geometrical configuration), S is the cross sectional area of the satellite, and $\frac{1}{2}\rho V^2 = \bar{q}$ is the dynamic pressure.

The deceleration experienced by a spacecraft parallel to the velocity vector can be written as:

$$\mathrm{d}V = -\frac{D}{m}\mathrm{d}t \tag{11.154}$$

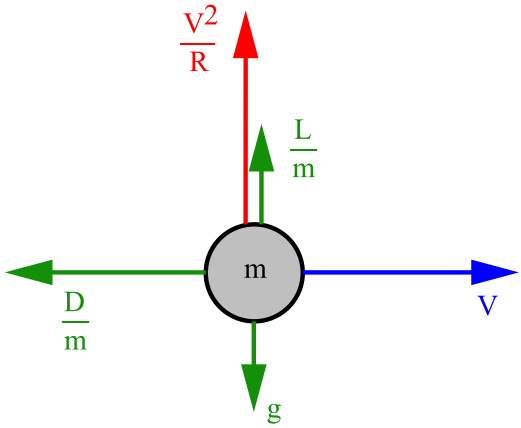


Fig. 11.49 Kinetic diagram of the atmospheric phase of an AGA with constant-altitude assumption

An important aerodynamic performance parameter of the spacecraft is the lift-to-drag ratio, L/D. Introducing this in Eq. (11.154) yields:

$$\mathrm{d}V = -\frac{L}{(L/D)\,m}\mathrm{d}t \tag{11.155}$$

The time of flight can be related to the arc length, θ, traveled at radius R in the atmosphere, since $V\mathrm{d}t = R\mathrm{d}\theta$ (this is identical to the change of longitude, Eq. (11.151), with $\gamma = 0°$). Thus

$$\mathrm{d}V = -\frac{L}{(L/D)\,m}\frac{R}{V}\mathrm{d}\theta \tag{11.156}$$

Note that this arc length is the same as the (differential) turn angle indicated in Fig. 11.48.

By pointing the lift vector towards the planet, the spacecraft can remain in the sphere of influence of the planet for an extended period of time: typically in the order of minutes. Therefore, an AGA can provide a larger deflection angle than a gravity assist (as the spacecraft will be following the curvature of the planet). Under the assumption of constant-altitude atmospheric flight, there must be a vertical force equilibrium between the downward "lift", the gravity force mg and the centrifugal force:

$$L = mg - \frac{mV^2}{R} \tag{11.157}$$

Substituting this, with $g = \frac{\mu_{target}}{R}$, into Eq. (11.156) results in

$$\mathrm{d}V = -\frac{mg - \frac{mV^2}{R}}{m}\frac{D}{L}\frac{R}{V}\mathrm{d}\theta = -\frac{D}{L}\left(V - \frac{\mu_{target}}{RV}\right)\mathrm{d}\theta \tag{11.158}$$

and, after rearranging:

$$\mathrm{d}\theta = -\frac{L}{D}\left(\frac{\mathrm{d}V}{V - \frac{\mu_{target}}{RV}}\right) \tag{11.159}$$

To facilitate the next step, this expression can be simplified by introducing the local circular velocity, $V_c = \sqrt{\frac{\mu_{target}}{R}}$ and some rearranging:

$$\mathrm{d}\theta = -\frac{L}{D}\frac{1}{V}\frac{\mathrm{d}V}{1 - \frac{V_c^2}{V^2}} \tag{11.160}$$

Integrating both sides, thereby assuming L/D to be constant,[20] yields a simple expression for the turn angle:

$$\theta = -\frac{L}{D}\int_{V_{s,i}}^{V_{s,f}} \frac{1}{V}\frac{\mathrm{d}V}{1-\frac{V_c^2}{V^2}} = -\frac{1}{2}\frac{L}{D}\ln\left(\frac{V_{s,f}^2 - V_c^2}{V_{s,i}^2 - V_c^2}\right) \tag{11.161}$$

or, with the introduction of the normalised velocity $\bar{V} = \frac{V_s}{V_c}$:

$$\theta = -\frac{1}{2}\frac{L}{D}\ln\left(\frac{1-\bar{V}_{s,f}^2}{1-\bar{V}_{s,i}^2}\right) \tag{11.162}$$

This normalisation is often done in AGA literature, and has also been done, in fact, in Sect. 5.4.4, where the flight-range expressions to establish the footprint for gliding flight were derived. Note that θ is the turn angle for the constant-altitude phase. The angle coverage during the descending and ascending legs should be added to this to get the complete atmospheric-turn angle, but this effect is usually small due to the steepness of the legs.

The atmospheric exit velocity, $V_{s,f}$, can be written in terms of the atmospheric entrance velocity, $V_{s,i}$ by rearranging terms in Eq. (11.161):

$$\ln\left(\frac{V_{s,f}^2 - V_c^2}{V_{s,i}^2 - V_c^2}\right) = -\frac{2\theta}{L/D} \quad \Rightarrow \tag{11.163}$$

$$\frac{V_{s,f}^2 - V_c^2}{V_{s,i}^2 - V_c^2} = e^{-\frac{2\theta}{L/D}} \quad \Rightarrow \tag{11.164}$$

$$V_{s,f}^2 = e^{-\frac{2\theta}{L/D}}(V_{s,i}^2 - V_c^2) + V_c^2 \quad \Rightarrow \tag{11.165}$$

$$V_{s,f} = \sqrt{e^{-\frac{2\theta}{L/D}}V_{s,i}^2 - (e^{-\frac{2\theta}{L/D}} - 1)V_c^2} \tag{11.166}$$

11.3.3.2 Total Bending Angle

To arrive at the total bending angle achieved by an AGA, the effect of a pure gravity assist has to be added to the atmospheric bending angle, θ, see also Fig. 11.50 that is a simplified version of Fig. 11.48. In this figure, the bending angle θ is taken as the result of a flight at constant altitude ($R = R_t + h$), where the descending and ascending legs have been ignored for the atmospheric flight. Continuing from Eq. (11.166), for

[20] In Intermezzo 11.4 the cases where either L/D is not constant, or when $L/D = L/D_{max}$ are treated.

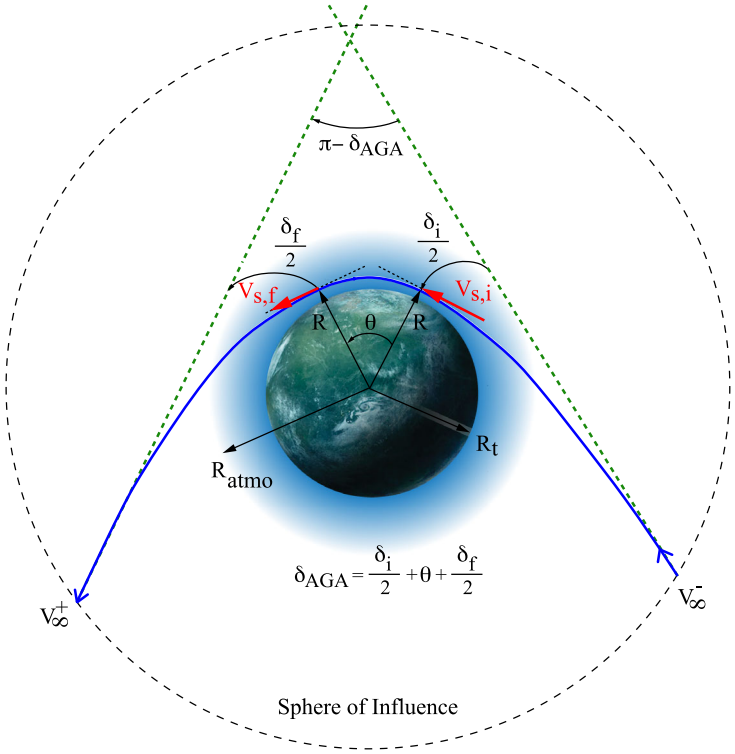


Fig. 11.50 Total bending angle for an aerogravity assist, consisting of three components δ_i, θ, and δ_f

the interplanetary mission we want to relate $V_{s,i}$ and $V_{s,f}$ to V_∞. A spacecraft with an incoming relative velocity V_∞ approaches the gravitational source, e.g., a planet, and accelerates to velocity

$$V(R) = \sqrt{V_\infty^2 + 2\frac{\mu_{target}}{R}} \tag{11.167}$$

in line with Eq. (11.107). With this equation, the incoming and outgoing (hyperbolic) velocities may be written as

$$V_{s,i}^2 = (V_\infty^-)^2 + 2\frac{\mu_{target}}{R} \qquad V_{s,f}^2 = (V_\infty^+)^2 + 2\frac{\mu_{target}}{R} \tag{11.168}$$

where it is noted that the velocity loss on the descending and ascending leg, i.e., the portions of flight deviating from the circular atmospheric flight, has been ignored. The above expressions are now substituted into Eq. (11.166), yielding

$$(V_\infty^+)^2 + 2V_c^2 = e^{-\frac{2\theta}{L/D}}\left[(V_\infty^-)^2 + 2V_c^2\right] - (e^{-\frac{2\theta}{L/D}} - 1)V_c^2 \tag{11.169}$$

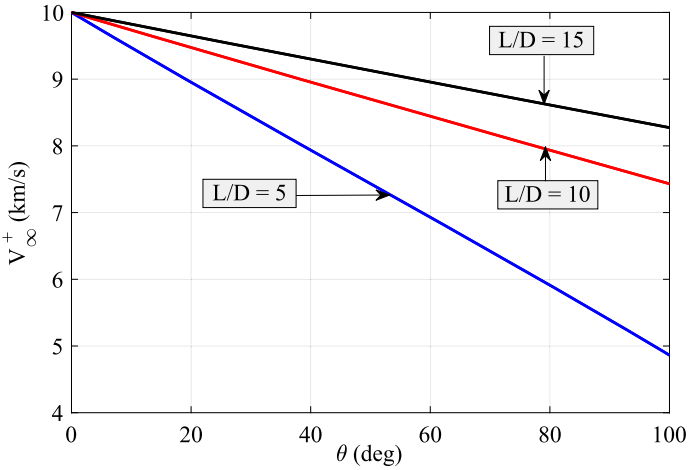


Fig. 11.51 The hyperbolic excess velocity as a function of the aerodynamic turn angle for a Venus aerogravity assist at $h = 150$ km with $V_\infty^+ = 10$ km/s

where we have used the notation $V_c^2 = \frac{\mu_{target}}{R}$ for convenience. Isolating V_∞^+ on the left side results in

$$(V_\infty^+)^2 = e^{-\frac{2\theta}{L/D}}\left[(V_\infty^-)^2 + 2V_c^2\right] - (e^{-\frac{2\theta}{L/D}} - 1)V_c^2 - 2V_c^2 \tag{11.170}$$

such that, after rearranging terms:

$$V_\infty^+ = \sqrt{e^{-\frac{2\theta}{L/D}}(V_\infty^-)^2 + (e^{-\frac{2\theta}{L/D}} - 1)V_c^2} \tag{11.171}$$

This expression gives the hyperbolic excess velocity as a function of the aerodynamic turn angle, and has been visualised in Fig. 11.51 for an approach of Venus with $V_\infty^- = 10$ km/s and passing through the atmosphere at $h = 150$ km. It is clear that for (relatively) low L/D the reduction in hyperbolic excess velocity can be substantial.

Concerning the bending angles δ_i and δ_f, it is clear from the notion $V_\infty^+ < V_\infty^-$ that $\delta_f > \delta_i$. An expression for δ_f can be derived in terms of V_∞^- and θ, starting from Eq. (11.102):

$$\sin\left(\frac{\delta_f}{2}\right) = \frac{1}{1 + \frac{(V_\infty^+)^2}{V_c^2}} \tag{11.172}$$

Rearranging terms in Eq. (11.171) provides us with

$$1 + \frac{(V_\infty^+)^2}{V_c^2} = e^{-\frac{2\theta}{L/D}}\left[\frac{(V_\infty^-)^2}{V_c^2} + 1\right] \tag{11.173}$$

which, after substitution into Eq. (11.197), leads to

$$\sin\left(\frac{\delta_f}{2}\right) = \frac{e^{\frac{2\theta}{L/D}}}{1+\frac{(V_\infty^-)^2}{V_c^2}} \tag{11.174}$$

Realising that

$$\sin\left(\frac{\delta_i}{2}\right) = \frac{1}{1+\frac{(V_\infty^-)^2}{V_c^2}} \tag{11.175}$$

Equation (11.200) can thus also be written as

$$\sin\left(\frac{\delta_f}{2}\right) = e^{\frac{2\theta}{L/D}} \sin\left(\frac{\delta_i}{2}\right) \tag{11.176}$$

With these results, one can express the total turn angle with respect to the hyperbolic asymptotes, Φ, as follows:

$$\Phi = \delta_{AGA} = \arcsin\left(\frac{\delta_i}{2}\right) + \theta + \arcsin\left(\frac{\delta_f}{2}\right) \tag{11.177}$$

and, with the results of Eqs. (11.200), and (11.175) and keeping θ as a dependent parameter, this results in

$$\Phi = \delta_{AGA} = 2\arcsin\left(\frac{1}{1+\frac{(V_\infty^-)^2}{V_c^2}}\right) + \theta + 2\arcsin\left(\frac{e^{\frac{2\theta}{L/D}}}{1+\frac{(V_\infty^-)^2}{V_c^2}}\right) \tag{11.178}$$

Written with the normalised velocity $\bar{V} = +\frac{V_\infty^-}{V_c}$:

$$\Phi = 2\arcsin\left(\frac{1}{1+\bar{V}^2}\right) + \theta + 2\arcsin\left(\frac{e^{\frac{2\theta}{L/D}}}{1+\bar{V}^2}\right) \tag{11.179}$$

a result that is also stated by, for instance, Lewis and McRonald (1992) and Sims et al. (2000).

11.3.3.3 Alternative Formulations for the Bending Angle

Casoliva et al. (2008) provide a relatively simple approach to derive the total bending angle, based on the geometry shown in Fig. 11.48. For an AGA trajectory, the bending angle δ_{AGA} is the angle between the two asymptotic velocities $\mathbf{V}_\infty^-$ and $\mathbf{V}_\infty^+$:

$$\delta_{AGA} = \arccos\left(\frac{\langle \mathbf{V}_\infty^-, \mathbf{V}_\infty^+ \rangle}{\|\mathbf{V}_\infty^-\|\|\mathbf{V}_\infty^+\|}\right) \tag{11.180}$$

with $\langle .., .. \rangle$ being the inner product. The angle δ_{AGA} would normally a mission design parameter with which one can specify either (or both) velocities; evaluation of this angle is done when $R(t_f) = R_{atmo}$.

Finally, the total bending angle for the AGA trajectory can also be defined as

$$\delta_{AGA} = \frac{\delta_{GA}(t_i)}{2} + \theta_p(t_i) + \theta(t_f) + \frac{\delta_{GA}(t_f)}{2} - \theta_p(t_f) \tag{11.181}$$

with $\theta(t_f)$ being the total atmospheric turning angle when the spacecraft leaves the atmosphere, and where the initial atmospheric angle is zero, i.e., $\theta(t_i) = 0$. The total gravity-assist bending angles at the entry and exit of the AGA hyperbolic trajectory legs are

$$\delta_{GA}(t_i) = 2\arcsin\left[\frac{1}{e(t_i)}\right] \qquad \delta_{GA}(t_f) = 2\arcsin\left[\frac{1}{e(t_f)}\right] \tag{11.182}$$

where $e(t_i)$ and $e(t_f)$ are the eccentricities at the initial and final AGA hyperbolic trajectory legs, respectively. These eccentricities are defined as a function of velocities, normalised by the local circular velocity. To avoid repeating equations, a general notation is introduced for both the initial and the final conditions that will be indicated with a subscript j ($j = i$ or f). In this general form, the normalised velocity is defined as

$$\bar{V}_j = \frac{V(t_j)}{V_{c,j}} \quad \text{with} \quad V_{c,j} = \sqrt{\frac{\mu_{target}}{R(t_j)}} \tag{11.183}$$

Using this, the eccentricities can be expressed as

$$e(t_j) = \sqrt{\bar{V}^2(t_j)\left[\bar{V}^2(t_j) - 2\right]\cos^2\gamma(t_j) + 1} \tag{11.184}$$

and the small parts of δ_{GA} that enter inside the atmosphere of the planet, $\theta_p(t_j)$, are given by

$$\theta_p(t_j) = \arctan\left[\frac{\bar{V}^2(t_j)\sin\gamma(t_j)\cos\gamma(t_j)}{\bar{V}^2(t_j)\cos^2\gamma(t_j) - 1}\right] \tag{11.185}$$

11.3.3.4 Maximum Bending Angle

At the end of the AGA, the spacecraft needs to escape from the sphere of influence of the planet such that it can continue its interplanetary trajectory and thus, $V_\infty^+ > 0$.

This translates into a maximum atmospheric turn angle. Starting from Eq. (11.171) and assuming the limit case $V_\infty^+ = 0$, this means that the square-root argument should be zero :

$$e^{-\frac{2\theta}{L/D}}(V_\infty^-)^2 + (e^{-\frac{2\theta}{L/D}} - 1)V_c^2 = 0 \tag{11.186}$$

Normalising this equation by dividing left and right by V_c^2 yields

$$e^{-\frac{2\theta}{L/D}}\bar{V}^2 + (e^{-\frac{2\theta}{L/D}} - 1) = 0 \tag{11.187}$$

or

$$e^{-\frac{2\theta}{L/D}}(1 + \bar{V}^2) = 1 \tag{11.188}$$

Taking the natural logarithm from either side yields

$$-\frac{2\theta}{L/D} + \ln(1 + \bar{V}^2) = 0 \tag{11.189}$$

which finally results in

$$\theta = \theta_{max} = \frac{1}{2}\frac{L}{D}\ln\left(1 + \bar{V}^2\right) \tag{11.190}$$

where $\bar{V}$ is the velocity at infinite distance from the planet normalised to the circular orbit velocity at periapsis passage, hence $\bar{V} = \frac{V_\infty}{V_c}$, with $V_c = \sqrt{\frac{\mu_{target}}{R_p}}$.

As was derived earlier, the total turn angle Φ, includes gravity-only components and an atmospheric effect. One way to interpret this total turn angle is the following: suppose that a regular gravity assist is performed and that a certain bending of the incoming velocity vector is required. Then, one can calculate the required flyby altitude with Eq. (11.102). The turn angle provided by a gravity assist is limited, because the altitude of periapsis flyby has to remain well outside the atmosphere and certainly above the surface of the planet. The aerodynamic turn angle makes up for the deficit in turn angle caused by moving up the flyby altitude to a realistic value, somewhere in the upper atmosphere. Another way to look at this is to define, for a given total turn angle Φ, an equivalent radius of closest approach that would be necessary to achieve the same angular deflection using just a normal gravity assist (Lewis 1993).

Applying Eq. (11.102) once more and assuming that the resulting angle is $\Phi/2$ we may write:

$$\sin\left(\frac{\Phi}{2}\right) = \frac{1}{1 + \bar{V}^2} = \frac{1}{1 + \frac{R_{equiv}(V_\infty^+)^2}{\mu}} \tag{11.191}$$

and isolating R_{equiv}:

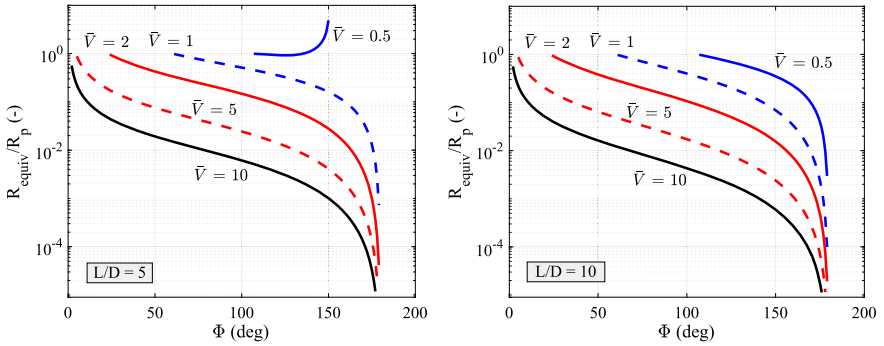


Fig. 11.52 R_{equiv}/R_p as a function of the total deflection angle and various values of the normalised approach velocity, $\overline{V}$, for a spacecraft with $L/D = 5$ (left) and $L/D = 10$ (right). Target planet is Venus

$$R_{equiv} = \frac{\mu_{target}}{(V_\infty^+)^2}\left(\frac{1}{\sin\frac{\Phi}{2}} - 1\right) \tag{11.192}$$

This equation effectively links the gravity and aerogravity assists: it shows that the trajectory after the atmospheric passage of the AGA is equal to the trajectory after a gravity assist with periapsis radius equal to R_{equiv}. Many existing orbital trajectory codes have already incorporated the gravity assist, hence the AGA can be included with minor modifications. One should bear in mind the corrections that have to be made for velocity loss due to atmospheric drag, which is to be discussed shortly.

Figure 11.52 visualises Eq. 11.192 for two different vehicle configurations, i.e., $L/D = 5$ and $L/D = 10$, approaching Venus. The complete (physically allowable) range of Φ is considered (0 to 180°), whereas the normalised hyperbolic excess velocity is discretely varied, with values of $\bar{V} = 0.5, 1, 2, 5$ and 10. The found R_{equiv} is normalised with a periapsis radius equivalent to an altitude of 200 km, to avoid crashing into the surface. However, changing this altitude to zero does not visibly affect the curves.

The figure shows that a vehicle with aerodynamic properties such that L/D is on the order of 5 to 10 can realise an equivalent flyby radius that is several orders of magnitude smaller than what would be possible with a regular gravity assist. Consider a spacecraft with $L/D = 5$ that performs an AGA manoeuvre with a deflection angle of 90°. The equivalent gravity assist periapsis with respect to the AGA periapsis is 60, 3 and 0.7% for $\bar{V}$ equal to 1, 5 and 10, respectively. For an even more aerodynamically shaped spacecraft with $L/D = 10$ making a 135° turn, the relative equivalent periapsis is 55% ($\bar{V} = 0.5$), 15% ($\bar{V} = 1$), 0.55% ($\bar{V} = 5$) and 0.15% ($\bar{V} = 10$). It should be noted that a regular gravity assist can only be performed down to a certain altitude, say R_0 plus a safety margin. The higher the approach velocity and bending angle, the larger the benefit. For very low approach

velocities ($\bar{V} = 0.5$ and $L/D = 5$), R_{equiv} can actually be larger than the AGA radius, meaning that the drag losses outweigh the benefit of the larger turning angle.

11.3.3.5 Drag Losses

So far we have discussed the benefit of using the atmosphere for increasing the bending angle, but the incurred drag losses have not yet been addressed. It is easily understood that diving to deep into the atmosphere will not just lead to excessive thermo-mechanical loads, but may also result in such a large velocity loss that the spacecraft will not be able to escape the atmosphere anymore. Equation (11.166) gives $V_{s,f}$, i.e., the velocity at leaving the atmosphere, as a function of the incoming velocity, $V_{s,i}$ and the atmospheric bending angle, θ. One approach can be that given the initial velocity, θ is varied from 0 to 180°, and after calculating $V_{s,f}$, it is easy to calculate $\Delta V = V_{s,f} - V_{s,i}$, and scaled with $V_{s,i}$ this gives the relative velocity loss as a function of θ. Keep in mind, though, that the descending and ascending leg are not included and also the local flight dynamics due to the continuous effect of drag is ignored.

Before we consider this in some more detail, an even simpler approach can be found in literature (McRonald et al. 1999), although it may be obvious upon closer inspection that this approach is inferior even to the one just discussed above. The expression for the decrease in velocity due to atmospheric drag ΔV with respect to V_∞ that is presented by McRonald et al. (1999) finds its origin in Eq. (11.160) as well:

$$\mathrm{d}\theta = -\frac{L}{D}\frac{1}{V}\frac{\mathrm{d}V}{1-\frac{V_c^2}{V^2}} \tag{11.193}$$

and rearranging this

$$\mathrm{d}V = -\frac{D}{L}V\left(1-\frac{V_c^2}{V^2}\right)\mathrm{d}\theta \tag{11.194}$$

To reach their result, McRonald et al. (1999) must have assumed that $V \approx$ constant, which will only hold for small θ, such that $\mathrm{d}V = \Delta V$. Following this, the relative velocity loss is

$$\frac{\Delta V}{V} = -\frac{\theta}{\frac{L}{D}}\left[1-\left(\frac{V_c}{V}\right)^2\right] \tag{11.195}$$

Further, with the relation between the velocity at the periapsis, $V = V_p$ and V_∞, Eq. (11.107), they claim without further proof that

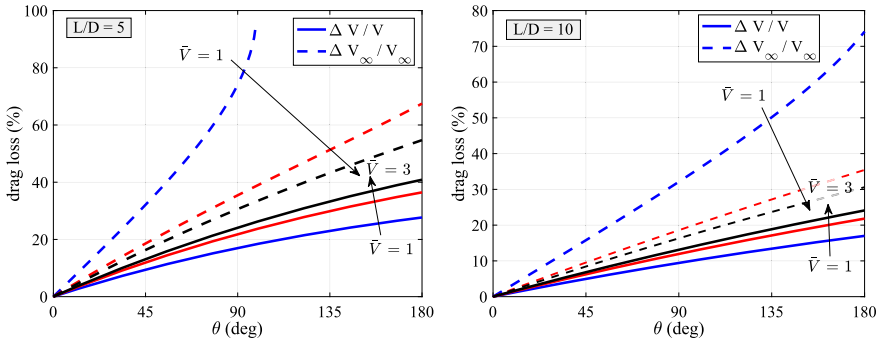


Fig. 11.53 The velocity loss relative to the interplanetary arrival velocity as a function of the bending angle

$$\frac{\Delta V_\infty}{V_\infty} = -\frac{\theta}{\frac{L}{D}}\left[1 + \left(\frac{V_c}{V_\infty}\right)^2\right] \tag{11.196}$$

However, a more accurate result can be obtained from Eq. (11.171), similarly to the approach discussed earlier to obtain $\Delta V = V_{s,f} - V_{s,i}$. Therefore, Eqs. (11.195)–(11.196) will not be considered any further.

As an example, an aerogravity assist at Mars will be considered, to anticipate on the case study introduced later (Sect. 11.3.5). Assuming a vehicle with $L/D = 5$ and a second one with $L/D = 10$, three initial hyperbolic excess velocities will be considered: $\bar{V} = 1$, 2 and 3, which means $V_\infty^- = 3.5$, 7,1 and 10.6 km/s and also $V_{s,i} = 6.1$, 8.7 and 11.7 km/s. Figure 11.53 shows the resulting velocity losses, both in terms of relative velocity and hyperbolic excess velocity. The obvious conclusion is that the velocity loss will be smaller the larger the L/D is. The larger lift will make it easier for the vehicle to bend its trajectory at a lower drag loss. It is also easily understood that a larger entry velocity reduces the relative velocity loss. Since the velocity with respect to the atmosphere, V_s, is significantly larger than V_∞, the maximum velocity reduction for the latter will be smaller before the vehicle can no longer escape the gravity well of the planet. This means that the maximum bending angle is a lot smaller compared to the situation where one would only consider the relative velocity. For the case with $L/D = 5$ and the low initial velocity of $\bar{V} = 1$ ($V_\infty^- = 3.5$ km/s), the maximum bending angle is about $\theta_{max} = 100°$. The corresponding relative velocity loss is only about 20%, whereas it does not even exceed 30% for $\theta = 180°$. The situation for the larger L/D is better, where θ_{max} reaches 180°, for the initial velocity of $V_\infty^- = 3.5$ km/s. However, the velocity loss exceeds 75%.

There are several ways to account for the drag loss during the atmospheric portion of an AGA, which are also discussed by Sims et al. (2000): before, during, or after the AGA. *Before* the AGA, one can increase the velocity with which the spacecraft is inserted into its interplanetary transfer trajectory. Obviously, this will change the geometry of the trajectory and increase the overall speed, but it does not require any

extra mass on the spacecraft itself.[21] *During* the AGA one can use some kind of thrusters (possibly the exhaust of an active cooling system), which has the advantage that optimal flight conditions can be better maintained. Finally, one can perform a corrective propulsive manoeuvre (directly) *after* the AGA. However, the required additional fuel leads to a potentially larger and heavier spacecraft.

To get more insight in the validity of the presented simplified model, the results of numerical simulations done by Hess and Mooij (2017) are briefly reviewed here. A more detailed discussion of similar results is presented in the earlier mentioned case study that concludes the discussion on aerogravity assists. This case study looks at several mission aspects benefitting from an AGA around Mars. Currently, the focus is on the relative velocity loss of a waverider passing through the Martian atmosphere, starting from a variety of initial velocities. The aerodynamic properties of this waverider will be presented later, but as a heads up, the maximum L/D at a Mach number of 40 is around 13 (Fig. 11.58). The vehicle mass is $m = 800$ kg.

The entry velocity range considered is $V_{s,i} = 8$ to 11 km/s with steps of 1 km/s (one of the terminal constraints in the optimisation was that the final eccentricity of the orbit was $e > 1$, thus guaranteeing that the vehicle can continue its hyperbolic trajectory). Figure 11.54 shows the velocity history as a function of atmospheric bending angle, both in absolute sense and as a velocity loss due to drag. The flown trajectory includes the descending and ascending legs, as well as a flight at a constant target altitude ($h_p = 30$ km). In the descending leg, initially the velocity is increasing when the gravitational pull is stronger than the aerodynamic deceleration.

What the results show is that the velocity loss is not as regular as the curves shown in Fig. 11.53. Several issues can explain the difference. Obviously, the velocity increase in the descending leg is one of them, albeit not the most important one. As we will see later while discussing the case study, the (optimised) trajectories are not as smooth as one wants them to be. Even though the optimiser tries to find a

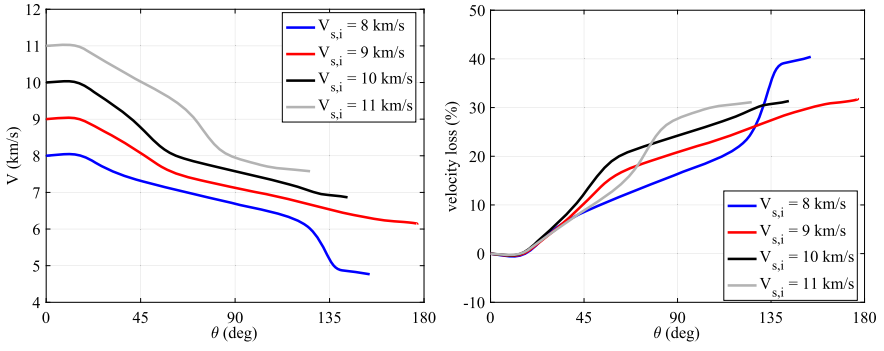


Fig. 11.54 Relative velocity (left) and velocity loss due to drag (right), for a waverider aerogravity assist at Mars, for a range of initial velocity

[21] Extra TPS mass could be necessary due to the larger speed in the atmosphere and thus an increased heat load.

solution with constant cruise altitude, due to the inertia of the vehicle it more or less undershoots the target altitude. That means that the vehicle dives deeper into the atmosphere and is encountering a higher density, which increases the dynamic pressure and thus the drag. And this lies at the heart of the problem: in the analytical model the actual energy dissipation due to drag is not considered, the model has only been derived under the assumption of constant L/D. This assumption also holds for the numerical solution, as the aerodynamics of the waverider is largely dependent on the solution obtained from the modified Newtonian method. This method will always give a Mach-independent, constant L/D, as long as the angle of attack is fixed, which is one of its characteristics. The moment the altitude has stabilised at its cruise value, the gradient has decreased to its equilibrium value for that particular altitude.

The curve for $V_{s,i} = 8$ km/s shows the largest deviation, as also towards the end of the cruise flight it dives to a lower altitude. This is mainly due to the fact that its velocity has reduced so much that it becomes difficult to maintain its cruise altitude (note that even when L/D remains constant, if the individual L and D have reduced in proportion, the lower lift may cause a problem to maintain force equilibrium). The small flattening part of the curve towards the end represents the ascending leg, where density is rapidly decreasing.

It may be clear that to understand the ins and outs of this non-linear simulation, more in-depth analysis is required. This goes beyond the current discussion, though. The main take-away of the current analysis is that the assumption of constant cruise altitude is an important one to keep in mind, especially when doing trajectory optimisation and/or guidance-system design. Also, even though only a smaller portion of the total mission, the descending and ascending legs may have a noticeable effect on the total bending angle. Finally, the actual energy dissipation due to drag, and thus the deceleration, should be carefully considered. The analytical model does not include this, in part because it is independent of vehicle mass.

As a concluding statement, Fig. 11.55 shows the difference between the analytical and numerical approach for a Mars aerogravity assist of the waverider with an entry velocity of $V_{s,i} = 9$ km/s (which happens to be the optimal velocity to maximise the

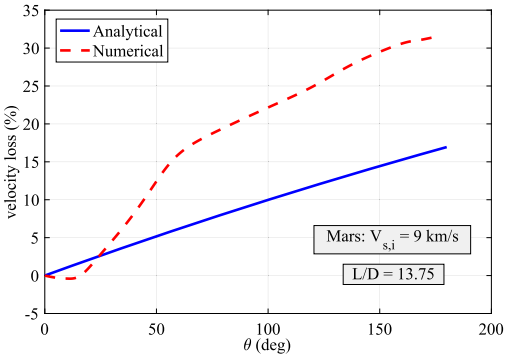


Fig. 11.55 Comparison between analytical and numerical approach: velocity loss due to drag for a waverider aerogravity assist at Mars ($V_{s,i} = 9$ km/s)

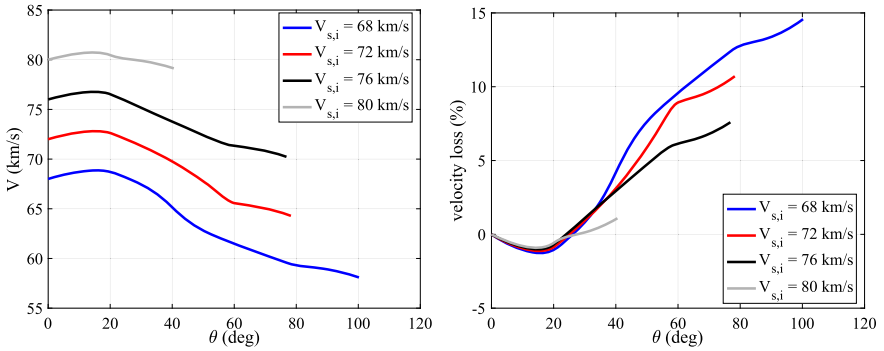


Fig. 11.56 Relative velocity (left) and velocity loss due to drag (right), for a waverider aerogravity assist at Jupiter and a range of initial velocity

bending angle). The effect of the descending leg (from $\theta = 0°$ till $\theta \approx 18°$), as well as the flight at lower altitude (from $\theta \approx 18°$ till $\theta \approx 60°$) and the ascending leg ($\theta > 162°$) can be clearly distinguished. What is interesting to note is that once the vehicle stays at its cruise altitude, the gradient of velocity loss is practically the same. So, this would mean that the analytical model would give a decent approximation for the cruise flight.

Finally, for the sake of comparison, Fig. 11.56 shows the velocity loss for an aerogravity assist around Jupiter. Due to the planet's large mass (and much larger gravity), the initial velocities that were used for Mars would result in elliptical capture orbits. Therefore, larger initial velocities are required to ensure hyperbolic orbits, because the aerogravity assist requires hyperbolic arrival and departure legs. The initial (relative) velocities that were selected for this comparison were 68.0 km/s, 72.0 km/s, 76.0 km/s and 80.0 km/s, at a starting altitude of 3200.0 km (where $h = 0$ m is defined as the point where the atmospheric pressure is 1 bar). The cruising altitude was set to $h_p = 250$ km. Two aspects are immediately evident: first, the effect of the descending and ascending leg is a lot larger, and, second, for this range of initial velocity the maximum bending angle is a lot smaller. It can also be observed that the velocity loss seems to occur more smoothly, which, after inspecting the altitude history, comes from a much more constant cruise altitude. No matter how interesting a more in-depth analysis would be, unfortunately, we have to leave it at this.

11.3.3.6 Equilibrium Altitude

Returning to Eq. (11.157), in Chap. 5 with Eq. (5.104), the same equation was found or the equilibrium glide, albeit written slightly differently:

$$L = mg\left(1 - \frac{V^2}{V_c^2}\right) \tag{11.197}$$

with $V_c = \sqrt{\frac{\mu_{target}}{R_p}}$ being the local circular velocity at the constant altitude the vehicle is flying (using index p for *periapsis*, which corresponds to the lowest point during the atmospheric flight). Thus, with the definition of the lift force, $L = C_L \frac{1}{2}\rho_p V_p^2 S$ it follows directly from Eq. (11.197) that a relationship exists between the air density and velocity:

$$\rho_p = 2\frac{W/S}{C_L}\left(\frac{1}{V_p^2} - \frac{1}{V_c^2}\right) \tag{11.198}$$

where, as in Chap. 5, $\frac{W/S}{C_L}$ is the lift parameter, also known as the wing load for aircraft-like vehicles, an important aerodynamic design parameter. Note that the circular velocity is the one at periapsis altitude, i.e., $V_c = \sqrt{\frac{\mu_{target}}{R_p}}$.

To establish the equilibrium altitude, one may assume an exponential atmosphere, such that

$$\rho_0 e^{-\frac{h}{H_s}} = 2\frac{W/S}{C_L}\left(\frac{1}{V_p^2} - \frac{1}{V_c^2}\right) \tag{11.199}$$

Rewriting this equation, and substituting (for a non-rotating planet) the expressions for circular velocity and periapsis velocity, $V_p^2 \approx 2\frac{\mu_{target}}{R_p} + (V_\infty^-)^2$ (Eq. (11.107), where the velocity loss due to drag has been neglected), yields:

$$\rho_0 e^{-\frac{h_p}{H_s}} = \frac{2}{R_p}\frac{m}{C_L}\left[1 - \left(\frac{\frac{\mu_{target}}{R_p}}{2\frac{\mu_{target}}{R_p} + (V_\infty^-)^2}\right)\right] \tag{11.200}$$

In this equation, the planetary-dependent parameters are ρ_0 and H_s (the atmosphere), R_{target} (the radius of the planet to calculate the altitude from the radial distance), and μ_{target} (gravity parameter of the target planet). Furthermore, there is a dependency on the incoming velocity, V_∞^-, and the periapsis distance, R_p, which obviously includes the altitude above the surface ($R_p = R_{target} + h_p$ for a spherical planet). Thus, the altitude to solve appears both on the right-hand and left-hand side of the equation, and can, in this form, only be solved iteratively.

Given a representative value for the wing load of $\frac{m}{C_L S} = 50$ kg/m^2 (McRonald and Randolph 1992), the altitude for vertical equilibrium is calculated for several planets, and two distinct velocities: $V_\infty^- = 0$ km/s (which is the theoretical minimum velocity), and $V_\infty^- = 50$ km/s. The results are listed in Table 11.8. It can be seen that with an increase of incoming velocity the equilibrium altitude decreases. This is easily explained when one realises that because the centrifugal force is larger at a higher velocity, also the lift should increase to counteract this. For given aerodynamics, this

Table 11.8 Equilibrium height during aerogravity assist for various planets

Planet	Exponential atmosphere[a]		$V_\infty^- = 0$ km/s		$V_\infty^- = 50$ km/s	
	ρ_0 (kg/m^3)	H_s (km)	ρ (kg/m^3)	h (km)	ρ (kg/m^3)	h (km)
Venus	65	15.9	$7.93{\cdot}10^{-6}$	253.1	$1.56{\cdot}10^{-5}$	242.4
Earth	1.217	8.5	$7.72{\cdot}10^{-6}$	101.7	$1.51{\cdot}10^{-5}$	96.0
Mars	0.020	11.1	$1.44{\cdot}10^{-5}$	80.3	$2.87{\cdot}10^{-5}$	72.7
Jupiter[b]	0.16	27	$6.96{\cdot}10^{-7}$	333.3	$9.85{\cdot}10^{-7}$	323.9
Saturn[b]	0.19	59.5	$8.20{\cdot}10^{-7}$	735.0	$1.37{\cdot}10^{-6}$	704.6
Uranus[b]	0.42	27.7	$1.93{\cdot}10^{-6}$	340.4	$3.57{\cdot}10^{-6}$	323.4
Neptune[b]	0.45	19.7	$2.00{\cdot}10^{-6}$	242.8	$3.64{\cdot}10^{-6}$	231.0

[a] https://nssdc.gsfc.nasa.gov/planetary/, date accessed: August 30, 2021
[b] The reference altitude for the gas planet is the 1-bar pressure level

Table 11.9 Acceleration as a function of altitude during an aerogravity assist at Venus; negative accelerations are in downward direction

h (km)	Acceleration (m/s^2) due to		
	Gravity	Lift force	Centrifugal force
260	−8.2	−5.3	+16.3
253.1 ($= R_{eq}$)	−8.2	8.2	+16.4
240	−8.2	−18.7	+16.4
200	−8.3	−232.7	+16.6
150	−8.5	−5,445.1	+16.9

means that the density should be larger and the altitude lower. The increase of gravity will only have a marginal effect on the equilibrium.

To conclude this discussion, suppose the vehicle will be flying at a different altitude than the equilibrium altitude. In this example, Venus will serve as the target planet, and with a velocity equal to the escape velocity (for this lower bound, $V_\infty^- = 0$) the prescribed height will be 253.1 km (as also shown in Table 11.8). In the case the altitude is slightly different from this target altitude, e.g., $h = 260$ km, the lift acceleration has reduced (in absolute sense) to -5.29 m/s^2, resulting in a positive acceleration of about 3 m/s^2 (see Table 11.9 for all relevant results). This would lead to an early exit from the atmosphere. If, on the other hand, the altitude is lower, $h = 240$ km, the resulting downwards lift acceleration is about -10 m/s^2, which will cause the vehicle to rapidly descend in the atmosphere and burn up if nothing is changed. As a limit case, for $h = 150$ km (with $\rho = 0.0052$ kg/m^3, the (downwards) lift acceleration is a staggering 5.5 km/s^2! Of course, the acceleration will not increase or decrease drastically when deviating from the desired altitude too much, but for sure some form of guidance system is required to maintain the altitude within bounds.

Note that for a more realistic incoming velocity of $V_\infty^- = 50$ km/s (and an equilibrium altitude of $h = 242.4$ km), the corresponding data are $g = -8.2$ m/s^2,

$L/m = -405.4$ m/s^2 and $a_c = 413.6$ m/s^2. For $h = 243.4$ km, the resulting acceleration is 25.0 m/s^2 upwards, and for $h = 241.4$ km it is 25.9 m/s^2 downwards. This indicates that the higher the entry velocity, the more sensitive the manoeuvre is for altitude errors (in this particular case just $\Delta h = \pm 1$ km).

Intermezzo 11.4 (Non-constant Lift-to-Drag Ratio)
Dropping the assumption of constant L/D, an exact differential equation for the turn angle θ associated with the aerodynamic portion of an AGA manoeuvre was derived by Bonfiglio et al. (2000):

$$\mathrm{d}\theta = -\frac{1}{2}\left(\frac{C_L}{C_D}\right)_{max} n\eta^{n-1}u^{n-1}[(n-1)\eta^n u^n + (u-1)^n]^{-1}\mathrm{d}u \tag{11.201}$$

where n is a parameter used in the drag polar: $C_D = C_{D_0} + K|C_L|^n$, and dependent on the aerodynamic shape of the vehicle, η is a dimensionless parameter specifying the glide altitude, and u is a dimensionless velocity variable.[22]

For the sake of argument, let us assume a parabolic lift-drag polar ($n = 2$). Then, after integrating Eq. (11.201), the expression for θ becomes:

$$\begin{aligned}\theta = &\left(\frac{C_L}{C_D}\right)_{max} \eta\left[2(1+\eta^2)\right]^{-1} \ln\left[\frac{\left(1+\eta^2\right)u_1^2 - 2u_1 + 1}{\left(1+\eta^2\right)u_2^2 - 2u_2 + 1}\right] \\ &+ \left(\frac{C_L}{C_D}\right)_{max} \left(1+\eta^2\right)^{-1} \tan\left[\frac{\eta(u_1-u_2)}{(1+\eta^2)u_1u_2 - (u_1+u_2) + 1}\right]\end{aligned} \tag{11.202}$$

with the subscripts 1 and 2 defining the begin and end of the atmospheric flight. For the hypersonic flight regime, a more realistic value for n is $n = \frac{3}{2}$, which makes the integration of Eq. (11.201) a little tedious, and the solution more elaborate:

$$\theta = 2k[I(x_2) - I(x_1)] \tag{11.203}$$

where $k = \frac{3}{4}\left(\frac{C_L}{C_D}\right)_{max} \eta^{\frac{1}{2}}$. The function $I(x)$ is defined as $I(x) = I_1(x) + I_2(x) + I_3(x)$, where:

$$I_1(x) = \frac{1}{2(1+a)} \ln(2x+1) \tag{11.204}$$

$$I_2(x) = \frac{1-\kappa+\kappa^2}{3\kappa(1-\kappa)(1+\kappa^3)} \ln\left(x + \frac{1}{1-\kappa}\right) \tag{11.205}$$

[22] The glide altitude η designates the dimensionless glide altitude, defined as $\eta = \left(\frac{\rho SR}{2m}\right) C_L^*$, where R is the radius from the centre of the planet to the spacecraft and C_L^* indicates the value of C_L for maximum L/D. The dimensionless velocity, u, is defined as $u = \frac{RV^2}{\mu}$.

$$I_3(x) = \frac{-(1+\kappa)(2\kappa^2+1)}{6\kappa(1+\kappa+\kappa^2)(1+\kappa^3)} \ln\left[(1-\kappa^3)x^2 + (1-\kappa)(2+\kappa)x + (1-\kappa)\right] - \left[\frac{1+\kappa}{\sqrt{3}\kappa(1+\kappa+\kappa^2)(1+\kappa^3)}\right] \times \tan^{-1}\left[\frac{2(1+\kappa+\kappa^2)x + (2+\kappa)}{\sqrt{3}\kappa}\right] \tag{11.206}$$

In the definitions of I_1, I_2, and I_3, $\kappa = a^{\frac{1}{3}}$ and $a = \frac{1}{2}\eta^{\frac{3}{2}}$.

Equations (11.202) and (11.203) define the turn angles during the atmospheric phase of the AGA in case of a parabolic and hypersonic drag polar, respectively. Assuming that the vehicle will fly with maximum lift-to-drag ratio, $E^* = \left(\frac{C_L}{C_D}\right)_{max}$, Bonfiglio et al. (2000) derived an expression for the outgoing velocity as a function of the incoming velocity and the aerodynamic turn angle:

$$u_\infty^+ = \left(u_\infty^- + 1\right) e^{-2\theta/E^*} - 1 \tag{11.207}$$

The assumption of maximum lift-to-drag ratio implies that the vehicle keeps flying very close to this maximum throughout the atmospheric flight. Obviously, a guidance system should be used to enforce this (in the next Sect. 11.3.4 a potential guidance system will be discussed).

In Eq. (11.207), the corresponding θ for maximum L/D is given by:

$$\theta = -\frac{1}{2}\left(\frac{C_L}{C_D}\right)_{max} \ln\left(\frac{u_2 - 1}{u_1 - 1}\right) \tag{11.208}$$

Lewis and McRonald (1992) provide an alternative formulation for the outgoing velocity, Eq. (11.207), but in this case in dimensional form:

$$V_{\infty,f}^2 = e^{-2\theta/E^*}(V_\infty^-)^2 + \left(e^{-2\theta/E^*} - 1\right)\frac{\mu_{target}}{R} \tag{11.209}$$

where we retained the notation E^*. Inspecting Eq. (11.209), the incoming velocity is decreased by an exponential scale factor $-2\theta/E^*$. In other words, for a given allowable energy loss due to the atmospheric passage, θ can be increased by enlarging the vehicle's L/D. ■

11.3.4 Feedback Guidance

All trajectories that came up so far, were nominal trajectories flown under nominal conditions. Any uncertainty encountered should be corrected for by a guidance system, fed with information from navigation sensors and filters. It was shown earlier that even a small discrepancy in the equilibrium altitude can lead to large perturbing accelerations. These discrepancies are commonly introduced by uncertainties in the atmospheric density, aerodynamic coefficients, or initial conditions. In this section an example guidance system is discussed, focussing on longitudinal guidance to control

the altitude. Typically, a lateral guidance system is added to allow for controlling the cross-range deviations, in a similar way as discussed in Sect. 7.6.3.2: once a certain performance limit is exceeded, the sign of the bank angle should be "flipped", and a bank reversal has to be executed. In the current section the focus is on the longitudinal guidance, but some remarks on the lateral guidance will be made.

For the longitudinal guidance, a drag-tracking algorithm is considered. Such an algorithm, as was proposed by Saraf et al. (2004), uses a reference trajectory from which a reference drag profile D_{ref} is obtained. The drag-tracking algorithm, which will be shown below, tries to match this reference profile by determining a value for the bank angle. The required reference trajectory could come from a mission optimisation that is assumed to be available, but will currently not be discussed any further. An application of the methodology will be discussed at a later stage, though, notably in Sect. 11.3.5.

Let $e = D_{ref} - D$ be the drag-acceleration error. Saraf et al. (2004) modelled the guidance law such that, in absence of modelling errors, second-order linear drag-error dynamics would be achieved. A standard second-order model for the error $e = D_{ref} - D$ is:

$$\left(D''_{ref} - D''\right) + 2\zeta\omega\left(D'_{ref} - D'\right) + \omega^2\left(D_{ref} - D\right) = 0 \tag{11.210}$$

where ω is the natural frequency and ζ is the damping ratio of the second order system. In Eq. (11.210), the derivatives are defined with respect to (specific) energy instead of time:

$$D' = \frac{\mathrm{d}D}{\mathrm{d}\mathcal{E}} \tag{11.211}$$

where the total (specific) energy $\mathcal{E}$ of the vehicle is defined as:

$$\mathcal{E} = \frac{V_I^2}{2} - \frac{\mu}{R_I} \tag{11.212}$$

in which V_I and R_I are the velocity radial distance of the vehicle in the inertial planetocentric reference frame, $\mathcal{F}_I$, and μ is the gravitational parameter of the central body.

In the case all variables in Eq. (11.210) are known, it can be used to determine the bank angle, σ. In this equation ω and ζ are design variables with which the response of the error dynamics can be controlled. They typically follow from response analyses, aimed at having a smooth and stable transient response of the vehicle. The variables D'_{ref} and D''_{ref} can be obtained from the reference trajectory, and should be evaluated in the mission-design process. Similar to the discussion in Sect. 7.6.2, where the Space-Shuttle guidance law was derived following the same principles of drag-tracking, both the drag, D, and its derivative of the drag with respect to energy, D', can be determined from analytical expressions. These expressions are derived from the equations of motion and will be shown in the intermezzo later in this section.

The second derivative of drag with respect to energy, D'', is a function of the bank angle, which is the variable that needs to be estimated. Saraf et al. (2004) provides an expression for D'':

$$D'' = a + bu_D \tag{11.213}$$

with the vertical lift-to-drag ratio, u_D, defined by

$$u_D = \frac{\tilde{L}}{\tilde{D}} \cos \sigma_0 \tag{11.214}$$

As was the case for the discusion of the Space-Shuttle guidance, $\tilde{L}$ and $\tilde{D}$ are the lift and drag *acceleration*, respectively, although for the definition of L/D this distinction is not relevant,[23] σ_0 is the required bank angle, and a and b are two variables that are a function of the current flight parameters. The variables a and b can be obtained by taking the derivative of $\tilde{D} = \bar{q} \frac{C_D S}{m}$ with respect to energy twice (assuming an exponential atmosphere). This derivation is rather lengthy and would disrupt the flow of the discussion. Therefore, only the final result is shown here, but the details of the derivation are found in the intermezzo. Thus, a and b are found to be:

$$\begin{aligned} a = \; & D\left(\frac{C_D''}{C_D} - \frac{C_D'^2}{C_D^2}\right) + D'\left(\frac{C_D'}{C_D} + \frac{2}{V^2}\right) - \frac{4D}{V^4} \\ & + \frac{1}{DV^2}\left(\frac{1}{H_s} + \frac{2g}{V^2}\right)\left(g - \frac{V^2}{R}\right) + \left(\frac{1}{H_s} + \frac{2g}{V^2}\right)\mathcal{C}_\gamma \end{aligned} \tag{11.215}$$

$$b = -\frac{1}{V^2}\left(\frac{1}{H_s} + \frac{2g}{V^2}\right) \tag{11.216}$$

in which

$$\mathcal{C}_\gamma = -\left[2\omega_p/(VD)\right] \cos\chi \cos\delta \tag{11.217}$$

By substituting Eq. (11.213) into (11.210), an expression for the control variable u_D can be found (which is, in essence, a proportional-derivative regulator):

$$u_D = \frac{1}{b}\left[-a + D''_{ref} + 2\zeta\omega\left(D'_{ref} - D'\right) + \omega^2\left(D_{ref} - D\right)\right] \tag{11.218}$$

Once u_D has been determined, the required magnitude of the bank angle can be determined using Eq. (11.214), whereas the sign will be decided by the lateral guidance. For a trajectory with the same initial conditions as the reference trajectory, the differences between the reference and the actual values of drag acceleration and

[23] Note that in the remainder of the discussion the tilde notation will be dropped for the sake of convenience.

its derivative will both be zero. This results in an identical bank-angle profile as was used to generate the reference trajectory. For deviations from the nominal trajectory, due to, for example, uncertainties in initial conditions or atmospheric density, a value for the bank angle is determined that takes these deviations into account. In case of persistent errors, one could decide to add an integral term to u_D, which is a common control solution. The control variable would then change according to $u_D = u_{D,0} + K_I \int \left(D_{ref} - D\right) \mathrm{d}t$, with $u_{D,0}$ given by Eq. (11.218). The gain K_I follows from response analyses, aided by techniques from classical PID control-system design.

A few remarks should be made about this guidance law. For very small values of the aerodynamic drag, this law does not work properly. The argument of the arccosine that is to be applied to extract σ_0 in Eq. (11.214), has to be in the interval $[-1, 1]$, so no solution exists when $\|u_D/(L/D)\| > 1$. This problem can be resolved by using open-loop control for small values of D, and use the bank angle computed using Eq. (11.218) once this drag threshold has been passed. It was found that a good value of the drag-acceleration threshold $D_{lim} = 0.6$ N/kg. At drag accelerations smaller than this drag threshold, deviations of the bank angle from the reference bank angle had hardly any effect on the trajectory. Therefore, during the arrival phase, the bank angle should be set to the value of the reference bank angle. For the exit phase, a different strategy will be used that will be discussed shortly.

A similar problem can also occur when the deviation from the reference trajectory becomes too excessive. If this occurs, it basically means that the vehicle is not capable of handling this deviation. To guarantee that the guidance algorithm still provides a solution, the bank angle will be either set to 0° (in case $u_D/(L/D) > 1$), or be set to 180° (for the situation that $u_D/(L/D) < -1$).

Equation (11.210) requires values for the natural frequency and the damping ratio. To find values for these parameters, there are several ways to do so. The formal way would be to specify the response characteristics of the vehicle, and, in combination with linearised system analysis, to resolve the values for ω and ζ. Alternatively, one could randomly select values and study the system's response and accept the values when this seems reasonable enough. However, to do a manual variation could be rather tedious and time consuming. In case a (numerical) performance objective can be defined, for instance, related to the atmospheric bending angle of the reference trajectory, an optimisation procedure could be performed. This was done by Hess and Mooij (2017), which provided the values $\omega = 0.126$ rad/s and $\zeta = 0.662$ for an 800-kg waverider performing an aerogravity assist around Mars.

Before we dive into the derivation of the drag-acceleration derivatives with respect to energy, a few words are justified to talk about the lateral guidance. A typical lateral-guidance scheme used by, for instance, the Space Shuttle is based on defining a so-called heading-error dead-band and execute a bank reversal once the heading towards the target exceeds the pre-defined limit values. In case of the aerogravity assist, a similar principle is used, but now related to the exit conditions of the manoeuvre, i.e., a maximum allowable inclination error with respect to the target orbital plane. Values between 1° and 5° were found to give acceptable results, depending on the required bending angle.

It is noted, though, that constant dead-band values do not always give the best results. Therefore, it is advised to treat the definition of the inclination-error dead-band as a design problem to find the best combination of values, for instance, as a function of local bending angle. One option could be to allow for a large inclination offset at the beginning of the assist, but then that the allowable offset decreases with increasing atmospheric bending angle. An example of a function that has this type of behaviour is a hyperbolic tangent function. Finally, controlling the inclination error will only work for a minimum drag-acceleration value. Thus, once the drag threshold is passed and the exit phase is initiated, a small inclination offset can still remain. The remaining error should be dealt with in a different manner, e.g., through a propulsive correction.

Implementation-wise, one should realise that a bank reversal cannot be executed in zero time, i.e., at an infinite rate. Casoliva et al. (2008) propose to change the magnitude of the bank angle during the reversal as follows:

$$\sigma = \text{sign}(i - i_t) \arcsin\left[\text{sat}\left(\kappa_0(i - i_f)\right)\right] \tag{11.219}$$

The above equation contains a saturation function $\text{sat}(x)$, which saturates the variable x if its value is outside of the interval $0 \leq x \leq 1$. The constant κ_0 is a proportional gain. If the difference $\kappa_0(i - i_f) \geq 1$, Eq. (11.219) will command a bank angle of 90°, which is the value for the bank angle that has the largest cross-track component. For values of $0 \leq \kappa_0(i - i_f) \leq 1$, the commanded bank angle will decrease until $i - i_f = 0°$, where the commanded bank angle will be 0°, such that the lift vector will not have a cross-track component. In turn, this will ensure that the inclination is not changed.

So far, it has been assumed that once a bank reversal was required, the sign of the commanded bank angle would be changed. In reality, however, an instantaneous change in bank angle is not possible, as this would imply an infinite bank angle-rate and acceleration. To overcome this simplification, a constant, but finite bank-angle rate $\dot{\sigma}_c$ should be used for a bank reversal (see Sect. 7.6.3.2 for more details on how the guidance commands should be issued). The bank angle is increased or decreased during a reversal according to Eq. (7.274), until the difference between the target bank angle and the actual bank angle is less than $\dot{\sigma}_c \Delta t_g$. Once this condition is true, the bank angle will be determined using Eq. (11.214), unless the next reversal has to be commanded.

Intermezzo 11.5 (Drag Derivatives with Respect to Energy)
The guidance algorithm discussed in the section requires some coefficients that include derivatives of the drag acceleration with respect to energy. For instance, to get expressions for the coefficients a and b, defined by Eqs. (11.215) and (11.216), one has to differentiate the drag acceleration twice with respect to the energy, after which expressions with known variables obtained from a set of equations of motion is substituted. These "normal" equations of motion as a function of time will have to be converted to functions with energy as independent variable. Once that step has been done, the first and second derivative of drag acceleration with respect to energy

can easily be derived. This intermezzo will show the derivations, which would have otherwise disrupted the discussion earlier.

As a refresher, the equations of motion that are relevant for the derivation are restated below. These have been mentioned several times before, e.g., in Sect. 8.2.1, and will not be discussed in detail here. So, for an Earth with a central gravity field, the dynamic and kinematic equations read:

$$\dot{V} = -\frac{D}{m} + g\sin\gamma + \omega_{cb}^2 R\cos\delta(\sin\gamma\cos\delta - \cos\gamma\sin\delta\cos\chi) \tag{11.220}$$

$$\begin{aligned} V\dot{\gamma} =& \frac{L\cos\sigma}{m} - g\cos\gamma + 2\omega_{cb}V\cos\delta\sin\chi + \frac{V^2}{R}\cos\gamma + \\ &+\omega_{cb}^2 R\cos\delta(\cos\delta\cos\gamma + \sin\gamma\sin\delta\cos\chi) \end{aligned} \tag{11.221}$$

$$\begin{aligned} V\cos\gamma\dot{\chi} =& \frac{L\sin\sigma}{m} + 2\omega_{cb}V(\sin\delta\cos\gamma - \cos\delta\sin\gamma\cos\chi) + \\ &+\frac{V^2}{R}\cos^2\gamma\tan\delta\sin\chi + \omega_{cb}^2 R\cos\delta\sin\delta\sin\chi \end{aligned} \tag{11.222}$$

$$\dot{R} = \dot{h} = V\sin\gamma \tag{11.223}$$

$$\dot{\tau} = \frac{V\sin\chi\cos\gamma}{R\cos\delta} \tag{11.224}$$

$$\dot{\delta} = \frac{V\cos\chi\cos\gamma}{R} \tag{11.225}$$

Consider now the total specific energy of a vehicle, $\mathcal{E}$:

$$\mathcal{E} = \frac{V^2}{2} + \frac{\mu}{R} \tag{11.226}$$

Using energy as independent variable allows to reduce the six equations shown above to five differential equations. Let $(\cdot)$ be an arbitrary variable, then its derivative with respect to energy, indicated by a *prime*, can be obtained as follows:

$$\frac{\mathrm{d}(\cdot)}{\mathrm{d}\mathcal{E}} = (\cdot)' = \frac{1}{\dot{\mathcal{E}}}\frac{\mathrm{d}(\cdot)}{\mathrm{d}t} = -\frac{1}{DV}\frac{\mathrm{d}(\cdot)}{\mathrm{d}t} \tag{11.227}$$

With energy as independent variable and the corresponding differentiation rule presented in Eq. (11.227), the equations of motion become:

$$\chi' = \frac{\sin\chi\tan\delta\cos\gamma}{R}\left(\frac{1}{D}\right) + \frac{1}{V^2\cos\gamma}\left(\frac{L}{D}\sin\sigma\right) + \mathcal{C}_\chi \tag{11.228}$$

$$\gamma' = \left(g - \frac{V^2}{R}\right)\frac{\cos\gamma}{V^2}\left(\frac{1}{D}\right) - \frac{1}{V^2}\left(\frac{L}{D}\cos\sigma\right) + \mathcal{C}_\gamma \tag{11.229}$$

in which $\mathcal{C}_\chi$ and $\mathcal{C}_\gamma$, also known as the Coriolis terms, are given by:

$$\mathcal{C}_\chi = -\left[2\omega_{cb}/(V\,D)\right](\tan\gamma\cos\chi\cos\delta - \sin\delta) \tag{11.230}$$

$$\mathcal{C}_\gamma = -\left[2\omega_{cb}/(V\,D)\right]\sin\chi\cos\delta \tag{11.231}$$

The resulting kinematic equations are:

$$R' = -\sin\gamma\left(\frac{1}{D}\right) \tag{11.232}$$

$$\tau' = -\frac{\cos\gamma\sin\chi}{R\cos\delta}\left(\frac{1}{D}\right) \tag{11.233}$$

$$\delta' = -\frac{\cos\gamma\cos\chi}{R}\left(\frac{1}{D}\right) \tag{11.234}$$

Equations (11.228) through (11.234) form a set of differential equations in which the independent variable is energy. The drag tracking algorithm presented by Saraf et al. (2004) uses these expressions in combination with the second derivative of the drag acceleration.

First Derivative

Consider the drag acceleration D (again, we will not use the tilde to distinguish the drag acceleration from the drag force):

$$D = \frac{1}{2}\rho V^2 C_D S/m \tag{11.235}$$

The first derivative of drag with respect to energy, D', is given by:

$$D' = \frac{\mathrm{d}D}{\mathrm{d}\mathcal{E}} = \underbrace{\frac{\partial D}{\partial\rho}\frac{\partial\rho}{\partial\mathcal{E}}}_{\text{Term 1}} + \underbrace{\frac{\partial D}{\partial V}\frac{\partial V}{\partial\mathcal{E}}}_{\text{Term 2}} + \underbrace{\frac{\partial D}{\partial C_D}\frac{\partial C_D}{\partial\mathcal{E}}}_{\text{Term 3}} \tag{11.236}$$

Because the above contains many partial derivatives, they will be treated individually for the sake of readability, after which they are combined to get the expression for D'.

Term 1

The derivative of D with respect to the atmospheric density ρ is given by:

$$\frac{\partial D}{\partial \rho} = \frac{1}{2} V^2 C_D S/m \tag{11.237}$$

The derivative of the atmospheric density ρ with respect with respect to the energy is obtained from:

$$\frac{\partial \rho}{\partial \mathcal{E}} = \frac{\partial \rho}{\partial R}\frac{\partial R}{\partial \mathcal{E}} + \frac{\partial \rho}{\partial V}\frac{\partial V}{\partial \mathcal{E}} \tag{11.238}$$

The derivative of the density with respect to velocity in Eq. (11.238) is known to be zero. From Eq. (11.232), the partial derivative $\partial R/\partial \mathcal{E}$ can be obtained, resulting in the expression for the derivative of the density with respect to the energy:

$$\frac{\partial \rho}{\partial \mathcal{E}} = -\frac{\partial \rho}{\partial R}\frac{\sin\gamma}{D} \tag{11.239}$$

Term 2

The derivative of D with respect to V is given by:

$$\frac{\partial D}{\partial V} = \rho V C_D S/m \tag{11.240}$$

and the one of V with respect to $\mathcal{E}$ by

$$\frac{\partial V}{\partial \mathcal{E}} = \frac{\partial V}{\partial t}\frac{\partial t}{\partial \mathcal{E}} \tag{11.241}$$

From Eq. (11.220), the partial derivative $\partial V/\partial t$ can be obtained. Combining this with Eq. (11.227), $\frac{\partial V}{\partial \mathcal{E}}$ is found to be

$$\frac{\partial V}{\partial \mathcal{E}} = \frac{D + g\sin\gamma}{DV} \tag{11.242}$$

Term 3

The partial derivative of D with respect to C_D is given by:

$$\frac{\partial D}{\partial C_D} = \frac{1}{2}\rho V^2 S/m \tag{11.243}$$

Now only the partial derivative of C_D with respect to $\mathcal{E}$ remains, but since there is no analytical expression for this derivative, only the definition will be provided:

$$\frac{\partial C_D}{\partial \mathcal{E}} = C'_D \tag{11.244}$$

This derivative much follow from the aerodynamic characteristics of the vehicle, in relation to its state dependency. See also the remarks made during the discussion of the analytical Space-Shuttle guidance in Sect. 7.6.2.

Substituting Eqs. (11.237), (11.239), (11.240), (11.242), (11.243), and (11.244) into Eq. (11.236) and combining terms, yields:

$$D' = D\left(\frac{2}{V^2} + \frac{C'_D}{C_D}\right) + \sin\gamma\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R} + \frac{2g}{V^2}\right) \tag{11.245}$$

Second Derivative

To obtain the second derivative of drag with respect to energy, one starts with differentiating Eq. (11.211) with respect to energy. This second derivative is found to be

$$\begin{aligned} D'' = \frac{\partial D}{\partial \mathcal{E}}\left(\frac{2}{V^2} + \frac{C'_D}{C_D}\right) + \overbrace{\frac{\partial}{\partial \mathcal{E}}\left(\frac{2}{V^2} + \frac{C'_D}{C_D}\right)}^{\text{Term 1}} D + \\ + \underbrace{\frac{\partial \sin\gamma}{\partial \mathcal{E}}}_{\text{Term 2}}\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R} + \frac{2g}{V^2}\right) + \underbrace{\frac{\partial}{\partial \mathcal{E}}\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R} + \frac{2g}{V^2}\right)}_{\text{Term 3}} \sin\gamma \end{aligned} \tag{11.246}$$

In Eq. (11.246), three terms have been indicated that require evaluation, which is done subsequently.

Term 1

The first term that needs to be expanded can be divided into two separate terms:

$$\frac{\partial}{\partial \mathcal{E}}\left(\frac{2}{V^2}+\frac{C_D'}{C_D}\right)=\underbrace{\frac{\partial}{\partial \mathcal{E}}\left(\frac{2}{V^2}\right)}_{A}+\underbrace{\frac{\partial}{\partial \mathcal{E}}\left(\frac{C_D'}{C_D}\right)}_{B} \tag{11.247}$$

Subterm A
With the help of Eq. (11.242), Subterm A evaluates to

$$\frac{\partial}{\partial \mathcal{E}}\left(\frac{2}{V^2}\right)=\frac{\partial}{\partial V}\left(\frac{2}{V^2}\right)\frac{\partial V}{\partial \mathcal{E}}=\frac{-4}{V^4}\left(\frac{D+g\sin\gamma}{D}\right) \tag{11.248}$$

Subterm B
Next, using the quotient rule, Subterm B is found to be

$$\frac{\partial}{\partial \mathcal{E}}\left(\frac{C_D'}{C_D}\right)=\frac{C_D''C_D-C_D'^2}{C_D^2}=\frac{C_D''}{C_D}-\frac{C_D'^2}{C_D^2} \tag{11.249}$$

Term 2

The second term that needs to be expanded contains a partial derivative of flight-path angle γ with respect to energy. By applying the chain rule, Term 2 can be expressed in terms of known quantities:

$$\frac{\partial \sin\gamma}{\partial \mathcal{E}}=\frac{\partial \sin\gamma}{\partial \gamma}\frac{\partial \gamma}{\partial \mathcal{E}}=\cos\gamma\frac{\partial \gamma}{\partial \mathcal{E}} \tag{11.250}$$

Substituting Eq. (11.229) into Eq. (11.250) yields

$$\frac{\partial \sin\gamma}{\partial \mathcal{E}}=\cos\gamma\left[\left(g-\frac{V^2}{R}\right)\frac{\cos\gamma}{V^2}\left(\frac{1}{D}\right)-\frac{1}{V^2}\left(\frac{L}{D}\cos\sigma\right)+\mathcal{C}_\gamma\right] \tag{11.251}$$

Term 3

The third term is divided into two separate terms:

$$\frac{\partial}{\partial \mathcal{E}}\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R}+\frac{2g}{V^2}\right)=\underbrace{\frac{\partial}{\partial \mathcal{E}}\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R}\right)}_{\text{A}}+\underbrace{\frac{\partial}{\partial \mathcal{E}}\left(\frac{2g}{V^2}\right)}_{\text{B}} \tag{11.252}$$

which will be evaluated separately.

Subterm A
The expansion of Subterm A is given by:

$$\frac{\partial}{\partial \mathcal{E}}\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R}\right)=\frac{\partial}{\partial \mathcal{E}}\left(-\frac{1}{\rho}\right)\frac{\partial \rho}{\partial R}-\frac{1}{\rho}\frac{\partial}{\partial \mathcal{E}}\left(\frac{\partial \rho}{\partial R}\right) \tag{11.253}$$

$$=\frac{\partial}{\partial R}\left(-\frac{1}{\rho}\right)\frac{\partial \rho}{\partial \mathcal{E}}\frac{\partial \rho}{\partial R}-\frac{1}{\rho}\frac{\partial^2 \rho}{\partial R^2}\frac{\partial R}{\partial \mathcal{E}} \tag{11.254}$$

Expressions for the terms $\partial\rho/\partial\mathcal{E}$ and $\partial r/\partial\mathcal{E}$ can be found in Eqs. (11.239) and (11.232), respectively.

Subterm B
The expansion of Subterm B is done as

$$\frac{\partial}{\partial \mathcal{E}}\left(\frac{2g}{V^2}\right)=\frac{\partial}{\partial V}\left(\frac{2g}{V^2}\right)\frac{\partial V}{\partial E} \tag{11.255}$$

With Eq. (11.242), Eq. (11.255) can be expressed as:

$$\frac{\partial}{\partial \mathcal{E}}\left(\frac{2g}{V^2}\right)=\frac{-4g}{V^3}\frac{D+g\sin\gamma}{DV} \tag{11.256}$$

Assuming that $\sin^2\gamma \approx 0$ and $D+g\sin\gamma \approx D$, combining terms A, B, and C, and substituting them into Eq. (11.246) provides an expression for D'':

$$\begin{aligned} D''=\; & D\left(\frac{C_D''}{C_D}-\frac{C_D'^2}{C_D^2}\right)+D'\left(\frac{C_D'}{C_D}+\frac{2}{V^2}\right)-\frac{4D}{V^4}+\frac{1}{DV^2}\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R}+\frac{2g}{V^2}\right)\left(g-\frac{V^2}{r}\right) \\ & +\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R}+\frac{2g}{V^2}\right)\mathcal{C}_\gamma+\frac{1}{V^2}\left(-\frac{1}{\rho}\frac{\partial \rho}{\partial R}+\frac{2g}{V^2}\right)\frac{L}{D}\cos\sigma \end{aligned} \tag{11.257}$$

Equation (11.257) can be written as

$$D'' = a + b\left(\frac{L}{D}\cos\sigma\right) \tag{11.258}$$

where

$$a = D\left(\frac{C_D''}{C_D} - \frac{C_D'^2}{C_D^2}\right) + D'\left(\frac{C_D'}{C_D} + \frac{2}{V^2}\right) - \frac{4D}{V^4} + \frac{1}{DV^2}\left(-\frac{1}{\rho}\frac{\partial\rho}{\partial R} + \frac{2g}{V^2}\right)\left(g - \frac{V^2}{r}\right) + \left(-\frac{1}{\rho}\frac{\partial\rho}{\partial R} + \frac{2g}{V^2}\right)\mathcal{C}_\gamma \tag{11.259}$$

and

$$b = -\frac{1}{V^2}\left(-\frac{1}{\rho}\frac{\partial\rho}{\partial R} + \frac{2g}{V^2}\right) \tag{11.260}$$

which was the final goal of this intermezzo. ■

11.3.5 Case Study: Aerogravity Assist at Mars

The discussion on aerogravity assists will be concluded with a summary of some results of a study conducted by Hess and Mooij (2017). It concerns a preliminary analysis of selecting a proper reference trajectory for using Mars as the fly-by planet. The second analysis focusses on the interplanetary trajectory of Rosetta on its way to comet 67P/Churyumov-Gerasimenko, and how an aerogravity assist around Mars could have benefitted the mission. Mars has been selected as a good candidate for an aerogravity assist, even though its atmosphere is relatively thin. This means, though, that the thermo-mechanical loads may be easier contained, compared to, for instance, an aerogravity assist around Venus, Earth, or Jupiter. This conclusion was also drawn by McRonald and Randolph (1992), which is one of the reasons we have selected Mars.

Typical re-entry vehicle shapes, such as capsules or even winged shapes like the Space Shuttle, do not have sufficient lifting capabilities to execute aerogravity assist. Even if they would have sufficient lift to perform a skipping flight, the drag would be equally large, and the energy dissipation would be too large. Therefore, in this study a high lift-to-drag ratio waverider was selected from a pool of generic waverider shapes to have the best aerodynamic characteristics, as has been suggested by many authors in the field (Lewis and McRonald 1992; Lohar et al. 1994; Lavagna et al. 2005; Armellin et al. 2007). The next part will introduce the waverider geometry and aerodynamics, after which the problem setup will be presented.

11.3.5.1 Model Description and Problem Setup

Starkey and Lewis (2007) define a generic waverider that is described by only five parameters, namely l, w, θ, δ, and n (Fig. 11.57). The parameters w and l correspond to the width and length of the vehicle, as is shown in Fig. 11.57(a). The parameters θ and δ are the wedge angle and the oblique-shock angle, respectively, shown in Fig. 11.57b. The parameter n is related to the curvature of the planform. Using these five parameters, the upper and lower surface can be described by simple functions that allow for a parametric optimisation of the shape. In the equations shown in the figure, $A = \frac{1}{2} w/l^n$ and $B = A/\tan^n \delta$. The curvature of the upper surface is given by z_u, while the lower surface is defined by z_l. Both are shown in Fig. 11.57c. With this parametrisation an optimal waverider shape was found, shown in the isometric view of Fig. 11.57d, with $w = 2.0$ m, $l = 5.0$ m, $n = 0.5$, $\theta = 5.0°$, and $\delta = 9.0°$.

This particular configuration has the aerodynamic characteristics shown in Fig. 11.58. The evaluation of the free-molecular flow properties was done with the open-source software SPARTA, see Intermezzo 3.5 in Sect. 3.5, where also some theoretical background of this flow regime is provided. The continuum-flow characteristics are obtained from the modified Newtonian method (Sect. 3.4.4). The transition region between the two flow regimes is covered by the so-called Space-Shuttle bridging function, Eqs. (3.249) and (3.250). From the results shown, it is clear that the waverider has indeed a large L/D (around 14 for $\alpha = -1°$). These results match well with the data provided by Sandlin and Pessin (1993).

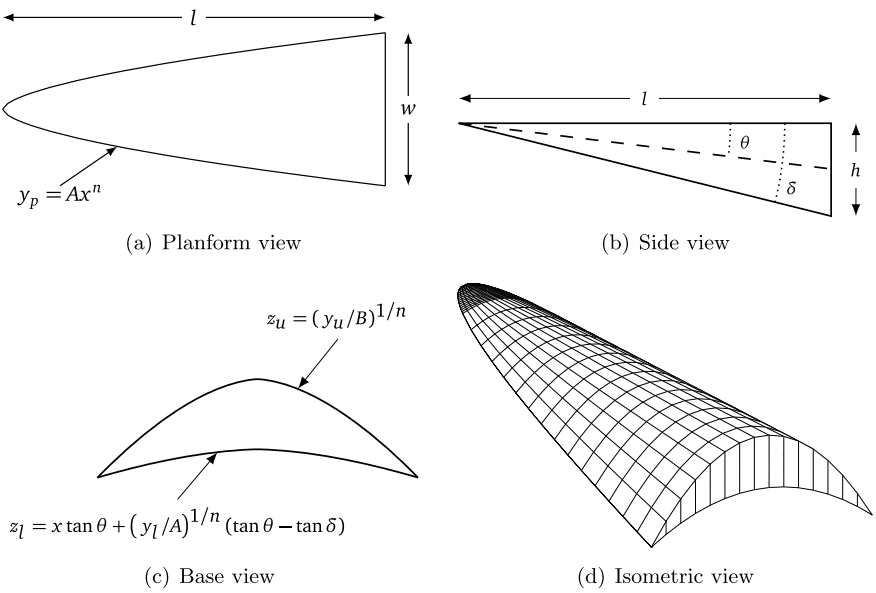


Fig. 11.57 Geometry definition of a generic waverider; the isometric view shows a waverider with $w = 2.0$ m, $l = 5.0$ m, $n = 0.5$, $\theta = 5.0°$, and $\delta = 9.0°$

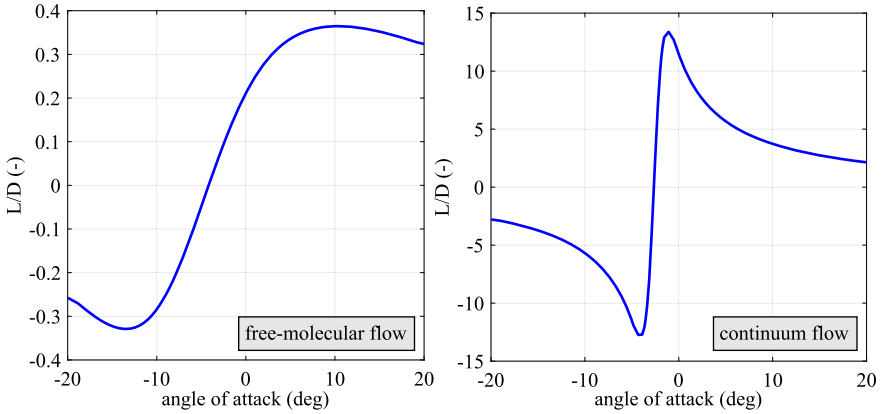


Fig. 11.58 Lift-to-drag ratio as a function of the angle of attack for the selected waverider in free molecular flow ($V = 8{,}000$ m/s), obtained with SPARTA, and continuum flow ($M = 40$), obtained with the modified Newtonian method

To evaluate the aerodynamic forces that are input to the equations of motion, it is necessary to determine the reference area and the vehicle mass. Based on values found in literature, a vehicle mass of 800 kg has been selected. The reference area required to determine the aerodynamic forces can be determined using

$$S_{ref} = \frac{wl}{n+1} \tag{11.261}$$

and is 6.67 m^2 for the selected waverider. Unfortunately, for an aerogravity assist by Mars, due to the very thin atmosphere, this reference area does not provide a sufficiently large lift to skip out of the atmosphere. Therefore, just for the sake of the current study, it was decided to increase the reference area by a factor of 10. This would, taking into account the same aspect ratio of 2:5, result in a waverider with $l = 15.8$ m and $w = 6.32$ m. As it turns out to be, such dimensions are beyond the current launcher nose-cone dimensions. However, redesign of the vehicle and/or the use of an inflatable aeroshell could increase the required lift to the required level as well. Such redesign would not serve the nature of the study, i.e., the analysis of mission aspects, though, and is therefore not considered.

The heat load on the vehicle is evaluated using a combination of convective and radiative heat flux. The former is the simplified cold-wall model of Chapman, which has been discussed at length in Chap. 5. The flux due to radiation of the gas is the model developed by Tauber and Sutton (1991), and has been described in more detail in Sect. 3.6.3. It is noted that no TPS design has been included in the waverider design. It may be obvious from the geometry definition in Fig. 11.57, that the nose and leading edges are sharp, and will lead to very high heat fluxes. It could be that some form of active cooling is required, but also this is not pursued any further either. Therefore, to have a realistic bound on the heat flux, a reference radius of 1 m is

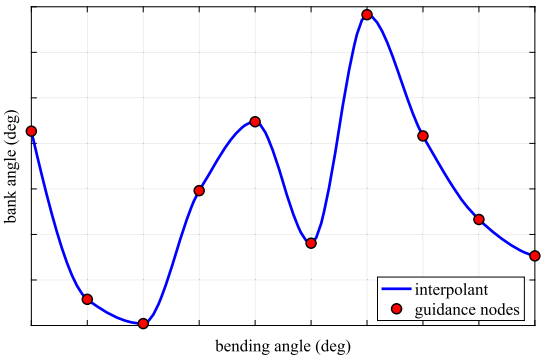


Fig. 11.59 Generic guidance profile (bank angle): nodes connected with Hermite-spline interpolant

considered during the mission analysis. This is in agreement with the value of nose radii that have been found in literature (Casoliva et al. 2008; Lohar et al. 1994).

The optimisation of the trajectory (i.e., translational motion only) is done using a global, heuristic optimisation algorithm based on differential evolution.[24] Originally, the objective of this optimisation was to find the guidance commands, here the commanded angle of attack, α_c, and bank angle, σ_c, but it was found that setting the angle of attack to the value for maximum L/D would give the best results, also in terms of design space. One way to define guidance commands for the bank angle is based on node control (Mooij and Hänninen 2009). In this concept, $\mathcal{N}$ nodes are considered as a function of an independent variable. For each node, there is a corresponding value for the guidance variable. To obtain continuous guidance commands, the nodes are connected using an interpolation method. To prevent overshooting the guidance nodes, a Hermite-spline interpolator was used (see Appendix C.4.3 for an explanation of this interpolation method). An example of the concept is shown in Fig. 11.59.

One option for the independent variable, used by Mooij and Hänninen (2009), is the specific total energy of the vehicle, which, in case of the studied entry mission, would be $\hat{E} = 1$ at the entry interface and close to $\hat{E} = 0$ towards the end. For the aerogravity assist it is not clear how much the energy will reduce during the atmospheric passage. Suppose now that the guidance nodes are spaced between $\hat{E} = 1$ and $\hat{E} = 0$, where $\hat{E} = 1$ is the specific energy at the start of the trajectory and $\hat{E} = 0$ is the energy level where the vehicle has no potential or kinetic energy. As the goal of an aerogravity assist is to leave the planet again, the energy level at the end of the trajectory will not be zero, which means that a part of the defined nodes is not used. It would, of course, be possible to space the control nodes between $\hat{E} = 1$ and the energy level at the end of the trajectory $\hat{E}_f$, but this would require to know the specific energy after the manoeuvre before the optimisation is started.

[24] The applied algorithm is the Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) (Liu et al. 2010), which is part of ESA's open-source Parallel Global Multiobjective Optimizer (PaGMO) toolbox. The latest version of this toolbox, PaGMO2 (Biscani and Izzo 2020), can be downloaded from https://esa.github.io/pagmo2/. Date last accessed: September 10, 2021.

Even though one could estimate $\hat{E}_f$ by simple approximations, this was found to be difficult nonetheless. Furthermore, if $\hat{E}$ would be estimated incorrectly, the specific energy could decrease below $\hat{E}_f$, for which no nodes are defined.

Therefore, it is decided not to use specific energy as an independent variable, but a more appropriate independent variable, i.e., the atmospheric bending angle, θ. This angle is defined to be the angle between the position vector at t_0, given by $\boldsymbol{X}_0^I$, and the position vector at a time $t > t_0$, given by $\boldsymbol{X}^I$:

$$\cos\theta = \frac{\boldsymbol{X}_0^I \cdot \boldsymbol{X}^I}{\|\boldsymbol{X}_0^I\| \, \|\boldsymbol{X}^I\|} \tag{11.262}$$

This variable will not change as drastically as the specific energy, and even if we do not know the final bending angle, it is clear that it will never be more than 180°, which would be a trajectory in exactly opposite direction as from where the spacecraft came (if this would be necessary, it would be better to approach the planet from the other side). Therefore, the end node will be better defined (constrained) than the specific energy.

In addition to the guidance nodes also the initial flight-path angle is used as decision variable. The selected initial flight-path angle will have a large impact on the resulting trajectory and the encountered loads. A shallow entry into the atmosphere can potentially result in a skipping flight, with no effect on the bending angle. Entering the atmosphere at a too steep angle, on the other hand, will result in an entry without the possibility to leave the atmosphere again. In fact, a similar discussion was held in Sect. 11.2.2, where the full lift-up (bank angle $\sigma = 0°$) and full lift-down ($\sigma = 180°$) trajectories were introduced as limit trajectories for aerocapture. Due to the large velocity in combination with the curvature of the trajectory, an aerogravity assist is typically flown with $\sigma = 180°$. However, depending on the magnitude of the lift force, it may be that a smaller bank angle would do, which would also allow for some manoeuvring to keep tracking the (constant) cruise altitude.

The division between a feasible (or *controllable*) and an infeasible entry is shown in Fig. 11.60. The area that would result in a feasible entry for $\sigma \leq 180°$ is shaded green, as these combinations of flight-path angle and velocity would guarantee a trajectory that is controllable. The combinations of flight-path angle and velocity that would result in a bank-angle profile of $\sigma = 180°$, but an even larger downwards lift force would be required to prevent the vehicle from skipping out, are shaded red. This region is infeasible as the vehicle is effectively uncontrollable. Intuitively, one would want to avoid the area below the curve in Fig. 11.60, because this would result in an entry. However, the figure was generated by keeping the bank angle set to 180°, such that the curve represents the constraint line. It also means, in the area below the curve, that even though the downward component of the lift vector pulls the vehicle to the surface of the planet, by reducing the bank angle, the downward component of the lift vector is also decreased, which will avoid the entry. Thus, for a combination of V and γ below the curve, the optimiser will be able to find a guidance profile with $\sigma \leq 180°$, which is exactly the sort of bank-angle profile that is desired.

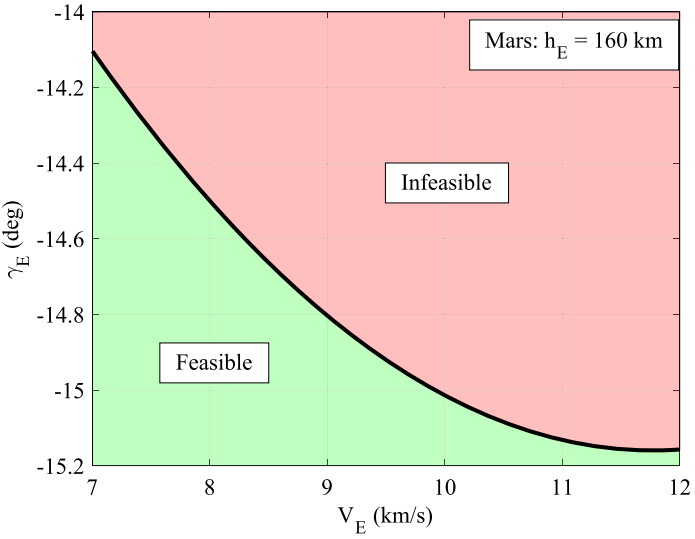


Fig. 11.60 Allowable velocity and flight-path angle combinations based on a Monte Carlo analysis with $\sigma = 180°$ and an initial altitude of 160 km for Mars

On a final note, it is also expected that small variations in the initial flight-path angle can greatly influence the trajectory, as we have seen in Sect. 5.5.5, while discussing the skipping flight of Apollo. Earlier, when we discussed the velocity loss due to drag, it was already shown that flying a constant altitude is not as easy as it may seem, and that the condition of $\gamma = 0°$ to guarantee that is not a trivial solution to be found by the optimiser. Careful tuning of the applied algorithms should be done to prevent excursions away from this conditions.

The objectives of the optimisation problem are to simultaneously maximise the atmospheric bending angle, θ, and minimise the offset between final inclination and target inclination, Δi. With the bank angle limited to the range between 0 and 180°, the lateral motion is controlled by bank reversals that are executed when $\Delta i > 3°$. Finally, operational (= path) constraints considered are a maximum mechanical load of 15 g and a maximum heat flux of 500 W/cm^2 (this heat flux is a combination of the radiative and convex heat flux, see for more details Sect. 3.6). Note that these thermo-mechanical constraints are very demanding, which is also the primary explanation why aerogravity assists have not yet been performed.

11.3.5.2 Results

In the original study two different analyses were performed, i.e., the first one to determine the largest achievable bending angle, and the second one to re-analyse the Rosetta gravity assist if it had been possible to replace it with an aerogravity assist. A summary of the outcome of either analysis will be presented below. The

focus of the analysis will be on aerogravity assists around Mars; even though this planet may not provide the best possible performance boost, it would be a good candidate when it concerns the maximum thermo-mechanical loads on the vehicle, a conclusion also drawn by McRonald and Randolph (1992).

Achievable Bending Angle

To study the maximum achievable bending angle will not directly give insight into the maximum possible heliocentric velocity or inclination change, but it provides insight in the possibilities of an aerogravity assist around Mars. In any case this may prove helpful for interplanetary mission designers. This part of the study has been performed for four different entry velocities (8.0 km/s, 9.0 km/s, 10.0 km/s and 11.0 km/s). Each optimisation run was performed four times with different initial random seeds to prevent that a solution would get stuck in a local optimum; the initial population size was set to 100. The four final populations were successively combined and per initial velocity a single Pareto front was obtained, the result of which is shown in Fig. 11.61.

What the figure does not show is that depending on the random seed all four analysed initial velocities suffered from premature convergence, exhibiting a significant underperformance and confirming the need for repetitive batches. Nonetheless, with the combination of the four initial seeds a decent Pareto front has been obtained for each of the initial velocities. Inspecting the results it can be observed that each initial velocity can achieve a zero inclination offset, albeit at different maximum bending angles. In terms of this angle, the aerogravity assist with $V_{s,i} = 9$ km/s is a clear winner. Apparently it has the right speed to overcome the drag loss during its extended stay in the atmosphere. The AGA with $V_{s,i} = 8$ km/s does not have sufficient kinetic energy to extend its atmospheric stay, whereas the other two, i.e., $V_{s,i} = 10$ and 11 km/s are flying too fast to achieve a longer equilibrium flight. This is particularly true for the AGA with $V_{s,i} = 11$ km/s, as its maximum bending angle is θ_{max} is only just over 125°, but with a penalty on the inclination offset. With zero offset, $\theta_{max} \approx 105°$. It is noted that in the Pareto fronts, there are multiple clustered points. Refining the optimisation may yield a better spread over the fronts; alternatively, different optimisation methods could be used. However, such a study is outside the scope of this case study.[25]

From each of the Pareto fronts one individual has been selected that achieved the maximum atmospheric bending angle for that initial velocity. For these individuals, the actual trajectories are shown in Figs. 11.62 through 11.64. The optimisation algorithm found trajectories that do not always have a constant altitude (Fig. 11.62,

[25] Note that with this kind of heuristic optimisation and no mathematical proof that the global optimum has been found, there is always a possibility that in some case the optimiser has converged to a local solution. Despite the fact, that the optimisation process was run four times with different random seeds, one should be careful to draw black and white conclusions. Future research should confirm this, for instance, by running the same problem with different optimisers, such as a particle-swarm or ant-colony optimiser. Because the used Pagmo optimisation toolbox includes these and more algorithms, this should not be too great an effort to do. Since it does not really serve the nature of the current discussion, it is left as a future direction of the research.

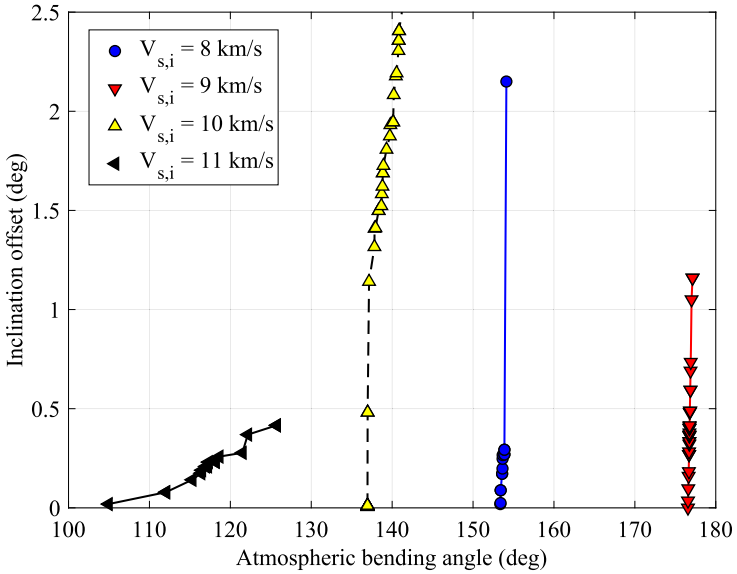


Fig. 11.61 Combined Pareto fronts (four random seeds) after optimisation for Mars, for four different $V_{s,i}$

which has already been identified earlier when discussing the velocity loss. These dips in the atmosphere, especially for initial velocities 8.0 km/s and 11.0 km/s, are the result of small changes in flight-path angle, as is shown in Fig. 11.64. These altitude excursions result in a larger density and also in an increase of the *g-load* and heat flux, clearly noticeable in Fig. 11.63. As these dips are undesirable, one should impose tighter constraints during the cruise phase to prevent this from happening. It is clear that due to the lower velocity and consequently lower dynamic pressure, the trajectory starting from $V_{s,i} = 8$ km/s suffers the most from this, because due to the lower aerodynamic lift correcting the altitude deviations is more difficult.

In general, though, the larger the entry velocity the larger the thermo-mechanical loads, due to their direct dependency on density and velocity. Thus, for practical purposes, one wants to keep the entry velocity as low as possible, for the vehicle to survive the high loads. It does mean that the vehicles need to have a dedicated thermal protection system. A low entry velocity also has a downside, though. In Fig. 11.62 the orbit's eccentricity is shown. To be able to escape the planet's gravity well, the condition of $e > 1$ should always be met (as $e < 1$ means that the orbit is elliptical, and thus closed, and the vehicle has been captured by the planet). For an initial velocity of 8.0 km/s, the maximum atmospheric bending angle is approximately 153°, but after the atmospheric manoeuvre, the eccentricity of the orbit is only $e = 1.029$. Thus, it can be concluded that for low arrival velocities the limiting constraint is $e > 1.0$, and not the heat-flux or *g-load* constraints. For the larger entry velocities, e_{final} increases to 2.260, 3.052 and 3.885, for $V_{s,i} = 9.0$ km/s, 10.0 km/s and 11.0

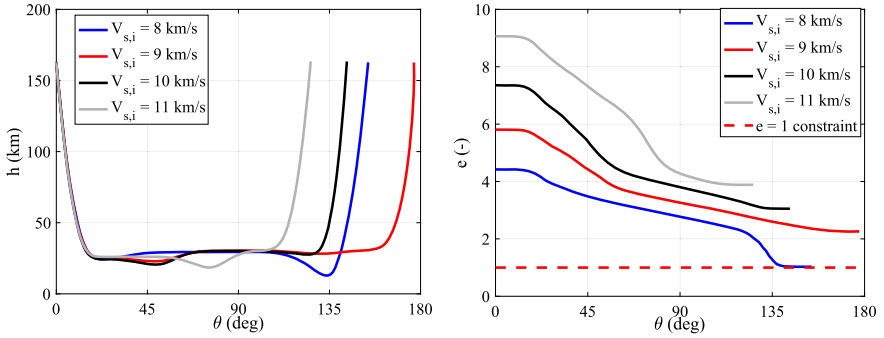


Fig. 11.62 Selected Pareto trajectories for four different $V_{s,i}$ around Mars: altitude and eccentricity as a function of bending angle

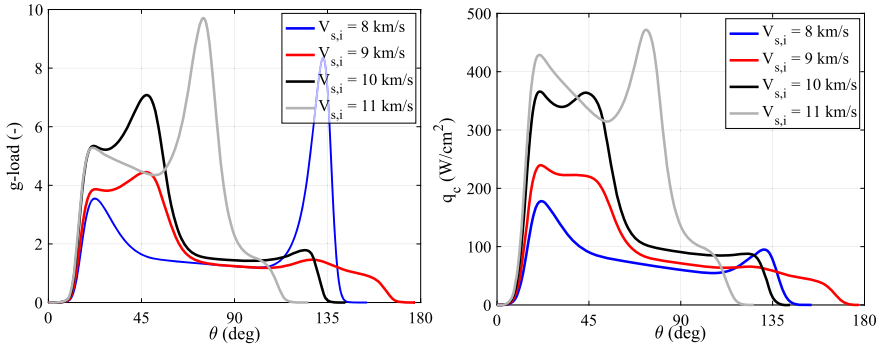
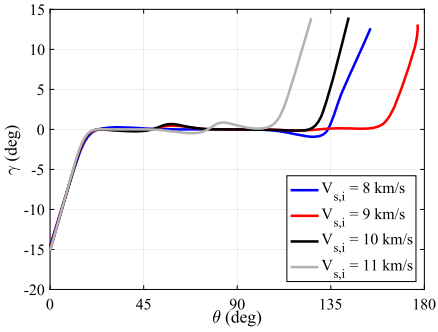


Fig. 11.63 Selected Pareto trajectories for four different $V_{s,i}$ around Mars: *g-load* and heat flux as a function of bending angle

Fig. 11.64 Selected Pareto trajectories for four different $V_{s,i}$ around Mars: flight-path angle as a function of bending angle



km/s, respectively. This confirms that the orbital energy is still sufficiently large to leave the planet. However, in that case the loads on the vehicle become the driving constraints.

The results found during the trajectory optimisation of the aerogravity assist are now compared with a gravity assist, using the same initial conditions as before. Of course, for any planet with an atmosphere such a GA cannot be executed, because at close passage the atmosphere would be entered. Yet, to be able to show the advantage of the AGA, for the sake of this comparison the atmosphere model will be disabled while simulating the GA. In that way, the contribution of only the atmosphere can be investigated. The results of the analysis, for each of the four initial velocities, is shown in Table 11.10, where it should be noted that the angle over which the velocity is bent is indicated by δ, rather than the atmospheric bending angle θ. This has been done, because the rotation of the velocity vector is the major purpose of an (aero)gravity assist.

For all four cases of the AGA, the hyperbolic excess velocity is decreased with respect to a GA due to energy dissipation as a result of drag. However, for an entry velocity of 8.0 km/s, the hyperbolic excess velocity is just over 600 m/s, which is barely enough to leave the gravity well of Mars. This is in line with the found eccentricity of the departure leg of the trajectory, mentioned above. In the worst case, the velocity bending angle is increased by 490%, while for an entry velocity of 11 km/s, this angle is increased by 818%. Deviations up to 6.5% in the final inclination of the trajectory can be observed. However, if necessary, this can be corrected by means of a propulsive manoeuvre once the vehicle has left the atmosphere. For instance, in case that for this worst-case scenario a propulsive manoeuvre would be performed at V_∞^+, a total ΔV of 250 m/s would be required to get back to the original orbital plane. If this ΔV would be considered too large, a smaller Δi constraint should

Table 11.10 Comparison of gravity assist (GA) and aerogravity assist (AGA) at Mars

	Variable	Unit	GA	AGA	% Difference
8.0 km/s	δ	deg	26.15	154.39	+490.48
	V_∞^+	km/s	6.54	0.61	−90.70
	i	deg	43.71	42.85	−1.97
9.0 km/s	δ	deg	19.83	178.53	+800.51
	V_∞^+	km/s	7.75	3.96	−48.86
	i	deg	43.85	44.66	1.84
10.0 km/s	δ	deg	15.63	140.57	+799.32
	V_∞^+	km/s	8.91	5.06	−43.20
	i	deg	43.96	41.12	−6.47
11.0 km/s	δ	deg	12.67	116.31	+818.20
	V_∞^+	km/s	10.04	6.00	−40.26
	i	deg	44.05	45.42	3.09

Table 11.11 Modified initial conditions for pseudo-Rosetta's swing-by around Mars

Parameter	Adjusted	Original
Altitude (km)	162.6	421.7
Velocity (km/s)	10.27	10.27
Longitude (deg)	−37.90	−33.50
Latitude (deg)	45.07	41.99
Flight-path angle (deg)	−14.24	−15.89
Heading angle (deg)	−78.05	−77.04

be introduced in the optimisation problem, but this would automatically result in a larger number of bank reversals. This trade-off will then enter the mission-design problem.

Rosetta Aerogravity Assist

The second analysis that was performed focussed on the benefit of an atmospheric manoeuvre for a more realistic mission scenario. For this analysis, the Rosetta mission was used as a baseline. Rosetta's journey took ten years, and required three gravity assists at Earth and one at Mars to obtain sufficient orbital energy to rendezvous with the comet. Once arrived there, the lander called Philae descended towards the surface of the comet to perform scientific measurements (Munoz et al. 2012).

In the actual mission, it would not have been possible for Rosetta to perform an aerogravity assist, because this would require a vehicle with a large lift-to-drag ratio. Therefore, instead of using Rosetta's original geometry, the same waverider as before has been used. Therefore, only the trajectory of the Rosetta mission is considered, such that realistic arrival conditions could be obtained. The trajectory of Rosetta can be downloaded from the ESA website,[26] and is contained in a so-called SPICE kernel. With a proper interface it is possible to determine the Cartesian state of a planetary body or vehicle with respect to any body in the Solar System, given that the appropriate kernels are loaded. In addition to Rosetta's SPICE kernel, the DE413 kernel was required to determine the position of Mars.[27] The initial conditions that were derived from these SPICE kernels, were slightly modified to ensure that the trajectory would enter the atmosphere. Table 11.11 shows the used initial conditions.

Rosetta's Mars gravity assist took place on 25 February, 2007, during which the spacecraft approached the Martian surface to approximately 250 km. The Martian gravity assist was used to *decelerate* Rosetta: the heliocentric velocity was reduced by 2.32 km/s. The heliocentric inclination remained more or less constant: 26.97° before the gravity assist and 25.13° after the gravity assist. To formulate the aerogravity assist of the *pseudo Rosetta*, based on the actual Rosetta swing-by, one can define

[26] SPICE for Rosetta: http://www.cosmos.esa.int/web/spice/spice-for-rosetta, last accessed: March 1, 2023.

[27] JPL Planetary and Lunar Ephemerides: https://ssd.jpl.nasa.gov/planets/eph_export.html, last accessed: March 1, 2023.

many different objectives. For example, a requirement could be put to achieve a specific atmospheric bending angle, to adjust the heliocentric inclination to a particular target, or to increase or decrease the heliocentric velocity. Within the scope of this case study, it is not feasible to investigate each of these objectives, so only one objective is selected to assess the influence an aerogravity assist. As the goal of Rosetta's swing-by at Mars was to reduce its heliocentric velocity, it will be investigated what the velocity decrease could have been if use was made of an aerogravity assist. To achieve this, instead of maximising the atmospheric bending angle, θ, the heliocentric velocity once the vehicle has left the atmosphere will be minimised.

It is noted, though, that the comparison is not completely fair, because of the earlier mentioned change in spacecraft geometry. In addition, since the initial conditions of the Rosetta trajectory are slightly modified, comparing the aerogravity assist trajectory with the original Rosetta trajectory might result in some differences that are not caused by the atmospheric manoeuvre. To make this a fair comparison, a gravity assist trajectory is computed with the modified initial conditions from Table 11.11, but ignoring the effect of the atmosphere. Where for a gravity assist, the velocity vector is rotated over an angle of almost 16°, the atmospheric manoeuvre increased this angle to 65° for an aerogravity assist. The atmospheric manoeuvre, however, comes at a price: the atmospheric drag results in a decrease in magnitude of the departure velocity. The magnitude of the hyperbolic excess velocity of the arriving trajectory is $V_\infty^- = 8.829$ km/s, whereas this magnitude is decreased to $V_\infty^+ = 7.558$ km/s. The aerogravity assist increased the velocity change from 2.308 km/s to 6.160 km/s, which is an increase of 167%.

To conclude this brief discussion, Figs. 11.65 through 11.67 show the corresponding trajectory and some associated parameters. Figure 11.65 shows that the vehicle will travel at a more or less constant altitude for approximately 300 s, without the large altitude excursions that we saw earlier. This is also visible in Fig. 11.67, where the flight-path angle remains close to 0° for the same time interval. The velocity (Fig. 11.65) decreases as a result of energy dissipation due to drag. With a maximum

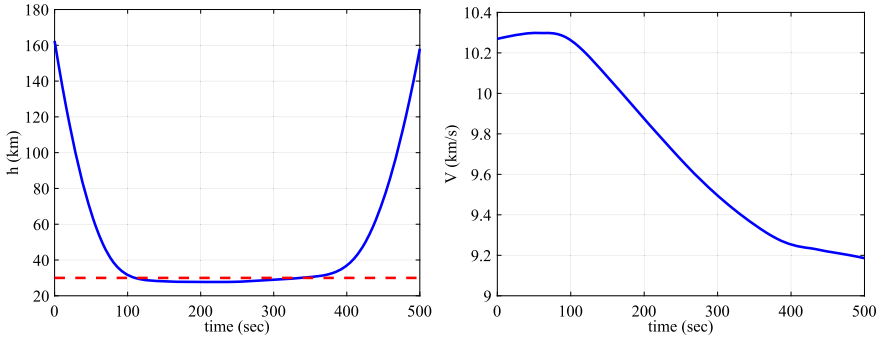


Fig. 11.65 Aerogravity assist trajectory for pseudo-Rosetta around Mars: altitude (left) and relative velocity (right) as a function of time

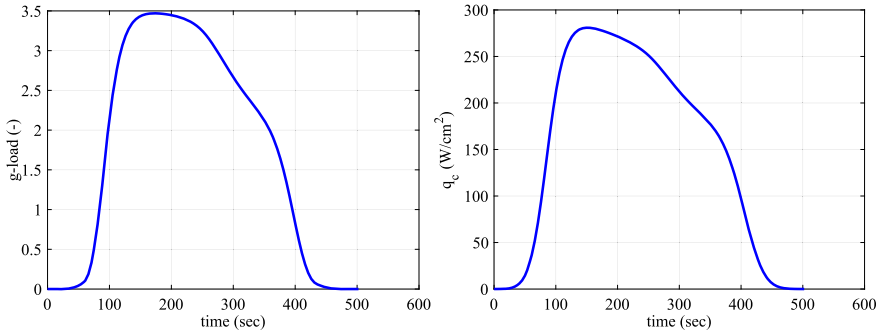


Fig. 11.66 Aerogravity assist trajectory for pseudo-Rosetta around Mars: *g-load* (left) and heat flux (right) as a function of time

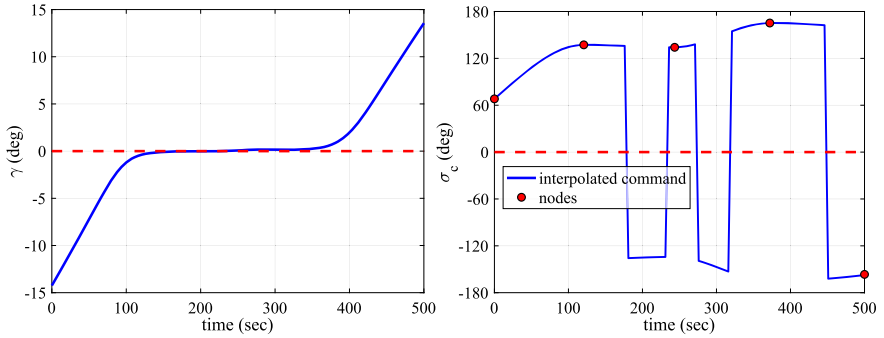


Fig. 11.67 Aerogravity assist trajectory for pseudo-Rosetta around Mars: flight-path angle (left) and commanded bank angle (right) as a function of time

mechanical load of 3.5 g and a maximum heat load of 280 W/cm^2 (Fig. 11.66), the found trajectory is well within the imposed constraints of 15 g and 500 W/cm^2, given, of course, that the structural design of the vehicle can sustain this (which may not be the case for the original Rosetta). Furthermore, it can be seen in Fig. 11.67 that the optimisation algorithm found five bank reversals to keep the trajectory in the same orbital plane. Ideally, only one bank reversal should be performed. This could be achieved by increasing the allowed Δi, which was the condition for a bank reversal. However, as was shown above, the optimisation algorithm was not always able to ensure that the final orbital inclination would be identical to the initial one. Therefore, a trade-off should be made between the number of bank reversals performed, the allowed orbital inclination offset and the required ΔV to correct for this offset. Alternatively, one could study and adapt the reversal analysis of , who has developed an algorithm for precision landing with carefully planned reversals for a gliding entry vehicle targeted to a runway on Earth, or the research by Miele et al. (1990), Miele and Wang (1996), who studied optimal trajectories and robust predictor-corrector guidance for aeroassisted orbital transfer.

Problems

Judge Whether the Following Statements are TRUE or FALSE (*Problems* 11.1 *to* 11.14).

11.1. Aerogravity assists use a combination of aerodynamic lift and gravitational forces to alter a spacecraft's trajectory.
11.2. Aerogravity assists are not suitable for precise trajectory adjustments and require significant margin for error.
11.3. During an aerocapture manoeuvre, the spacecraft enters the planet's atmosphere at a shallow angle, using aerodynamic drag to decelerate and capture into orbit.
11.4. Aerogravity assists are only applicable to spacecraft traveling within Earth's atmosphere.
11.5. Aerocapture missions require precise navigation and guidance systems to ensure the spacecraft enters the atmosphere at the correct angle and trajectory for a successful capture.
11.6. Because of the shallow entry angles, aerocapture spacecraft do not require heat-resistant thermal-protection systems capable of withstanding the high temperatures generated during atmospheric entry.
11.7. Aerogravity assists are primarily used to conserve propellant and reduce mission duration.
11.8. Aerogravity assists are primarily used to decrease a spacecraft's velocity and facilitate its capture into orbit around a celestial body.
11.9. Gravity and aerogravity assists can only be performed once during a spacecraft's mission trajectory.
11.10. A gravity assist is a safer manoeuvre compared to an aerogravity assist.
11.11. Of the two, a gravity assist is more efficient than an aerogravity assist, because no additional fuel is required to correct for errors induced by the atmosphere.
11.12. Aerocapture missions are often designed to take advantage of multiple aerocapture passes to gradually lower the spacecraft's orbit or adjust its trajectory as needed.
11.13. Aerocapture manoeuvres are typically conducted autonomously by the spacecraft's onboard systems, with minimal intervention or control from ground-based operators.
11.14. Aerogravity assists are solely dependent on the gravitational pull of the celestial body and do not consider atmospheric conditions.

Hohmann Trajectory (*Problems* 11.15 *to* 11.19).

Consider a Hohmann transfer from an inner Planet 1 to an outer Planet 2 (later to be substituted by Earth and Neptune; data to be used: $\mu_{Sun} = 1.3271 \cdot 10^{11}$ km^3/s^2, distance Earth-Sun = 1 au = $149.6 \cdot 10^6$ km, and distance Neptune-Sun = 30.1 au).

11.15. Derive the following general equations for the epoch of departure, t_1, and the epoch of arrival, t_2:

$$t_1 = t_0 + \frac{\theta_2(t_0) - \theta_1(t_0) + n_2 T_H - \pi}{n_1 - n_2} \tag{11.263}$$

where t_0 is a common reference epoch, T_H is the transfer time in a Hohmann orbit (equivalent to $T_{transfer}$ in Intermezzo 11.3), n_1 and n_2 are the mean motion of the two planets, and θ_1 and θ_2 are the true anomalies of the planetary positions, respectively. Assume circular, coplanar orbits for both planets. Make a drawing for clarity.

11.16. Consider a Hohmann transfer from Earth to Neptune. What is the transfer time?

11.17. Assuming that on January 1, 2010, $\theta_{Earth} = 70°$ and $\theta_{Neptune} = 240°$, what would be the epoch of departure (expressed in days with respect to this January 1?

11.18. What would be the arrival epoch?

11.19. (*) Can we change the launch date and still fly a perfect Hohmann transfer? If so, how? A qualitative answer is sufficient.

Gravity Assist (*Problems* 11.20 *to* 11.25)

Consider a gravity assist along Jupiter. The following data may be used: $\mu_{Sun} = 1.3271{\cdot}10^{11}$ km^3/s^2, $\mu_{Jupiter} = 1.267{\cdot}10^8$ km^3/s^2, and the distance Jupiter-Sun = 5.20 au, with 1 au = $149.6{\cdot}10^6$ km.

11.20. Assuming a relative velocity (i.e., the hyperbolic excess velocity) when entering the sphere of influence of 10 km/s, and a nearest passing distance with respect to the centre of Jupiter of 200,000 km, compute the value of the pericentre velocity of this hyperbolic orbit.

11.21. Compute the value of the semi-major axis and the eccentricity of this orbit.

11.22. Compute the heliocentric velocity of Jupiter, considering that it is in circular orbit around the Sun.

11.23. Draw the diagram with the velocities of Jupiter and the spacecraft with respect to a heliocentric system, and the velocity of the spacecraft with respect to Jupiter. Do this for the situation before and after the gravity assist.

11.24. Using the relation between the bending angle, δ, and the true-anomaly limit, θ_{lim}, compute the heliocentric velocity of the satellite before and after the encounter, respectively. Assume that the orientation of the hyperbola around Jupiter is symmetric with respect to Jupiter's velocity around the Sun.

11.25. Compute the energy gain caused by the gravity assist.

Aerogravity Assist (*Problems* 11.26 *to* 11.33)

A vehicle with an L/D of 3 enters the sphere of influence of Venus with incoming $V_\infty^- = 8$ km/s, and performs an aerogravity assist at $h_p = 200$ km. The following data may be used: $\mu_{Sun} = 1.3271{\cdot}10^{11}$ km^3/s^2, $\mu_{Venus} = 324{,}858.592$ km^3/s^2, $R_e = 6{,}051.8$ km, and the distance Venus-Sun = 0.723332 au (= $108.21 \cdot 10^6$ km).

11.26. Calculate Venus's orbital velocity and the (local) circular velocity at h_p.

11.27. Given V_∞^-, calculate the relative velocity with which the vehicle enters the atmosphere, $V_{s,i}$, as well as the exit velocity, $V_{s,f}$. Assume an atmospheric bending angle, θ, of 60°.
11.28. To complete the set of velocities, calculate the outgoing hyperbolic excess velocity.
11.29. Compute the velocity loss, both in terms of relative velocity and hyperbolic excess velocity.
11.30. Repeat the calculations for $L/D = 4$ and $L/D = 5$. What can you conclude?
11.31. Calculate the maximum bending angle for the three values of L/D.
11.32. How does the situation change when $V_\infty = 6$ or 10 km/s?
11.33. Explain the concept of the *equivalent radius*, and calculate its value for the different cases mentioned above.

References

Alpert HS, Saunders DA, Mahzari M, Monk JD, White TR (2022) Inverse Estimation of Mars 2020 entry aeroheating environments using MEDLI2 flight data, AIAA-2022-0550, SciTech Forum, San Diego, CA & Virtual, 3–7 Jan 2022

Armellin R, Lavagna M, Starkey RP, Lewis MJ (2007) Aerogravity-Assist Maneuvers: coupled trajectory and vehicle shape optimization. J Spacecr Rocket 44(5):1051–1059

Atkins KL, Brownlee DE, Duxbury TC, Yen C, Tsou P, Vellinga JM (1997) STARDUST: discovery's inter-stellar dust and cometry sample return mission. In: IEEE Aerospace conference, vol 4, Piscataway, NJ, United States, Feb 1997, pp 229–245

Belbruno E (2004) Capture dynamics and chaotic motions in celestial mechanics: with applications to the construction of low energy transfers. Princeton University Press

Biscani F, Izzo D (2020) A parallel global multiobjective framework for optimization: PaGMO. Journal of Open Source Software 5, Nr 53:2338

Bonfiglio EP, Logunski JM, Vinh NX (2000) Automated design of Aerogravity-Assist trajectories. J Spacecr Rocket 37(6):768–775

Braun

Casoliva J, Lyons DT, Wolf AA, Mease KD (2008) Robust guidance via a predictor-corrector algorithm with drag tracking for Aero-Gravity assist maneuvers, AIAA-2008-6816. In: AIAA, guidance, navigation and control conference, Honolulu, Hawaii, 18–21 Aug 2008

Cassel AM, Winter MW, Allen Jr GA, Grinstead JH, Antimisiaris ME, Albers J, Jenniskens P (2011) Hayabusa re-entry: trajectory analysis and observation mission design, AIAA-2011-3330. In: 42nd AIAA thermophysics conference, Honolulu, Hawaii, 27–30 June 2011

Cerimele CJ, Gamble JD (1985) A simplified guidance algorithm for lifting Aeroassist orbital transfer vehicles, AIAA-1985-0348. In: AIAA 23rd aerospace sciences meeting, Reno, NV, 14–17 Jan 1985

Chabot T, Bonnéry G, Val Serra S, Kemble S (2013) Robust autonomous Aerobraking strategies. In: 5th European conference for Aeronautics and space sciences (EUCASS), Munich, Germany, 1–5 July 2013

Chambre PA, Schaaf SA (2017) Flow of rarefied gases. Princeton University Press, NJ, 2017, Chap 2

Colin L (1980) The Pioneer Venus program. J Geophys Res 85(A13):7575–7598

Croom CA, Tolson RH (1994) Using Magellan attitude control data to study the Venusian atmosphere and various spacecraft properties. Adv Astronaut Sci 87:451–467
Cruz MI (1979) The Aerocapture vehicle mission design concept, AIAA-1979-0893. In: Conference on advanced technology for future space systems, Hampton, VA, 8–10 May 1979
Damiani S, Garcia J, Guilanyà R, Muñoz P, Müller M (2015) Flight dynamics operations for Venus express aerobraking campaign: a successful end of life experiment. In: 25th international symposium on space flight dynamics, Munich
Dec J, Amundsen R (2003) A thermal analysis approach for the Mars Odyssey spacecraft's Solar array, AIAA-2003-3764. In: 36th AIAA thermophysics conference, Orlando, FL, 23–26 June 2003
Denis M, Schmitz P, Sangiorgi S, Guilanyà R, Mitschdoerfer P, Montagna M, Renault S, Kutrowski N (2018) Thousand times through the atmosphere of Mars: Aerobraking the ExoMars trace gas orbiter, AIAA-2018-2713. SpaceOps Conference, Marseille, France, 28 May–1 June 2018
Doody D (1995) Aerobraking the Magellan spacecraft in Venus orbit. Acta Astronautica 35:475–480
Dutta S, Karlgaard CD, Kass D, Mischna M, Villar III GG (2022) Post-flight analysis of atmospheric properties from Mars 2020 entry, descent, and landing, AIAA-2022-0422. SciTech Forum, San Diego, CA & Virtual, 3–7 Jan 2022
Engelsma J, Mooij E (2020) Aerocapture mission analysis, AIAA-2020-1740. SciTech Forum: AIAA atmospheric flight mechanics conference, Orlando, FL, 6–10 Jan 2020
Gamble JD, Cerimele CJ, Moore TE, Higgins J (1988) Atmospheric guidance concepts for an Aeroassist flight experiment. J Astronaut Sci 36:45–71
Gurley JG (1993) Guidance for an Aerocapture maneuver. J Guid Control Dyn 16(3):505–510
Hess JR, Mooij E (2017) Node control and numerical optimization of Aerogravity-Assist trajectories, AIAA-2017-0471. SciTech Forum: AIAA atmospheric flight mechanics conference, Grapevine, TX, 9–13 Jan 2017
Johnson WR (2002) Optimization of atmospheric flythrough for Aerogravity-Assist trajectories, AAE508 Report
Johnson M, Willcockson W (2003) Mars Odyssey Aerobraking: the first step towards autonomous Aerobraking operations. IEEE Aerosp Conf 8:3503–3510
Kawaguchi J, Fujiwara A, Uesugi T (2006) Hayabusa—its technology and science accomplishment summary and Hayabusa-2, IAC-06-A3.5.02. In: 57th international astronautical congress, Valencia, Spain, 2–6 Oct 2006
Karlgaard CD, Schoenenberger M, Dutta S, Way DW (2022) Mars entry, descent, and landing instrumentation 2 trajectory, aerodynamics, and atmosphere reconstruction, AIAA-2022-0423. SciTech Forum, San Diego, CA & Virtual, 3–7 Jan 2022
Lavagna M, Povoleri A, Finzi A (2005) Interplanetary mission design with Aero-Assisted maneuvers multi-objective evolutive optimization. Acta Astronautica 57:498–509
Lewis MJ, McRonald AD (1992) Design of hypersonic waveriders for Aeroassisted interplanetary trajectories. J Spacecr Rocket 29(5):653–660
Lewis MJ (1993) The use of hypersonic waveriders for Aeroassisted orbital manoeuvring. J British Interplanet Soc 46:11–20
Liechty DS (2007) Aeroheating analysis for the Mars reconnaissance orbiter with comparison to flight data. J Spacecr Rocket 44(6):1226–1231
Liu B, Fernandez FV, Zhang Q, Pak M, Sipahi S, Gielen G (2010) An enhanced MOEA/D-DE and its application to multiobjective analog cell sizing. In: Proceeding of the IEEE congress on evolutionary computation, Barcelona, Spain, 18–23 July 2010
Lohar FA, Mateescu D, Misra AK (1994) Optimal atmospheric trajectory for Aero-Gravity assist. Acta Astronautica 32(2):89–96
Lohar FA, Misra AK, Mateescu D (1995) Optimal atmospheric trajectory for Aerogravity assist with heat constraint. J Guid Control Dyn 18(4):723–730
Lohar FA, Misra AK, Mateescu D (1996) Optimal Aerogravity assist with heliocentric plane change. Acta Astronautica 38(4–8):445–456

Long SM, You T-H, Halsell CA, Bhat RS, Demcak SW, Graat EJ, Higa ES, Highsmith DE, Mottinger NA, Jah MK (2008) Mars Reconnaissance Orbiter Aerobraking navigation operation, AIAA-2008-3349. In: SpaceOps 2008 conference, Heidelberg, Germany, 12–16 May 2008

Lu P (2014) Entry guidance: a unified method. J Guid Control Dyn 37(3):713–728

Lu P, Cerimele CJ, Tigges MA, Matz DA (2015) Optimal Aerocapture guidance. J Guid Control Dyn 38(4):553–565

Lyons DT (1999) Aerobraking at Venus and Mars: a comparison of the Magellan and Mars Global Surveyor Aerobraking phases, AAS-99-358. In: AAS/AIAA astrodynamics conference, Girdwood, Alaska, 16–19 Aug 1999

Lyons DT (July2001) Aerobraking automation options, AAS-01-385. AIAA/AAS Astrodynamics Meeting, Quebec City, Quebec, Canada

Lyons DT (2002) Mars Reconnaissance Orbiter: Aerobraking reference trajectory, AIAA-2002-4821. In: AIAA/AAS astrodynamics specialist conference and exhibit, Monterey, CA, 5–8 Aug 2002

Lyons DT (2012) Aerobraking challenges for a combined Mars orbiter and lander in 2018, AIAA-2012-5068. In: AIAA/AAS astrodynamics specialist conference, Minneapolis, MN, Aug 2012

Lyons DT, Saunders RS, Griffith DG (1995) The Magellan Venus mapping mission: Aerobraking operations. Acta Astronautica 35:669–676

Martin JJ (1966) Atmospheric reentry: introduction to its science and engineering. Prentice-Hall Series in Space Technologies, Prentice-Hall, NJ

Marcos FA, Garrett HB, Champion KSW, Forbes JM (1977) Density variations in lower thermosphere from analysis of AE-C accelerometer measurements. Planet Space Sci 25(5):499–507

Masciarelli J, Rousseau S, Fraysse H, Perot E (2000) An analytic aerocapture guidance algorithm for the Mars Sample Return Orbiter, AIAA-2000-4116. In: AIAA atmospheric flight mechanics conference, Denver, CO, 14–17 Aug 2000

McRonald AD, Nock K (1977) Use of atmospheric drag to circularize an elliptic spacecraft orbit at Venus. In: AIAA/AAS astrodynamics specialist conference, Jackson Hole, Wyoming, pp 760–784, 7–9 Sept 1977

McRonald AD, Randolph JE (1992) Hypersonic maneuvering for augmenting planetary gravity assist. J Spacecr Rocket 9(2):216–222

McRonald AD, Randolph JE, Lewis MJ, Bonfiglio EP, Longuski J, Kolodziej P (1999) From LEO to the planets using waveriders, AIAA-1999-4803. In: 9th international space planes and hypersonic systems and technologies conference and 3rd weakly ionized-gases workshop, Norfolk, VA, 1–5 Nov 1999

Miele A, Wang T, Lee WY, Zhao ZG (1990) Optimal trajectories for the Aeroassisted flight experiment. Acta Astronautica 21(11–12):735–747

Miele A, Wang T (1996) Robust predictor-corrector guidance for Aeroassisted orbital transfer. J Guid Control Dyn 19(5):1134–1141

Mooij E, Hänninen G (2009) Distributed global trajectory optimization of a moderate lift-to-drag re-entry vehicle, AIAA-2009-5770. In: AIAA guidance, navigation, and control and co-located conferences, Chicago, Illinois, 10–13 Aug 2009

Munoz P, Budnik F, Godard B, Morley T, Companys V, Herfort U, Cacas C (2012) Preparations and strategy for navigation during Rosetta comet phase. In: Proceedings 23rd international symposium on space flight dynamics, Pasadena, CA

Murri DG, Powell RW, Prince JL (2012) Development of autonomous Aerobraking—Phase 1, NASA-TM-2012-217328. NASA Langley Research Center, Hampton, Virginia

Murri DG (2013) Development of autonomous Aerobraking—Phase 2, NASA-TM-2012-218032. NASA Langley Research Center, Hampton, Virginia

North American Aviation, Inc., Aerodynamic Data Manual for Project Apollo, NASA-CR-82907, 1 Jan 1965

Perot E, Fraysse H, Rousseau S, Berges JC (2001) Comparison of an analytical predictor corrector and a terminal point controller for the Mars sample return Aerocapture, AAAF-14-67. In: 2nd international symposium atmospheric re-entry vehicles and systems, Arcachon, 26–29 Mar 2001

Rao AV, Tang S, Hallman WP (2002) Numerical optimization study of multiple-pass Aeroassisted orbital transfer. Opt Control Appl Methods 23:215–238
Rao AV, Scherich A III, Cox S, Mosher T (2008) A concept for operationally responsive space mission planning using Aeroassisted orbital transfer, AIAA-RS6-2008-1001. In: AIAA 6th responsive space conference, Los Angeles, CA, 28 Apr–1 May 2008
Revelle DO, Edwards WN (2007) Stardust—an artificial, low-velocity meteor fall and recovery. Meteorit Planet Sci 42(3):271–299
Robinson JS, Wurster KE, Mills JC (2009) Entry trajectory and Aeroheating environment definition for capsule-shaped vehicles. J Spacecr Rocket 46(1):74–86
Rousseau S, Perot E, Graves C, Masciarelli JP, Queen E (2002) Aerocapture guidance algorithm comparison campaign, AIAA-2002-4822. In: AIAA/AAS astrodynamics specialist conference and exhibit, Monterey, CA, 5–8 Aug 2002
Sandlin D, Pessin D (1993) Aerodynamic analysis of hypersonic waverider aircraft, NASA-CR-192981, NASA Ames Research Center
Saraf A, Leavitt J, Chen D, Mease K (2004) Design and evaluation of an acceleration guidance algorithm for entry. J Spacecr Rocket 41(6):986–996
Schaub H, Junkins JL (2014) Analytical mechanics of space systems, 3rd ed, AIAA education series
Scott CF Jr et al (2022) SCIFLI airborne observation of the Hayabusa2 sample return capsule re-entry, AIAA-2022-3798. AIAA AVIATION Forum, Chicago, IL and Virtual, 27 June–1 July 2022
Siddiqi AA (2018) Beyond earth. A chronicle of deep space exploration, 1958–2016, NASA SP-2018-4041
Sigal E, Guelman M (2001) Optimal aerocapture with minimum total heat load, IAF-01-A.5.05. In: Proceedings of the 52nd international astronautical congress
Sims JA, Longuski JM, Patel MR (2000) Aerogravity-Assist trajectories to the outer planets and the effect of drag. J Spacecr Rocket 37(1):406–415
Smith JC Jr, Bell JL (2005) 2001 Mars Odyssey Aerobraking. J Spacecr Rocket 42(3):406–415
Spencer DA, Tolson RH (2007) Aerobraking cost and risk decisions. J Spacecr Rocket 44(6):1285–1293
Starkey RP, Lewis MJ (1999) Simple analytical model for parametric studies of hypersonic waveriders. J Spacecr Rocket 36(4):516–523
Strangeway RJ (1993) The Pioneer Venus Orbiter entry phase. Geophys Res Lett 20:2715–2717
Tauber M, Sutton K (1991) Stagnation-point radiative heating relations for Earth and Mars entries. J Spacecr Rocket 28(3):43–49
Tolson RH, Prince JLH, Konopliv AA (2013) An atmospheric variability model for Venus Aerobraking missions, AIAA-2013-4830. In: AIAA modeling and simulation technologies conference, Boston, MA, 19–22 Aug 2013
Uesugi K (1996) Results of the MUSES-A Hiten Mission. Adv Space Res 18(11):69–72
Vallado DA (2013) Fundamentals of astrodynamics and applications, 4th ed. Microcosm Press
Val Serra S, Bonnamy O, Witasse O, Camino O (2011) Venus Express Aerobraking. IFAC Proc Vol 44(1):715–720
Way DW, Dutta S, Zumwalt CH, Blette DJ (2022) Assessment of the Mars 2020 entry, descent, and landing simulation, AIAA-2022-0421. SciTech Forum, San Diego, CA and Virtual, 3–7 Jan 2022
Wertz JR (2009) Mission Geometry. Second Printing, Kluwer Academic Publishers, Orbit and Constellation Design and Management
Zucchelli EM (2016) Aerocapture guidance. Numerical and artificial intelligence solutions, MSc thesis, Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands, Nov 2016
Zucchelli EM, Mooij E (2018) Minimum radiative heat-load aerocapture guidance with attitude-kinematics constraints, AIAA-2018-1319. In: Guidance, navigation, and control conference (SciTech 2018), Kissimmee, FL, Jan 8–12 2018

Answers

Chapter 1—Introduction

TRUE/FALSE questions (Problems 1.1–1.10).

1.1. FALSE. The first American in space was Alan Shepard (May 5, 1961), albeit in a sub-orbital orbit only. He was followed by "Gus" Grissom on the Mercury-Redstone 4 (July 21, 1961), albeit again in a sub-orbital orbit. John Glenn was the third American (and fourth world-wide) to reach space and the first American fully orbiting Earth (February 20, 1962). For completeness: Yuri Gagarin was the first human in space, who, on April 12, 1961, completed one orbit of Earth in the Vostok 1. Gherman Titov became the second human to orbit the Earth (and the third one in space), on August, 6, 1961.

1.2. FALSE. Korolev was the chief designer of the Soviet space program in the 1950s and 1960s. The "inventor" of the rocket equation was Konstantin Tsiolkovsky.

1.3. TRUE. Explorer 1 was launched on January 31, 1958, local time.

1.4. TRUE. As of February 2024, Curiosity is the only one of the three still active.

1.5. FALSE. The V-2 (as a first stage) was combined with the 7.3 m long WAC Corporal sounding rocket (as a second stage), such that it could achieve hypersonic flight, both during launch and re-entry back into the atmosphere.

1.6. FALSE. Isaac Newton defined the physical principles of force, motion, and gravitational attraction. Kepler postulated the laws of planetary motion.

1.7. TRUE. On August 10, 1960, mid-air recovery of Discoverer-13, a small re-entry capsule with a photo canister originating from a spy satellite, was successfully achieved.

1.8. TRUE. Apollo 10's Lunar Module was fully equipped to perform a soft landing on the Moon. However, the amount of fuel onboard would not have been sufficient to take-off again and dock with the Service Module orbiting the Moon.

1.9. FALSE. Von Braun was one of the driving forces behind the Apollo Moon Program and the chief architect of the Saturn V rocket.

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1

1.10. TRUE. Goddard was one of the four "Founding Fathers of Modern Rocketry", the others being Robert Esnault-Pelterie, Konstantin Tsiolkovsky, and Hermann Oberth.

Chapter 2—Planetary Entry Environment

TRUE/FALSE questions (Problems 2.1–2.12).

2.1. FALSE. Both density and pressure follow an exponential function.
2.2. FALSE. Besides an accurate shape model of an asteroid also the density is required to provide the user with an accurate gravity field using a polyhedra model. And, since the composition may vary throughout the asteroid, this density may very well change with three-dimensional position in the interior.
2.3. FALSE. The use of a spherical-harmonics model is restricted to altitudes outside the Brillouin sphere. Some locations on the Moon are located inside this sphere, and thus the spherical-harmonics model cannot be used there.
2.4. TRUE. The atmospheric density can be affected by local (spatial) effects, as well as, for instance, the activity of the Sun. The latter is typically time-dependent, amongst others. Also an interaction with the planetary magnetic field, if present, exists.
2.5. FALSE. The difference between geocentric and geodetic latitude increases with increasing flattening; the central body becomes more of an ellipsoid, and deviates more from the "ideal" spheroid shape.
2.6. FALSE. The gravity field is not only dependent on the height variation of the Moon's surface, but also on the density variation of its interior. The combination of the two makes the gravity field very non-linear.
2.7. TRUE. The hydrostatic equation is given by $\mathrm{d}p = \rho g \mathrm{d}h$, or in words: the pressure is depending on the weight of a column air per unit area.
2.8. FALSE. The "surface" of a gas planet is defined in terms of pressure, i.e., there, where the pressure is 1 bar.
2.9. TRUE. In a general sense the atmosphere is co-rotating with the planetary body. At higher altitudes, the tangential velocity will be larger.
2.10. TRUE. The gravity field is not only a function of the shape, but also of the mass distribution. More variation in this distribution gives more local effects on the gravity, which can be captured by terms of higher degree and order.
2.11. TRUE. Saturn rotates once around its spin axis in about 10 hrs. This relatively large planetary rotation causes this gas planet to flatten significantly.
2.12. FALSE. The underlying assumption of an exponential atmosphere is that it is isothermal, i.e., temperature is constant. Since the speed of sound is a function of temperature that is constant with altitude, it remains constant.

Exponential model of the atmosphere (Problems 2.13–2.17).

2.13. The physical interpretation of the scale height, H_s, is that if the altitude increases by an amount equal to H_s, the density reduces by a factor of $e = 2.7181...$ This can be easily shown numerically by evaluating different densities. The density follows the profile

$$\rho = \rho_0 e^{-\frac{h}{H_s}} \tag{P.2.1}$$

Let $h_1 = 0$ m and $h_2 = H_s$, then

$$\rho_1 = \rho_0 e^0 = \rho_0 \quad \text{and} \quad \rho_2 = \rho_0 e^{-\frac{H_s}{H_s}} = \rho_0 e^{-1} \tag{P.2.2}$$

The ratio $\frac{\rho_1}{\rho_2} = \frac{1}{e^{-1}} = e = 2.7183...$, which means that the sea-level density is 2.7183 times larger than the density at h_2. Suppose $h_3 = 2H_s$, then the ratio $\frac{\rho_3}{\rho_2} = \frac{e^{-2}}{e^{-1}}$ is found to be e^{-1}. So, at the higher altitude h_3 the density is 2.7183 times lower than the one at h_2.

2.14. The data in Table 2.18 have been generated using an exponential atmosphere model for the density, with $\rho_0 = 1.430$ kg/m^3 and $H_s = 4{,}915$ m, after which a $\pm 5\%$ random error (uniformly distributed) has been added to ρ. Fitting these data can be done in several ways, with trial and error being the simplest one (but probably the most time consuming). Suppose one wants to use linear regression. In that case, the exponential model is rewritten as

$$\ln \rho = \ln \left(\rho_0 e^{-\frac{h}{H_s}} \right) = \ln \rho_0 - \frac{1}{H_s} h \tag{P.2.3}$$

With a new variable $\hat{\rho}^* = \ln \hat{\rho}$ the problem has changed to find coefficients for

$$\hat{\rho}^* = b_0 + b_1 h \tag{P.2.4}$$

with $b_0 = \ln \rho_0$ and $b_1 = -\frac{1}{H_s}$. To find a solution for b_0 and b_1, the problem is written as $\mathbf{y} = \mathbf{X}\mathbf{b}$ with $\mathbf{y}$ is the vector with ρ, and $\mathbf{X}$ the matrix with independent variables, in this case is the first column a vector filled with 1 and the second column the linear term, h. More details on how to solve such a problem can be found in any book on regression analysis. Once the estimator of $\mathbf{b}$ is solved for, they are transformed back to the required parameters: $\rho_0 = e^{b_0}$ and $H_s = -\frac{1}{b_1}$. The found results are: $\rho_0 = 1.4228$ kg/m^3 and $H_s = 4{,}906$ m, which are quite close to their nominal counterparts.

2.15. The assumption underlying the exponential atmosphere model is that the temperature is constant, in this case $T = 220$ K. The speed of sound is given by the relation

$$a = \sqrt{\gamma R T} \tag{P.2.5}$$

with γ is the ratio of specific heats and R the gas constant. For the Earth atmosphere, commonly-used average values are $\gamma = 1.4$ and $R = 287$ J/(kg K).

(a) For the given temperature of 220 K, these values substituted gives $a = 297.3$ m/s. This value is constant for the complete altitude range for which the model is defined.

(b) If the temperature increases by 50 K, then for $T = 270$ K the speed of sound becomes $a = 329.4$ m/s.

(c) For the altitude range of the entry, [0, 100] km, it is important to know the temperature range the vehicle encounters. As an indication, Fig. 2.30 shows a minimum temperature of $T_{min} = 187$ K and a maximum temperature $T_{max} = 288$ K (modelled with the United States Standard Atmosphere 1976). This gives a range for the speed of sound of about [274,340] m/s, a variation of about 20%. Since the speed of sound is used to calculate the Mach number, which in turn is used to calculate the aerodynamic properties, and further used to determine the thermo-mechanical loads, one can imagine that using a constant speed of sound is perhaps not the best choice.

2.16. In case the density is modelled with two separate exponential models, it is important that there is no discontinuity at the junction, because this may give rise to spikes in the acceleration when using such a model in a simulation. This does not only apply to the density itself, but also to the gradient. However, since each exponential model has two "unknowns", i.e., ρ_0 and H_s, this means that four conditions are required to solve this.

2.17. The wrong approach to setup a bridging function: Let us analyse this for two layers with Model 1 and Model 2 that join at an altitude h. Continuity in density at $h_1 = h_2 = h$:

$$\rho_1 = \rho_2 \quad \Rightarrow \quad \rho_{0,1} e^{-h/H_{s,1}} = \rho_{0,2} e^{-h/H_{s,2}} \tag{P.2.6}$$

Continuity of first gradient, which is given by

$$\frac{\mathrm{d}\rho}{\mathrm{d}h} = -\frac{\rho_0}{H_s} e^{-h/H_s} = -\frac{\rho_0}{H_s}\rho \tag{P.2.7}$$

so

$$-\frac{\rho_{0,1}}{H_{s,1}} = -\frac{\rho_{0,2}}{H_{s,2}} \tag{P.2.8}$$

If we include also the second and third derivatives (to get the four conditions), then the only solution to this problem is: $\rho_{0,1} = \rho_{0,2}$ and $H_{s,1} = H_{s,2}$. And that is contradicting the fact that there are two different models.

A good approach is to use some form of interpolation, e.g., Hermite-spline interpolation (see Appendix C, Sect. C.4.3). Define an altitude boundary of $\Delta h = 1$ km (below and above the junction point) and determining ρ and $\frac{\mathrm{d}\rho}{\mathrm{d}h}$ for those two points, the Hermite spline is determined using Eq. (C.66). The resulting bridging function is shown in Fig. P.2.1. Depending on accuracy requirements one can change Δh; most important is, though, that the two profiles are merged into a single one.

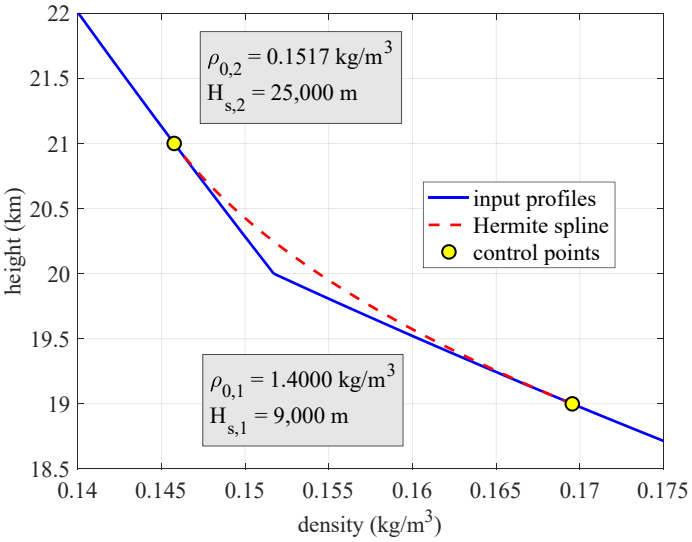


Fig. P.2.1 Linear interpolation in one dimension, including a bridging function using Hermite splines

Gravity-field modelling, GOCE's orbit around Earth (Problems 2.18–2.23).

Data for the Earth that can be used: $\mu = 398600.4415$ km^3/s^2, $R_e = 6378.137$ km, $J_{3,1} = -1.72{\cdot}10^{-6}$, and $\tau_{3,1} = -1.0°$.

2.18. The J_n and $J_{n,m}$ terms are scaling factors of the irregularities in the gravity field, whereas $\tau_{n,m}$ defines the orientation of a particular irregularity.

2.19. (r, δ, τ) defines the position of the satellite in spherical coordinates: r is the radial distance between the position of interest and the centre of Earth, δ is the latitude with respect to the Earth's equator, and τ is the longitude with respect to the Earth-fixed reference meridian.

2.20. The general expression to derive the radial acceleration (in m/s^2) from the potential formulation for the gravity field is given by

$$a_r = -\frac{\partial U}{\partial r} \tag{P.2.9}$$

Note to pay attention to the sign; it is common to define all three acceleration components in NED (north-east-down) direction. Sometimes the z-component is flipped in up direction (where the NEU-frame is a left-handed frame).

2.21. The general equation for the radial acceleration due to the term $J_{3,1}$ is derived as follows. To begin with, $J_{3,1}$ indicates that $n = 3$ and $m = 1$. The corresponding Legendre polynomial to evaluate is

$$P_{3,1}(x) = (1-x^2)^{1/2}\frac{\mathrm{d}^1 P_3(x)}{\mathrm{d}x^1} \quad \text{with} \quad P_3(x) = \frac{1}{(-2)^3 3!}\frac{\mathrm{d}^3}{\mathrm{d}x^3}(1-x^2)^3 \tag{P.2.10}$$

To begin with, we expand $(1-x^2)^3$ and write $P_3(x)$ as

$$P_3(x) = -\frac{1}{48}\frac{\mathrm{d}^3}{\mathrm{d}x^3}(1-3x^2+3x^4-x^6) \tag{P.2.11}$$

such that

$$\frac{\mathrm{d}}{\mathrm{d}x}\ldots = -6x+12x^3-6x^5 \quad \frac{\mathrm{d}^2}{\mathrm{d}x^2}\ldots = -6+36x^2-30x^4 \quad \frac{\mathrm{d}^3}{\mathrm{d}x^3}\ldots = 72x-120x^3 \tag{P.2.12}$$

and thus

$$P_3(x) = -\frac{3}{2}x+\frac{5}{2}x^3 = \frac{1}{2}(5x^3-3x) \tag{P.2.13}$$

This result substituted in the expression for $P_{3,1}$ yields

$$P_{3,1}(x) = \sqrt{1-x^2}\frac{1}{2}(15x^2-3) \tag{P.2.14}$$

and with $x = \sin\delta$

$$P_{3,1}(\sin\delta) = \sqrt{1-\sin^2\delta}\left(\frac{15}{2}\sin^2\delta - \frac{3}{2}\right) \tag{P.2.15}$$

This equation can be simplified when one realises that $\sqrt{1-\sin^2\delta} = \sqrt{\cos^2\delta} = \cos\delta$.
The last step is to evaluate $a_r = -\frac{\partial U}{\partial r}$; doing so results in

$$a_r = -4\mu J_{3,1} R_e^3 r^{-5}\cos\delta\left(\frac{15}{2}\sin^2\delta - \frac{3}{2}\right)\cos(\tau-\tau_{3,1}) \tag{P.2.16}$$

2.22. For arbitrary latitude and longitude (δ and τ), the radial acceleration follows from substituting the values for μ, R_e, $J_{3,1}$, and $\tau_{3,1}$:

$$a_r = 5.562\cdot 10^{-5}\cos\delta\left(\frac{15}{2}\sin^2\delta - \frac{3}{2}\right)\cos(\tau+1.0\frac{\pi}{180}) \text{ m/s}^2 \tag{P.2.17}$$

2.23. The radial acceleration changes sign at the trivial solution of $\cos\delta = 0$ ($\delta = \pm 90°$), which do not play a role. Other zero values are found when

$$\frac{15}{2}\sin^2\delta - \frac{3}{2} = 0 \quad \Rightarrow \quad \sin\delta = \pm\sqrt{0.2} \tag{P.2.18}$$

so $\delta = \pm 26.6°$.

Gravity-field modelling, the effect of J_2 on entry flight (Problems 2.24–2.26).

2.24. The central-field model provides the radial or down component of the gravitational acceleration, $g_d = \mu/(R_e + h_E)^2$, with, for Earth, $R_e = 6{,}378.137$ km, $\mu = 398{,}600.4415$ km^3/s^2, so at $h_E = 120$ km and $h = 25$ km the values for g_d are 9.440 m/s^2 and 9.722 m/s^2, respectively.

2.25. The error in acceleration when the J_2-effect is included is shows in the north (meridional) and down (radial) components:

$$g_n = -3\frac{\mu}{R^2}\left(\frac{R_e}{R}\right)^2 \sin\delta\cos\delta J_2 \tag{P.2.19}$$

$$g_d = \frac{\mu}{R^2}\frac{3}{2}J_2\left(\frac{R_e}{R}\right)^2 (3\sin^2\delta - 1) \tag{P.2.20}$$

The error in east component is zero because of rotational mass symmetry by dealing with only the harmonics.

The error in g_n is zero for $\delta = 0°$, which easily follows by substituting this into the equation for g_n. For g_d, the errors are -0.0148 m/s^2 ($h_E = 120$ km) and -0.0157 m/s^2 ($h = 25$ km). For a latitude of $\delta = 40°$, the corresponding errors in g_n are -0.0145 m/s^2 ($h_E = 120$ km) and -0.0154 m/s^2 ($h = 25$ km), and those in g_d are 0.0035 m/s^2 ($h_E = 120$ km) and 0.0038 m/s^2 ($h = 25$ km). Finally, for $\delta = 60°$ the errors in g_d are 0.0185 m/s^2 and 0.0196 m/s^2, and for g_n they are -0.0128 m/s^2 and -0.0136 m/s^2, respectively. The dependency on altitude and latitude can clearly be seen.

2.26. For a gliding entry with a duration of 20 min, the *average* error in (radial) position and velocity can be calculated as follows. Assume that the entry begins at $h_E = 120$ km and ends at $h = 25$ km. In principle it is unknown how far north the vehicle has flown, but it could be assumed it is as far as $\delta = 40°$. The average errors in gravitational acceleration could be calculated from the values derived before, in its simplest form simply the average of the two values (although one should realise that gravity does not vary inversely linear with altitude, but quadratically). So $\Delta g_d = (0.0038 + 0.0185)/2 = 0.01115$ m/s^2. Assuming that $\Delta V = \Delta g_d t_{flight}$ and $\Delta R = \frac{1}{2}\Delta g_d t_{flight}^2$, it follows that $\Delta V = 13.38$ m/s and $\Delta R = 8.028$ km.

These errors may seem large, but could, in fact, even be conservative. There are a few things to consider, though. In case of a guided flight, any small deviation from the nominal path will immediately be corrected by guidance, such that the final error will actually be quite small, without much of a guidance effort. In case of unguided flight, one should realise that any change in velocity will affect dynamic pressure and Mach number, which, in turn, affect the drag and lift force. The error in radial position (and thus altitude) affects density and temperature, thus speed of sound and then again the Mach number. One can

therefore not isolate the effect of an error in gravitational acceleration, but should always do an integrated (non-linear) analysis of all aspects involved.

2.27. 42

Chapter 3—Aerothermodynamics

TRUE/FALSE questions (Problems 3.1–3.13).

3.1. TRUE. Attached shockwaves are commonly observed in configurations like the (sharp) nose of a supersonic aircraft or the leading edge of a wing. Detached shockwaves are often seen in configurations with highly streamlined shapes or at higher angles of attack.

3.2. TRUE. As the Reynolds number increases, the flow becomes more turbulent, causing the boundary layer to grow in thickness due to enhanced mixing and momentum transfer between the fluid layers.

3.3. FALSE. Hypersonic and supersonic flow are aspects of continuum flow, not rarefied flow.

3.4. FALSE. Supersonic pockets can occur at Mach numbers lower than 1. Typically, at a Mach number of 0.85 the sonic point on an airfoil is reached, marking the formation of a shock wave.

3.5. FALSE. In hypersonic flow, the pressure distribution over the surface of a vehicle is influenced by a variety of factors, including shock waves, boundary-layer behaviour, flow separation, and geometric features. While the Mach independence principle may hold true for some aspects of the pressure distribution, such as the overall pressure coefficient (C_p) on certain sections of the vehicle, it may not hold universally across all conditions. For the above reasons, also the moment coefficients are in general not constant with Mach number.

3.6. FALSE. Turbulent boundary layers generally have higher skin friction due to increased mixing and velocity fluctuations within the boundary layer, leading to higher drag compared to laminar boundary layers.

3.7. FALSE. While compressibility effects play a significant role in hypersonic flow, viscosity can still have a notable impact, especially in regions close to the surface where boundary layer effects become important.

3.8. TRUE. Boundary layers are primarily a consequence of viscous effects near a solid surface, where the velocity of the fluid changes from zero at the surface to the free stream velocity away from the surface.

3.9. FALSE. Hypersonic flows can feature complex shock wave patterns, including detached shock waves, as the flow transitions through various Mach regimes, leading to significant changes in flow properties and aerodynamic forces.

3.10. FALSE. While hypersonic flows are often assumed to be nearly isentropic, there is still some entropy generation due to shock waves and viscous effects, particularly near surfaces or in regions of high gradients.

3.11. TRUE. The momentum thickness represents the distance from the body surface to a hypothetical plane where, if the boundary layer's velocity profile were

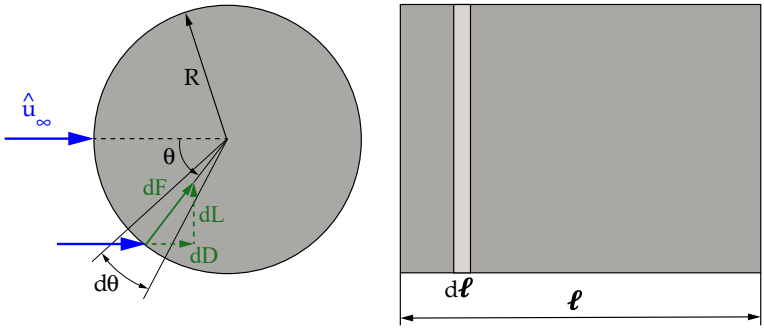


Fig. P.3.1 Cylinder in a transverse flow

extended, the average velocity would be equal to the free stream velocity. This concept relates to the loss in momentum due to the presence of the boundary layer.

3.12. TRUE. Hypersonic vehicles experience intense aerodynamic heating due to the compression of air ahead of them as they travel at extremely high speeds.

3.13. FALSE. The Mach independence principle also holds for the force coefficients, which are more or less constant for a large range of hypersonic Mach numbers.

Shock waves (Problems 3.14–3.16).

3.14. For a semi-cone-angle $\theta = 35°$, the figure shows that there are two solutions, $\beta_1 = 50°$ (weak shock) and $\beta_2 = 75.3°$ (strong shock). As the wedge angle is smaller than the maximum angle, both shocks are attached.

3.15. The flow velocity will be subsonic, i.e., $M_2 < 1$.

3.16. The shockwave will become detached, as according to the figure, $\theta = 43°$ is larger than the maximum wedge angle.

Newtonian Flow (Problems 3.17–3.21).

3.17. An advantage of high-temperature effects on vehicle design and/or performance is that at high temperatures the hypersonic flow forms a so-called "swallowing" entropy layer, which absorbs a lot of heat. This layer prevents the heat from reaching the vehicle wall, such that the TPS is loaded less heavily compared to what it would be without this layer. However, a disadvantage is that the high (remaining) temperature may still require the design of a heavier TPS to be able to absorb more heat.

3.18. In case of an axisymmetric entry capsule like Apollo, at zero angle of attack the pressure force at top and bottom (and left and right) cancel each other out, which means that the resulting lift and side force are zero.

3.19. In high Mach number continuum flow it is expected that there will be viscous-inviscid interaction due to real-gas effects. Since modified Newtonian flow

assumes only inviscid flow, this theory will be less accurate at high Mach numbers. The value of "2" is unrealistic for a sphere, but a zero value for a flat plate is realistic, since theoretically it has no interaction with the flow.

3.20. The object to consider is a cylinder with radius, R, and length, ℓ, which enters the atmosphere transverse to flow, see Fig. P.3.1 for the situation at hand and the variables that will be used to find an expression for the drag coefficient of the cylinder at zero angle of attack.
For a ring with width $\mathrm{d}\ell$, and a circle segment per ring with angle $\mathrm{d}\theta$, the element area exposed to the flow with velocity $\hat{\mathbf{u}}_\infty$ is

$$\mathrm{d}S = R\mathrm{d}\theta\mathrm{d}\ell \tag{P.3.1}$$

Following the method discussed in Example 3.1, the mass flow entering the control volume is

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \rho u_\infty \mathrm{d}S\cos\theta = \rho u_\infty \cos\theta R\mathrm{d}\theta\mathrm{d}\ell \tag{P.3.2}$$

restricting ourselves to a one-dimensional problem with scalar speed u_∞. The change in momentum generates an incremental force, $\mathrm{d}F$:

$$\mathrm{d}F = \frac{\mathrm{d}m}{\mathrm{d}t}\Delta u_\infty = 2\rho u_\infty^2 \cos^2\theta R\mathrm{d}\theta\mathrm{d}\ell \tag{P.3.3}$$

With the decomposition in drag and lift, as shown in the figure, the force components are:

$$\mathrm{d}D = \mathrm{d}F\cos\theta = 2\rho u_\infty^2 \cos^3\theta R\mathrm{d}\theta\mathrm{d}\ell \tag{P.3.4}$$

$$\mathrm{d}L = \mathrm{d}F \in \theta = 2\rho u_\infty^2 \cos^2\theta\sin\theta R\mathrm{d}\theta\mathrm{d}\ell \tag{P.3.5}$$

Integrating the drag component from $\theta = -\frac{\pi}{2}$ to $+\frac{\pi}{2}$—note that in the "shadow" of the cylinder for θ beyond these boundaries, the momentum exchange is zero—yields:

$$D = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^\ell \mathrm{d}D = 2\rho u_\infty^2 R \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^\ell \cos^3\theta\mathrm{d}\theta\mathrm{d}\ell \tag{P.3.6}$$

which results in

$$D = 2\rho u_\infty^2 R\ell \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^3\theta\mathrm{d}\theta \tag{P.3.7}$$

Applying the solution to the standard integral

$$\int \cos^n ax\mathrm{d}x = \frac{\cos^{n-1} ax \sin ax}{na} + \frac{n-1}{n}\int \cos^{n-2} ax\mathrm{d}x + C \qquad (n > 0) \tag{P.3.8}$$

with $n = 3$ and $a = 1$, this gives

$$\int \cos^3 \theta \mathrm{d}\theta = \frac{\cos^2 \theta \sin \theta}{3} + \frac{2}{3}\int \cos\theta \mathrm{d}\theta = \frac{\cos^2 \theta \sin \theta}{3} + \frac{2}{3}\sin\theta + C \tag{P.3.9}$$

Filling in the boundaries, this yields

$$\int \cos^3 \theta \mathrm{d}\theta = \frac{2}{3} - -\frac{2}{3} = \frac{4}{3} \tag{P.3.10}$$

Thus:

$$D = \frac{8}{3}\rho u_\infty^2 R\ell \tag{P.3.11}$$

With the definition $D = C_D \frac{1}{2}\rho u_\infty^2 S_{ref}$ and $S_{ref} = 2R\ell$ for a cylinder, the drag coefficient follows easily: $C_D = \frac{8}{3}$. Finding an expression for C_L is left as an exercise. Hint: use the standard integral

$$\int \cos^n ax \sin ax\mathrm{d}x = -\frac{1}{a(n+1)}\cos^{n+1} ax + C \tag{P.3.12}$$

3.21. In case the analysis is done for continuum Newtonian flow, the effect of the shock-wave passage is to decelerate the flow and turn it parallel to the vehicle surface. In this case all momentum perpendicular to the wall is absorbed by it (so no elastic bouncing of particles). This means that the momentum change is only half of that of the Newtonian analysis. For instance, in Example 3.1, ΔV for the flat plate was $\Delta V = 2u_\infty \sin\alpha$, which will now be $\Delta V = u_\infty \sin\alpha$. With the same mass flux, the total force (and thus drag and lift) become half the value, which also holds for C_D and C_L. The lift-to-drag ratio remains unchanged.

Chapter 4—Fundamentals of Motion

TRUE/FALSE questions (Problems 4.1–4.12).

4.1. FALSE. A transformation between two reference frames can also be based on a single axis rotation. For instance, the transformation between the inertial and rotating Earth-fixed reference frame will, in most applications, only include the Earth's rotation about the Earth's spin axis (the Z-axis of either frame). In general, depending on the angles that are available, a frame transformation can include any number of axis rotations.

4.2. TRUE. There are two types of Coriolis force: one originating from a rotating, mass-varying body, and one because of the rotation of the reference frame in which the motion is defined.

4.3. FALSE. A rigid body is non-flexible *and* has a constant mass. A launch vehicle will burn through a lot of fuel and oxidiser mass, and the mass variation is therefore significant.

4.4. TRUE. State variables for entry vehicles commonly include parameters that describe the vehicle's position and velocity in space, its orientation (attitude) relative to a reference frame, and its mass, for instance when the mass loss of an ablative heat shield must be taken into account. The latter can also be replaced by the fuel quantity, when, for instance, a lunar descent is considered.

4.5. FALSE. The selection of a state-variable set for position and velocity depends on the application. For instance, for guidance-system design it is preferred to use spherical components, as they give a much more physical interpretation to position and velocity. Also, the numerical stability of equations of motion depends on the choice of state variables, and should be considered in combination with numerical-integration techniques.

4.6. FALSE. The general formulation of Newton's Second Law is force equals mass times the time derivative of linear momentum, rather than the time derivative of velocity (acceleration). In this way, the Second Law is also applicable to mass-varying systems.

4.7. FALSE. Despite the fact that it is true that the cross product of each combination of matrix columns should be zero, it also holds for any row combination. Just these conditions makes the matrix orthogonal. The last condition to be checked is that the length of each column (row) vector should be 1, which makes the matrix ortho*normal*.

4.8. TRUE. This can easily be verified by applying the inverse transformation order and comparing the resulting matrix with the original's transpose.

4.9. TRUE. This concept is known as the *Solidification Principle*.

4.10. FALSE. The angular momentum is a vector property of a rotating body, equal to the product of its inertia tensor (a 3×3 matrix) and the body's angular velocity (a three-dimensional vector).

4.11. TRUE. Any weight-referenced ballistic parameter includes the gravitational acceleration of the corresponding celestial body, which will be different for each one of them:

$$K = \frac{m g_0}{C_D S} \quad \text{versus} \quad K_m = \frac{m}{C_D S}$$

4.12. FALSE. Euler angles are by definition not singularity free, as implied by Euler's rotation theorem. This theorem states that for a three-dimensional attitude description, three angles will suffice, but at the cost of a singularity. There are 12 sets of Euler angles, depending on the rotation order, and in each case the singularity will be present in the angle defining the second rotation. For instance, in the commonly applied 3-2-1 sequence used in the aerospace industry (yaw-pitch-roll), the singularity is linked to the pitch angle, θ ($\pm90°$).

Frame transformations (Problems 4.13–4.14).

4.13. The complete transformation matrix from frame B to frame C, $\mathbf{C_{C,B}}$, is

$$\mathbf{C_{C,B}} = \mathbf{C_y}(d\beta) = \begin{bmatrix} \cos d\beta & 0 & -\sin d\beta \\ 0 & 1 & 0 \\ \sin d\beta & 0 & \cos d\beta \end{bmatrix} \tag{P.4.1}$$

The inverse is

$$\mathbf{C_{B,C}} = \mathbf{C_{C,B}^{-1}} = \mathbf{C_{C,B}^{T}} = \mathbf{C_y}(-d\beta) = \begin{bmatrix} \cos d\beta & 0 & \sin d\beta \\ 0 & 1 & 0 \\ -\sin d\beta & 0 & \cos d\beta \end{bmatrix} \tag{P.4.2}$$

The inverse transformation matrix is the transpose, since the transformation matrices are orthonormal matrices.
Similarly, $\mathbf{C_{D,B}}$ and the inverse $\mathbf{C_{B,D}}$ are found:

$$\mathbf{C_{D,B}} = \mathbf{C_z}(d\gamma)\mathbf{C_y}(d\beta) = \begin{bmatrix} \cos d\gamma & \sin d\gamma & 0 \\ -\sin d\gamma & \cos d\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos d\beta & 0 & -\sin d\beta \\ 0 & 1 & 0 \\ \sin d\beta & 0 & \cos d\beta \end{bmatrix} \tag{P.4.3}$$

Matrix multiplication carried out:

$$\mathbf{C_{D,B}} = \begin{bmatrix} \cos d\gamma \cos d\beta & \sin d\gamma & -\cos d\gamma \sin d\beta \\ -\sin d\gamma \cos d\beta & \cos d\gamma & \sin d\gamma \sin d\beta \\ \sin d\beta & 0 & \cos d\beta \end{bmatrix} \tag{P.4.4}$$

The inverse matrix is:

$$\mathbf{C_{B,D}} = \mathbf{C_y}(-d\beta)\mathbf{C_z}(-d\gamma) = \begin{bmatrix} \cos d\gamma \cos d\beta & -\sin d\gamma \cos d\beta & \sin d\beta \\ \sin d\gamma & \cos d\gamma & 0 \\ -\cos d\gamma \sin d\beta & \sin d\gamma \sin d\beta & \cos d\beta \end{bmatrix} \tag{P.4.5}$$

4.14. The angular rate of frame D with respect to frame A, $\boldsymbol{\omega}_{D/A}$, can be obtained from the corresponding sequence of transformations. The transformation matrix $\mathbf{C_{D,A}}$ is written first, but with the intermediate frames. These indicators help us to identify the axis for the angular rate:

$$\mathbf{C_{D,A}} = |_D\mathbf{C_z}(d\gamma)\ \big|_C\mathbf{C_y}(d\beta)\big|_B\ \mathbf{C_x}(d\alpha)|_A \tag{P.4.6}$$

From this we can simply find:

Step 1: $\boldsymbol{\omega}_{D/A} = \boldsymbol{\omega}_{D/C} + \boldsymbol{\omega}_{C/B} + \boldsymbol{\omega}_{B/A}$
Step 2: Assume that angular displacement takes place in dt seconds, so the rates are: $\boldsymbol{\omega}_{D/C} = \dot{\gamma}$, $\boldsymbol{\omega}_{C/B} = \dot{\beta}$, and $\boldsymbol{\omega}_{B/A} = \dot{\alpha}$, with $\dot{\gamma} = \frac{d\gamma}{dt}$, $\dot{\beta} = \frac{d\beta}{dt}$, and $\dot{\alpha} = \frac{d\alpha}{dt}$.

Step 3: The corresponding unit axes follow from the intermediate frames and the axis of rotation: from $\mathbf{C_x}(d\alpha)|_A$ we find the X-axis of the A-frame, $\mathbf{x_A}$. However, we note that $\mathbf{x_A} = \mathbf{x_B}$. Similarly, from $\left|_C\mathbf{C_y}(d\beta)\right|_B$ we find $\mathbf{y_B} = \mathbf{y_C}$ and from $|_D\mathbf{C_z}(d\gamma)|_C$ we find $\mathbf{z_C} = \mathbf{z_D}$. So:

$$\boldsymbol{\omega}_{D/A} = \dot{\gamma}\mathbf{z_D} + \dot{\beta}\mathbf{y_C} + \dot{\alpha}\mathbf{x_B} \tag{P.4.7}$$

Step 4: What remains to be done is to transform $\mathbf{x_B}$ and $\mathbf{y_C}$ to the D-frame. The equations to solve are

$$\left[\mathbf{x_B}\ \mathbf{y_B}\ \mathbf{z_B}\right] = \mathbf{C_{B,D}}\left[\mathbf{x_D}\ \mathbf{y_D}\ \mathbf{z_D}\right] \tag{P.4.8}$$

with transformation matrix

$$\mathbf{C_{B,D}} = \begin{bmatrix} \cos d\gamma \cos d\beta & -\sin d\gamma \cos d\beta & \sin d\beta \\ \sin d\gamma & \cos d\gamma & 0 \\ -\cos d\gamma \sin d\beta & \sin d\gamma \sin d\beta & \cos d\beta \end{bmatrix} \tag{P.4.9}$$

Also:

$$\left[\mathbf{x_C}\ \mathbf{y_C}\ \mathbf{z_C}\right] = \mathbf{C_{C,D}}\left[\mathbf{x_D}\ \mathbf{y_D}\ \mathbf{z_D}\right] \tag{P.4.10}$$

with transformation matrix

$$\mathbf{C_{C,D}} = \begin{bmatrix} \cos d\gamma & -\sin d\gamma & 0 \\ \sin d\gamma & \cos d\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{P.4.11}$$

Doing the multiplication leads to

$$\mathbf{x_B} = \cos d\gamma \cos d\beta \mathbf{x_D} - \sin d\gamma \cos d\beta \mathbf{y_D} + \sin d\beta \mathbf{z_D} \tag{P.4.12}$$

and

$$\mathbf{y_C} = \sin d\gamma \mathbf{x_D} + \cos d\gamma \mathbf{y_D} \tag{P.4.13}$$

So, the final answer is:

$$\begin{aligned} \boldsymbol{\omega}_{D/A} = {} & \left(\cos d\gamma \cos d\beta \dot{\alpha} + \sin d\gamma \dot{\beta}\right)\mathbf{x_D} + \left(\cos d\gamma \dot{\beta} - \sin d\gamma \cos d\beta \dot{\alpha}\right)\mathbf{y_D} + \\ & + \left(\sin d\beta \dot{\alpha} + \dot{\gamma}\right)\mathbf{z_D} \end{aligned} \tag{P.4.14}$$

Coordinate transformations (Problems 4.15–4.19).

4.15. The task at hand is to derive $\mathbf{C_{B,V_1}}$, using the 3-2-1 rotation sequence:

$$\mathbf{C_{B,V_1}} = \mathbf{C_x}(\varphi_1)\mathbf{C_y}(\theta_1)\mathbf{C_z}(\psi_1) \tag{P.4.15}$$

Fully written out, this results in

$$\mathbf{C}_{\mathbf{B},\mathbf{V_1}} = \begin{bmatrix} \mathrm{c}\theta_1\mathrm{c}\psi_1 & \mathrm{c}\theta_1\mathrm{s}\psi_1 & -\mathrm{s}\theta_1 \\ \mathrm{s}\varphi_1\mathrm{s}\theta_1\mathrm{c}\psi_1 - \mathrm{s}\psi_1\mathrm{c}\varphi_1 & \mathrm{s}\psi_1\mathrm{s}\theta_1\mathrm{s}\varphi_1 + \mathrm{c}\psi_1\mathrm{c}\varphi_1 & \mathrm{s}\varphi_1\mathrm{c}\theta_1 \\ \mathrm{c}\psi_1\mathrm{c}\varphi_1\mathrm{s}\theta_1 + \mathrm{s}\psi_1\mathrm{s}\varphi_1 & \mathrm{s}\psi_1\mathrm{c}\varphi_1\mathrm{s}\theta_1 - \mathrm{c}\psi_1\mathrm{s}\varphi_1 & \mathrm{c}\varphi_1\mathrm{c}\theta_1 \end{bmatrix} \tag{P.4.16}$$

using the short-hand notation "c" for *cosine* and "s" for *sine*. It is noted that the singularity occurs for $\theta_1 = \pm\frac{1}{2}\pi$; this can be generalised to the fact that the singularity always occurs in the second (= middle) rotation. The reader should be aware that there is no singularity in $\mathbf{C}_{\mathbf{B},\mathbf{V_1}}$ itself, it only occurs in the corresponding kinematic attitude equations.

4.16. An Euler rotation-sequence following the order 1-3-2, with the singularity in $\psi_2 = \pm\frac{1}{2}\pi$ (as mentioned above: the middle rotation). The transformation matrix $\mathbf{C}_{\mathbf{B},\mathbf{V_2}}$ can be obtained from:

$$\mathbf{C}_{\mathbf{B},\mathbf{V_2}} = \mathbf{C}_{\mathbf{x}}(\varphi_2)\mathbf{C}_{\mathbf{z}}(\psi_2)\mathbf{C}_{\mathbf{y}}(\theta_2) \tag{P.4.17}$$

which yields:

$$\mathbf{C}_{\mathbf{B},\mathbf{V_2}} = \begin{bmatrix} \mathrm{c}\theta_2\mathrm{c}\psi_2 & -\mathrm{c}\theta_2\mathrm{s}\psi_2\mathrm{c}\varphi_2 - \mathrm{s}\theta_2\mathrm{s}\varphi_2 & \mathrm{c}\theta_2\mathrm{s}\psi_2\mathrm{s}\varphi_2 - \mathrm{s}\theta_2\mathrm{c}\varphi_2 \\ \mathrm{s}\psi_2 & \mathrm{c}\psi_2\mathrm{c}\varphi_2 & -\mathrm{c}\psi_2\mathrm{s}\varphi_2 \\ \mathrm{s}\theta_2\mathrm{c}\psi_2 & \mathrm{c}\theta_2\mathrm{s}\varphi_2 - \mathrm{s}\theta_2\mathrm{s}\psi_2\mathrm{c}\varphi_2 & \mathrm{c}\theta_2\mathrm{c}\varphi_2 + \mathrm{s}\theta_2\mathrm{s}\psi_2\mathrm{s}\varphi_2 \end{bmatrix} \tag{P.4.18}$$

4.17. Fundamental principle is that given a *known* and an *unknown* matrix, representing the same transformation, the unknown coefficients can be extracted from the known matrix. Explained with a simple example, from

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \tag{P.4.19}$$

it would be easy to solve for a, b, c and d. For the current problem, $\mathbf{C}_{\mathbf{B},\mathbf{V_1}}$ is the known matrix, as the attitude φ_1, θ_1, and ψ_1 has been obtained from propagating the kinematic attitude equations. Thus, since both $\mathbf{C}_{\mathbf{B},\mathbf{V_1}}$ and $\mathbf{C}_{\mathbf{B},\mathbf{V_2}}$ describe exactly the same transformation,

$$\mathbf{C}_{\mathbf{B},\mathbf{V_2}} = \mathbf{C}_{\mathbf{B},\mathbf{V_1}} \tag{P.4.20}$$

For each of the angles φ_2, θ_2, and φ_2 we find, after inspecting the matrix elements in Eq. (P.4.18):

$$\varphi_2 = -\arctan\left(\frac{\mathbf{C}_{\mathbf{B},\mathbf{V_1}}(2,3)}{\mathbf{C}_{\mathbf{B},\mathbf{V_1}}(2,2)}\right) \tag{P.4.21}$$

$$\theta_2 = \arctan\left(\frac{\mathbf{C}_{\mathbf{B},\mathbf{V}_1}(3,1)}{\mathbf{C}_{\mathbf{B},\mathbf{V}_1}(1,1)}\right) \tag{P.4.22}$$

$$\psi_2 = \arcsin\left(\mathbf{C}_{\mathbf{B},\mathbf{V}_1}(2,1)\right) \tag{P.4.23}$$

taking quadrant information into account. The angles are confined to the regions $-180° \leq \varphi_2 < 180°$, $-180° \leq \theta_2 < 180°$, and $-90° \leq \psi_2 \leq 90°$, which uniquely describe meaningful vehicle orientations.

4.18. The coordinate transformation discussed in the previous problem breaks down for the singularity $\psi_2 = \pm\frac{1}{2}\pi$, when φ_2 and θ_2 can no longer be calculated. For instance, for $\psi_2 = \frac{1}{2}\pi$:

$$\mathbf{C}_{\mathbf{B},\mathbf{V}_2} = \begin{bmatrix} 0 & -\mathrm{c}\theta_2\mathrm{c}\varphi_2 - \mathrm{s}\theta_2\mathrm{s}\varphi_2 & \mathrm{c}\theta_2\mathrm{s}\varphi_2 - \mathrm{s}\theta_2\mathrm{c}\varphi_2 \\ 1 & 0 & 0 \\ 0 & \mathrm{c}\theta_2\mathrm{s}\varphi_2 - \mathrm{s}\theta_2\mathrm{c}\varphi_2 & \mathrm{c}\theta_2\mathrm{c}\varphi_2 + \mathrm{s}\theta_2\mathrm{s}\varphi_2 \end{bmatrix} \tag{P.4.24}$$

which is equal to

$$\mathbf{C}_{\mathbf{B},\mathbf{V}_2} = \begin{bmatrix} 0 & -\cos(\theta_2 - \varphi_2) & \sin(\varphi_2 - \theta_2) \\ 1 & 0 & 0 \\ 0 & \sin(\varphi_2 - \theta_2) & \cos(\theta_2 - \varphi_2) \end{bmatrix} \tag{P.4.25}$$

This condition is called the *gimbal lock*, where θ_2 and φ_2 can no longer be independently determined. This confirms the need to do the switch ahead of reaching the singularity.

4.19. The solution based on two sets of Euler angles is not fail-safe, when the body is (arbitrarily) tumbling. It may be that at the moment of switching, the singularity of the second set of Euler angles has been reached. Even if the switch is made far away from the singularity of the first set, a check should be made to confirm the second singularity is not approached yet. This complicates matters, as in that case the coordinate switch has to be postponed. But there is never a 100% guarantee that if the switch is made at the last minute, this problem does not persevere. A better solution would be to use quaternions; even though they form a redundant set (with four parameters instead of three, and the so-called quaternion norm should be enforced as a constraint), quaternions are singularity-free and numerically very stable.

Chapter 5—Planar Motion

TRUE/FALSE questions (Problems 5.1–5.9).

5.1. FALSE. In the 1960s the Russian Zond missions performed several skipping entries on their return flights from the Moon. Also more recently, the Orion capsule of the Artemis-1 mission successfully flew a skipping flight.

5.2. FALSE. While it is true that a negative angle of attack can provide a negative lift, the same can also be achieved with a positive angle of attack and a bank

angle larger than 90°. In that way, the lift force is rotated to a downwards orientation.

5.3. TRUE. To move away from the heat-flux constraint, the vehicle requires more lift. At zero bank angle, one cannot increase the lift by banking more. An alternative would be to increase the angle of attack, provided this angle is not at its maximum as well.

5.4. FALSE. As a first-order approximation it was derived that $\frac{\tilde{a}_{max}}{g} = \frac{D}{L}\left(1 - 2\sqrt{\frac{2}{\beta R_e}}\right)$. So with *increasing* L/D the maximum *g-load decreases*. In conclusion, the maximum *g-load* increases inversely proportional with the lift-to-drag ratio.

5.5. FALSE. Since a flight through the atmosphere will dissipate energy due to the induced drag, it depends on the entry velocity whether there is sufficient kinetic energy (thus velocity) left when leaving the atmosphere. An entry velocity close to the escape velocity of 11.2 km/s may be reduced to a value lower than this limit. So it is not always guaranteed.

5.6. TRUE. To first order, this has been proven for ballistic, gliding, and skipping flight. However, also for actual flights this has been shown to be true, which can be understood by looking at the dependencies on velocity (mechanical load $\propto V^2$ and heat load $\propto V^3$). Even though density appears as ρ for mechanical load and ρ^n ($n = 0.2$ or 0.5) for thermal load, the effect of velocity is driving.

5.7. FALSE. During skipping flight, the normal acceleration can be a lot larger than the axial (or tangential) load, depending on the values of γ_E and/or L/D. See Fig. 5.68.

5.8. FALSE. Banking is the process of rolling around the X-axis of the trajectory frame, effectively around the velocity vector. Only when the angles of attack and sideslip are zero, this axis is aligned with the X-axis of the body frame.

5.9. FALSE. The constraint line for heat flux as well as the actual profile are a function of density (thus altitude) and velocity. Therefore, a shift in altitude-velocity profile will show as a shift in heat-flux profile.

Equations of motion (Problem 5.10).

5.10. For many of the derivations, the equations of planar motion are the starting point. Knowing these equations by heart would allow you to start these derivations, without having to look for them. As they are fundamental to this and other chapters, e.g., Chap. 11 on aeromanoeuvring, it is strongly recommended that you know these by heart.
The two dynamic equations are:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg\sin\gamma \tag{P.5.1}$$

$$mV\frac{\mathrm{d}\gamma}{\mathrm{d}t} = L - mg\cos\gamma\left(1 - \frac{V^2}{V_c^2}\right) \tag{P.5.2}$$

The kinematic equation is:

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} = V \sin\gamma \tag{P.5.3}$$

Ballistic entry (Problems 5.11–5.14).

5.11. The derivation can be found in Chap. 5, Eqs. 5.19–5.32. Main assumptions applied are $D >> mg\sin\gamma$ and $\gamma = \gamma_E =$ constant. Following the steps, it is shown that for a ballistic flight that

$$\frac{V}{V_E} = \exp\left(\frac{1}{2}\frac{g\rho_0 e^{-\beta h}}{K\beta\sin\gamma_E}\right) \tag{P.5.4}$$

5.12. The conditions for parachute deployment are determined as follows, to begin with the altitude.

- From Mach number $M = 2.2$ and speed of sound $a = 240$ m/s we find the velocity: $V = Ma = 2.2 \times 240 = 528$ m/s
- Equation (P.5.4) gives the density:

$$\ln\frac{V}{V_E} = \frac{1}{2}\frac{g\rho}{K\beta\sin\gamma_E} \Rightarrow \rho = 2K\frac{\beta}{g}\sin\gamma_E \ln\frac{V}{V_E} \tag{P.5.5}$$

- With the values for $m = 2{,}400$ kg, $C_D = 1.45$ and $S = 16.6$ m^2, the ballistic parameter is calculated, $K = 367.9$ N/m^2. Then, with the scale height of $H_s = 12{,}100$ m it follows that $\beta = 8.2645{\cdot}10^{-5}$ m^{-1}. Further, using $\rho_0 = 0.02$ kg/m^3, $V = 528$ m/s, $V_E = 5{,}900$ m/s and $\gamma_E = -12.2°$ we find for the density $\rho = 0.00841$ kg/m^3.
- Finally, from the exponential atmosphere the altitude follows: $h = -H_s \ln(\rho/\rho_0) = 10{,}488$ m.

Next, the dynamic pressure is calculated. This follows from $\bar{q} = \frac{1}{2}\rho V^2$ with $\rho = 0.00841$ kg/m^3 and $V = 528$ m/s from above, giving $\bar{q} = 1{,}172$ N/m^2.

5.13. The parachute drag area is calculated assuming that the parachute-payload system is in vertical equilibrium, close to final conditions, and that the drag area of the vehicle is much smaller than the parachute drag area. Vertical equilibrium gives:

$$\dot{V} = D_{para} - mg = 0 \tag{P.5.6}$$

or

$$C_D\frac{1}{2}\rho_0 V_f^2 S_{ref} = mg_0 \tag{P.5.7}$$

Thus

$$C_D S_{ref}\big|_p = \frac{mg}{\frac{1}{2}\rho_0 V_f^2} \tag{P.5.8}$$

Substituting $\rho_0 = 0.02$ kg/m^3, $V_f = 30$ m/s, $m = 2{,}400$ kg, $g_0 = 3.69$ m/s^2 gives $C_D S_{ref}\big|_p = 984$ m^2.
Note that if the drag area of the vehicle is not ignored:

$$C_D S_{ref}\big|_p = \frac{mg_0 - D_{vehicle}}{\frac{1}{2}\rho_0 V_f^2} \tag{P.5.9}$$

the answer will be 959.3 m^2. A check shows that with $C_D = 1.45$ and $S_{ref} = 16.6$ m^2, the drag area of the vehicle is 24.1 m^2, which is indeed much smaller than the parachute drag area (so the assumption to ignore the vehicle drag area is valid).

5.14. The calculations to determine the diameter of the (circular) aeroshell to reduce the dynamic pressure by 30% proceed as follows.

- A reduction of 30% in dynamic pressure ($\bar{q} = 1{,}172$ N/m^2 becomes 820 N/m^2) gives for the velocity—with the altitude fixed ($\rho = 0.00841$ kg/m^3): $V = \sqrt{2\bar{q}/\rho} = 441.8$ m/s.
- From $\frac{V}{V_E} = \exp\left(\frac{1}{2}\frac{g\rho}{K\beta \sin\gamma_E}\right)$ the corresponding ballistic parameter (= the only unknown) should be:

$$K = \frac{1}{2}\frac{g\rho}{\ln\frac{V}{V_E}\beta \sin\gamma_E} \tag{P.5.10}$$

and all numbers substituted yields: $K = 342.6$ N/m^2.
- With the data of the aeroshell configuration ($m = 4{,}000$ kg and $C_D = 1.35$), the reference area can be extracted from the ballistic parameter:

$$S = \frac{mg_0}{C_D K} \tag{P.5.11}$$

which gives $S = 31.91$ m^2. Thus, the diameter becomes $d = \sqrt{4S/\pi} = 6.37$ m.

Thermo-mechanical loads (Problems 5.15–5.26).

5.15. The circular velocity is defined as:

$$V_c = \sqrt{gR} = \sqrt{\frac{\mu}{R}} \tag{P.5.12}$$

Table P.5.1 Ballistic parameters K in (N/m^2)

Planet	$K_m = 50.0$ kg/m^2	$K_m = 100.0$ kg/m^2	$K_m = 150.0$ kg/m^2
Venus	443.2	886.3	1,329.5
Earth	489.6	980.0	1,468.8
Mars	186.2	372.4	558.6

Table P.5.2 Maximum deceleration, $\bar{a}_{max}$, and corresponding altitude, h', for ballistic entry

Planet	$K_m = 50.0$ kg/m^2		$K_m = 100.0$ kg/m^2		$K_m = 150.0$ kg/m^2	
	$\bar{a}_{max}$ (m/s^2)	h' (km)	$\bar{a}_{max}$ (m/s^2)	h' (km)	$\bar{a}_{max}$ (m/s^2)	h' (km)
Venus	63.0	190.3	63.0	179.3	63.0	172.8
Earth	169.6	50.7	169.6	45.8	169.6	42.9
Mars	27.3	27.5	27.3	21.4	27.3	17.8

With the planetary data provided in Table 5.21 and $R = R_e + h_E$ gives $V_c =$ 7,177.2 m/s (Venus), 7,841.4 m/s (Earth), and 3,512.0 m/s (Mars).

5.16. To calculate the weight-referenced ballistic parameters, i.e., $K = g_0 K_m$, the surface gravitational accelerations as provided in Table 5.21 are used. The computed K are listed in Table P.5.1.

5.17. The maximum deceleration and the altitude at which this occurs are calculated with:

$$\bar{a}_{max} = -\left(\frac{\mathrm{d}V}{\mathrm{d}t}\right)_{max} = -\frac{\beta \sin\gamma_E}{2e} V_E^2 \tag{P.5.13}$$

and

$$h' = \frac{1}{\beta} \ln\left(-\frac{\rho_0 g}{K\beta \sin\gamma_E}\right) \tag{P.5.14}$$

Results for the three planets and three capsules are listed in Table P.5.2.

5.18. Inspecting the above results shows that the maximum deceleration is independent from K, which is obvious from Eq. (P.5.13). The altitude at which it occurs *is* dependent, Eq. (P.5.14): with larger K the altitude decreases. Concerning $\bar{a}_{max}$, the much denser atmosphere of Venus causes the capsule to decelerate more at higher altitudes, which reduces the peak value. On the other end of the spectrum, during a Mars entry there is not much deceleration at all: the thin atmosphere does not dissipate so much energy, such that the terminal velocity will be very large. It is for this reason that (large) parachutes are deployed at high Mach number, to have as much effect as possible. Retro rockets are usually required to remove the last bit of excess terminal velocity.

5.19. A Monte Carlo analysis of a simple equation with two unknowns, i.e., a perturbed ρ_0 and H_s, can usually be done with a limited number of runs, say about 50. However, the commonly used random generators show clustering of data points and "white" areas in the design space, such that many more runs are

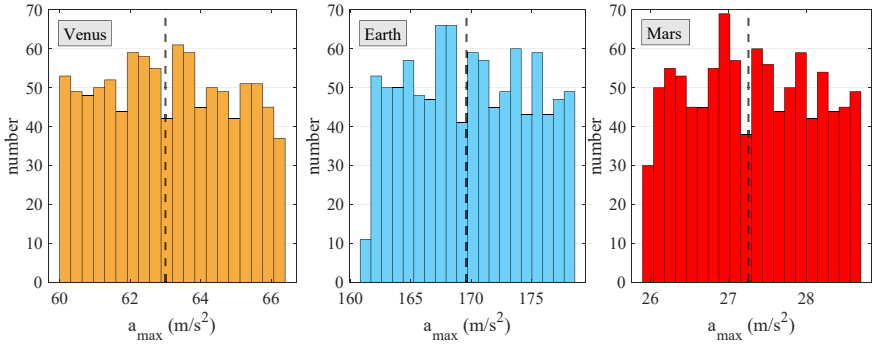


Fig. P.5.1 Maximum deceleration for Venus, Earth, and Mars, given uncertainties of 5% in ρ_0 and H_s ($N = 1{,}000$)

required. For this problem, a total of 100, 1000, and 10,000 runs were executed. Even though 100 seemed to be enough, while looking at $\bar{a}_{max}$ the distribution seems to converge to a uniform distribution for $N > 1,000$, whereas for $N = 100$, one could not discern such a distribution. Figure P.5.1 shows the results for the three planets (reminder: the ballistic parameter has no effect on $\bar{a}_{max}$).[1] A similar reasoning as before can be given. For the dense atmosphere of Venus, more energy is absorbed at higher altitudes, and thus the atmospheric uncertainties have less of an effect on the peak values: the spread is about $\pm$ 3 m/s^2. The thin atmosphere of Mars cannot decelerate the capsule so well, so out of the three planets, the perturbations have the least effect on the peak values: just over $\pm$ 1 m/s^2. Finally, the medium atmosphere of Earth shows the largest spread: $\pm$ 10 m/s^2. Because of the lesser deceleration at higher altitudes (thinner atmosphere than Venus), the effect of the perturbations is larger when the capsule hits the denser layers of the atmosphere.

Figure P.5.2 shows the corresponding change in altitude, h', at which the maximum deceleration occurs. An important observation is that for each of the three capsules, the histograms show very similar distributions, which means that K has only a small effect. In terms of spread, Venus exhibits the largest and Mars the smallest. For Venus, the explanation comes from the fact that it has the largest scale height, about twice the value of that of Earth. Since H_s determines the gradient of the density profile (see also Problem 2.17), and for each of the three planets the relative variation is the same, this means that the effect on density is largest for Venus, followed by Mars, then Earth (albeit that H_s for Earth and Mars are quite similar). Thus, this causes the larger shift in h' for Venus. Even though the effect for Mars is larger than for Earth, given

[1] Different random generators or even a different starting point in the random sequence (called the "seed") will give a different histogram. Therefore, compare your results only at a global level. The current results have been obtained with Matlab's default Mersenne Twister random generator, initialised with seed 0.

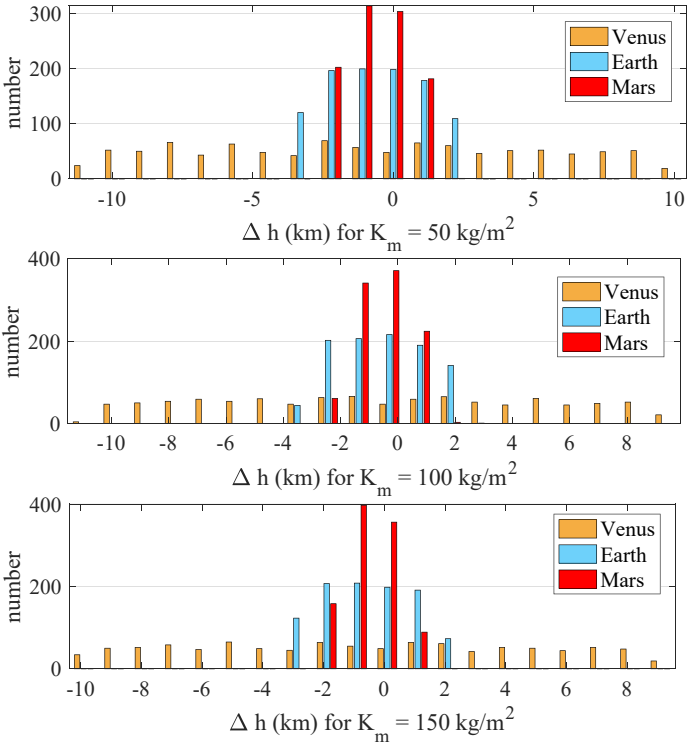


Fig. P.5.2 Variation in altitude for maximum deceleration (Venus, Earth, and Mars, for three capsules), given uncertainties of 5% in ρ_0 and H_s ($N = 1{,}000$)

Table P.5.3 Altitude for maximum heat flux (laminar and turbulent boundary layer, h'', for ballistic entry

	$K_m = 50.0$ kg/m^2		$K_m = 100.0$ kg/m^2		$K_m = 150.0$ kg/m^2	
	h'' ($n = 0.5$)	h'' ($n = 0.2$)	h'' ($n = 0.5$)	h'' ($n = 0.2$)	h'' ($n = 0.5$)	h'' ($n = 0.2$)
Venus	207.8	200.3	196.8	189.3	190.3	182.8
Earth	58.4	55.1	53.5	50.2	50.7	47.4
Mars	37.1	33.0	31.0	26.9	27.5	23.3

the much lower density in general (compare ρ_0 for both planets), the $\Delta h'$ for Mars is smaller.

5.20. The nominal altitude at which the heat flux is a maximum, h'', follows from Eq. (5.68). The results for both laminar ($n = 0.5$) and turbulent ($n = 0.2$) boundary layer are shown in Table P.5.3. As expected, each of the altitudes for the heat flux with a turbulent boundary layer is lower, i.e., closer to the altitude of maximum deceleration. In general, the altitude decreases with increasing K.

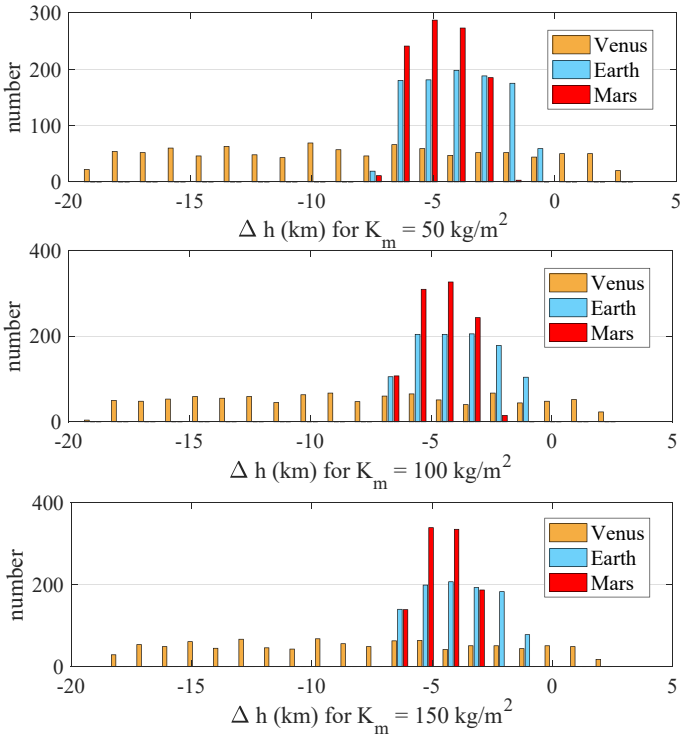


Fig. P.5.3 Variation in altitude for maximum heat flux given a laminar boundary layer (Venus, Earth, and Mars, for three capsules), given uncertainties of 5% in ρ_0 and H_s ($N = 1{,}000$)

5.21. The results of a Monte Carlo analysis ($N = 1{,}000$, random seed 0) for both laminar ($n = 0.5$) and turbulent ($n = 0.2$) boundary layer are shown in Figs. P.5.3 and P.5.4. The patterns are not readily explained from the single equation representing h''. What can be observed is that the spread is the largest for Venus, for the same reason as discussed earlier: a relative error with larger ρ_0 and H_s will have a stronger effect than the same relative error on lower values. The ballistic parameter has only a marginal effect. For the laminar boundary layer, the results are a lot more asymmetric, which comes from the smaller scaling due to the term $1 - n$ inside the logarithm. For the turbulent boundary layer, $n = 0.2$, this effect is stronger.

If only H_s is varied (and taking $N = 100{,}000$ samples to get a proper representation of the output distribution), a similar pattern as before emerges, but with a much more uniform distribution: for the laminar boundary layer the range -20 km to 5 km for Venus, -7 km to -1.5 km for Earth, and -7 km to -3 km for Mars, and for the turbulent boundary layer -11 km to 10 km for Venus, -3.5 km to 2.5 km for Earth, and -2.5 km to 1.5 km for Mars. Finally, if only ρ_0 is perturbed, the effect on h'' is a lot smaller: for the laminar

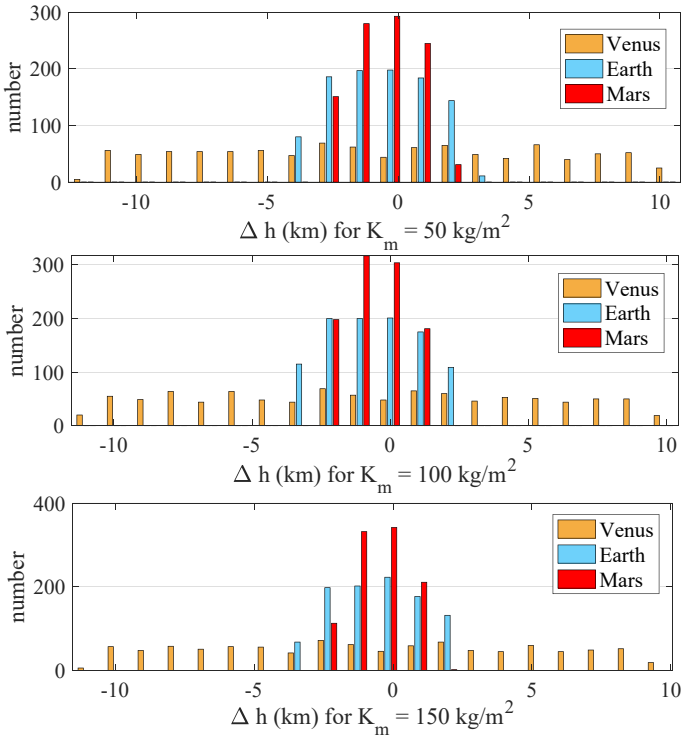


Fig. P.5.4 Variation in altitude for maximum heat flux given a turbulent boundary layer (Venus, Earth, and Mars, for three capsules), given uncertainties of 5% in ρ_0 and H_s ($N = 1{,}000$)

boundary layer the range is -8.5 km to -7 km for Venus, -4 km to -3 km for Earth, and -5 km to -4 km for Mars, and for the turbulent boundary layer -0.8 km to 0.8 km for Venus, -0.4 km to 0.3 km for Earth, and -0.5 km to 0.4 km for Mars. Whereas the results for $n = 0.5$ are a lot more asymmetric, for $n = 0.2$, they are a lot more symmetric around 0.

To conclude: for a physical interpretation of these results, a much more detailed analysis is required. Without a proper understanding of the physics involved, the simplified equations will not provide this insight.

5.22. The deviation of the perturbed maximum heat flux with respect to the nominal value is shown in Fig. P.5.5. Since the expression for $q_{c,max}$ is only valid for Earth (and a laminar boundary layer), only a single plot for the three capsules is given. For each K, the range in flux is the same, as well as the distribution. This is logical, of course, since K is a constant in both nominal and perturbed flux values. When ρ_0 and H_s are individually perturbed, the distribution is almost uniform, with a range of $\pm 2.5\%$. Combined, the range is double, but with a triangular distribution, where the effect each parameter seems to be countering one another, hence the majority of the points centred around the nominal value.

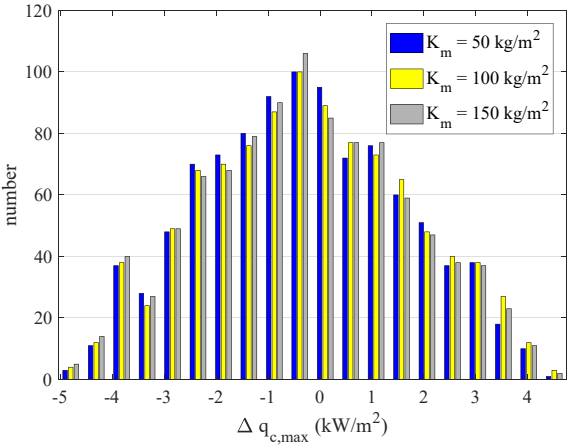


Fig. P.5.5 Maximum heat-flux deviation (laminar boundary layer) for three capsules (ballistic Earth entry), given uncertainties of 5% in ρ_0 and H_s ($N = 1{,}000$)

5.23. To generalise the expression for maximum heat flux, Eq. (5.81), to one for arbitrary n (type of boundary layer) and m (velocity exponent), we begin once more from:

$$q_c = c^* \frac{1}{R_N^n} \left(\frac{\rho}{\rho_0}\right)^{1-n} \left(\frac{V}{V_c}\right)^m \tag{P.5.15}$$

With $\rho = \frac{2K\beta \sin\gamma_E}{g} \ln \frac{V}{V_E}$ substituted, the velocity ratio at which the maximum flux occurs is derived from the generic equivalent of Eq. (5.57):

$$q_c \propto \left(\ln \frac{V}{V_E}\right)^{1-n} \left(\frac{V}{V_E}\right)^m \tag{P.5.16}$$

Following the same steps as in Sect. 5.3.2:

$$\frac{V''}{V_E} = e^{-\frac{1-n}{m}} \quad \Rightarrow \quad (V'')^m = V_E^m e^{n-1} \tag{P.5.17}$$

and for the corresponding altitude the density follows by substituting the above velocity ratio into Eq. (5.61):

$$\rho'' = -\frac{2}{m}(1-n)\frac{K\beta}{g} \sin\gamma_E \tag{P.5.18}$$

so

$$q_{c,max} = c^* \frac{1}{R_N^n} \left[-\frac{2}{m}(1-n)\frac{K\beta}{g\rho_0} \sin\gamma_E\right]^{1-n} \left(\frac{V_E}{V_c}\right)^m e^{n-1} \tag{P.5.19}$$

It is noted that both c^* is dependent on the atmospheric composition, so one has to be careful when using this model; also m may vary from atmosphere to atmosphere, and even though for Earth $m = 3$, in literature one can also find $m = 3.15$.

5.24. Starting point is the notion that the entry is a ballistic one, which means that the density-velocity relation is given by:

$$\rho = \frac{2K\beta \sin\gamma_E}{g} \ln \frac{V}{V_E} \tag{P.5.20}$$

This substituted in the expression for $q_{rad,gas}$, with $\bar{V} = \frac{V}{V_E}$,

$$q_{rad,gas} = C R_N \rho^{1.6} V^{8.5} = C V_E^{8.5} \rho^{1.6} \left(\frac{V}{V_E}\right)^{8.5} = C' \rho^{1.6} \bar{V}^{8.5} \tag{P.5.21}$$

yields

$$q_{rad,gas} = C' R_N \left(\frac{2K\beta \sin\gamma_E}{g} \ln \bar{V}\right)^{1.6} \bar{V}^{8.5} \tag{P.5.22}$$

or

$$q_{rad,gas} = C' R_N \left(\frac{2K\beta \sin\gamma_E}{g}\right)^{1.6} \left(\ln \bar{V}\right)^{1.6} \bar{V}^{8.5} \tag{P.5.23}$$

$$q_{rad,gas} = C'' R_N \left(\ln \bar{V}\right)^{1.6} \bar{V}^{8.5} \tag{P.5.24}$$

$$q_{rad,gas} \propto \left(\ln \bar{V}\right)^{1.6} \bar{V}^{8.5} \tag{P.5.25}$$

The peak value is found with

$$\frac{\mathrm{d}q_{rad,gas}}{\mathrm{d}\bar{V}} = 1.6 \left(\ln \bar{V}\right)^{0.6} \frac{1}{\bar{V}} \bar{V}^{8.5} + 8.5 \left(\ln \bar{V}\right)^{1.6} \bar{V}^{7.5} = 0 \tag{P.5.26}$$

The above expression simplifies to

$$1.6 \left(\ln \bar{V}\right)^{0.6} + 8.5 \left(\ln \bar{V}\right)^{1.6} = 0 \tag{P.5.27}$$

or

$$\ln \bar{V} = -\frac{1.6}{8.5} \quad \text{or} \quad \bar{V} = \frac{V''''}{V_E} = e^{-\frac{1.6}{8.5}} = 0.8284 \tag{P.5.28}$$

The corresponding altitude follows from substituting this result into the density equation:

$$\rho'''' = -\frac{2K\beta \sin\gamma_E}{g} \left(\frac{1.6}{8.5}\right) \tag{P.5.29}$$

and

$$h'''' = -H_s \ln\left[-\frac{2K\beta \sin\gamma_E}{g\rho_0}\left(\frac{1.6}{8.5}\right)\right] \tag{P.5.30}$$

5.25. For a gliding flight, the density-velocity relation is given by

$$\rho = 2\frac{W/S}{C_L}\left(\frac{1}{V^2} - \frac{1}{V_c^2}\right) \tag{P.5.31}$$

such that

$$q_{rad,gas} = CR_N\left[2\frac{W/S}{C_L}\left(\frac{1}{V^2} - \frac{1}{V_c^2}\right)\right]^{1.6} V^{8.5} \tag{P.5.32}$$

Normalising the velocity with V_c, i.e., $\bar{V}^* = \frac{V}{V_c}$ gives

$$q_{rad,gas} = CR_N CV_c^{8.5}\left[2\frac{W/S}{C_L}\frac{1}{V_c^2}\left(\frac{V_c^2}{V^2} - 1\right)\right]^{1.6}\left(\frac{V}{V_c}\right)^{8.5} \tag{P.5.33}$$

$$q_{rad,gas} = CR_N V_c^{8.5}\left[2\frac{W/S}{C_L}\frac{1}{V_c^2}\left(\frac{1}{(\bar{V}^*)^2} - 1\right)\right]^{1.6}(\bar{V}^*)^{8.5} \tag{P.5.34}$$

Isolating the constants from the velocity terms results in

$$q_{rad,gas} = CR_N V_c^{8.5}\left(2\frac{W/S}{C_L}\frac{1}{V_c^2}\right)^{1.6}\left(\frac{1}{(\bar{V}^*)^2} - 1\right)^{1.6}(\bar{V}^*)^{8.5} \tag{P.5.35}$$

$$q_{rad,gas} \propto \left(\frac{1}{(\bar{V}^*)^2} - 1\right)^{1.6}(\bar{V}^*)^{8.5} \tag{P.5.36}$$

Taking the derivative yields

$$\frac{\mathrm{d}q_{rad,gas}}{\mathrm{d}\bar{V}^*} = -3.2\left(\frac{1}{(\bar{V}^*)^2} - 1\right)^{0.6}\frac{1}{(\bar{V}^*)^3}(\bar{V}^*)^{8.5} + 8.5\left(\frac{1}{(\bar{V}^*)^2} - 1\right)^{1.6}(\bar{V}^*)^{7.5} = 0 \tag{P.5.37}$$

Dividing by $\left(\frac{1}{(\bar{V}^*)^2} - 1\right)^{0.6}(\bar{V}^*)^{5.5}$ gives

$$-3.2 + 8.5\left(\frac{1}{(\bar{V}^*)^2} - 1\right)(\bar{V}^*)^2 = 0 \tag{P.5.38}$$

$$\sqrt{\bar{V}^*} - (\bar{V}^*)^2 = \frac{3.2}{8.5} \tag{P.5.39}$$

Table P.5.4 Key points thermo-mechanical loads for an Earth ballistic entry

Parameter	V/V_E	h (km)
$q_{c,max,lam}$	0.847	53.5
$q_{rad,gas,max}$	0.828	54.1
$q_{c,max,turb}$	0.766	50.2
a_{max}	0.607	45.8

With standard root finding, the solution is $\bar{V''''}^* = \frac{V}{V_c} = 0.6605$. Note that the solution is normalised with the circular velocity, rather than the entry velocity that is used for normalisation. For the case considered in this series of questions: the entry velocity for the ballistic flight is $V_E = 0.9V_c$. In the previous question it was found that $\frac{V''''}{V_E} = 0.8284$, or $V = 0.8284 \times 0.9\ V_c = 0.7456\ V_c$. With $V'''' = 0.6605\ V_c$, the values are much closer together than, for instance, those for maximum deceleration (which are very much apart).

5.26. The key points for the thermo-mechanical loading during a ballistic flight are to be defined for maximum deceleration, convective heat flux (both type of boundary layers), and radiative heat flux. To be determined for an Earth entry with $K = 100$ kg/m^2, $V_E = 0.9V_c$, and $\gamma_E = -7.5°$. Other data to be found in Table 5.21. The answers are listed in Table P.5.4.
In Fig. 5.10, the curves for deceleration, and the two types of convective heat flux are shown. In Fig. P.5.6, the curve for the radiative heat flux has been added. It is clearly seen that the shape of the curve is changed, because of the large exponent in V (= 8.5) (Fig. P.5.6).

Ballistic, gliding, and skipping flight (Problems 5.27–5.30).

5.27. Since the maximum deceleration is given we need to use the expressions that give the corresponding altitude, since altitude is related to density and thus the atmosphere model. However, there are two unknowns to be estimated, so that a second equation is required. The maximum deceleration is given by:

$$\bar{a}_{\max} = -\frac{\beta \sin\gamma_E}{2e} V_E^2 \quad \Rightarrow \quad H_s = \frac{1}{\beta} = -\frac{\sin\gamma_E}{2e} \frac{V_E^2}{\bar{a}_{\max}} \tag{P.5.40}$$

such that H_s can be evaluated. Furthermore, the density-velocity relation is given by:

$$\rho = \frac{2K\beta \sin\gamma_E}{g} \ln \frac{V}{V_E} \tag{P.5.41}$$

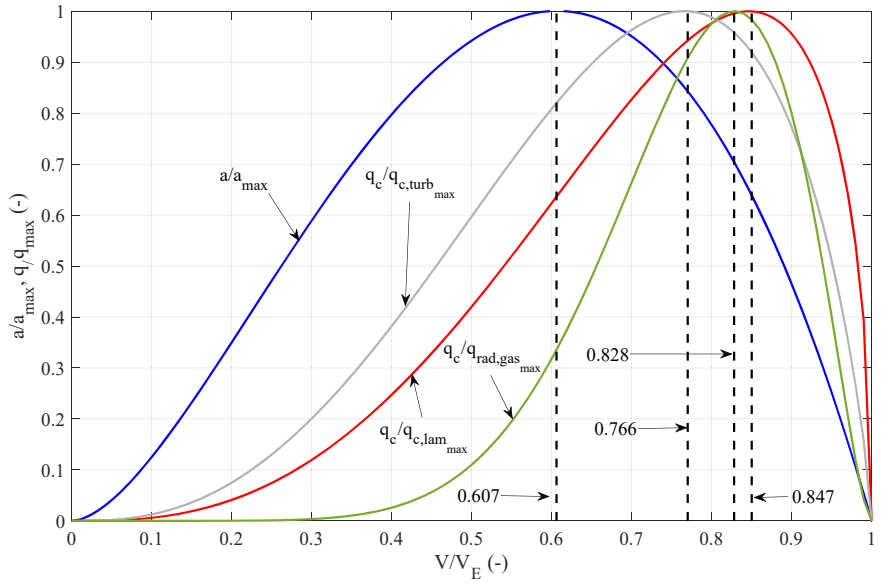


Fig. P.5.6 Generalised curves for the ballistic flight, with the radiative heat flux added

and the deceleration peaks at

$$\ln \frac{V/}{V_E} = -0.5 \tag{P.5.42}$$

This result substituted yields for the corresponding density and altitude:

$$\rho' = -\frac{K\beta \sin\gamma_E}{g} = \rho_0 e^{-\beta h'} \quad \Rightarrow \quad \rho_0 = -\frac{K \sin\gamma_E e^{\frac{h'}{H_s}}}{H_s g} \tag{P.5.43}$$

Substituting $\gamma_E = -7.5°$ and $\bar{a}_{\max} = 169.4$ m/s^2 into Eq. (P.5.40) gives $H_s = 15.5$ km, whereas substituting $\gamma_E = -7.5°$, $H_s = 15.5$ km, $g = 15.4$ m/s^2, $h' = 162.2$ km, and $K = 326.8$ N/m^2 into Eq. (P.5.43) results in $\rho_0 = 7.9$ kg/m^3.

5.28. Given is the altitude difference between peak heat flux and peak deceleration. The latter is known to occur at an equivalent density of

$$\rho' = -\frac{K\beta \sin\gamma_E}{g} \tag{P.5.44}$$

What remains is to derive the corresponding equation for when maximum heat flux occurs, starting from $q_c = c_1 \rho^{1-n} V^3$. This expression is given by:

$$\rho'' = \rho_0 e^{-\beta h'} = -\frac{2}{3}(1-n)\frac{K\beta \sin\gamma_E}{g} \tag{P.5.45}$$

Combining this with the expression for ρ':

$$\frac{\rho''}{\rho'} = \frac{2}{3}(1-n) \tag{P.5.46}$$

and with $\rho = \rho_0 e^{-\beta h}$ this yields

$$\frac{e^{-h''\beta}}{e^{-h'\beta}} = \frac{2}{3}(1-n) \tag{P.5.47}$$

Thus[2]:

$$\Delta h = h'' - h' = -H_s \ln\left[\frac{2}{3}(1-n)\right] \tag{P.5.48}$$

Isolating n yields

$$n = 1 - \frac{3}{2}e^{-\frac{\Delta h}{H_s}} \tag{P.5.49}$$

Given that $\Delta h = 9743.4$ m gives $n = 0.2$. This value indicates that the boundary-layer flow was turbulent. Note: for this (artificial) atmosphere we have assumed that the same conditions as on Earth apply for laminar and turbulent flow. For other planets, like Mars or Venus, the value for n could be (slightly) different.

5.29. At the beginning of the ballistic phase, the vehicle leaves the atmosphere with $V_F = 8{,}635.1$ m/s and $\gamma_F = \gamma_E = +8.5°$. In this flight phase kinetic and potential energy are exchanged but total energy remains constant.
With a horizontal and vertical velocity of $V_h = V_F \cos\gamma_F$ and $V_v = V_F \sin\gamma_F$ we find at the top of the ballistic flight that $V_{v,top} = 0 = V_F \sin\gamma_F - g t_{top}$, so

$$t_{top} = \frac{V_F \sin\gamma_F}{g} \tag{P.5.50}$$

The altitude difference covered is $\Delta h = V_F \sin\gamma_F t_{top} - \frac{1}{2} g t_{top}^2$. Further, $\gamma = 0$ and $V = V_h$, and with t_{top} given the horizontal distance flown is

$$X_{top} = V_F \cos\gamma_F t_{top} \tag{P.5.51}$$

Of course, this distance will be twice as much when the vehicle enters the atmosphere again. Thus, $X_{h,2} = 2X_{top}$. With V_F and γ_F substituted, it is found that $t_{top} = 73.8$ s and $X_{top} = 637.6$ km. Thus, $X_{h,2} = 1275.2$ km.

[2] One can also calculate the numerical value of h'', substitute it in the above equation and then calculate n.

5.30. The range for a gliding flight has been derived from the $\dot{V}$ and $\dot{\gamma}$ equations of motion, under the assumptions of $V_c^2 \approx gR_e$ and $V_F << V_E$. However, the former assumption does not hold here (see later), and the value for V_c has therefore been provided. So:

$$R_f = -\frac{1}{2}\frac{V_c^2}{g}\frac{L}{D}\ln\left(1 - \frac{V_E^2}{V_c^2}\right) \tag{P.5.52}$$

The derivation is not repeated here; see the appropriate section in Chap. 5, leading to Eq. (5.128). For a given flight range, the lift-to-drag ratio follows by re-arranging the above equation:

$$\frac{L}{D} = -\frac{2}{\ln\left(1 - \frac{V_E^2}{V_c^2}\right)}\frac{R_f g}{V_c^2} \tag{P.5.53}$$

Required flight range for gliding flight is the given 15,900 km minus $X_{h,1}$ and $X_{h,2}$, with $X_{h,1-2} = 1.25X_{h,2} = 1.25 \times 1275.2 = 1594.1$ km. So, $R_f = 14{,}306.4$ km. Finally, with $V_E = 8{,}635.1$ m/s and the given $V_c = 11{,}498.8$ m/s it follows that $\frac{V_E}{V_c} = 0.7510$ so $\frac{V_E^2}{V_c^2} = 0.5639$ Thus, for a range of $R_f = 14{,}306.4$ km the lift-to-drag ratio is: $L/D = 3.96$. Note that if one would use the approximation $V_c \approx V_{c,0} = \sqrt{g_0 R_e}$ (with $g_0 = 15.4$ m/s^2 and $R_e = R_e = 8385.9$ km, this gives $V_c = 11{,}364.1$ km/s) $L/D = 4.06$. But, as one can judge from the values for V_c, this approximation is not really justified.

Skipping flight (Problems 5.31–5.35).

5.31. Starting from the equations of planar motion, with the assumptions $D >> W$ (not required) and $L >> W$, the axial and normal acceleration are given by:

$$\frac{a_V}{g} = -\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{D}{mg} \tag{P.5.54}$$

$$\frac{a_n}{g} = \frac{V}{g}\frac{d\gamma}{dt} = \frac{L}{mg} \tag{P.5.55}$$

In skipping flight, both components should be considered. The normal acceleration can be rewritten as

$$\frac{a_n}{g} = \frac{L}{mg} = \frac{L}{D}\frac{D}{mg} = \frac{L}{D}\frac{a_V}{g} \tag{P.5.56}$$

For the resulting acceleration vectorially add the two components a_V and a_n:

$$\frac{a_{res}}{g} = \sqrt{\left(\frac{a_V}{g}\right)^2 + \left(\frac{a_n}{g}\right)^2} = \frac{a_V}{g}\sqrt{1 + \left(\frac{L}{D}\right)^2} \tag{P.5.57}$$

With $L/D = 2$ and $a_V = 15$ g it follows that $a_{res} = 15\sqrt{1+4}$ g = 33.54 g

5.32. From the equations of motion and realising that at the lowest point $\gamma = 0°$ (horizontal flight), one can derive

$$1 - \cos\gamma_E = \frac{g}{2\beta}\frac{C_L}{W/S}\rho_p = \frac{g}{2\beta}\frac{C_L}{W/S}\rho_0 e^{-h_p\beta} \tag{P.5.58}$$

or, with the use of $\cos 2\alpha = \cos^2\alpha - \sin^2\alpha = 1 - 2\sin^2\alpha$

$$\rho_p = \rho_0 e^{-h_p\beta} = \frac{4\beta}{g}\frac{W/S}{C_L}\sin^2\frac{\gamma_E}{2} \tag{P.5.59}$$

Isolating h_p:

$$h_p = -H_s\frac{4\beta}{g\rho_0}\frac{W/S}{C_L}\sin^2\frac{\gamma_E}{2} \tag{P.5.60}$$

From this equation it is understood that if the lift parameter, $\frac{W/S}{C_L}$, is constant (which it is by approximation for the hypersonic flight), and that the entry angle for each skip is the same (which it is, since $\gamma_{E,2} = -\gamma_{F,1} = \gamma_{E,1}$, the lowest point in each skip is the same for all. With the provided data $W/(SC_L)$) = 0.2 N/cm^2, $H_s = 7$ km and $\rho_0 = 1.225$ kg/m^3, $\gamma_E = -15°$, and $g_0 = 9.81$ m/s^2, it follows that $h_p = 44.98$ km.

5.33. The reduction in velocity for the first skip can be calculated with

$$\frac{V_F}{V_E} = e^{\frac{2\gamma_E}{L/D}} \tag{P.5.61}$$

and for n skips

$$\frac{(V_F)_n}{(V_E)_1} = e^{n\frac{2\gamma_E}{L/D}} = \left\{\frac{(V_F)_1}{(V_E)_1}\right\}^n \tag{P.5.62}$$

For a vehicle with $L/D = 3$, the exit velocity after 1 skip is $V_F = 6{,}718.8$ m/s, a reduction of 16.0%. The next skip reduces V_F to 5,642.8 m/s, a reduction with respect to V_E of 29.5%. Since no fractional skips can be executed, two skips are required to reduce the velocity with 25%.

5.34. For an exact velocity reduction of 25% with a single skip, the L/D should be 1.82. This follows from

$$\frac{L}{D} = \frac{2\gamma_E}{\ln\left(\frac{V_F}{V_E}\right)} \tag{P.5.63}$$

with $\frac{V_F}{V_E} = 0.75$.

5.35. The maximum deceleration in a skipping flight is calculated in three steps. Equation (5.284) gives $\gamma_{a,max}$ where it occurs, Eq. (5.278) provides the axial deceleration, $\frac{a_V}{g}$, and, finally, Eq. (5.275), defines the resulting acceleration, $\frac{a_{res}}{g}$:

$$\gamma_{a,max} = 2 \arctan \left[\frac{\frac{L}{D}}{2\left(1+\cos\gamma_E\right)} \left(1 - \sqrt{1 + \frac{4\sin^2\gamma_E}{\left(\frac{L}{D}\right)^2}} \right) \right] \tag{P.5.64}$$

$$\frac{a_V}{g} = \frac{\beta V_E^2}{g} \frac{D}{L} \left(\cos\gamma_{a,max} - \cos\gamma_E \right) e^{-\frac{2\left(\gamma_{a,max}-\gamma_E\right)}{L/D}} \tag{P.5.65}$$

$$\frac{a_{res}}{g} = \sqrt{\left(\frac{a_V}{g}\right)^2 + \left(\frac{a_n}{g}\right)^2} = \frac{a_V}{g}\sqrt{1+\left(\frac{L}{D}\right)^2} \tag{P.5.66}$$

For $L/D = 1.5$: $\gamma_{a,max} = -2.52°$, $a_V = 11.78$ g, $a_n = 17.67$ g, and $a_{res} = 21.24$ g.
For $L/D = 3$: $\gamma_{a,max} = -1.29°$, $a_V = 6.86$ g, $a_n = 20.57$ g, and $a_{res} = 21.69$ g.
For $L/D = 4.5$: $\gamma_{a,max} = -0.86°$, $a_V = 4.83$ g, $a_n = 21.72$ g, and $a_{res} = 22.25$ g.
With increasing L/D the point of maximum acceleration shifts towards the lowest point (where $\gamma_{a,max}$ would be 0°). The normal acceleration starts dominating, since the lift curves the trajectory more and more. Drag will have a lesser (relative) effect, because the axial deceleration is predominantly inversely proportional with L/D. It is curious to see, though, that the resulting acceleration does not change much. The one component's increase is compensated by the other component's decrease.

Chapter 6—Characteristic Motion of Entry Vehicles

TRUE/FALSE questions (Problems 6.1–6.9).

6.1. TRUE. Eigenvalues and eigenvectors can by definition only be determined for linear systems, as they are based on the (constant) nominal state of the system. Time dependency would therefore require a re-evaluation of the linearised state for the current time. The combination of theses aspects makes the system Linear Time Invariant.
6.2. FALSE. Supersonic and hypersonic winged vehicles often exhibit eigenmotions that are distinct from those of subsonic aircraft due to the significant differences in aerodynamic behaviour, flow regimes, and control mechanisms at high speeds.
6.3. FALSE. While longitudinal modes such as pitch and heave may indeed be significant, supersonic and hypersonic vehicles can also exhibit lateral modes

such as roll and yaw, especially considering the complex aerodynamic interactions and control requirements at high speeds.

6.4. TRUE. Eigenmotions, also known as eigenmodes or natural modes, are the inherent modes of oscillation or motion of a vehicle that arise due to its dynamic characteristics and without external disturbances.

6.5. TRUE. Due to the complex aerodynamic interactions and control mechanisms inherent in supersonic and hypersonic flight, eigenmotions of winged vehicles at these speeds can involve coupling between different degrees of freedom, leading to coupled modes of oscillation or motion.

6.6. FALSE. Eigenmotions of supersonic and hypersonic winged vehicles are heavily influenced by aerodynamic forces, including shock waves, boundary layer interactions, and control-surface effectiveness, which can significantly affect the vehicle's dynamic behaviour and stability.

6.7. FALSE. Analysing eigenmotions of hypersonic winged vehicles is crucial for understanding both structural dynamics and aerodynamic performance, as these aspects are tightly coupled and significantly impact the vehicle's overall flight behaviour and stability.

6.8. TRUE. Eigenmotions of supersonic and hypersonic winged vehicles can be inherently unstable or exhibit complex dynamic behaviour, requiring sophisticated control systems and stability augmentation to maintain desired flight characteristics and ensure safety.

6.9. FALSE. The characteristic motion of a system is defined by the combination of eigenvalues and eigenvectors, not just one of the components.

Linearisation and eigenmotion of entry capsules (Problems 6.10–6.14).

6.10. Studying the given equations gives for the state vector $\Delta\mathbf{x} = (\Delta V, \Delta p, \Delta q, \Delta r, \Delta\alpha)^T$ and for the control vector $\Delta\mathbf{u} = M_{T,y}$. To linearise, we need to evaluate the partials (= first derivatives) to each state and control.

6.11. Preparing for the linearisation, the pitch moment and lift force need special care. Both $\mathcal{M}$ and L are a function of Mach and α. Since the Mach number is not directly a state, but the aerodynamic coefficients have a dependency on Mach, the chain rule needs to be applied to go from Mach to velocity. Mach is defined as

$$M = \frac{V}{a} \tag{P.6.1}$$

with $a = \sqrt{\gamma RT}$ being the speed of sound. Even though the speed of sound is dependent on altitude (a state) we have assumed that the atmospheric properties are constant for the type of motion we are considering. Therefore:

$$\frac{\partial M}{dV} = \frac{1}{a} = \frac{V}{V}\frac{1}{a} = \frac{M}{V} \tag{P.6.2}$$

We have removed the speed of sound from the above partial, as this is a parameter usually unknown.

Both forces and moments are driven by the dynamic pressure, which is a function of altitude (through density ρ) and velocity. Since (again) density is constant, the partial for dynamic pressure to velocity is:

$$\frac{\partial \bar{q}}{dV} = \frac{\partial(0.5\rho V^2)}{dV} = \rho V = \frac{0.5V}{0.5V}\rho V = 2\frac{\bar{q}}{V} \tag{P.6.3}$$

We have removed the atmospheric density from the above partial, as this is a parameter usually unknown. Now that we have the above partials we can derive the partials to velocity of $\mathcal{M}$ and L:

$$\left.\frac{\partial \mathcal{M}}{\partial V}\right|_0 = \frac{\partial C_m}{\partial M}\frac{\partial M}{\partial V}\bar{q}_0 S_{ref} c_{ref} + C_m|_0 \frac{\partial \bar{q}}{\partial V} S_{ref} c_{ref} = \frac{\partial C_m}{\partial M}\frac{\partial M}{\partial V}\bar{q}_0 S_{ref} c_{ref} \tag{P.6.4}$$

Note that $C_m|_0 = 0$ because the equilibrium state is trimmed (no remaining pitching moment). However, if this is not stated and $C_m|_0$ appears in the equations do not subtract points. Substituting $\frac{\partial M}{dV}$ and $\frac{\partial \bar{q}}{dV}$ yields:

$$\left.\frac{\partial \mathcal{M}}{\partial V}\right|_0 = \frac{\partial C_m}{\partial M}\frac{M_0}{V_0}\bar{q}_0 S_{ref} c_{ref} \tag{P.6.5}$$

Similarly:

$$\left.\frac{\partial L}{\partial V}\right|_0 = \left(\frac{\partial C_L}{\partial M}M_0 + 2C_L\right)\frac{\bar{q}_0}{V_0} S_{ref} \tag{P.6.6}$$

Of course, since we also have an angle of attack dependency we also find in a straightforward manner:

$$\left.\frac{\partial \mathcal{M}}{\partial \alpha}\right|_0 = \frac{\partial C_m}{\partial \alpha}\bar{q}_0 S_{ref} c_{ref} \tag{P.6.7}$$

$$\left.\frac{\partial L}{\partial \alpha}\right|_0 = \frac{\partial C_L}{\partial \alpha}\bar{q}_0 S_{ref} \tag{P.6.8}$$

Summarised, with the given state vector we find for all partials of state matrix **A**:

$$\left.\frac{\partial \dot{q}}{\partial V}\right|_0 = \frac{1}{I_{yy}}\frac{M_0}{V_0}\frac{\partial C_m}{\partial M}\bar{q}_0 S_{ref} c_{ref} \tag{P.6.9}$$

$$\left.\frac{\partial \dot{q}}{\partial p}\right|_0 = -2\frac{I_{xz}}{I_{yy}}p_0 + \frac{I_{zz} - I_{xx}}{I_{yy}}r_0 \tag{P.6.10}$$

$$\left.\frac{\partial \dot{q}}{\partial r}\right|_0 = \frac{I_{zz} - I_{xx}}{I_{yy}}p_0 + 2\frac{I_{xz}}{I_{yy}}r_0 \tag{P.6.11}$$

$$\left.\frac{\partial \dot{q}}{\partial \alpha}\right|_0 = \frac{1}{I_{yy}} \frac{\partial C_m}{\partial \alpha} \bar{q}_0 S_{ref} c_{ref} \tag{P.6.12}$$

$$\left.\frac{\partial \dot{\alpha}}{\partial V}\right|_0 = -\frac{1}{mV_0^2} \left(M_0 \frac{\partial C_L}{\partial M} + C_L \right) \bar{q}_0 S_{ref} \tag{P.6.13}$$

$$\left.\frac{\partial \dot{\alpha}}{\partial q}\right|_0 = 1.0 \tag{P.6.14}$$

$$\left.\frac{\partial \dot{\alpha}}{\partial \alpha}\right|_0 = -\frac{1}{mV_0} \frac{\partial C_L}{\partial \alpha} \bar{q}_0 S_{ref} \tag{P.6.15}$$

and for **B**:

$$\left.\frac{\partial \dot{q}}{\partial M_y}\right|_0 = \frac{1}{I_{yy}} \tag{P.6.16}$$

The final answer should be given in scalar form:

$$\Delta \dot{q} = \left.\frac{\partial \dot{q}}{\partial V}\right|_0 \Delta V + \left.\frac{\partial \dot{q}}{\partial p}\right|_0 \Delta p + \left.\frac{\partial \dot{q}}{\partial r}\right|_0 \Delta r + \left.\frac{\partial \dot{q}}{\partial \alpha}\right|_0 \Delta \alpha \tag{P.6.17}$$

$$\Delta \dot{\alpha} = \left.\frac{\partial \dot{\alpha}}{\partial V}\right|_0 \Delta V + \left.\frac{\partial \dot{\alpha}}{\partial q}\right|_0 \Delta q + \left.\frac{\partial \dot{\alpha}}{\partial \alpha}\right|_0 \Delta \alpha \tag{P.6.18}$$

6.12. The eigenvalue problem is stated as $\mathbf{A}\boldsymbol{\mu} = \lambda \boldsymbol{\mu}$. This equation can be solved by

$$\mathbf{A}\boldsymbol{\mu} - \lambda \boldsymbol{\mu} = (\mathbf{A} - \lambda \mathbf{I})\, \boldsymbol{\mu} = 0 \tag{P.6.19}$$

The condition for a non-trivial solution is

$$|\mathbf{A} - \lambda \mathbf{I}\,| = 0 \tag{P.6.20}$$

$$\mathbf{A} - \lambda \mathbf{I} = \begin{bmatrix} -\lambda & \frac{1}{I_{yy}} \frac{\partial C_m}{\partial \alpha} \bar{q}_0 S_{ref} c_{ref} \\ 1.0 & -\frac{1}{mV_0} \frac{\partial C_L}{\partial \alpha} \bar{q}_0 S_{ref} - \lambda \end{bmatrix} \tag{P.6.21}$$

Note that the equations can be reversed, and instead of $|\mathbf{A} - \lambda \mathbf{I}\,| = 0$ it can be written as $|\lambda \mathbf{I} - \mathbf{A}\,| = 0$. Calculating the determinant of this matrix gives second-order polynomial:

$$\lambda^2 + \frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref} \lambda - \frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref} = 0 \tag{P.6.22}$$

Solution to this quadratic equation is:

$$\lambda_{1,2} = \frac{-\frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref} \pm \sqrt{\left(\frac{1}{mV_0} C_{L_\alpha} \bar{q}_0 S_{ref}\right)^2 + 4 \frac{1}{I_{yy}} C_{m_\alpha} \bar{q}_0 S_{ref} c_{ref}}}{2} \quad \text{(P.6.23)}$$

6.13. To be used in calculations: real part eigenvalue: $\mathrm{Re}(\lambda) = -1.9744 \cdot 10^{-2}$, imaginary part eigenvalue: $\mathrm{Im}(\lambda) = 15.534$. The damping factor, ζ, is defined as:

$$\zeta = -\frac{Re(\lambda)}{\sqrt{Re(\lambda)^2 + Im(\lambda)^2}} \quad \text{(P.6.24)}$$

which gives $\zeta = 1.2710 \cdot 10^{-3}$. The period, P, is expressed as:

$$P = \frac{2\pi}{Im(\lambda)} \quad \text{(P.6.25)}$$

which gives $P = 0.4\,\mathrm{s}$. The amplitude half time $T_{\frac{1}{2}}$ (because $\mathrm{Re}(\lambda) < 0$) is defined as:

$$T_{\frac{1}{2}} = \frac{\ln \frac{1}{2}}{Re(\lambda)} \quad \text{(P.6.26)}$$

which gives $T_{\frac{1}{2}} = 35.1\,\mathrm{s}$.

6.14. The mode under Problem 6.13 is stable, because the real part of the eigenvalue is negative (or: the damping ratio ζ is positive).

Kinematic attitude equations formulated with Euler angles φ, θ, and ψ (Problems 6.15–6.16).

6.15. Linearisation of the first equation $\dot{\varphi} = p + \sin\varphi \tan\theta q + \cos\varphi \tan\theta r$:

$$\left.\frac{\partial \dot{\varphi}}{\partial p}\right|_0 = 1 \quad \text{(P.6.27)}$$

$$\left.\frac{\partial \dot{\varphi}}{\partial q}\right|_0 = \sin\varphi_0 \tan\theta_0 \quad \text{(P.6.28)}$$

$$\left.\frac{\partial \dot{\varphi}}{\partial r}\right|_0 = \cos\varphi_0 \tan\theta_0 \quad \text{(P.6.29)}$$

$$\left.\frac{\partial \dot{\varphi}}{\partial \varphi}\right|_0 = \cos\varphi_0 \tan\theta_0 q_0 - \sin\varphi_0 \tan\theta_0 r_0 \quad \text{(P.6.30)}$$

$$\left.\frac{\partial \dot{\varphi}}{\partial \theta}\right|_0 = \frac{\sin\varphi_0}{\cos^2\theta_0} q_0 + \frac{\cos\varphi_0}{\cos^2\theta_0} r_0 \quad \text{(P.6.31)}$$

Second equation $\dot{\theta} = \cos\varphi q - \sin\varphi r$:

$$\left.\frac{\partial\dot{\theta}}{\partial q}\right|_0 = \cos\varphi_0 \tag{P.6.32}$$

$$\left.\frac{\partial\dot{\theta}}{\partial r}\right|_0 = -\sin\varphi_0 \tag{P.6.33}$$

$$\left.\frac{\partial\dot{\theta}}{\partial\varphi}\right|_0 = -\sin\varphi_0 q_0 - \cos\varphi_0 r_0 \tag{P.6.34}$$

Third equation $\dot{\psi} = \frac{\sin\varphi}{\cos\theta}q + \frac{\cos\varphi}{\cos\theta}r$:

$$\left.\frac{\partial\dot{\psi}}{\partial q}\right|_0 = \frac{\sin\varphi_0}{\cos\theta_0} \tag{P.6.35}$$

$$\left.\frac{\partial\dot{\psi}}{\partial r}\right|_0 = \frac{\cos\varphi_0}{\cos\theta_0} \tag{P.6.36}$$

$$\left.\frac{\partial\dot{\psi}}{\partial\varphi}\right|_0 = \frac{\cos\varphi_0}{\cos\theta_0}q_0 - \frac{\sin\varphi_0}{\cos\theta_0}r_0 \tag{P.6.37}$$

$$\left.\frac{\partial\dot{\psi}}{\partial\theta}\right|_0 = \frac{\sin\varphi_0}{\cos\theta_0}\tan\theta_0 q_0 + \frac{\cos\varphi_0}{\cos\theta_0}\tan\theta_0 r_0 \tag{P.6.38}$$

6.16. The state-space formulation requires the **A** and **B** matrices. Since the given equations are kinematic attitude equations, i.e., there are no moments present and thus no control variables can be discerned, **B** is a zero matrix of dimension 3×1 (which effectively means it can be removed from the state-space formulation). The system matrix **A** can only be partially filled, since the dynamic equations, i.e., the *Euler equations of rotational motion* were not asked to linearise. For the full system, the linearised state would thus be:

$$\Delta\mathbf{x} = \left(\Delta p\ \Delta q\ \Delta r\ \Delta\varphi\ \Delta\theta\ \Delta\psi\right)^{\mathrm{T}} \tag{P.6.39}$$

which means that only the bottom three rows of the 6×6 matrix can be filled:

$$\mathbf{A} = \begin{bmatrix} .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. \\ 1 & \mathrm{s}\varphi_0\mathrm{t}\theta_0 & \mathrm{c}\varphi_0\mathrm{t}\theta_0 & \mathrm{c}\varphi_0\mathrm{t}\theta_0 q_0 - \mathrm{s}\varphi_0\mathrm{t}\theta_0 r_0 & \frac{\mathrm{s}\varphi_0}{\mathrm{c}^2\theta_0} q_0 + \frac{\mathrm{c}\varphi_0}{\mathrm{c}^2\theta_0} r_0 & 0 \\ 0 & \mathrm{c}\varphi_0 & -\mathrm{s}\varphi_0 & -\mathrm{s}\varphi_0 q_0 - \mathrm{c}\varphi_0 r_0 & 0 & 0 \\ 0 & \frac{\mathrm{s}\varphi_0}{\mathrm{c}\theta_0} & \frac{\mathrm{c}\varphi_0}{\mathrm{c}\theta_0} & \frac{\mathrm{c}\varphi_0}{\mathrm{c}\theta_0} q_0 - \frac{\mathrm{s}\varphi_0}{\mathrm{c}\theta_0} r_0 & \frac{\mathrm{s}\varphi_0}{\mathrm{c}\theta_0}\mathrm{t}\theta_0 q_0 + \frac{\mathrm{c}\varphi_0}{\mathrm{c}\theta_0}\mathrm{t}\theta_0 r_0 & 0 \end{bmatrix} \tag{P.6.40}$$

using the short-hand notation "c" for *cosine*, "s" for *sine*, and "t" for *tangent*.

Chapter 7—Guidance, Navigation, and Control

TRUE/FALSE questions (Problems 7.1–7.15).

7.1. FALSE. The bias error is a "constant" deviation from the actual acceleration, independent of its magnitude. However, the bias itself may behave like a random walk, driven by white noise. In that sense, the bias can be considered to be the standard deviation of the white noise, with mean value zero.

7.2. TRUE. The heading-angle error may build up to a certain maximum value, and with careful planning the reduction to a value within the heading-error dead band can be achieved after the reversal. This planning does not only require an onboard optimisation process, but also a solid knowledge of the aerodynamic performance of the vehicle in the second half of the flight.

7.3. FALSE. Both accelerometers and gyroscopes suffer from bias errors. Bias can also be called drift, but is in essence the same.

7.4. TRUE. The navigation system can estimate any other variable for which there exist a relation between the available measurements and on-board models. A simple example could be that from acceleration measurements one can derive velocity and position, with a model of the Earth (angular rate, equatorial radius, flattening, *etc.*), the altitude can be evaluated. Finally, with an onboard model of the atmosphere (density versus altitude? More complex?) the (nominal) density can be derived.

7.5. TRUE. The integral term of a PID controller will keep growing in the presence of persistent errors. Multiplied with a (small/large) gain, this will give a (small/large) contribution to the actuator command. The larger this command, the faster the system will move. If this is in the direction of the set point, this faster motion will give rise to a larger maximum overshoot. So, a smaller gain value in that case is preferred.

7.6. While ground-based command and control centres may provide overall mission supervision and strategic guidance, hypersonic vehicles often incorporate onboard autonomy and decision-making capabilities to enable rapid response and real-time adjustments to changing flight conditions.

7.7. TRUE. To enhance reliability and accuracy, navigation systems of entry vehicles often incorporate redundant sensors, such as multiple accelerometers or

gyroscopes, and use sensor fusion algorithms to combine data from different sensors and improve navigation performance, particularly during critical phases of flight.

7.8. FALSE. While precomputed trajectories are often used as a baseline for entry vehicle guidance, real-time adjustments are typically necessary to account for deviations from the expected trajectory due to atmospheric uncertainties or other factors. Closed-loop guidance systems continuously monitor and adjust the vehicle's trajectory to optimise landing accuracy and ensure mission objectives are met.

7.9. FALSE. While characteristics of passive stabilisation, such as aerodynamic stability derivatives (with the proper signs), contribute to stability, hypersonic vehicles will often need active control systems to adjust the vehicle's attitude and trajectory. Active control systems use actuators to provide real-time adjustments in response to changing flight conditions.

7.10. TRUE. Adaptive control algorithms can be employed in hypersonic vehicles to continuously adjust control strategies in response to variations in flight conditions, such as changes in aerodynamic forces, structural dynamics, or atmospheric disturbances. These algorithms may use online parameter estimation techniques to adapt controller gains and dynamics to maintain stability and performance.

7.11. FALSE. The elevator is used for longitudinal control. To initiate a bank reversal the primary surface will be the aileron. Due to the roll-yaw coupling, the rudder may be used to correct for the induced angle of sideslip. This coupling is stronger for large angle of attack.

7.12. FALSE. In a typical implementation of a GNC system, the navigation system is the input interface with the real world, i.e., through sensors it measures the state of the vehicle and provides this information to the guidance system, the "black box" in the middle that provides commands to the control system. This system must guarantee that the commanded attitude is achieved by commanding the actuators, the output interface to the real world.

7.13. TRUE. The lateral guidance has to make sure the vehicle will remain targeted to the landing site (or the terminal-area interface). In other words, it controls the flight heading. The heading error, i.e., the difference between the actual and the reference heading angle, should remain within a so-called dead band, and exceeding this initiates a bank reversal.

7.14. TRUE. Reinforcement-learning algorithms offer a promising approach for autonomous control of hypersonic vehicles by allowing them to learn control policies through interaction with the environment. These algorithms can adaptively learn optimal control strategies for various mission objectives, including trajectory tracking, obstacle avoidance, and optimal resource usage, without relying on explicit mathematical models.

7.15. FALSE. Even though, in principle, the 3D position can be determined from three GPS signals, these signals contain errors and should also be related to the position and velocity of the corresponding satellites. Therefore, a time stamp is required. Thus, not just position should be estimated, also the time

of reception, as the clock in the receiver is not accurate enough to provide this directly. With three satellites only, the position estimate would be very inaccurate; it will be a lot more accurate with four or more satellites.

Navigation system: IMU measurement errors (Problems 7.16–7.17).

7.16. The measurement equation for the acceleration is given by:

$$\mathbf{a}_m = \mathbf{a} + \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} + \begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix} \mathbf{a} + \begin{bmatrix} 0 & m_{xy} & m_{xz} \\ m_{yx} & 0 & m_{yz} \\ m_{zx} & m_{zy} & 0 \end{bmatrix} \mathbf{a} = \mathbf{b_a} + (\mathbf{I} + \mathbf{S_a})\,\mathbf{a} \quad \text{(P.7.1)}$$

with

$$\mathbf{S_a} = \begin{bmatrix} s_x & m_{xy} & m_{xz} \\ m_{yx} & s_y & m_{yz} \\ m_{zx} & m_{zy} & s_z \end{bmatrix} \quad \text{(P.7.2)}$$

7.17. The calculation of acceleration, velocity, and position errors proceeds as follows. In the measurement equation we have:

$$\mathbf{a} = \begin{pmatrix} -11.8 \\ 1 \\ 4.3 \end{pmatrix} \text{m/s}^2 \quad \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = \begin{pmatrix} 0.0025 \\ -0.018 \\ 0.0025 \end{pmatrix} \text{m/s}^2 \quad \begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix} = \begin{pmatrix} 0.02 \\ -0.01 \\ 0.04 \end{pmatrix} \quad \text{(P.7.3)}$$

$$\begin{bmatrix} 0 & m_{xy} & m_{xz} \\ m_{yx} & 0 & m_{yz} \\ m_{zx} & m_{zy} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1.3 \cdot 10^{-1} & 1.3 \cdot 10^{-1} \\ 1.3 \cdot 10^{-1} & 0 & 1.3 \cdot 10^{-1} \\ 1.3 \cdot 10^{-1} & 1.3 \cdot 10^{-1} & 0 \end{bmatrix} \quad \text{(P.7.4)}$$

Substituting all the numbers, we get for the measured acceleration, $\mathbf{a_m}$:

$$\begin{aligned} \mathbf{a_m} &= \begin{pmatrix} -11.8 \\ 1 \\ 4.3 \end{pmatrix} + \begin{pmatrix} 0.0025 \\ -0.018 \\ 0.0025 \end{pmatrix} + \begin{bmatrix} 1 + 0.02 & 1.3 \cdot 10^{-1} & 1.3 \cdot 10^{-1} \\ 1.3 \cdot 10^{-1} & 1 - 0.01 & 1.3 \cdot 10^{-1} \\ 1.3 \cdot 10^{-1} & 1.3 \cdot 10^{-1} & 1 + 0.04 \end{bmatrix} \begin{pmatrix} -11.8 \\ 1 \\ 4.3 \end{pmatrix} \\ &= \begin{pmatrix} -11.3445 \\ -0.0030 \\ 3.0705 \end{pmatrix} \text{m/s}^2 \end{aligned} \quad \text{(P.7.5)}$$

The error in acceleration is:

$$\Delta\mathbf{a} = \mathbf{a_m} - \mathbf{a} = \begin{pmatrix} -11.3445 \\ -0.0030 \\ 3.0705 \end{pmatrix} - \begin{pmatrix} -11.8 \\ 1 \\ 4.3 \end{pmatrix} = \begin{pmatrix} 0.4555 \\ -1.0030 \\ -1.2295 \end{pmatrix} \text{m/s}^2 \quad \text{(P.7.6)}$$

In terms of velocity error (acceleration error is constant):

$$\Delta\mathbf{V} = \Delta\mathbf{a}\int_0^{65} dt = 65\begin{pmatrix}0.4555\\-1.0030\\-1.2295\end{pmatrix} = \begin{pmatrix}29.6075\\-65.1950\\-79.9175\end{pmatrix}\ \text{m/s} \tag{P.7.7}$$

The position error follows from integrating the velocity error:

$$\Delta\mathbf{r} = \Delta\mathbf{a}\int_0^{65}\int_0^{65} dt = 0.5\ \Delta\mathbf{a}\ 65^2 = 4225\begin{pmatrix}0.4555\\-1.0030\\-1.2295\end{pmatrix} = \begin{pmatrix}962.2\\-2118.8\\-2597.3\end{pmatrix}\ \text{m} \tag{P.7.8}$$

Control-system design—Inverted pendulum (Problems 7.18–7.25).

7.18. The given dynamics equation of the inverted pendulum is given by:

$$\ddot{\theta}(t) = \frac{g}{L}\sin\theta(t) + \frac{T(t)}{mL^2} \tag{P.7.9}$$

To split this into two first-order differential equations, we introduce the pitch velocity $q = \dot{\theta}$. Thus (and dropping the (t) notation):

$$\begin{aligned}\dot{q} &= \frac{g}{L}\sin\theta + \frac{T}{mL^2}\\ \dot{\theta} &= q\end{aligned} \tag{P.7.10}$$

Linearisation is easily done, by assuming small θ, such that $\sin\theta \approx \theta$:

$$\begin{aligned}\dot{q} &= \frac{g}{L}\theta + \frac{T}{mL^2}\\ \dot{\theta} &= q\end{aligned} \tag{P.7.11}$$

which can be written in matrix form:

$$\mathbf{\dot{x}} = \begin{pmatrix}\dot{q}\\ \dot{\theta}\end{pmatrix} = \overbrace{\begin{bmatrix}0 & \frac{g}{L}\\ 1 & 0\end{bmatrix}}^{\mathbf{A}}\overbrace{\begin{pmatrix}q\\ \theta\end{pmatrix}}^{\mathbf{x}} + \overbrace{\begin{bmatrix}\frac{1}{mL^2}\\ 0\end{bmatrix}}^{\mathbf{B}}\overbrace{T}^{u} \tag{P.7.12}$$

With the indicated **A** and **B** matrices, the above equation is in state-space form.

7.19. The eigenvalues of the open-loop system are the eigenvalues of **A**, and can be obtained by solving the characteristic equation (see Example 6.1 in Chap. 6). The condition for a non-trivial solution is $|\mathbf{A} - \lambda\mathbf{I}| = 0$, so:

$$\det\left(\begin{bmatrix}0 & \frac{g}{L}\\ 1 & 0\end{bmatrix}-\begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}\right)=0 \quad \Rightarrow \quad \begin{vmatrix}-\lambda & \frac{g}{L}\\ 1 & -\lambda\end{vmatrix}=0 \tag{P.7.13}$$

which gives: $\lambda^2-\frac{g}{L}=0$, with solutions $\lambda_{1,2}=\pm\sqrt{\frac{g}{L}}$. Conclusion is that the system has a stable (negative real eigenvalue) and an unstable mode (positive real eigenvalue), both representing an aperiodic motion.

7.20. Typical requirements that one can use to design the controller are the delay and rise time, maximum overshoot of a set point, and the settling time, given a certain tolerance, see, for instance, Fig. 7.45. Some of these criteria can be linked to the eigenvalues of the system (or the "poles"), and shifting the poles by applying feedback control is a process called *pole placement*. More details on this can be found in the many books on classic control-system design and analysis, for instance, the book by Ogata (2010).

7.21. A proportional feedback controller has the form $u=-K_pe$, with K_p a feedback gain[3] and e the error, i.e., the difference between the command (= set point) and the feedback signal. Taking the error on θ as input, with the set point being zero (so $e=0-\theta=-\theta$), u becomes

$$T=K_p^*\theta=-K_p\theta \tag{P.7.14}$$

which is of the form $u=-Kx$. Substituting a general control law $\mathbf{u}=-\mathbf{Kx}$ in the state space equation, Eq. (P.7.12), results in

$$\dot{\mathbf{x}}=\mathbf{Ax}+\mathbf{Bu}=\mathbf{Ax}-\mathbf{BKx}=(\mathbf{A}-\mathbf{BK})\mathbf{x}=\boldsymbol{A^*x} \tag{P.7.15}$$

which shows the earlier mentioned concept of pole placement: selecting a value of $\mathbf{K}$ affects the eigenvalues of $\mathbf{A}^*$. Substituting Eq. (P.7.14) into Eq. (P.7.12) yields:

$$\begin{pmatrix}\dot{q}\\ \dot{\theta}\end{pmatrix}=\begin{bmatrix}0 & \frac{g}{L}-\frac{K_p}{mL^2}\\ 1 & 0\end{bmatrix}\begin{pmatrix}q\\ \theta\end{pmatrix} \tag{P.7.16}$$

which means that the characteristic equation becomes

$$\lambda^2-\left(\frac{g}{L}-\frac{K_p}{mL^2}\right)=0 \tag{P.7.17}$$

with the solution $\lambda_{1,2}=\pm\sqrt{\frac{g}{L}-\frac{K_p}{mL^2}}$. It is clear that for any real eigenvalue, i.e., $\frac{g}{L}>\frac{K_p}{mL^2}$ there is always a positive solution, so the system is always unstable. For

$$\frac{g}{L}=\frac{K_p}{mL^2} \quad \Rightarrow \quad K_p=mgL \tag{P.7.18}$$

[3] The minus sign in the control law indicates that the error signal is a feedback signal. However, removing the minus sign will not change the concept; the only change will be, in that case, that the gain changes sign.

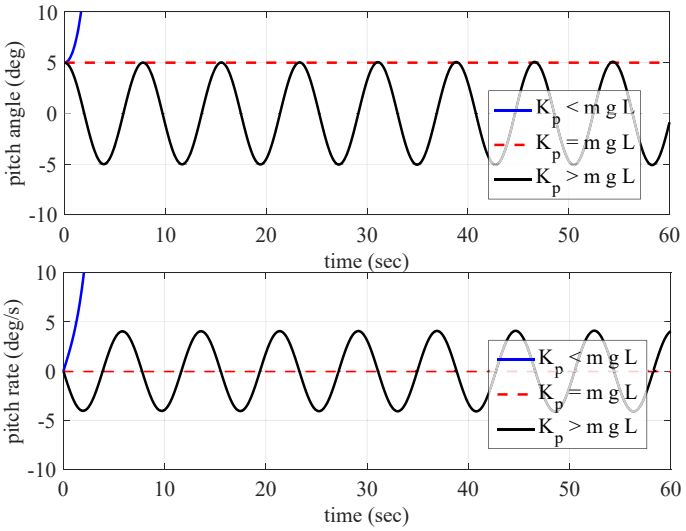


Fig. P.7.1 Response of the inverted pendulum to an initial perturbation of $\theta = 5°$. Three cases considered: $K_p = 0.9mgL$, $K_p = mgL$, and $K_p = 1.1mgL$

the two eigenvalues are zero. So, if the pendulum were to stand perfectly vertical, it would not fall. However, the slightest angular perturbation away from the vertical position will grow very quickly, causing the pendulum to fall from its vertical position. Because of its extreme sensitivity to perturbations the inverted pendulum is a classical example of an unstable system. Finally, in case $K_p > mgL$ the pendulum will be oscillating in between the negative and positive values of the initial perturbation. This oscillation is undamped. In conclusion, the inverted pendulum cannot be stabilised by proportional feedback control alone. Figure P.7.1 summarises the above discussion; initial perturbation is $\theta = 5°$.

7.22. After a bit of trial and error, the gains selected are $K_p = 28.0$ and $K_d = 7.4$. The resulting response is shown in Fig. P.7.2. The initial error works as a step error, and the control torque discretely reaches a very large value ($T > 600$ Nm, resulting in a large angular rate of almost $-14°$/s). However, the torque immediately reduces to more realistic values. The transient response is very smooth and the inverted pendulum is stabilised after about 3 s.

7.23. It is not necessary to add an integral component to the controller, as the steady-state error is very small. Of course, in case there is a strict requirement on this, this decision may have to be reviewed.

7.24. In case limiters to the control are added, i.e., instead of having an infinitely large torque available it is limited to $|T_{min}| = T_{max} = 5$ Nm and 1 Nm, respectively, the responses as shown in Fig. P.7.3 are obtained. Saturation limiters on the torque reduce the maximum angular rate, which in turn avoids an overshoot of θ. So, a less "violent" control is the consequence. Last, but not least, with a 1-Nm saturation the inverted pendulum cannot be stabilised, the control demands a larger torque than which is available.

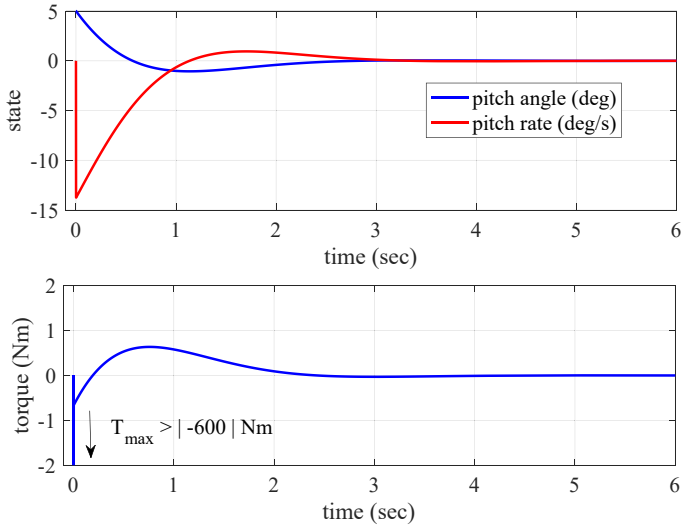


Fig. P.7.2 Response of the inverted pendulum to an initial perturbation of $\theta = 5°$, stabilised with a PD controller (unsaturated control)

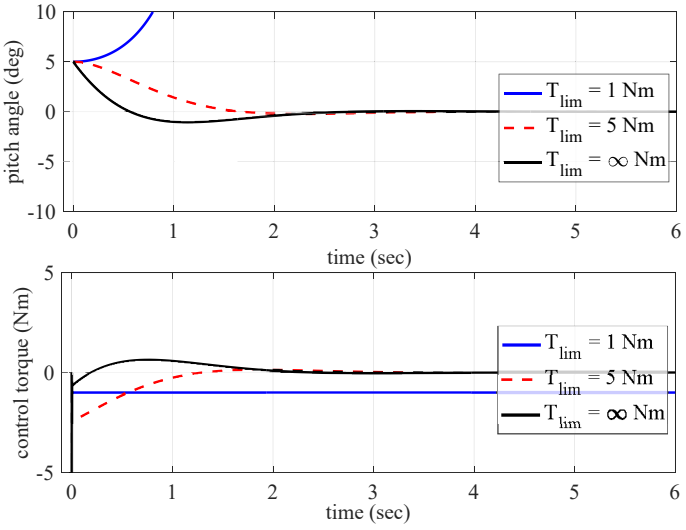


Fig. P.7.3 Response of the inverted pendulum to an initial perturbation of $\theta = 5°$, stabilised with a PD controller and limiters on the commanded torque

7.25. The stability and controllability of the system when either L or m is changed, can be found from inspecting the eigenvalues of the open-loop system. An increase of L (or decrease of m) will bring the unstable pole closer to the origin, which means that the system becomes less unstable. This will also

benefit the controllability, as the transient response becomes slower and easier to control. *Vice versa*, when the pendulum is shorter, or the tip mass heavier, the system becomes more unstable and harder to control.

As an example, the following Matlab® code may help you to get started with your own implementation.

```
%%
% Configuration
%
m    = 1.2;
L    = 1.5;
g    = 9.81;
Tmin = -5000;
Tmax = 5000;
%
% State space model
%
A = [0 g/L;1 0];
B = [1/(m*L^2);0];
C = eye(2);
D = [0;0];
%
% PID controller
%
Kd = 7.4;
Ki = 0;
Kp = 28;
%%
% Initialisation
%
x0     = [0;deg2rad(5)];
u      = 0;
eint   = 0; % zero-initialise the integral control error
eold   = 0; % previous error value
t0     = 0;
tfinal = 6;
dt     = 0.01;
nrtot  = (tfinal-t0)/dt+1;
t      = zeros(nrtot,1);
q      = zeros(nrtot,1);
theta  = zeros(nrtot,1);
T      = zeros(nrtot,1);
%
% Store initial conditions
%
t(1)     = t0;    % time
q(1)     = x0(1); % pitch rate
theta(1) = x0(2); % pitch angle
T(1)     = u;     % commanded torque
%
x        = x0;
```

```
nr        = 1;        % index variable (nr of integration steps)
t1        = t0;
t2        = t1 + dt;  % [t1 t2] is integration interval (single step)
while t2 <= tfinal+1e-6*dt
    nr    = nr + 1;
%
% State error, input to PID controller
%
    error = -x(2);
    dedt  = (error-eold)/dt;
%
% Calculate (unsaturated) control input (torque).
%
    u = max(min(Kp*error+Ki*eint+Kd*dedt,Tmax),Tmin);
%
% Evaluate eom (Euler integration)
%
    xdot = A*x + B*u;
    x    = x + xdot*dt;
%
% Store data
%
    t(nr)     = t2;    % time
    q(nr)     = x(1);  % pitch rate
    theta(nr) = x(2);  % pitch angle
    T(nr) = u;         % torque
%
% Update integral error (PID) and prepare for next step
%
    eint = eint + error*dt;
    eold = error;
    t1 = t2;
    t2 = t1 + dt;
end
```

(*) *Optimal control-system design—bank-reversal control (Problems* 7.26 *and* 7.29*).*

7.26. To design an LQR, the required input consists of **A** and **B**, which have been provided, as well as the weighing matrices **Q** and **R**, associated with the maximum allowable state deviation and the maximum available control effort, respectively. They can be specified with "Bryson's Rule":

$$\mathbf{Q} = \text{diag}\left\{\frac{1}{\Delta p_{max}^2}, \frac{1}{\Delta r_{max}^2}, \frac{1}{\Delta \beta_{max}^2}, \frac{1}{\Delta \sigma_{max}^2}\right\}$$

$$\mathbf{R} = \text{diag}\left\{\frac{1}{\Delta \delta_{a_{max}}^2}, \frac{1}{\Delta \delta_{r_{max}}^2}, \frac{1}{\Delta M_{T,z_{max}}^2}\right\}$$

where the actuator limits of HORUS will be applied: $\Delta\delta_{a_{max}} = \Delta\delta_{r_{max}} = 40°$, and $\Delta M_{T,z_{max}} = 7{,}600$ Nm. For the controller design to respond to a step com-

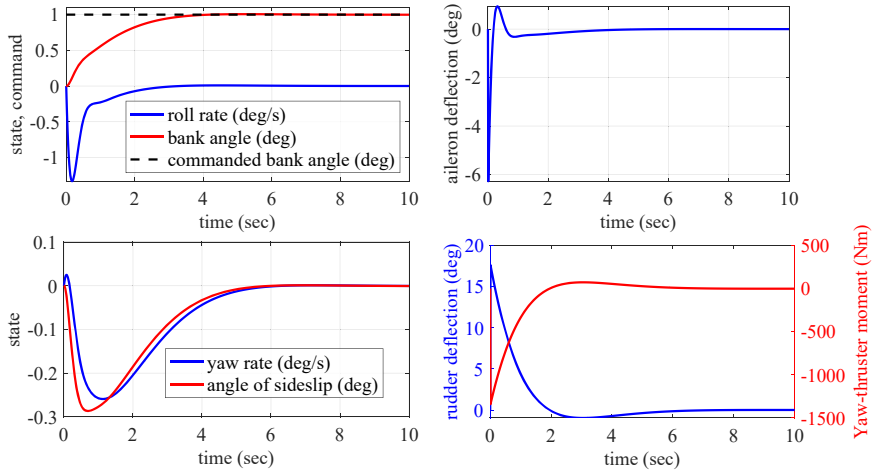


Fig. P.5.4 Optimal control of a bank-angle step response ($\sigma_c = 1°$)

mand of $\sigma_c = 1°$, the following values have been selected: $\Delta p_{max} = \Delta r_{max} = \infty$ °/s, and $\Delta\beta_{max} = \Delta\sigma_{max} = 2°$. Effectively, this means that if the angle error is larger than the specified values, the control will be saturated. However, it will also result in a tight control when the errors are well below these limit values.

Figure P.5.4 shows the transient response to $\sigma_c = 1°$. The set point is reached in about 4 s, without overshoot. Due to the roll-yaw coupling, the induced angle of sideslip peaks at $-0.3°$, but is smoothly reduced to zero. All three actuators, i.e., aileron, rudder, and yaw thruster, are active, and exhibit a moderate peak value: $|\delta_{a,max}|$ is just over 6°, $|\delta_{r,max}|$ is close to 18°, and $|M_{T,z,max}|$ is just below 1,400 Nm. It is clear that despite the small induced sideslip, yaw control is relatively strong. For the vehicle under consideration this can be explained that the moment of inertia around the Z-axis is significantly larger than the one around the X-axis. This may lead to control issues when the induced angle is larger.

7.27. For $\sigma_c = 5°$ and $\sigma_c = 10°$, the responses are shown in Figs. P.5.5 and P.5.6. The controller in both cases is the same as under Problem 7.26. Since the initial error (the step at $t = 0$ s is discrete), one finds that the controls are saturated, most notably the rudder and yaw thruster. With increasing commands in σ, the induced sideslip error is larger. In conclusion of these step responses, whenever large errors in the angles can be expected, $\mathbf{Q}$ has to be relaxed to avoid actuator saturation; however, this means a poorer control when the errors are small. In case a tight control in either situation is required, one may decide to change the controller to a non-linear one, or to add some form of integral control that kicks in when the errors are below a certain level.

7.28. After initialising the system with $\Delta\beta = 5°$, the response as shown in Fig. P.5.7 is obtained. As the initial error is larger than the specified $\Delta\beta_{max} = 2°$ for the

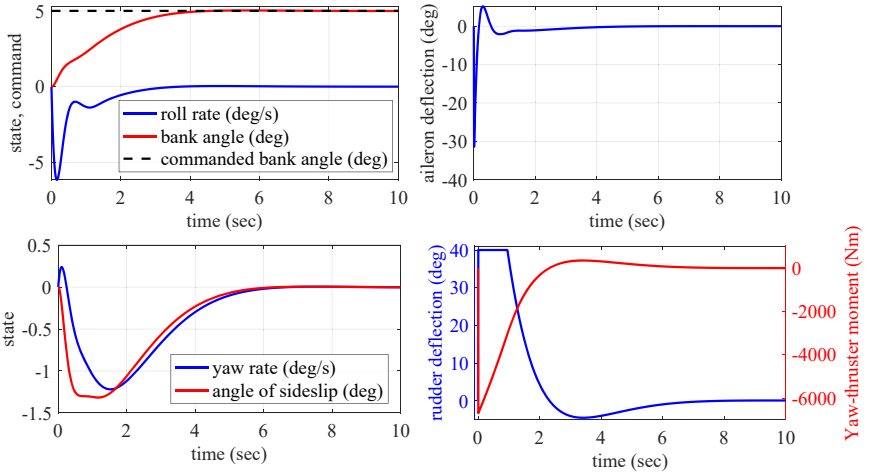


Fig. P.5.5 Optimal control of a bank-angle step response ($\sigma_c = 5°$)

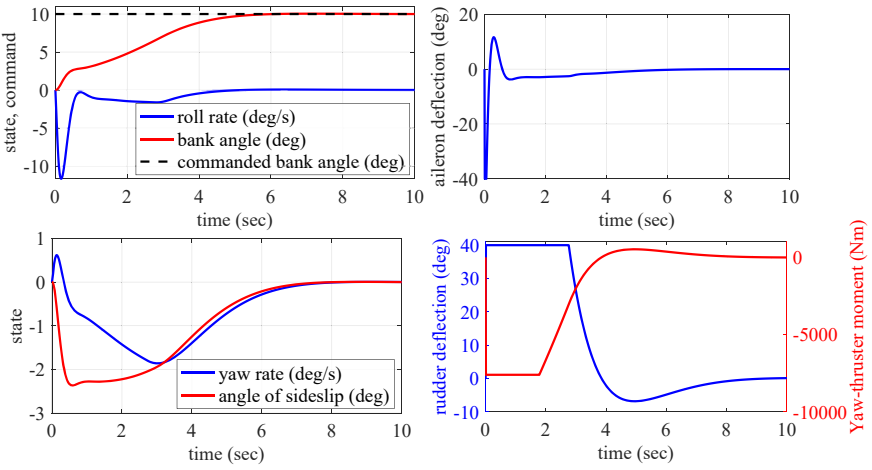


Fig. P.5.6 Optimal control of a bank-angle step response ($\sigma_c = 10°$)

controller design, the rudder and yaw thruster immediately saturate, although the controller remains stable. The induced bank angle is significant, due to the strong roll-yaw coupling. Note that a combined roll/yaw rate is required to control both angles. The results are consistent with what was found in the previous question(s).

7.29. The last response test concerns a bank reversal over 130°, with a duration of 13 s ($\dot{\sigma}_c = 10°/s$). The controller, as designed before, was not able to produce a stable transient response, so $\mathbf{Q}$ was adapted with $\Delta\sigma_{max} = 15°$ and $\Delta\beta_{max} = 10°$. The resulting response is found in Figs. P.5.8 and P.5.9. There is a slight

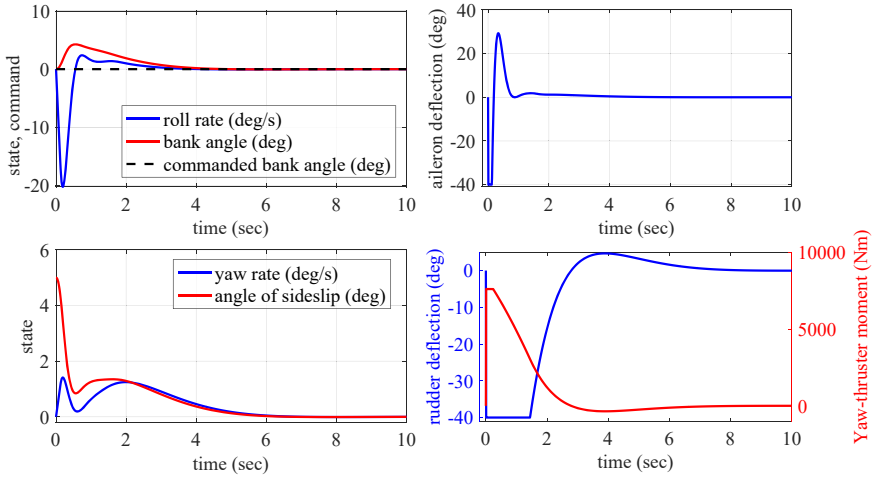


Fig. P.5.7 Optimal control due to an initial sideslip error ($\beta_0 = 5°$)

delay before the reversal begins, due to the fact that induced sideslip angle takes priority on being controlled. This introduces a counter command on the aileron, hence the delay in bank-angle response. The reader is referred to Sect. 7.5.7 for a more in-depth discussion on how to correct for this.

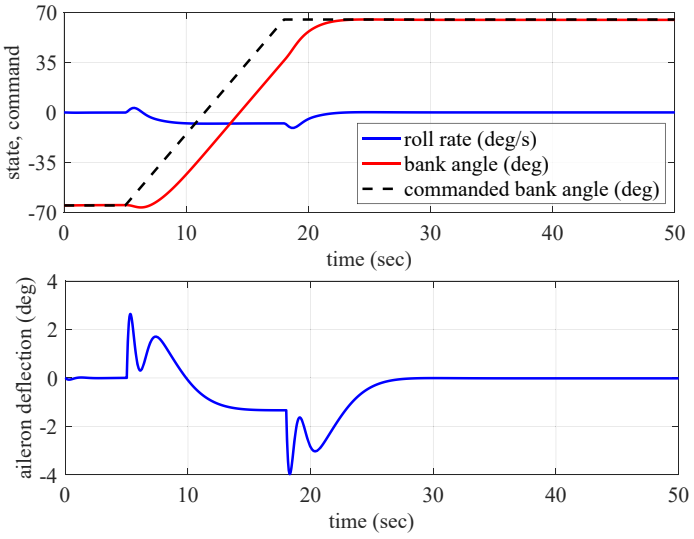


Fig. P.5.8 Optimal control of a bank reversal over 130°

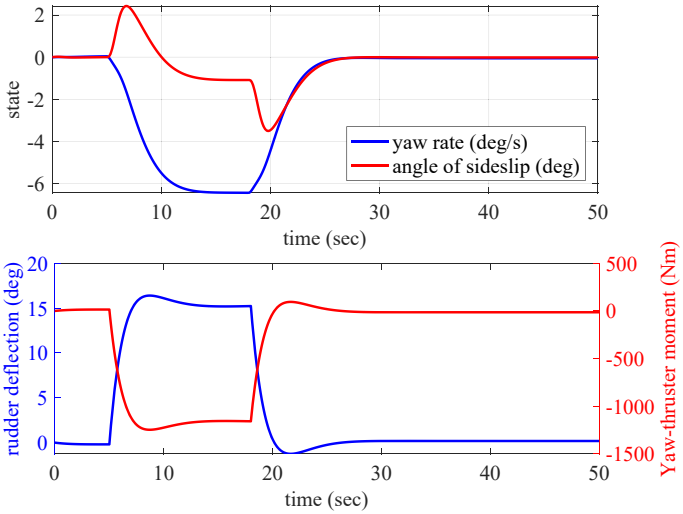


Fig. P.5.9 Optimal control of a bank reversal over 130°—concluded

Guidance-system design—g-load tracking (Problems 7.30–7.34).

7.30. The two given equations are the ones for flight duration (gliding flight) and mechanical load, the former derived with the assumption that L/D is constant, and for the latter it is specified that both C_D and C_L are constant. This is in line with the Mach independence principle for hypersonic flight.

$$t_{flight} = \frac{1}{2}\frac{V_c}{g}\frac{L}{D}\ln\left(\frac{1+V_E/V_c}{1-V_E/V_c}\right) \tag{P.7.19}$$

$$n_g = \frac{\sqrt{D^2+L^2}}{mg_0} \tag{P.7.20}$$

Equation (P.7.19) can be rewritten as

$$t_{flight} = c_1\frac{L}{D} \quad \Rightarrow \quad t_{flight}^2 \propto \left(\frac{L}{D}\right)^2 \tag{P.7.21}$$

assuming that g and V_c are approximately constant. For Eq. (P.7.20) we may write:

$$c_2 n_g^2 = \left(D^2+L^2\right) \quad \Rightarrow \quad D^2 = c_2 n_g^2 - L^2 \tag{P.7.22}$$

Substituting this into Eq. (P.7.21)

$$t_{flight}^2 \propto \frac{L^2}{c_2 n_g^2 - L^2} \tag{P.7.23}$$

With $c_2 n_g^2$ larger than L^2 the ratio on the right-hand side is positive (and smaller than 1). It can also be concluded that when n_g increases, this ratio decreases, which means that t_{flight} will reduce. The longer a flight along a constraint $n_g =$ constant will last, the shorter, therefore, the total flight duration will be.

7.31. The approach to find a tracking law for n_g is similar to the one derived in Sect. 7.6.4 for heat-flux tracking. First step is to find a nominal value for the drag, then to add a PI corrective control, and finally to extract a commanded angle of attack from the commanded drag. To find the nominal drag, D_0, the following steps are taken.
The *g-load* is defined as:

$$n_g = \frac{\sqrt{D^2 + L^2}}{m g_0} = \frac{\sqrt{C_D^2 + C_L^2}}{m g_0} \frac{1}{2} \rho V^2 S = c_1 \rho V^2 \tag{P.7.24}$$

where it is assumed that the aerodynamic properties of the vehicle are (for a large portion of the flight) constant.
For tracking a constant n_g, its time derivative must be zero, so

$$\dot{n}_g = c_1 \left(\dot{\rho} V^2 + 2 \rho V \dot{V} \right) = 0 \tag{P.7.25}$$

which means that

$$\dot{\rho} V^2 + 2 \rho V \dot{V} = 0 \Rightarrow \frac{\mathrm{d}V}{\mathrm{d}t} = -\dot{\rho} \frac{V}{2\rho} \tag{P.7.26}$$

Here, $\frac{\mathrm{d}V}{\mathrm{d}t} = -D - mg \sin\gamma$ is the dynamic equation of motion. This means that if we can find an alternative expression for $\dot{\rho}$ and substitute this in the above equation, an explicit expression for D_0 can be found.
The partial derivative $\frac{\mathrm{d}\rho}{\mathrm{d}t}$ is obtained from the exponential-atmosphere model $\rho = \rho_0 e^{-\beta h}$, with $\beta = \frac{1}{H_s}$ being the inverted scale height. So:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{\mathrm{d}\rho}{\mathrm{d}h} \frac{\mathrm{d}h}{\mathrm{d}t} = -\beta \rho \frac{\mathrm{d}h}{\mathrm{d}t} \tag{P.7.27}$$

With the kinematic altitude relation, $\frac{\mathrm{d}h}{\mathrm{d}t} = V \sin\gamma$ substituted, this yields

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\beta \rho V \sin\gamma \tag{P.7.28}$$

Thus,

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \beta \rho V \sin\gamma \frac{V}{2\rho} = \beta \sin\gamma \frac{V^2}{2} \tag{P.7.29}$$

and, together with the the dynamic equation of motion for $\dot{V}$:

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \beta \sin\gamma \frac{V^2}{2} = -D - mg\sin\gamma \tag{P.7.30}$$

Finally,

$$D_{cmd,0} = -mg\sin\gamma - \beta \sin\gamma \frac{V^2}{2} \tag{P.7.31}$$

7.32. The definition of the *g-load* error is $n_{g,err} = n_{g,setpoint} - n_g$, with $n_{g,setpoint}$ being some maximum constraint value. Thus, the three components of a PID controller can be written as:

P-term: $u(t) = K_p n_{g,err}(t)$, with $K_p =$ proportional gain,
I-term: $u(t) = K_i \int_0^t n_{g,err}(\tau)\mathrm{d}\tau$, with $K_i =$ integral gain, and
D-term: $u(t) = K_d \frac{\mathrm{d}n_{g,err}(t)}{\mathrm{d}t}$, with $K_d =$ derivative gain.

7.33. The PI-regulator is given by

$$\Delta D_{cmd} = -K_P n_{g,err} - K_I \int_0^t n_{g,err}\mathrm{d}t \tag{P.7.32}$$

such that the commanded drag can be written as:

$$D_{cmd} = D_{0,cmd} + \Delta D_{cmd} \tag{P.7.33}$$

7.34. Once D_{cmd} is known, given the definition of the drag ($D = C_D \frac{1}{2}\rho V^2 S$) and assuming a dependency of C_D on the Mach number and angle of attack, the commanded drag easily translates into a commanded angle of attack, α_{cmd}, which is the actual guidance command:

$$D_{cmd} = C_{D,cmd}\frac{1}{2}\rho V^2 S \Rightarrow C_{D,cmd} = C_{D,cmd}(\alpha_{cmd}, M) = \frac{D_{cmd}}{\frac{1}{2}\rho V^2 S} \tag{P.7.34}$$

When the (theoretical) drag coefficient is stored in the on-board computer as a polynomial or a multi-dimensional data table, through an inverse procedure, α_{cmd} can be retrieved for a given flight condition (i.e., combination of h and V).

Chapter 8—Terminal Area Energy Management

TRUE/FALSE questions (Problems 8.1–8.10).

8.1. TRUE. The TAEM guidance system of the Space Shuttle incorporates feedback control mechanisms that continuously adjust the vehicle's trajectory and descent rate based on real-time sensor feedback. This allows the guidance system to adapt to changing environmental conditions, such as wind gusts or

atmospheric disturbances, and adhere to landing constraints, such as runway length and glide slope angle.

8.2. FALSE. While minimising descent rate is one aspect of TAEM guidance, the primary goal is to regulate the vehicle's energy state, including airspeed and altitude, to ensure a safe and precise touchdown on the runway within specified landing constraints.

8.3. FALSE. The terminal area for winged entry vehicles is typically beginning some 80–50 km from the landing site.

8.4. FALSE. To remove excess energy the vehicle can also perform a zoom climb, or, as the terminal area has been designed as such, to fly to the heading-alignment cylinder across the landing direction, such that the flight duration is extended and thus more energy can be dissipated (see also Figs. 7.76 and 8.1).

8.5. FALSE. The targeting of the terminal area is initiated at entry, where the lateral guidance system will keep estimating the heading error towards the target, and checking whether it is still within the allowable dead band. The moment the dead-band limit is exceeded, a bank reversal is initiated.

8.6. TRUE. The key parameter in targeting the terminal area is indeed the heading error, i.e., the difference between the current flight heading and the nominal one aimed at the terminal-area targeting point. See also the answer to the previous question.

8.7. FALSE. The performance curve for a gliding flight relates the horizontal to the vertical flight speed, see also Fig. 8.4.

8.8. TRUE. This follows directly from Eq. (8.25).

8.9. FALSE. The unit of energy height is meter, which follows from the definition, Eq. (8.10): $\mathcal{E}_h = \frac{\mathcal{E}_{tot}}{mg_0} = \frac{\mathcal{E}_{tot}}{W}$.

8.10. TRUE. TAEM guidance of the Space Shuttle is specifically designed to control the vehicle's energy state during the final phase of descent and landing, ensuring that the Shuttle reaches the runway with the correct speed and altitude for a safe touchdown.

Theoretical aspects of terminal-area flight (Problems 8.11–8.15).

The vehicle is flying in a stationary glide, so the assumption: $\dot{V} = \dot{\gamma} = 0$

8.11. Given the assumption, the equations of motion in a stationary glide are:

$$-D - W\sin\gamma = 0 \quad \text{and} \quad L - \cos\gamma = 0 \tag{P.8.1}$$

or

$$-C_D\frac{1}{2}\rho V^2 S = W\sin\gamma \quad \text{and} \quad C_L\frac{1}{2}\rho V^2 S = W\cos\gamma \tag{P.8.2}$$

8.12. The airspeed in a stationary glide is obtained by re-writing the expression for C_L:

$$V = \sqrt{\frac{W}{S}\frac{2}{\rho}\frac{1}{C_L}\cos\gamma} \tag{P.8.3}$$

Dividing the expressions for C_D and C_L gives an expression for the flight-path angle:

$$-\tan\gamma = \frac{C_D}{C_L} \tag{P.8.4}$$

The rate of descent is equal to the vertical component of the flight velocity.

$$RD = V_v = \dot{h} = V\sin\gamma \tag{P.8.5}$$

The horizontal range is given by

$$x = \int V\cos\gamma dt \tag{P.8.6}$$

8.13. See Fig. 8.4.

8.14. The maximum ratio between horizontal and vertical velocity (negative RD) gives

$$-\tan\gamma = \frac{C_D}{C_L} \tag{P.8.7}$$

so $\frac{C_L}{C_D}_{max}$ gives yields a minimum flight-path angle

8.15. In a steady turn, a flight with constant bank angle gives:

$$V = \sqrt{\frac{W}{S}\frac{2}{\rho}\frac{1}{C_L}\frac{\cos\gamma}{\cos\sigma}} \tag{P.8.8}$$

$$-\tan\gamma = \frac{C_D}{C_L}\frac{1}{\cos\sigma} \tag{P.8.9}$$

Performance calculations for terminal-area flight (Problems 8.16–8.21).

8.16. The groundspeed can be derived from the equivalent airspeed. From the relation

$$\frac{V_{h1}}{V_{h0}} = \sqrt{\frac{\rho_{h0}}{\rho_{h1}}} \tag{P.8.10}$$

one finds $V = V_{eq}\sqrt{\frac{\rho_0}{\rho}}$. At sea level, i.e., h_0, the density is $\rho = \rho_0 = 1.225$ kg/m^3, whereas at $h = 10$ km, $\rho = 0.290(57)$ kg/m^3. With $V_{eq} = 114.93$ m/s it follows that $V = 235.98$ m/s.

8.17. The equilibrium glide angle follows from

$$\gamma_{eq} = -\arctan\left(\frac{C_D}{C_L}\right) \tag{P.8.11}$$

so, with $C_D = 0.068$ and $C_L = 0.272$ (i.e., the lift-to-drag ratio is 4.0), it follows that $\gamma_{eq} = -14.04°$. The rate of descent is given by

$$RD = -V_{h0} \sin \gamma_{eq} \tag{P.8.12}$$

With the previous found values for $V_{h0} = 235.98$ m/s and $\gamma_{eq} = -14.04°$, the rate of descent is $RD = 57.23$ m/s.

8.18. At 5 km altitude, the vehicle's velocity can be calculate from the velocity and density at $h_0 = 10$ km that was found earlier (see Answer 8.16): $V_{h0} = 235.98$ m/s and $\rho_{h0} = 0.290(57)$ kg/m^3. With a density at $h_1 = 5$ km of $\rho_{h1} = 0.59661$ kg/m^3. Then, with

$$\frac{V_{h1}}{V_{h0}} = \sqrt{\frac{\rho_{h0}}{\rho_{h1}}} \tag{P.8.13}$$

it follows that $V_{h1} = 164.69$ m/s. With $\dot{\chi} = 4°$/s and Eq. (8.43),

$$\dot{\chi} = \frac{V \cos \gamma}{R_{turn}} \quad \Rightarrow \quad R_{turn} = \frac{V_{h1} \cos \gamma_{eq}}{\dot{\chi}} \tag{P.8.14}$$

the turn radius is $R_{turn} = 2{,}288.5$ m. At this moment, we need to assume that γ_{eq} is not affected by the turn, or, in other words, the vehicle can somehow maintain its lift-to-drag ratio. This will be addressed in the next question, though. Finally, the required bank angle to fly this turn follows from Eq. (8.42):

$$\frac{L \sin \sigma}{m \cos \gamma} = V\dot{\chi} \quad \Rightarrow \quad \sigma = \arcsin\left(\frac{m V_{h1} \cos \gamma_{eq} \dot{\chi}}{L}\right) \tag{P.8.15}$$

The lift force is evaluated with $L = C_L \bar{q} S_{ref}$; with a dynamic pressure of $\bar{q} = 8.09$ kN/m^2 (kPa) it follows that $L = 24.207$ kN. Thus, $\sigma = 60.6°$.
In case $\dot{\chi} = 5°$/s, there is not sufficient lateral acceleration (= horizontal (lift) force) available to sustain this turn rate. This means that such a turn is beyond the capabilities of the vehicle. Of course, under the assumption that the equilibrium glide angle remains the same, as well as the available lift.

8.19. In a turning flight the equilibrium glide angle decreases according to Eq. (8.47):

$$-\tan \gamma = \frac{C_D}{C_L} \frac{1}{\cos \sigma} \tag{P.8.16}$$

so $\gamma_{eq,turn} = -27.01°$. This means that the assumption of constant γ_{eq} does not hold.

8.20. The turn radius for the newly found equilibrium glide angle is $R_{turn} = 2{,}101.6$ m.

8.21. The load factor is defined as (Eq. (8.45)):

$$n = \frac{\cos \gamma_{eq,turn}}{\cos \sigma} \tag{P.8.17}$$

With $\gamma_{eq,turn} = -27.01°$ and $\sigma = 60.6°$, it follows that $n = 1.82$ g.

Chapter 9—Entry, Descent, and Landing Systems

TRUE/FALSE questions (Problems 9.1–9.15).

9.1. FALSE. During the first reefing stage, the velocity (and thus dynamic pressure) is the largest, such that the drag deceleration due to the partially inflated parachute is largest. Prior to the last reefing stage, the velocity (and dynamic pressure) has reduced significantly, such that the drag deceleration is a lot smaller. Of course, if the first reefing step is deploying the parachute to 5% and a few seconds later the remaining 95% is deployed, then the peak load occurs during the second (= last) step, but this is an unrealistic scenario.

9.2. TRUE. In addition to parachutes, planetary landing systems may incorporate additional deceleration technologies such as retrorockets, airbags, or crushable materials to further reduce landing velocity and absorb impact energy, especially for larger or more massive spacecraft, or, in the case of Mars landings, when the atmosphere is very thin.

9.3. FALSE. Parachute clustering can be used for all kinds of payloads, either to divide the loads over multiple parachutes, or simply for reliability (independent of the payload mass).

9.4. FALSE. Parachute reefing can be used for both drogue and main parachutes, independent of the presence of a pilot chute.

9.5. TRUE. When a descent guidance system is present, it can issue commands to steer the vehicle in the right direction of the landing site, thereby reducing the landing ellipse. Of course, the guidance system can only work well, when also a navigation system is present.

9.6. TRUE. Environmental factors such as wind shear, atmospheric turbulence, and variations in atmospheric density can impact the stability and descent trajectory of parachutes during planetary landing, requiring careful planning and design considerations to mitigate their effects.

9.7. TRUE. E-guidance is a linear function of the time-to-go, where the two constants in the guidance law are obtained by solving the two-point boundary-value problem, i.e., both the initial and final position and velocity are prescribed. During execution of the guidance law, based on feedback signals, the time-to-go is continuously updated.

9.8. TRUE. In case the surface pressure is large, also the density is proportionally larger. This means that the drag efficiency of parachutes will increase, and for

the same final velocity one can make do with a smaller parachute, or for a given parachute land a larger payload mass. This is particularly beneficial for Mars landings, because the atmoshere is very thin.

9.9. FALSE. Precision landing requires the combination of an advanced navigation and guidance system. A hazard detection and avoidance system, as the name implies, detects hazards and based on predefined thresholds it can send redirection commands to the guidance system.

9.10. TRUE. See also the previous question. Once hazards have been detected and a redirection command has been sent to the guidance system, it is up to this system to guide the vehicle to the new landing location, with initially the same precision requirements as for the original lading site.

9.11. FALSE. Despite the fact that the final altitude and speed are input to the gravity-turn algorithm, activation is not based on a predetermined state, but rather a switch point of maximum available thrust deceleration. Finally, once activated there is no feedback loop providing information whether this final state can be achieved.

9.12. TRUE. The landing ellipses of both Mars Science Laboratory and Mars2020 are in the range of 10×5 km, which are the smallest achieved till date. This accuracy could only be achieved by the use of an active guidance system.

9.13. FALSE. Parachute systems can be designed and scaled to accommodate a wide range of spacecraft sizes and masses, from small robotic landers to large crewed capsules. The effectiveness of the parachute system depends on various factors, including atmospheric density, parachute design, and spacecraft mass.

9.14. TRUE. Parachute deployment-altitude and timing are critical parameters that must be carefully planned and coordinated to ensure that the parachute opens at the appropriate altitude and velocity to achieve a safe and controlled descent to the planetary surface.

9.15. TRUE. Terrain Absolute Navigation requires an overview of the landing area with sufficient visibility of (complete) landmarks. This can only be achieved in the early phase of descent at higher altitudes, as at lower altitudes it may be that only (part of) a single landmark is visible, insufficient for determining the position.

Parachute-system design (Problems 9.16–9.20).

Given data: lander mass $m = 400$ kg; Titan: $\rho_0 = 5.44$ kg/m^3, $g_0 = 1.354$ m/s^2. Mars: $\rho_0 = 0.016$ kg/m^3, $g_0 = 3.727$ m/s^2. $C_{D_{para}} = 0.512$.

9.16. Vertical equilibrium gives us:

$$\dot{V} = D_{para} - mg = 0 \tag{P.9.1}$$

or

$$C_D \frac{1}{2} \rho V^2 S = mg \tag{P.9.2}$$

With $S = \frac{1}{4}\pi d^2$ this becomes

$$\frac{1}{4}\pi d^2 = \frac{mg}{C_D \frac{1}{2}\rho V^2} \tag{P.9.3}$$

or

$$d = \sqrt{\frac{8mg}{C_D \rho V^2}} \tag{P.9.4}$$

With $\rho_0 = 5.44\,\text{kg/m}^3$, $V_f = 5\,\text{m/s}$, $m = 400\,\text{kg}$, $C_D = 0.512$ and $g_0 = 1.354\,\text{m/s}^2$ this gives $S = 15.55\,\text{m}^2$, or $d = 4.45\,\text{m}$.

9.17. The parachute does not change, because the second one is added for redundancy and thus the impact velocity should not change in case one parachute fails.

9.18. With $\rho_0 = 0.016\,\text{kg/m}^3$, $V_f = 5\,\text{m/s}$, $m = 400\,\text{kg}$, $C_D = 0.512$ and $g_0 = 3.727\,\text{m/s}^2$ this gives $S = 14{,}558.6\,\text{m}^2$, or $d = 136.1\,\text{m}$.

9.19. It is not feasible, the parachute would be too large. Retro rockets should be added to the design

9.20. On Titan, a 6-m parachute gives an impact velocity of 3.7 m/s. On Mars this would be 113.5 m/s. Reefing helps to lower the *g-load* peak, not the impact velocity, so the answer is "no".

Powered descent (Problems 9.21–9.26).

Given data: the gravitational acceleration can be assumed to be constant and is $g_{Mars} = 3.69\,\text{m/s}^2$ and $g_{Earth} = 9.81\,\text{m/s}^2$. The initial vehicle mass is 2,196 kg, whereas the 8 engines provide a thrust of 3,000 N each and have a specific impulse of 200 s (Earth referenced).

9.21. For a vertical flight, the equations of motion for the successive powered flight are:

$$\dot{V} = -\frac{D}{m} + g - \frac{T}{m} \tag{P.9.5}$$

$$\dot{h} = -V \tag{P.9.6}$$

9.22. In case the landing engines fail to ignite for 5 seconds, the velocity at the beginning of the powered phase can be calculated as follows. During these 5 s, constant vehicle drag can be assumed. So, for constant deceleration, the equation of motion Eq. (P.9.5) gives for the velocity $V = V_0 + g\Delta t - \frac{D}{m}t$. With $\Delta t = 5\,\text{s}$, $V_0 = 100\,\text{m/s}$, $g = 3.69\,\text{m/s}^2$, $D = 2{,}500\,\text{N}$ and $m = 2{,}196\,\text{kg}$, it follows that $\frac{D}{m} = 1.1384\,\text{m/s}^2$ and $V = 100 + 3.69{\cdot}5 - 1.1384{\cdot}5 = 112.76\,\text{m/s}$.

9.23. The corresponding altitude (at the beginning of the powered phase) follows from $h = h_0 - V_0\Delta t - \frac{1}{2}a\Delta t^2$, given the same assumption of constant deceleration. With $a = g - \frac{D}{m} = -3.69 + 1.14 = -2.55\,\text{m/s}^2$, it follows that $h = 1{,}800\text{–}500 - \frac{1}{2}2.55 \cdot 5^2 = 1{,}268.1\,\text{m}$.

9.24. To lower the rover to the surface of Mars all 8 engines are used simultaneously for 13 seconds. Assuming $m = 1{,}200$ kg at the beginning of this phase, and realising that the thrust needs to equate the weight, i.e., $T = mg$. With the numbers filled in it follows that $T = 1{,}200{\cdot}3.69 = 4{,}428$ N. The throttle setting follows from the maximum thrust: 8 engines, total $8{\cdot}3{,}000 = 24{,}000$ N. Required is only 4,428 N, so: $\delta_T = \frac{4{,}428}{24{,}000} = 0.1845 = 18.45\%$.
To calculate the amount of fuel used during that timespan, first the fuel-mass flow is calculated with:

$$\dot{m} = -\frac{T}{g_E I_{sp}} \tag{P.9.7}$$

Filling in the numbers gives $\dot{m} = -\frac{4{,}428}{9.81{\cdot}200} = -2.2569$ kg/s. With $t_{burn} = 13$ s, this gives $\Delta m_{fuel} = -29.34$ kg.

9.25. For $\Delta m_{fuel} = 29.34$ kg it follows that $m_{final} = 1170.7$ kg. The weight is $W = mg = 4319.7$ N. As the thrust is 4428 N, there is a surplus of $\Delta T = 108.26$ N. The maximum perturbing acceleration is therefore $\frac{\Delta T}{m_{final}} = \frac{108.26}{1170.7} = 0.0925$ m/s^2 (= 2.43%).

9.26. If the engines are not throttled to compensate for this error, the vehicle would move upwards.

Chapter 10—Thermal Protection Systems

TRUE/FALSE questions (Problems 10.1–10.10).

10.1. TRUE. While actively-cooled TPS systems are effective at managing heat, ablative heat shields are often better suited for withstanding extremely high thermal loads encountered during atmospheric re-entry. This is due to their ability to dissipate heat through ablation and the formation of insulating char layers, making them a preferred choice for missions with particularly demanding re-entry conditions.

10.2. FALSE. While reusable thermal protection systems aim to withstand multiple re-entry cycles without significant degradation, they still require inspection, maintenance, and potential refurbishment between missions to ensure continued effectiveness and safety.

10.3. FALSE. The selection of a thermal protection system depends on various factors, including re-entry velocity, mission duration, vehicle geometry, expected heating rates, and material properties. Each of these factors influences the thermal loads experienced by the vehicle and dictates the appropriate TPS design.

10.4. TRUE. Ablative thermal protection systems operate by gradually eroding or ablating away as they absorb and dissipate heat during re-entry. This process involves the material melting and vaporising, carrying away heat energy and protecting the underlying structure.

10.5. TRUE. Thermal protection systems are specifically engineered to shield the vehicle from the intense heat generated during re-entry into the (planetary) atmosphere, ensuring the integrity of the vehicle's structure and components.

10.6. FALSE. Thermal protection systems are essential for all re-entry vehicles, including those returning from orbit around Earth, interplanetary missions, or even suborbital flights. The heat generated during re-entry can pose a significant risk to the vehicle's integrity regardless of the mission profile.

10.7. TRUE. Non-ablative thermal protection systems typically use materials with high thermal conductivity or reflective properties to either reflect or refract heat away from the vehicle's surface, thereby minimising heat transfer to the underlying structure.

10.8. FALSE. While the primary focus of TPS is on protecting the vehicle's exterior surfaces from the intense heat of re-entry, internal components, such as avionics, propulsion systems, and crew compartments, may also require thermal insulation or cooling mechanisms to withstand the thermal environment.

10.9. FALSE. In the design of the thermal protection system one should also pay attention to the structural design, The interface layer between the TPS and the underlying cold structure has to prevent the heat going in, and also the connection of the TPS to the structure will affect the structural design. In some cases the TPS is both thermal-load and mechanical-load carrying, in which case it is even more evident that the two designs go hand in hand.

10.10. TRUE. Looking at the formulation of convective heat flux, for instance, Eq. (3.303),

$$q_{c,lam} = \frac{c^*_{lam}}{\sqrt{R_N}} \sqrt{\rho_\infty} u^3_\infty \left(1 - \frac{T_w}{T_{aw}}\right) \tag{P.10.1}$$

one can easily see that the term between brackets is smaller than 1 (hot-wall condition). When $T_w << T_{aw}$, this term is 1, leading to a larger q_c and thus a higher equilibrium wall temperature.

TPS design and analysis (Problems 10.11–10.15).

10.11. There is radiation equilibrium, thus:

$$q_c = q_{rad} \tag{P.10.2}$$

or

$$\alpha \left(T_{aw} - T_w\right) = \varepsilon \sigma T^4_w \tag{P.10.3}$$

From this equation the wall temperature of the leading edge can be determined. The heat transfer coefficient α can be determined with the heat flux for a wall temperature of 300 K with

$$q_c = \alpha \left(T_{aw} - T_w\right) \tag{P.10.4}$$

or

$$\alpha = \frac{q_{c_{T_w=300)}}}{(T_{aw} - 300)} \tag{P.10.5}$$

which yields

$$\alpha = \frac{200}{(2500 - 300)} = 0.909 \text{ kW/(Km}^2\text{)}. \tag{P.10.6}$$

10.12. For radiation equilibrium there is

$$\Delta q = q_c - q_{rad} = 0 \tag{P.10.7}$$

From the table we see that the requested wall temperature has a value between 1250 K and 1260 K. Linearly interpolating between two table entries gives

$$T_w = T_n + \frac{T_{n+1} - T_n}{\Delta q_{n+1} - \Delta q_n} \Delta q_n \tag{P.10.8}$$

with $T_n = 1250$ K, $T_{n+1} = 1260$ K, $\Delta q_n = 113.46 - 110.75 = 2.71$ kW/m^2, and $\Delta q_{n+1} = 112.73 - 114.34 = -1.61$ kW/m^2. Substitution yields:

$$T_w = 1256.27 \text{ K}$$

This temperature is acceptable, because it is lower than the maximum allowable

10.13. To determine the maximum allowable convective heat flux for the leading edge, one cannot use the equation for the convective flux, $q_c = \alpha\,(T_{aw} - T_w)$, because we enforce a maximum allowable temperature of $T_w = 1300$ K, such that the flux will be different from 200 kW/m^2. Therefore, α will have a different value as it was computed for $T_w = 300$ K (albeit that the change in α is small).
The flux should be calculated from radiation equilibrium, i.e.,

$$q_c = \varepsilon \sigma T_w^4 \tag{P.10.9}$$

with $T_w = 1300$ K substituted. This leads to $q_c = 130$ kW/m^2.

10.14. The emission coefficient changes from $\varepsilon = 0.8$ to $\varepsilon = 0.4$. So, for equal temperature the radiative heat flux is half. We use the same interpolation scheme as under Problem 10.12, but need to half the values in column 3. We find for the numerical values $T_n = 1430$ K, $T_{n+1} = 1450$ K, $\Delta q_n = 97.27 - 94.85 = 2.42$ kW/m^2, and $\Delta q_{n+1} = 96.36 - 97.53 = -1.17$ kW/m^2.
Substitution yields:

$$T_w = 1436.74 \text{ K}$$

This is threatening the structural integrity of the wing leading edge, because the temperature is larger than the maximum allowable.

10.15. To determine the maximum allowable convective heat flux for the leading edge with the removed surface coating, the same calculation procedure as for Problem 10.13 is used. The answer is half the flux value of Problem 10.13 (because of half the emission coefficient), so $q_c = 65$ kW/m^2.

TPS design with Mass Estimation Relationships (Problems 10.16–10.20).

Given: an Earth entry capsule, approximated by a truncated cone with a base diameter of $d_{front} = 4.2$ m, an aft diameter of $d_{aft} = 1.2$ m, and a height of 1.4 m. Other data to be used: $V_E = 11$ km/s, $q_{c,front} = 8$ MW/m^2, $Q_{front} = 100$ MJ/m^2, $q_{c,aft} = 100$ kW/m^2, and $Q_{aft} = 3$ MJ/m^2.

10.16. The MER for PICA is given in Table 10.4: $t_{MER} = 1.8696 \left(\frac{Q}{V_E^2}\right)^{0.1879}$ cm. Substituting $Q_{front} = 10$ kJ/cm^2 and $V_E = 11$ km/s, so $\frac{Q}{V_E^2} =$ and thus $t_{front} = 4.29$ cm (0.0429 m).

10.17. Assuming that the front of the heat shield is approximated by a flat, circular disk (diameter 4.2 m) with a 40% correction to account for the curvature of the shield, the surface area is 1.4 times $0.25\pi d^2 = 19.4$ m^2. With $\rho = 266$ kg/m^3, the mass is then $S_{front} t_{front} \rho_{pica} = 221.1$ kg.

10.18. The TPS of the aftbody is evaluated for three options, with $Q_{aft} = 300$ J/cm^2: SIRCA, Acusil II, and SLA-561V. The required thickness in cm for each of these materials follows from their corresponding MERs. For SIRCA: $t_{MER} = 0.5281 \left(\frac{Q}{V_E^2}\right)^{0.5416}$, which yields $t_{SIRCA} = 0.86$ cm, for Acusil II: $0.623 \left(\frac{Q}{V_E^2}\right)^{0.5697}$, resulting in $t_{AcusilII} = 1.05$ cm, and, finally, for SLA-561V: $t_{MER} = 0.0064 \left(\frac{Q}{V_E^2}\right)^2 + 0.0961 \left(\frac{Q}{V_E^2}\right) + 0.3322$, giving $t_{SLA-561V} = 0.61$ cm. Each of these thicknesses is above the minimum thickness for the MER to be valid.

10.19. With the provided densities, i.e., SIRCA ($\rho = 300$ kg/m^3), Acusil II ($\rho = 250$ kg/m^3), and SLA-561V ($\rho = 224.3$ kg/m^3), as well as the given surface area of 32.4 m^2, it follows that $m_{SIRCA} = 83.9$ kg, $m_{AcusilII} = 84.6$ kg, and $m_{SLA-561V} = 44.3$ kg. These masses are comparable for the former two, but the SLA-561V has a significant mass advantage. Note that this material was the material of choice for the aftbody of MSL and Mars2020.

10.20. The total TPS mass is 221.1+83.9 = 305 kg for the SIRCA aftbody, 305.6 kg for the Acusil II aftbody, and 265.4 kg for the SLA-561V aftbody. It is difficult to say whether these masses are realistic, for several reasons. First, it is hard to find mass breakdowns for entry capsules, where the TPS mass is isolated as a subsystem. Second, the TPS usually includes more elements than just the two covered here. There may be local protections for instruments or antennae, non-uniform thicknesses, and particular sub structures that are

included in the mass. Third, concerning the front shield, the assumption about the frontal area may not be accurate enough for a fair comparison.
However, the found thicknesses match well with the simulated data published by Sepka and Samareh (2015), who created the MERs.[4] Some other comparisons can be found in other literature, although it is difficult to find specific data on (parts) of the TPS designs, when it comes to geometry and mass. The thickness of the PICA front shield of Mars2020 was just over 3 cm, but that shield would be experiencing a lower heat load. The nominal, uniform thickness of the SLA-561V back shell was 1.27 cm, the same as for MSL (Mahzari et al. 2022). The PICA heat shield of Stardust had a pre-flight thickness of 6.45 cm (Beerman et al. 2009), designed for an Earth entry with $q_{c,max} = 9.42$ MW/m^2 and a higher integrated heat load.

Chapter 11—Aero-assisted Manoeuvres

TRUE/FALSE questions (Problems 11.1–11.14*).*

11.1 TRUE. Aerogravity assists involve leveraging the gravitational pull of a celestial body, such as a planet or moon, along with aerodynamic lift generated by the spacecraft's interaction with the body's atmosphere to modify the spacecraft's velocity and trajectory.

11.2 FALSE. While aerogravity assists involve complex interactions with celestial bodies and their atmospheres, they can be precisely planned and executed to achieve specific trajectory adjustments with high accuracy. Mission planners use advanced trajectory optimisation tools to maximise the effectiveness of aerogravity assists while minimising uncertainties. Knowledge about the atmosphere is a crucial input to the manoeuvre planning, though.

11.3 TRUE. In an aerocapture manoeuvre, the spacecraft typically enters the planet's atmosphere at a shallow angle, allowing the aerodynamic drag to gradually slow down the spacecraft while maintaining control, ultimately leading to orbital capture.

11.4 FALSE. Aerogravity assists can be employed during planetary flybys or encounters with any celestial body that possesses an atmosphere, such as Mars, Venus, or Titan. The interaction between the spacecraft and the atmosphere of these bodies allows for the use of aerodynamic lift and gravitational forces to modify the spacecraft's trajectory.

11.5 TRUE. Precise navigation and guidance systems are crucial for aerocapture missions to ensure that the spacecraft enters the planet's atmosphere at the correct angle and trajectory for a successful capture. Small deviations in entry parameters can result in mission failure or suboptimal orbital insertion.

11.6 FALSE. Even with shallow entry angles the heat load on the spacecraft can be significant, especially when the entry velocity is large. Therefore, thermal protection is always required, especially when capture is to be guaranteed (a deeper dive into the atmosphere is usually required).

[4] This and following references are included in the reference list for Chap. 11.

11.7 TRUE. By leveraging gravitational and aerodynamic forces, aerogravity assists can enable significant changes in a spacecraft's velocity without the need for large propulsive manoeuvres. This can lead to significant fuel savings and allow for shorter mission durations compared to purely propulsive manoeuvres.

11.8 FALSE. While aerogravity assists can indeed be used to decrease a spacecraft's velocity, they can also be used to increase velocity and alter the trajectory's direction, depending on the specific mission requirements. For example, aerogravity assists can be used to accelerate a spacecraft for interplanetary missions or to modify its trajectory for planetary exploration.

11.9 FALSE. Depending on the mission objectives and available opportunities, multiple (aero)gravity assist manoeuvres can be planned and executed throughout a spacecraft's trajectory. Sequential flybys of different celestial bodies can be used to further adjust the spacecraft's trajectory or achieve complex mission objectives.

11.10 TRUE. Due to the uncertainties in atmospheric knowledge of the target body, in combination with the usually large speed, small errors in the entry conditions can lead to (very) large thermo-mechanical loads. This makes the aerogravity assist a more complex and riskier manoeuvre compared to a gravity assist.

11.11 FALSE. With a gravity assist the velocity change (magnitude and direction) will not be as large as with an aerogravity assist. Any error induced by atmospheric uncertainty can be compensated for at the expense of a small amount of fuel, which will be far less compared to the benefits the atmosphere has provided.

11.12 TRUE. Aerocapture missions may use multiple passes through the planet's atmosphere to gradually lower the spacecraft's orbit or fine-tune its trajectory as needed, providing flexibility and optimisation in orbital insertion or change.

11.13 TRUE. Aerocapture manoeuvres are often conducted autonomously by the spacecraft's onboard systems, which are preprogrammed with the necessary trajectory and guidance commands. Ground-based operators may provide oversight and monitoring, but generally have minimal intervention during the manoeuvre itself, as communication delays and time-critical responses prevent that.

11.14 FALSE. Atmospheric conditions play a significant role in aerogravity assists, as the density, composition, and structure of the atmosphere directly impact the magnitude of aerodynamic forces experienced by the spacecraft. Mission planners must account for these atmospheric factors when designing aerogravity assist manoeuvres.

Hohmann trajectory (Problems 11.15–11.19)

Data to be used: $\mu_{Sun} = 1.3271 \cdot 10^{11}$ km^3/s^2, distance Earth-Sun = 1 au = $149.6 \cdot 10^6$ km, and distance Neptune-Sun = 30.1 au).

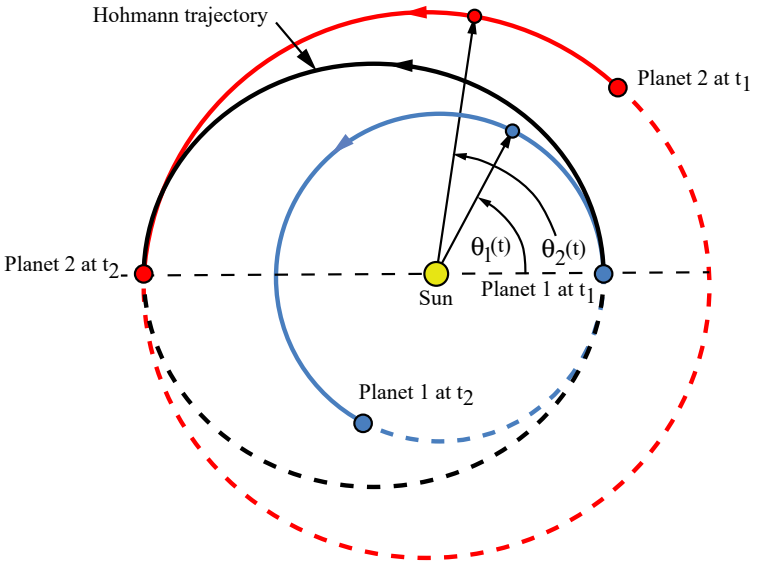


Fig. P.11.1 Hohmann transfer from Planet 1 to Planet 2, with definition of the true anomaly, θ

11.15. A Hohmann transfer from an inner Planet 1 to an outer Planet 2 is schematically shown in Fig. P.11.1. For the positions at Epoch 1 we may write:

$$\theta_1(t_1) = \theta_1(t_0) + n_1(t_1 - t_0) \quad \theta_2(t_1) = \theta_2(t_0) + n_2(t_1 - t_0) \quad \theta_{sat}(t_1) = \theta_1(t_1) \tag{P.11.1}$$

whereas for Epoch 2 the positions of the planets are given by

$$\theta_1(t_2) = \theta_1(t_1) + n_1 T_H \quad \theta_2(t_2) = \theta_2(t_1) + n_2 T_H \tag{P.11.2}$$

with T_H the duration of the Hohmann transfer. The satellite's position at Epoch 2 is

$$\theta_{sat}(t_2) = \theta_{sat}(t_1) + \pi = \theta_1(t_1) + \pi = \theta_2(t_2) \tag{P.11.3}$$

from which it follows that

$$\theta_1(t_0) + n_1(t_1 - t_0) + \pi = \theta_2(t_0) + n_2(t_1 - t_0) + n_2 T_H \tag{P.11.4}$$

From this, we can isolate t_1

$$t_1 = t_0 + \frac{\theta_2(t_0) - \theta_1(t_0) + n_2 T_H - \pi}{n_1 - n_2} \tag{P.11.5}$$

which is the equation to prove. And, of course

$$t_2 = t_1 + T_H \tag{P.11.6}$$

11.16. To calculate the time for a Hohmann transfer from Earth to Neptune, we need to know the semi-major axis of the transfer ellipse. From Fig. P.11.1 it can be found that the major axis of the transfer ellipse is the sum of the orbital radii of Earth (1 au) and Neptune (30.1 au). Obviously, the semi-major axis is half of that: $a_H = 15.55$ au (= $2.32628 \cdot 10^9$ km). With Eq. (11.139)

$$T_H = \pi \sqrt{\frac{a_H^3}{\mu_{Sun}}} \tag{P.11.7}$$

and a_H and μ_{Sun} substituted, it follows that $T_H = 9.6759 \cdot 10^8$ s (= 11,198.9 days, or 30.68 years, ignoring leap years).

11.17. On January 1, 2010, it is assumed that $\theta_{Earth} = 70°$ and $\theta_{Neptune} = 240°$. From Eq. (P.11.5) we find that we require the mean orbital motion of Earth and Neptune, n_1 and n_2, respectively. The orbital periods of Earth and Neptune are 365.2 and 59,800 days,[5] which yields $n_1 = 0.986$ °/day and $n_2 = 6.02 \cdot 10^{-3}$ °/day. So, with $T_H = 11{,}198.9$ days from the previous question, $t_1 - t_0 =$ -64.5 days, i.e., about October 28, 2009 (so before the given epoch t_0).

11.18. The arrival epoch is given by $t_2 = t_1 + T_H$, so $t_2 = 11{,}134.5$ days, i.e., 30.485 years after January 1, 2010: about June 26, 2040.

11.19. (*) The launch date can be shifted by the so-called synodic period (or an integer times this period), i.e., the time it takes for the two planets to be aligned with the same relative positions as shown in Fig. P.11.1. The synodic period, T_{syn} is easily derived, starting with the positions of planets 1 and 2:

$$\theta_1(t) = \theta_{1,0} + n_1(t - t_0) \qquad \theta_2(t) = \theta_{2,0} + n_2(t - t_0) \tag{P.11.8}$$

which gives for the difference

$$\Delta\theta(t) = \theta_2(t) - \theta_1(t) = (\theta_{2,0} - \theta_{1,0}) + (n_2 - n_1)(t - t_0) \tag{P.11.9}$$

With the requirement that the geometry should repeat after T_{syn}, it follows that

$$\Delta\theta(t_2) - \Delta\theta(t_1) = 2\pi = (n_2 - n_1)(t_2 - t_1) = (n_2 - n_1)T_{syn} \tag{P.11.10}$$

or

$$\frac{1}{T_{syn}} = \left| \frac{1}{T_2} - \frac{1}{T_2} \right| \tag{P.11.11}$$

For instance, the synodic period for Earth and Neptune is about 1 year, determined by the fact that the further away from the Sun one of the planets is,

[5] See, for instance, https://nssdc.gsfc.nasa.gov/planetary/factsheet/, date accessed: December 30, 2022.

the less of an effect on the shift in T_{syn} it will have. For the Earth-Neptune case, with orbital periods of 365.2 and 59,800 days, it is easily verified that the reciprocal period for Neptune is close to zero.

Gravity assist (Problems 11.20–11.25)

Consider a gravity assist along Jupiter. The following data may be used: $\mu_{Sun} = 1.3271{\cdot}10^{11}$ km^3/s^2, $\mu_{Jupiter} = 1.267{\cdot}10^8$ km^3/s^2, and the distance Jupiter-Sun = 5.20 au, with 1 au = $149.6{\cdot}10^6$ km. This question closely follows the steps shown in Example 11.3. Therefore, the answers here are concise; details can be found in said example.

11.20. The velocity relation for a hyperbolic orbit is

$$V^2 = V_{esc}^2 + V_\infty^2 \tag{P.11.12}$$

with the escape velocity given by $V_{esc} = \sqrt{\frac{2\mu}{r}}$. For the pericentre, with $R_p = 200{,}000$ km, it follows that $V_{esc} = 35.59$ km/s. Then, with the above equation and $V_\infty = 10$ km/s, $V_p = 36.97$ km/s.

11.21. The characteristics of the orbit, semi-major axis, a, and eccentricity, e, are calculated as follows. The semi-major axis follows from the energy at $r \to \infty$, with $V \to V_\infty = 10$ km/s:

$$\frac{V_\infty^2}{2} = -\frac{\mu}{2a} \quad \Rightarrow \quad a = -\frac{\mu_{Jupiter}}{V_\infty^2} \tag{P.11.13}$$

so $a = -1{,}269{,}000$ km. Then, from the definition of pericentre radius,

$$R_p = a(1 - e) \tag{P.11.14}$$

it follows that $e = 1.158$.

11.22. The heliocentric velocity of Jupiter in a circular orbit around the Sun is calculated with $V_{Jupiter} = \sqrt{\frac{\mu_{Sun}}{R_{Jupiter}}}$. The orbital radius is provided as $R_{Jupiter} = 5.20$ au ($= 777.92{\cdot}10^6$ km). Thus, $V_{Jupiter} = 13.061$ km/s.

11.23. The diagram with the velocities of Jupiter and the spacecraft with respect to a heliocentric system, and the velocity of the spacecraft with respect to Jupiter is shown in Fig. P.11.2. V_s^- and V_s^+ are the spacecraft's velocity before and after the gravity assist.

11.24. To establish the bending angle, the limit value of the true anomaly, θ_{lim}, is required. The relation between θ_{lim} and e is given by

$$\cos\theta_{lim} = -\frac{1}{e} \tag{P.11.15}$$

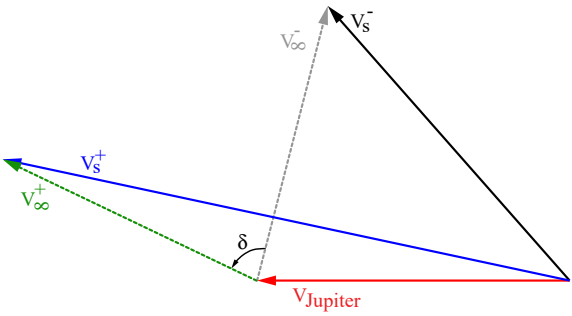


Fig. P.11.2 Velocity diagram for a gravity assist along Jupiter. Note that $V_\infty^- = V_\infty^+ = V_\infty$

such that with $e = 1.158$ it follows that $\theta_{lim} = 149.72°$, and therefore $\delta = 2\theta_{lim} - \pi = 119.44°$. To compute the vehicle's velocity before and after the gravity assist, i.e., V_s^- and V_s^+, use can be made of the vector diagram of Fig. P.11.2 (see for an alternative formulation, Eqs. (11.118) through (11.120)):

$$(V_s^-)^2 = V_{Jupiter}^2 + V_\infty^2 - 2V_{Jupiter}V_\infty \cos\left(\frac{\pi}{2} - \frac{\delta}{2}\right) \tag{P.11.16}$$

$$(V_s^+)^2 = V_{Jupiter}^2 + V_\infty^2 - 2V_{Jupiter}V_\infty \cos\left(\frac{\pi}{2} + \frac{\delta}{2}\right) \tag{P.11.17}$$

$$\Delta V = 2V_\infty \sin\frac{\delta}{2} \tag{P.11.18}$$

Substituting the relevant data results in $V_s^- = 6.707$ km/s and $V_s^+ = 22.276$ km/s ($\Delta V = 15.569$ km/s).

11.25. The energy gain caused by the gravity assist is calculated with

$$\Delta\mathcal{E} = \frac{(V_s^+)^2}{2} - \frac{(V_s^-)^2}{2} \tag{P.11.19}$$

yielding $\Delta\mathcal{E} = 225.611$ km^2/s^2.

Aerogravity assist (Problems 11.26–11.33)

A vehicle with $L/D = 3$ enters the sphere of influence of Venus with (relative) $V_\infty = 8$ km/s, and performs an aerogravity assist at $h_p = 200$ km. The following data may be used: $\mu_{Venus} = 324{,}858.592$ km^3/s^2, $R_e = 6{,}051.8$ km, and the distance Venus-Sun = $a = 0.723332$ au ($= 108.21 \cdot 10^6$ km).

Table P.11.1 Velocity values for $L/D = 3$, 4, and 5; aerogravity assist along Venus

L/D	$V_{s,i}$	$V_{s,f}$	ΔV	V_∞^-	V_∞^+	ΔV
(–)	(km/s)	(km/s)	(km/s)	(km/s)	(km/s)	(km/s)
3	8.000	7.613	1.387	8.000	2.394	5.606
4	8.000	7.687	0.313	8.000	4.091	3.901
5	8.000	7.738	0.262	8.000	4.931	3.069

11.26. The orbital velocity of Venus is 35.02 km/s and calculated with $V_{Venus} = \sqrt{\frac{\mu_{Sun}}{a}}$. At $R_p = R_e + h_p$, the circular velocity is given by $\sqrt{\frac{\mu_{Venus}}{R_p}}$, which yields $V_{c,p} = 7.2085$ km/s.

11.27. The (relative) velocity with which the vehicle enters the atmosphere, $V_{s,i}$, is, of course dependent on the rotation of the planet (where it is assumed that the atmosphere is co-rotating with the planet).[6] Venus's rotation is very slow (one rotation per 243 days), such that the effect can safely be ignored. Thus, $V_{s,i} \approx V_\infty^- = 8$ km/s.
Given the atmospheric bending angle, θ, of 60°, one can compute the (relative) velocity when exiting the atmosphere using Eq. (11.166):

$$V_{s,f} = \sqrt{e^{-\frac{2\theta}{L/D}} V_{s,i}^2 - (e^{-\frac{2\theta}{L/D}} - 1) V_c^2} \tag{P.11.20}$$

Substituting the known values for L/D, θ, $V_{s,i}$, and V_c results in $V_{s,f} = 7.613$ km/s.

11.28. The outgoing hyperbolic excess velocity is given by

$$V_\infty^+ = \sqrt{e^{-\frac{2\theta}{L/D}} (V_\infty^-)^2 + (e^{-\frac{2\theta}{L/D}} - 1) V_c^2} \tag{P.11.21}$$

such that $V_\infty^+ = 2.394$ km/s.

11.29. The velocity loss in the atmosphere is $\Delta V = V_{s,i} - V_{s,f} = 1.387$ km/s. Concerning the hyperbolic excess velocity, the difference is $\Delta V = V_\infty^- - V_\infty^+ = 5.606$ km/s.

11.30. The calculations for $L/D = 4$ and $L/D = 5$ are shown in Table P.11.1. An increasing L/D reduces the velocity loss. With an active guidance system it is possible to extend the atmospheric flight by banking the vehicle. To avoid large inclination changes, though, bank reversals should be included in the mission design.

11.31. The maximum bending angle for the three values of L/D are 67.5°, 89.9° and 112.4°, respectively. The obvious conclusion is that a larger L/D will allow

[6] Note that the motion of Venus in its orbit around the Sun should only be considered when determining the heliocentric velocity (see Fig. 11.46).

for a longer atmospheric flight, if properly executed. See also the previous question.

11.32. The situation changes when $V_\infty = 6$ km/s or 10 km/s? A lower V_∞ leads to a lower entry velocity than the local circular velocity: this means aerocapture, rather than being able to do an aerogravity assist. Even with $V_\infty = 8$ km/s it could be that the margin with respect to $V_{c,p} = 7.2085$ km/s is (too) small for a real-life scenario. Obviously, with $V_\infty = 10$ km/s, the margin is a lot larger, and it will allow for larger bending angles and thus heliocentric velocity changes. The reader is advised to check this themselves, by going through the previous questions with the larger V_∞ substituted.

11.33. The *equivalent radius*, R_{equiv}, represents the closest approach in a normal gravity assist that would lead to the same angular deflection (the combination of the atmospheric turn and the effect of the gravity-assist component). Equation (11.192) gives the expression for this radius:

$$R_{equiv} = \frac{\mu_{target}}{(V_\infty^+)^2}\left(\frac{1}{\sin\frac{\Phi}{2}} - 1\right) \tag{P.11.22}$$

with the total deflection angle in the aerogravity assist, Φ, given by Eq. (11.178):

$$\Phi = 2\arcsin\left(\frac{1}{1+\frac{(V_\infty^-)^2}{V_c^2}}\right) + \theta + 2\arcsin\left(\frac{e^{\frac{2\theta}{L/D}}}{1+\frac{(V_\infty^-)^2}{V_c^2}}\right) \tag{P.11.23}$$

For the nominal case, $V_\infty = 8$ km/s and $L/D = 3$, $\Phi = 153.6°$, which leads to $R_{equiv} = 1537$ km. Obviously, a gravity assist to achieve the same deflection as the aerogravity assist is impossible, as this would require a sub-surface passage ($R_e = 6{,}051.8$ km). The other combinations of input are left as exercise. The reader is referred to Fig. 11.52 for more insight on this. Conclusions are that the higher the hyperbolic exit velocity is (smaller $\bar{V}$), the larger R_{equiv} will be, and that for a given Φ the actual value of L/D doe not influence the outcome so much.

Appendix A
General Formulation of the Equations of Motion

A.1 Introduction

A vehicle, when moving in the sphere of influence of a celestial body, is subjected to a number of external forces, some of them may be atmospheric, gravitational, magnetic or perturbing forces caused by a third body, e.g., the Sun or another planet. This description generally holds for satellites orbiting the Earth or aircraft flying in the atmosphere of the Earth, where the application usually determines which forces to consider. The motion of the vehicle can be divided into the translation *of* the centre of mass (c.o.m.) and the rotation *about* its c.o.m.. The equations, which describe the corresponding motion are therefore called the equations of translational motion and the equations of rotational motion.

The basis for the equations motion is formed by Newton's Laws of Motion. For a point mass with constant mass, they read:

1. Every point mass continues in its state of uniform motion in a straight line (or rest), unless compelled to change that state by forces acting upon it.
2. The time rate of change of linear momentum of a point mass is proportional to the force acting upon that point mass and is collinear with and in the direction of that force.
3. The mutual forces of two point masses acting upon each other are equal in magnitude and opposite in direction (action = reaction).

The first and second law hold for a certain reference frame with respect to which a point mass, free of all external forces, moves in a straight line with a constant speed. Such a frame is called an inertial frame. However, Newton's laws can be modified to derive the equations of motion in a (rotating) non-inertial frame. In this chapter, we will derive the general formulation of the equations of motion for an elastic, mass-varying body with respect to a rotating (non-inertial) frame. In Sect. A.2 we will first discuss the basic mathematics of the time derivative of a vector in a rotating frame, and use this in the next Sect. A.3 to derive expressions for the velocity and

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1

acceleration in a rotating frame. Sect. A.4 will present the equations of translational motion, whereas Sect. A.5 will give the derivation of the equations of rotational motion. In that section, also the angular momentum with respect to the c.o.m. of a solid body is discussed, and used to formulate the so-called Euler equations.

A.2 Time Derivative of a Vector in a Rotating Frame

Let us consider two frames: a fixed (inertial) frame with index I and a moving frame with index R which is rotating with an angular velocity $\boldsymbol{\omega}$ with respect to the inertial frame.[7] Suppose we want to take the time derivative of a vector $\mathbf{r}$ expressed in the R-frame and we want to define the result in the I-frame. The begin situation is shown in Fig. A.1.

For the vector $\mathbf{r}$ we can write:

$$\mathbf{r} = r_x \mathbf{x_R} + r_y \mathbf{y_R} + r_z \mathbf{z_R} \tag{A.1}$$

in which $\mathbf{x_R}$, $\mathbf{y_R}$ and $\mathbf{z_R}$ are the unit axes in the R-frame. If we take the derivative of $\mathbf{r}$ with respect to time, we get:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(r_x \mathbf{x_R} + r_y \mathbf{y_R} + r_z \mathbf{z_R}\right) \\ &= \frac{\mathrm{d}r_x}{\mathrm{d}t}\mathbf{x_R} + \frac{\mathrm{d}r_y}{\mathrm{d}t}\mathbf{y_R} + \frac{\mathrm{d}r_z}{\mathrm{d}t}\mathbf{z_R} + r_x \frac{\mathrm{d}\mathbf{x_R}}{\mathrm{d}t} + r_y \frac{\mathrm{d}\mathbf{y_R}}{\mathrm{d}t} + r_z \frac{\mathrm{d}\mathbf{z_R}}{\mathrm{d}t} \end{aligned} \tag{A.2}$$

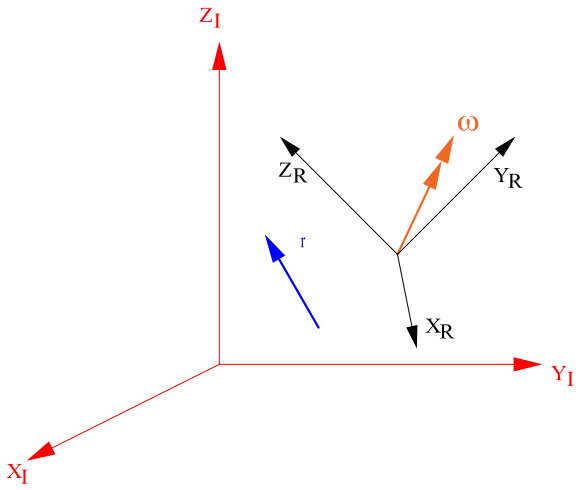


Fig. A.1 Relative angular motion. The index I denotes the inertial frame, whereas the R-index stands for the rotating frame

[7] As a starting point we consider one of the two frames to be inertial. However, as this assumption is not actively used, it means that the derived equations will also be valid for two frames, which are rotating relatively to each other.

The end p of a vector $\mathbf{r}$, which is a function of time and which rotates with an angular velocity $\boldsymbol{\omega}$ with respect to the inertial frame, will describe a curved path if the other end remains at a fixed point. The linear velocity of p with respect to inertial is given by (Cornelisse et al. 1979):

$$\mathbf{V_p} = \boldsymbol{\omega} \times \mathbf{r} \tag{A.3}$$

For the last three terms of Eq. (A.2) we can write, using Eq. (A.3):

$$r_x \frac{\mathrm{d}\mathbf{x_R}}{\mathrm{d}t} + r_y \frac{\mathrm{d}\mathbf{y_R}}{\mathrm{d}t} + r_z \frac{\mathrm{d}\mathbf{z_R}}{\mathrm{d}t} = r_x \boldsymbol{\omega} \times \mathbf{x_R} + r_y \boldsymbol{\omega} \times \mathbf{y_R} + r_z \boldsymbol{\omega} \times \mathbf{z_R} = \boldsymbol{\omega} \times \mathbf{r} \tag{A.4}$$

Finally, if we consider the first three terms of the right-hand side of Eq. (A.2) as the time derivative of $\mathbf{r}$ expressed in the rotating frame, $\frac{\delta \mathbf{r}}{\delta t}$, we obtain:

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\delta \mathbf{r}}{\delta t} + \boldsymbol{\omega} \times \mathbf{r} \tag{A.5}$$

A.3 Velocity and Acceleration in a Rotating Frame

We will now use Eq. (A.5) to derive expressions for the velocity and acceleration in a point p as given in Fig. A.2. If $\mathbf{R}$ is the position of the rotating frame with respect to the inertial frame, $\mathbf{r}$ and $\tilde{\mathbf{r}}$ the position vectors of p relative to the origins of the inertial and rotating frame, respectively, then:

$$\mathbf{r} = \mathbf{R} + \tilde{\mathbf{r}} \tag{A.6}$$

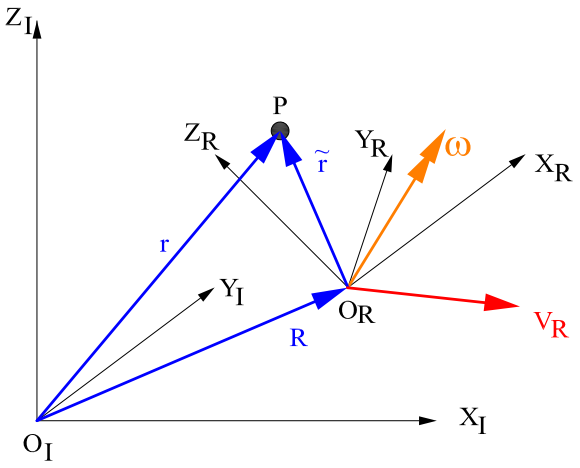


Fig. A.2 Motion of a point p with respect to a fixed and a moving frame

The velocity of p with respect to the inertial frame is:

$$\mathbf{V_I} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} + \frac{\mathrm{d}\tilde{\mathbf{r}}}{\mathrm{d}t} \tag{A.7}$$

Note that all derivatives are expressed in the inertial frame. Applying Eq. (A.5) yields:

$$\mathbf{V_I} = \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} + \frac{\delta\tilde{\mathbf{r}}}{\delta t} + \boldsymbol{\omega} \times \tilde{\mathbf{r}} \tag{A.8}$$

The second term on the right-hand side is the relative velocity of p.

The acceleration of p with respect to the inertial frame is:

$$\mathbf{a_I} = \frac{\mathrm{d}\mathbf{V}_I}{\mathrm{d}t} = \frac{\mathrm{d}^2\mathbf{R}}{\mathrm{d}t^2} + \frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \tilde{\mathbf{r}} + \boldsymbol{\omega} \times \frac{\mathrm{d}\tilde{\mathbf{r}}}{\mathrm{d}t} + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta\tilde{\mathbf{r}}}{\delta t}\right) \tag{A.9}$$

Applying Eq. (A.5) again gives us

$$\mathbf{a_I} = \frac{\mathrm{d}^2\mathbf{R}}{\mathrm{d}t^2} + \frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \tilde{\mathbf{r}} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) + 2\boldsymbol{\omega} \times \frac{\delta\tilde{\mathbf{r}}}{\delta t} + \frac{\delta^2\tilde{\mathbf{r}}}{\delta t^2} \tag{A.10}$$

The different terms can be defined as:

$\frac{\mathrm{d}^2\mathbf{R}}{\mathrm{d}t^2}$ = absolute acceleration of the moving frame with respect to to the inertial frame

$\frac{\delta\boldsymbol{\omega}}{\delta t} \times \mathbf{r}$ = apparent (tangential) acceleration of p due to angular acceleration of moving frame

$\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})$ = apparent (centripetal) acceleration of p due to angular motion of moving frame. Together with the first two terms, the resulting acceleration is called the dragging acceleration.

$2\boldsymbol{\omega} \times \frac{\delta\mathbf{r}}{\delta t}$ = Coriolis acceleration due to the motion of p in the moving frame

$\frac{\delta^2\mathbf{r}}{\delta t^2}$ = relative acceleration of p in the moving frame

Concluding, the absolute acceleration with respect to the inertial frame can be written as the summation of three terms, the dragging, Coriolis and relative acceleration.

A.4 Translational Motion

A.4.1 General Formulation for an Inertial Frame

We consider a mass system with N particles. Each particle has a constant mass m_i and a position vector $\mathbf{r_i}$ with respect to the origin of the inertial frame. The number of particles is varying with time, so $N = N(t)$. The linear momentum of a particle with respect to the inertial frame is given by

$$\mathbf{J_i} = m_i \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} \tag{A.11}$$

According to Newton's second law the summation of the external and internal forces acting on the particle equals the change in linear momentum:

$$\mathbf{F}_i + \sum_{k=1}^{N} \mathbf{F}_{ik} = \frac{\mathrm{d}\mathbf{J}_i}{\mathrm{d}t} \tag{A.12}$$

Summing over all particles yields

$$\sum_{i=1}^{N} \mathbf{F_i} + \sum_{\substack{i=1 \\ i \neq k}}^{N} \sum_{k=1}^{N} \mathbf{F_{ik}} = \sum_{i=1}^{N} \frac{\mathrm{d}\mathbf{J_i}}{\mathrm{d}t} \tag{A.13}$$

The second term on the left-hand side is zero, since Newton's third law says $\mathbf{F_{ik}} = -\mathbf{F_{ki}}$ and a particle cannot exert a force on itself.

Defining $\mathbf{F_I} = \sum_{i=1}^{N} \mathbf{F_i}$ as the total external force acting on the system, and applying Eq. (A.11) we obtain

$$\mathbf{F_I} = \sum_{i=1}^{N} m_i \frac{\mathrm{d}^2\mathbf{r_i}}{\mathrm{d}t^2} \tag{A.14}$$

When we consider a body, then the summation is replaced by an integration which gives us

$$\mathbf{F_I} = \int_m \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} \mathrm{d}m \tag{A.15}$$

with $\mathbf{r}$ as the location of a mass element with respect to the inertial frame. In Eq. (A.15), $\int_m$ stands for integration of the full body (volume), i.e., it can be seen as a volume integral with three variables of integration. Defining the location of the c.o.m. of the body to be

$$\mathbf{r_{cm}} \int_m \mathrm{d}m = \int_m \mathbf{r}\mathrm{d}m \tag{A.16}$$

and $\tilde{\mathbf{r}}$ as the position vector of the mass element with respect to the c.o.m., then

$$\mathbf{r} = \mathbf{r_{cm}} + \tilde{\mathbf{r}} \tag{A.17}$$

Using the expression for the acceleration, which we found in the previous section (Eq. A.10), realising that $\frac{\mathrm{d}^2\mathbf{R}}{\mathrm{d}t^2} = \frac{\mathrm{d}^2\mathbf{r_{cm}}}{\mathrm{d}t^2}$ and $\boldsymbol{\omega}$ is the rotation of the body with respect to inertial space, and taking into account that $\int_m \tilde{\mathbf{r}}\mathrm{d}m = 0$, which follows from substitution of Eq. (A.17) into Eq. (A.16), we can write for Eq. (A.15):

$$\mathbf{F_I} = m\frac{\mathrm{d}^2\mathbf{r_{cm}}}{\mathrm{d}t^2} + 2\boldsymbol{\omega} \times \int_m \frac{\delta\tilde{\mathbf{r}}}{\delta t}\mathrm{d}m + \int_m \frac{\delta^2\tilde{\mathbf{r}}}{\delta t^2}\mathrm{d}m \tag{A.18}$$

with

$\mathbf{F_I}$ = total of the external forces expressed in components of the inertial frame (N)

$\frac{\mathrm{d}^2\mathbf{r_{cm}}}{\mathrm{d}t^2}$ = acceleration of the c.o.m. with respect to the inertial frame (m/s^2)

$\mathbf{F_C} = -2\boldsymbol{\omega} \times \int_m \frac{\delta\tilde{\mathbf{r}}}{\delta t}\mathrm{d}m$ = Coriolis force, due to time variations in mass distribution (N)

$\mathbf{F_{rel}} = -\int_m \frac{\delta^2\tilde{\mathbf{r}}}{\delta t^2}\mathrm{d}m$ = relative force, due to time variations in mass distribution (N)

$\boldsymbol{\omega} = (p, q, r)^{\mathrm{T}}$ = the rotation vector of the body frame with respect to the inertial frame, expressed in components along the body axes

$\tilde{\mathbf{r}}$ = the location of a mass element with respect to the c.o.m. of the vehicle (m)

Equation B.18 describes the translational motion of an arbitrary body with variable mass distribution.

If we bring the two apparent forces due to the variation of the mass distribution to the left-hand side of Eq. (B.18), and designate the resulting force as $\tilde{\mathbf{F}}_I$, then the equation has the same form as the one for a rigid body (with a constant mass distribution), where both the Coriolis and the relative force are zero. This can be summarised by the *Principle of Solidification*, which states (Cornelisse et al. 1979) that *in general, equations of translational and rotational motion of an arbitrary variable mass system at time* t *can be written as the translational and rotational equations for a rigid body with mass* M *equal to the mass of the system at time* t*, while in addition to the true external forces and moments, two apparent forces and moments are applied: the Coriolis and relative forces and moments, respectively.*

A.4.2 General Formulation for a Rotating Frame

Let us now consider a vehicle moving in the sphere of influence of a celestial body. In Fig. A.3 we have established an inertial and a rotating planetocentric reference frame. The moving frame is rigidly attached to the central body, with its origin in its c.o.m., and hence rotates with the same rate as the central body, $\boldsymbol{\omega}_{\mathbf{R}}$. We assume that this rotational rate is constant. The vehicle has a variable mass m and its c.o.m. is at a distance $\mathbf{r}_{\mathbf{cm}}$ from the c.o.m.mon origin of the two reference frames. It is moving with a velocity $\mathbf{V}_{\mathbf{R}}$ with respect to the moving frame.

In accordance with the vehicle location definition in the previous sub-section, we may write

$$\mathbf{r}_{\mathbf{cm}_I} = \mathbf{R} + \mathbf{r}_{\mathbf{cm}_R} = \mathbf{r}_{\mathbf{cm}_R}$$

since the two frames are coinciding ($\mathbf{R} = \mathbf{0}$). Note that we have added the indices I and R to distinguish between the two position vectors.

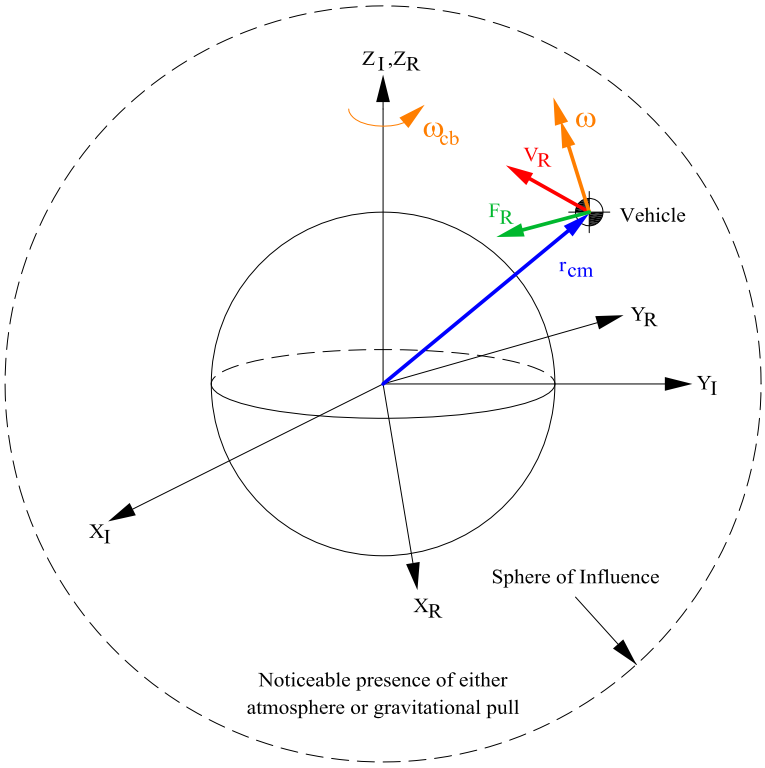


Fig. A.3 The vehicle with respect to the two related reference frames. The index I indicates the inertial planetocentric reference frame, whereas R stands for the rotating planetocentric frame

Starting with Eq. (B.18), which describes the motion of the vehicle with respect to the inertial frame, we have

$$\mathbf{F_I} + \mathbf{F_C} + \mathbf{F_{rel}} = \tilde{\mathbf{F}}_I = m\frac{\mathrm{d}^2\mathbf{r_{cm_I}}}{\mathrm{d}t^2} \tag{A.19}$$

Expressing the acceleration of the c.o.m. with respect to the inertial frame in the rotating frame, we apply again Eq. (A.5) which gives us, equivalent to Eq. (A.10):

$$\begin{aligned}\mathbf{a_R} = \frac{\mathrm{d}^2\mathbf{r_{cm_R}}}{\mathrm{d}t^2} = \frac{\mathrm{d}^2\mathbf{R}}{\mathrm{d}t^2} + \frac{\mathrm{d}\boldsymbol{\omega}_\mathbf{R}}{\mathrm{d}t} \times \mathbf{r_{cm_R}} + \boldsymbol{\omega}_\mathbf{R} \times (\boldsymbol{\omega}_\mathbf{R} \times \mathbf{r_{cm_R}}) + \\ + 2\boldsymbol{\omega}_\mathbf{R} \times \frac{\delta\mathbf{r_{cm_R}}}{\delta t} + \frac{\delta^2\mathbf{r_{cm_R}}}{\delta t^2}\end{aligned} \tag{A.20}$$

The first term on the right-hand side disappears, because $\mathbf{R} = \mathbf{0}$. Furthermore, we assumed that the central body is rotating with a constant angular rate, so also the second term is zero. Substitution of Eq. (A.20) into (A.19), and omitting the indices R for the sake of convenience, we get the equation of translational motion of a mass-varying vehicle with respect to the rotating planetocentric frame:

$$\mathbf{F_R} = m\frac{\mathrm{d}^2\mathbf{r_{cm}}}{\mathrm{d}t^2} + 2m\boldsymbol{\omega}_\mathbf{R} \times \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} + m\boldsymbol{\omega}_\mathbf{R} \times (\boldsymbol{\omega}_\mathbf{R} \times \mathbf{r_{cm}}) \tag{A.21}$$

In the above equation, $\mathbf{F_R}$ is the sum of the external forces *and the Coriolis and relative forces due to the variation of mass*, expressed in components of the R-frame. The other terms have the following meaning:

$\frac{\mathrm{d}^2\mathbf{r_{cm}}}{\mathrm{d}t^2}$ = apparent acceleration of the vehicle in the rotating frame

$2\boldsymbol{\omega}_\mathbf{R} \times \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t}$ = Coriolis acceleration due to the motion of the vehicle in the rotating frame

$\boldsymbol{\omega}_R \times (\boldsymbol{\omega}_R \times \mathbf{r}_{cm}) =$ apparent (or transport) acceleration of the vehicle due to angular rate of the rotating frame

Note that the derivatives $\frac{\delta}{\delta t}$ and $\frac{\delta^2}{\delta t^2}$ with respect to the rotating frame have been written with a conventional d, since there is no confusion about the frame.

A.5 Rotational Motion

A.5.1 General Formulation

We start the derivation, by looking at a single particle again. We multiply Eq. (A.12) vectorial with $\mathbf{r_i}$, yielding

$$\mathbf{r_i} \times \mathbf{F_i} + \sum_{k=1}^{N} \mathbf{r_i} \times \mathbf{F_{ik}} = \mathbf{r_i} \times \frac{\mathrm{d}\mathbf{J_i}}{\mathrm{d}t} \tag{A.22}$$

The moment of $\mathbf{F_i}$ with respect to the origin of the inertial frame is defined as

$$\mathbf{M_i} = \mathbf{r_i} \times \mathbf{F_i} \tag{A.23}$$

and the angular momentum of the particle is

$$\mathbf{B_i} = \mathbf{r_i} \times \mathbf{J_i} = m_i \mathbf{r_i} \times \frac{\mathrm{d}\mathbf{r_i}}{\mathrm{d}t} \tag{A.24}$$

In the next sub-section we will discuss the angular momentum in more detail.

Substituting Eqs. (A.23)–(A.24) into Eq. (A.22) gives us

$$\mathbf{M_i} + \sum_{k=1}^{N} \mathbf{r_i} \times \mathbf{F_{ik}} = \mathbf{r_i} \times \frac{\mathrm{d}\mathbf{J_i}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r_i} \times \mathbf{J_i}) = \frac{\mathrm{d}\mathbf{B_i}}{\mathrm{d}t} \tag{A.25}$$

Summing over all particles yields

$$\sum_{i=1}^{N} \mathbf{M_i} + \sum_{i=1}^{N} \sum_{k=1}^{N} \mathbf{r_i} \times \mathbf{F_{ik}} = \sum_{i=1}^{N} \frac{\mathrm{d}\mathbf{B_i}}{\mathrm{d}t} \tag{A.26}$$

The first term on the left-hand side is the total moment about the origin of the I-frame, $\mathbf{M_I}$. The second term is the total internal moment, which is equal to zero (which can be seen by taking the terms in pairs together). Equation (A.26) can then be rewritten as

$$\mathbf{M_I} = \sum_{i=1}^{N} \frac{\mathrm{d}\mathbf{B_i}}{\mathrm{d}t} = \sum_{i=1}^{N} m_i \mathbf{r_i} \times \frac{\mathrm{d}^2\mathbf{r_i}}{\mathrm{d}t^2} \tag{A.27}$$

For a solid body, summation is replaced by integration:

$$\mathbf{M_I} = \frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t} = \int_m \mathbf{r} \times \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} \mathrm{d}m \tag{A.28}$$

Suppose now we have a vehicle, which is rotating with an angular velocity $\boldsymbol{\omega}$ with respect to the inertial frame, and we want to express the rotational motion of the vehicle using the external moment about the vehicle's c.o.m.. $\mathbf{M_I}$ is the moment about the origin of the I-frame due to the external force $\mathbf{F_I}$. If the point, on which $\mathbf{F_I}$ is acting is defined as $\mathbf{r_I}$, and $\tilde{\mathbf{r}}_\mathbf{I}$ is the corresponding position with respect to the c.o.m. of the vehicle, then

$$\mathbf{M_I} = \mathbf{r_I} \times \mathbf{F_I} = (\mathbf{r_{cm}} + \tilde{\mathbf{r}}) \times \mathbf{F_I} \tag{A.29}$$

so

$$\mathbf{M_{cm}} = \tilde{\mathbf{r}} \times \mathbf{F_I} = \mathbf{M_I} - \mathbf{r_{cm}} \times \mathbf{F_I} \tag{A.30}$$

With $\mathbf{M_I}$ and $\mathbf{F_I}$ given by Eqs. (A.28) and (A.15), respectively, then we obtain with Eq. (A.17)

$$\mathbf{M_{cm}} = \int_m \tilde{\mathbf{r}} \times \frac{\mathrm{d}^2 \mathbf{r_I}}{\mathrm{d}t^2} \mathrm{d}m \tag{A.31}$$

Thus, with Eq. (A.10) we finally get for the equation of rotational motion for an arbitrary flexible body with variable mass:

$$\begin{aligned} \mathbf{M_{cm}} = & \int_m \tilde{\mathbf{r}} \times \left(\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \tilde{\mathbf{r}} \right) \mathrm{d}m + \int_m \tilde{\mathbf{r}} \times \left[\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \right] \mathrm{d}m + \\ & + 2 \int_m \tilde{\mathbf{r}} \times \left(\boldsymbol{\omega} \times \frac{\delta \tilde{\mathbf{r}}}{\delta t} \right) \mathrm{d}m + \int_m \tilde{\mathbf{r}} \times \frac{\delta^2 \tilde{\mathbf{r}}}{\delta t^2} \mathrm{d}m \end{aligned} \tag{A.32}$$

Here, the several terms on the right-hand side have the following meaning:

$2 \int_m \tilde{\mathbf{r}}_\mathbf{I} \times \left(\boldsymbol{\omega} \times \frac{\delta \tilde{\mathbf{r}}_\mathbf{I}}{\delta t} \right) \mathrm{d}m = -\mathbf{M_C} =$ Coriolis moment due to time variations in mass distribution (Nm)

$\int_m \tilde{\mathbf{r}}_\mathbf{I} \times \frac{\delta^2 \tilde{\mathbf{r}}_\mathbf{I}}{\delta t^2} \mathrm{d}m = -\mathbf{M_{rel}} =$ relative moment due to time variations in mass distribution (Nm)

$\int_m \tilde{\mathbf{r}}_\mathbf{I} \times \left(\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \tilde{\mathbf{r}}_\mathbf{I} \right) \mathrm{d}m =$ apparent moment due to the angular acceleration of the vehicle with respect to the inertial frame (Nm)

$\int_m \tilde{\mathbf{r}}_\mathbf{I} \times \left[\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}_\mathbf{I}) \right] \mathrm{d}m =$ apparent moment due to the angular velocity of the vehicle with respect to the inertial frame (Nm)

A.5.2 The Angular Momentum with Respect to the Centre of Mass of a Rigid Body

The angular momentum of a particle was previously defined as

$$\mathbf{B_i} = \mathbf{r_i} \times \mathbf{J_i} = m_i \mathbf{r_i} \times \frac{\mathrm{d}\mathbf{r_i}}{\mathrm{d}t} \tag{A.33}$$

So the angular momentum $\mathbf{B_I}$ of a solid body with respect to to the origin of an inertial frame can be expressed as

$$\mathbf{B_I} = \int_m \mathbf{r} \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \mathrm{d}m \tag{A.34}$$

with $\mathbf{r}$ given by Eq. (A.17). In Fig. A.4, the situation is schematically shown. We can distinguish two reference frames, an inertial one with index I and one, which is fixed to the body (index b), with the origin in the c.o.m., and therefore shares the rotational rates of the body (denoted by $\boldsymbol{\omega}$). $\mathbf{r}$ is the position of the mass element dm with respect to the inertial frame, $\tilde{\mathbf{r}}$ describes the same position but now measured from the c.o.m.. The location of the c.o.m. in the I-frame is indicated by $\mathbf{r_{cm}}$.

To compute the angular momentum with respect to the c.o.m. of a *rigid* body, we proceed as follows.

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r_{cm}} + \tilde{\mathbf{r}}) = \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} + \frac{\delta\tilde{\mathbf{r}}}{\delta t} + \boldsymbol{\omega} \times \tilde{\mathbf{r}} = \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} + \boldsymbol{\omega} \times \tilde{\mathbf{r}} \tag{A.35}$$

Here, we have used Eqs. (A.17) and (A.5).

Substituting Eqs. (A.17) and (A.35) in Eq. (A.34) yields

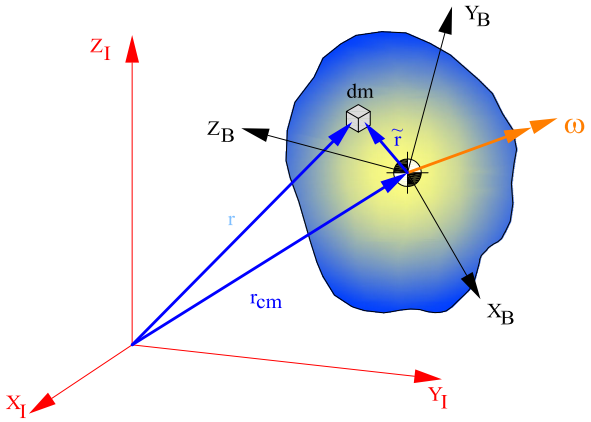


Fig. A.4 The solid body with the two reference frames, the rotating body frame (index B) with the origin in the centre of mass and the inertial frame (index I)

$$\mathbf{B_I} = \int_m (\mathbf{r_{cm}} + \tilde{\mathbf{r}}) \times \left(\frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} + \boldsymbol{\omega} \times \tilde{\mathbf{r}} \right) \mathrm{d}m \tag{A.36}$$

Writing out the terms in Eq. (A.36) results in:

$$\begin{aligned} \mathbf{B_I} = & \int_m \mathbf{r_{cm}} \times \frac{\mathbf{r_{cm}}}{\mathrm{d}t} \mathrm{d}m + \int_m \mathbf{r_{cm}} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \mathrm{d}m + \\ & + \int_m \tilde{\mathbf{r}} \times \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} \mathrm{d}m + \int_m \tilde{\mathbf{r}} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \mathrm{d}m \end{aligned} \tag{A.37}$$

The first term on the right-hand side can be written as

$$\int_m \mathbf{r_{cm}} \times \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} \mathrm{d}m = \mathbf{r_{cm}} \times \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} \int_m \mathrm{d}m = m\mathbf{r_{cm}} \times \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} \tag{A.38}$$

with m as the total mass of the solid body. We were allowed to place the cross product outside the integral because $\mathbf{r_{cm}}$ is independent of the variables of integration.

For the second term, we obtain, since $\boldsymbol{\omega}$ is also independent of the variables of integration (each mass element in a rigid body has the same rotation vector):

$$\int_m \mathbf{r_{cm}} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \mathrm{d}m = \mathbf{r_{cm}} \times \left(\boldsymbol{\omega} \times \int_m \tilde{\mathbf{r}} \mathrm{d}m \right) = 0 \tag{A.39}$$

since the definition of the c.o.m. prescribes that $\int_m \tilde{\mathbf{r}} \mathrm{d}m$ equals zero.

The third term leads to:

$$\int_m \tilde{\mathbf{r}} \times \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} \mathrm{d}m = -\frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} \times \int_m \tilde{\mathbf{r}} \mathrm{d}m = 0 \tag{A.40}$$

Equations (A.38)–(A.40) substituted in Eq. (A.37) gives

$$\mathbf{B_I} = m\mathbf{r_{cm}} \times \frac{\mathrm{d}\mathbf{r_{cm}}}{\mathrm{d}t} + \int_m \tilde{\mathbf{r}} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \mathrm{d}m \tag{A.41}$$

We see that the angular momentum with respect to the origin of the inertial frame consists of two terms. The first one describes the angular momentum of the body if its mass is thought to be concentrated in the c.o.m. (point mass), expressed in components of the inertial frame. The second one is a correction term, which takes into account the finite dimensions of the body (instead of being a point mass). This correction term represents a contribution of all mass elements with respect to the c.o.m., i.e., the angular momentum of the body with respect to the c.o.m., and is

denoted by $\mathbf{B}_{\mathbf{cm}}$.

$$\mathbf{B}_{\mathbf{cm}} = \int_m \tilde{\mathbf{r}} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \mathrm{d}m \tag{A.42}$$

In the remainder of this sub-section we will derive expressions for each of the components of $\mathbf{B}_{\mathbf{cm}}$. The following vector identity holds:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}\,(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}\,(\mathbf{b} \cdot \mathbf{a}) \tag{A.43}$$

If we write

$$\tilde{\mathbf{r}} = (x, y, z)^{\mathrm{T}}$$

$$\boldsymbol{\omega} = (p, q, r)^{\mathrm{T}}$$

then

$$\tilde{\mathbf{r}} \cdot \tilde{\mathbf{r}} = x^2 + y^2 + z^2$$

$$\boldsymbol{\omega} \cdot \tilde{\mathbf{r}} = px + qy + rz$$

Applying Eqs. (A.43)–(A.42) results in

$$\int_m \tilde{\mathbf{r}} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \mathrm{d}m = \int_m \left[\boldsymbol{\omega}(\tilde{\mathbf{r}} \cdot \tilde{\mathbf{r}}) - \tilde{\mathbf{r}}(\boldsymbol{\omega} \cdot \tilde{\mathbf{r}})\right] \mathrm{d}m$$

and substituting the above results yields

$$\int_m \Big[p(x^2 + y^2 + z^2)\mathbf{x_b} + q(x^2 + y^2 + z^2)\mathbf{y_b} + r(x^2 + y^2 + z^2)\mathbf{z_b} + \\ -x(px + qy + rz)\mathbf{x_b} - y(px + qy + rz)\mathbf{y_b} - z(px + qy + rz)\mathbf{z_b}\Big] \mathrm{d}m$$

with $\mathbf{x_b}$, $\mathbf{y_b}$ and $\mathbf{z_b}$ being the unit vectors of the b-frame.

Writing out the three scalar equations gives the following expressions for the components of the angular momentum:

$$\begin{aligned} B_x &= \int_m \left[p(y^2 + z^2) - qxy - rxz\right] \mathrm{d}m \\ B_y &= \int_m \left[-pyx + q(x^2 + z^2) - ryz\right] \mathrm{d}m \\ B_z &= \int_m \left[-pzx - qzy + r(x^2 + y^2)\right] \mathrm{d}m \end{aligned} \tag{A.44}$$

If we define

$$I_{xx} = \int_m (y^2 + z^2)\mathrm{d}m$$
$$I_{yy} = \int_m (x^2 + z^2)\mathrm{d}m$$
$$I_{zz} = \int_m (x^2 + y^2)\mathrm{d}m$$

as the so-called moments of inertia, and

$$I_{xy} = I_{yx} = \int_m xy\mathrm{d}m$$
$$I_{xz} = I_{zx} = \int_m xz\mathrm{d}m$$
$$I_{yz} = I_{zy} = \int_m yz\mathrm{d}m$$

as the products of inertia, Eqs. (A.44) can also be written as

$$\begin{aligned} B_x &= pI_{xx} - qI_{xy} - rI_{xz} \\ B_y &= -pI_{xy} + qI_{yy} - rI_{yz} \\ B_z &= -pI_{xz} - qI_{yz} + rI_{zz} \end{aligned} \tag{A.45}$$

or, as a matrix equation:

$$\mathbf{B_{cm}} = \mathbf{I} \cdot \boldsymbol{\omega} \tag{A.46}$$

with

$$\mathbf{I} = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{xy} & I_{yy} & -I_{yz} \\ -I_{xz} & -I_{yz} & I_{zz} \end{bmatrix}$$

called the inertia tensor.

A.5.3 Derivation of the Euler Equation of Rotational Motion

Let us commence with the general formulation of the equation of rotational motion for a mass-varying body, Eq. (A.32)

$$\mathbf{M_{cm}} = \int_m \tilde{\mathbf{r}} \times \left(\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \tilde{\mathbf{r}} \right) \mathrm{d}m + \int_m \tilde{\mathbf{r}} \times \left[\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \right] \mathrm{d}m + \\ + 2 \int_m \tilde{\mathbf{r}} \times \left(\boldsymbol{\omega} \times \frac{\delta \tilde{\mathbf{r}}}{\delta t} \right) \mathrm{d}m + \int_m \tilde{\mathbf{r}} \times \frac{\delta^2 \tilde{\mathbf{r}}}{\delta t^2} \mathrm{d}m$$

The last two terms on the right-hand side are the apparent moments due to elasticity and variation of mass. Bringing these terms to the other side, we get the same form of the equation as for a rigid body:

$$\mathbf{M_{cm}} + \mathbf{M_C} + \mathbf{M_{rel}} = \tilde{\mathbf{M}_{cm}} = \int_m \tilde{\mathbf{r}} \times \left(\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \tilde{\mathbf{r}} \right) \mathrm{d}m + \\ + \int_m \tilde{\mathbf{r}} \times \left[\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \tilde{\mathbf{r}}) \right] \mathrm{d}m \tag{A.47}$$

For a rigid body, we can derive the equation of rotational motion also in a different way. In the previous sub-section we found

$$\mathbf{B_{cm}} = \mathbf{I} \cdot \boldsymbol{\omega} \tag{A.48}$$

With Eqs. (A.5) and (A.28) we then find

$$\mathbf{M_{cm}} = \frac{\mathrm{d}\mathbf{B_{cm}}}{\mathrm{d}t} = \frac{\delta \mathbf{B_{cm}}}{\delta t} + \boldsymbol{\omega} \times \mathbf{B_{cm}} \tag{A.49}$$

The first term on the right-hand side of Eq. (A.49) can be written as

$$\frac{\delta \mathbf{B_{cm}}}{\delta t} = \frac{\delta \mathbf{I}}{\delta t} \cdot \boldsymbol{\omega} + \mathbf{I} \cdot \frac{\delta \boldsymbol{\omega}}{\delta t} = \mathbf{I} \cdot \frac{\delta \boldsymbol{\omega}}{\delta t} \tag{A.50}$$

since mass is invariant with time for a rigid body, and therefore also the inertia properties. Substituting Eq. (A.50) into Eq. (A.49), and applying Eq. (A.48) one more time we finally get

$$\mathbf{M_{cm}} = \mathbf{I} \cdot \dot{\boldsymbol{\omega}} + \boldsymbol{\omega} \times \mathbf{I} \cdot \boldsymbol{\omega} \tag{A.51}$$

According to the Principle of Solidification we can use the above equation for a mass-varying (non-rigid) body too, thereby taking the two apparent moments into

account and at time t substituting the current mass properties, i.e., mass, and moments and products of inertia in the equations.

Reference

Cornelisse JW, Schöyer HFR, Wakker KF (1979) Rocket propulsion and spaceflight dynamics. Pitman, London

Appendix B
Reference Vehicles

B.1 Entry Capsule: Apollo

B.1.1 General

The Apollo program was the spaceflight effort carried out by the United States' National Aeronautics and Space Administration (NASA) to land the first humans on Earth's Moon. Initiated in the early sixties of the previous century, Apollo followed the Mercury and Gemini programs, and ran from 1961 to 1972. Apollo was a semi-ballistic capsule that flew with a low lift-to-drag ratio ($L/D \approx 0.3$), in a trajectory characterised by a high heat and mechanical load. The high heat load required a thermal protection system (TPS) with an outer layer consisting of ablative materials that burn away and in doing so rejected the heat from the vehicle wall.

Apollo entered the atmosphere with the heat shield facing the flow. Let us define a body frame with the origin in the theoretical apex of the vehicle. This point is obtained when the smallest portion of the vehicle would be extended to a full cone (Fig. B.1). The x_B-axis is positive in the direction of the heat shield and intersects it in the centre of the circular cross section. The z_B-axis is positive pointing downwards for a nominal entry and the y_B-axis completes the right-handed system. In this frame of reference, the location of the centre of mass is given by $\mathbf{r_{cm}} = (2.57, 0.0, -0.137)^{\mathrm{T}}$. Due to the (significant) shift of the centre of mass in z-direction, the capsule is not rotationally symmetric in mass, which means that one of the products of inertia is non-zero. The corresponding numerical values of moments and products of inertia are $I_{xx} = 5617.6\ \mathrm{kg/m^2}$, $I_{yy} = 4454.6\ \mathrm{kg/m^2}$, $I_{zz} = 4454.8\ \mathrm{kg/m^2}$, $I_{xy} = I_{yz} = 0.0\ \mathrm{kg/m^2}$ and $I_{xz} = 1752.0\ \mathrm{kg/m^2}$.

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1

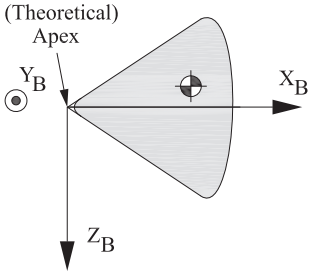


Fig. B.1 Definition of the body-fixed reference frame

B.1.2 Aerodynamic Database

The aerodynamic database presented here is based on extensive wind-tunnel measurements done in the mid sixties as part of the Apollo-capsule design (NAA 1965; Mosely et al. 1967). Typically, for axisymmetric vehicles like capsules these wind-tunnel measurements are done for global angle sweeps, defined by the so-called total angle of attack, rather than separate angle of attack and sideslip. As this definition is very common for entry capsules, we will include a discussion on the relation beween these "axisymmetric coefficients" and the more general full definition. The actual aerodynamic database is provided in Tables B.1, B.2, B.3 and B.4.

The coefficients are defined with respect to the theoretical apex, as indicated in Fig. B.1. This means that when the moments around the centre of mass are required, it is necessary to account for the distance between the two points and include the induced moments to to the aerodynamic-force components.

Table B.1 Apollo total drag coefficient, C_{D_T}, as a function of Mach number and total angle of attack (NAA 1965)

α_T (°) → Mach (–) ↓	0	5	10	15	20	25	30	35	40
0.90	1.130	1.124	1.107	1.080	1.039	0.981	0.919	0.856	0.777
1.10	1.305	1.296	1.284	1.261	1.225	1.177	1.109	1.048	0.981
1.20	1.340	1.335	1.321	1.298	1.267	1.220	1.162	1.092	1.015
1.35	1.394	1.391	1.385	1.369	1.340	1.295	1.235	1.164	1.075
1.65	1.445	1.432	1.420	1.385	1.342	1.297	1.240	1.150	1.035
2.00	1.472	1.465	1.447	1.413	1.369	1.310	1.218	1.110	0.983
2.40	1.480	1.479	1.458	1.416	1.347	1.264	1.165	1.065	0.950
3.00	1.468	1.454	1.422	1.378	1.304	1.219	1.128	1.021	0.911
4.00	1.454	1.440	1.408	1.349	1.270	1.184	1.094	0.973	0.861
6.00	1.486	1.470	1.420	1.350	1.270	1.175	1.064	0.956	0.843
10.00	1.510	1.492	1.450	1.390	1.300	1.195	1.085	0.965	0.844

Table B.2 Apollo total lift coefficient, C_{L_T}, as a function of Mach number and total angle of attack (NAA 1965)

α_T (°) → Mach (–) ↓	0	5	10	15	20	25	30	35	40
0.90	0.0	0.098	0.178	0.250	0.321	0.383	0.440	0.488	0.528
1.10	0.0	0.099	0.200	0.295	0.381	0.471	0.553	0.615	0.630
1.20	0.0	0.127	0.225	0.319	0.403	0.486	0.558	0.611	0.615
1.35	0.0	0.119	0.228	0.325	0.424	0.517	0.597	0.610	0.600
1.65	0.0	0.115	0.235	0.345	0.437	0.518	0.570	0.583	0.568
2.00	0.0	0.130	0.241	0.342	0.438	0.508	0.540	0.557	0.548
2.40	0.0	0.118	0.220	0.336	0.425	0.485	0.525	0.540	0.533
3.00	0.0	0.123	0.221	0.333	0.415	0.482	0.530	0.540	0.530
4.00	0.0	0.110	0.218	0.320	0.405	0.480	0.527	0.540	0.531
6.00	0.0	0.118	0.220	0.312	0.400	0.480	0.525	0.540	0.530
10.00	0.0	0.122	0.228	0.318	0.405	0.480	0.526	0.540	0.529

Table B.3 Apollo total pitch-moment coefficient, C_{m_T}, as a function of Mach number and total angle of attack (NAA 1965)

α_T (°) → Mach (–) ↓	0	5	10	15	20	25	30	35	40
0.90	0.0002	0.0108	0.0106	0.0063	0.0063	0.0068	0.0102	0.0124	0.0221
1.10	0.0000	0.0004	0.0028	0.0025	−0.0011	0.0019	0.0082	0.0068	−0.0166
1.20	0.0001	0.0166	0.0137	0.0098	0.0019	−0.0021	−0.0086	−0.0168	−0.0440
1.35	−0.0003	0.0075	0.0089	0.0023	−0.0012	−0.0038	−0.0076	−0.0452	−0.0806
1.65	0.0003	0.0022	0.0085	0.0109	0.0051	−0.0046	−0.0264	−0.0574	−0.0868
2.00	−0.0001	0.0105	0.0105	0.0050	0.0009	−0.0136	−0.0382	−0.0591	−0.0766
2.40	−0.0002	0.0018	−0.0056	0.0021	0.0018	−0.0151	−0.0306	−0.0522	−0.0700
3.00	0.0000	0.0057	−0.0012	0.0061	0.0023	−0.0045	−0.0142	−0.0335	−0.0532
4.00	−0.0002	−0.0012	0.0006	0.0028	0.0037	0.0065	−0.0011	−0.0135	−0.0294
6.00	0.0002	0.0033	0.0008	−0.0017	0.0014	0.0086	0.0069	−0.0056	−0.0219
10.00	0.0001	0.0040	0.0024	−0.0030	−0.0006	0.0039	0.0014	−0.0087	−0.0228

Intermezzo B.1 (The re-entry vehicle as an axisymmetric body)

If we consider the re-entry vehicle to be an axisymmetric body, we can simplify the definition of the aerodynamic coefficients, since only three independent coefficients exist (related with symmetrical motion). They only depend on the angle between the body axis of revolution, X_B, and the relative flow velocity, which is collinear with the X-axis of the aerodynamic frame. This angle is called the total angle of attack, α_T.

The total angle of attack is defined as the arc summation of α and β (see also Fig. B.2). From this figure, we can derive, using spherical trigonometry:

Table B.4 Apollo total pitch-damping moment coefficient in s/radian, $C_{m,qT}$, as a function of Mach number and total angle of attack (Mosely et al. 1967)

α_T (°) → Mach (–) ↓	0	10	20	30	40
2.5	−0.6	−0.6	−0.6	−0.6	−0.4
3.0	−0.6	−0.6	−0.6	−0.6	−0.4
4.0	−0.6	−0.6	−0.6	−0.6	−0.4
6.0	−0.6	−0.6	−0.8	−1.6	−0.4
10.2	−0.4	−0.4	−0.4	−0.4	−0.4

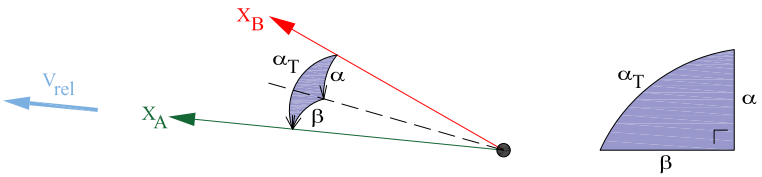


Fig. B.2 Definition of the total angle of attack, α_T

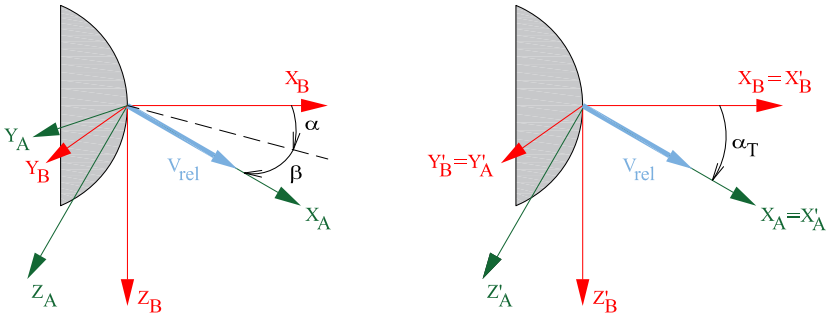


Fig. B.3 Body frame and aerodynamic frame based on α and β (left), and those based on α_T (right)

$$\cos\alpha_T = \cos\alpha\cos\beta \quad \Rightarrow \quad \alpha_T = \arccos(\cos\alpha\cos\beta) \tag{B.1}$$

with α_T in radians.

The three independent coefficients to be considered, are:

C_{D_T}: total drag coefficient, negative along positive $\mathbf{x}'_\mathbf{A}$,
C_{L_T}: total lift coefficient, negative along positive $\mathbf{Z}'_\mathbf{A}$, and
C_{m_T}: total pitch-moment coefficient, creating a positive moment about $\mathbf{y}'_\mathbf{B}$.

Here, we used two primed frames (body primed and aerodynamic primed), which we need to derive the non-axisymmetrical coefficients. The relation between the primed and unprimed frames is shown in Fig. B.3. The orientation of the aerodynamic

frame with respect to the body frame can be defined with the angle of attack α and the angle of sideslip β (left figure) or with the total angle of attack α_T (right figure). With the former definition we need at least five aerodynamic coefficients, whereas in the latter case three are sufficient.

The meaning of these primed frames becomes clear, if we look at the forces and moments in both primed and unprimed frames. In the (unprimed) aerodynamic frame, we have separate components S and L along the Y- and Z-axis. In the primed frame, however, these components are combined in the total lift force L_T, along the Z-axis. In both frames, however, the drag force is directed along the X-axis (collinear with the velocity vector). This means that the X-axes of both frames are the same.

For the moments, a similar reasoning holds. The total pitch moment about the Y-axis of the primed body frame, M_T is the combined effect of the pitch and yaw moment, M and N about Y_B and Z_B. Also in this case, the X-axis of both the primed and the unprimed frame must be the same, due to the definition of α_T. So

$$\mathbf{x_B} = \mathbf{x'_B} \qquad \text{and} \qquad \mathbf{x_A} = \mathbf{x'_A}$$

Furthermore, we consider the $\mathbf{x'_B}\,\mathbf{z'_B}$- and $\mathbf{x'_A}\,\mathbf{z'_A}$-planes to be in one plane, although rotated with respect to one another. As a result, $\mathbf{y'_B}$ and $\mathbf{y'_A}$ must be collinear.

The equations we need to solve are now:

$$C_{D_T}\mathbf{x'_A} + C_{L_T}\mathbf{z'_A} = C_D\mathbf{x_A} + C_S\mathbf{y_A} + C_L\mathbf{z_A} \tag{B.2}$$

$$C_{m_T}\mathbf{y'_B} = C_m\mathbf{y_B} + C_n\mathbf{z_B} \tag{B.3}$$

On the right-hand side of Eq. (B.3), the roll coefficient C_l is omitted, since it is equal to C_{L_T}, and this coefficient has not been defined for Apollo, which means no roll damping is implied.

From Fig. B.3, we can directly see that

$$\mathbf{x_B} = \cos\alpha_T\mathbf{x'_A} - \sin\alpha_T\mathbf{z'_A} \tag{B.4}$$

or

$$\mathbf{z'_A} = \frac{1}{\sin\alpha_T}\left(\cos\alpha_T\mathbf{x_A} - \mathbf{x_B}\right) \tag{B.5}$$

since $\mathbf{x'_A} = \mathbf{x_A}$.

In Sect. 4.3, several transformation matrices are introduced, and the transformation between the aerodynamic and the body frame is easily derived from the inverse of Eq. (4.59):

$$\mathbf{C_{B,A}} = \begin{bmatrix} \cos\alpha\cos\beta & -\cos\alpha\sin\beta & -\sin\alpha \\ \sin\beta & \cos\beta & 0 \\ \sin\alpha\cos\beta & -\sin\alpha\sin\beta & \cos\alpha \end{bmatrix} \tag{B.6}$$

which can be used to set up the expressions for each of the unit vectors of the body frame expressed in components of the aerodynamic frame. The expression for $\mathbf{x_B}$ follows from the first row:

$$\begin{aligned}\mathbf{x_B} &= \cos\alpha\cos\beta\mathbf{x_A} - \cos\alpha\sin\beta\mathbf{y_A} - \sin\alpha\mathbf{z_A} \\ &= \cos\alpha_T\mathbf{x_A} - \cos\alpha\sin\beta\mathbf{y_A} - \sin\alpha\mathbf{z_A}\end{aligned} \tag{B.7}$$

where the definition of α_T, Eq. (B.1), has been substituted. It thus follows that

$$\mathbf{z}'_\mathbf{A} = \frac{1}{\sin\alpha_T}\left(\cos\alpha\sin\beta\mathbf{y_A} + \sin\alpha\mathbf{z_A}\right) \tag{B.8}$$

$\mathbf{y}'_\mathbf{A}$ can be obtained from the cross product $\mathbf{y}'_\mathbf{A} = \mathbf{z}'_\mathbf{A} \times \mathbf{x}'_\mathbf{A}$, leading to

$$\mathbf{y}'_\mathbf{A} = \frac{\sin\alpha}{\sin\alpha_T}\mathbf{y_A} - \frac{\cos\alpha\sin\beta}{\sin\alpha_T}\mathbf{z_A} \tag{B.9}$$

Using the expressions for $\mathbf{x}'_\mathbf{A}$ and $\mathbf{z}'_\mathbf{A}$, the first two *total* aerodynamic coefficients can be expressed in components of the aerodynamic frame:

$$C_{D_T}\mathbf{x}'_\mathbf{A} + C_{L_T}\mathbf{z}'_\mathbf{A} = C_{D_T}\mathbf{x_A} + C_{L_T}\left(\frac{\cos\alpha\sin\beta}{\sin\alpha_T}\mathbf{y_A} + \frac{\sin\alpha}{\sin\alpha_T}\mathbf{z_A}\right) \tag{B.10}$$

Substituting this result in Eq. (B.2) and solving for C_D, C_S and C_L yields

$$C_D = C_{D_T} \tag{B.11}$$

$$C_S = \frac{\cos\alpha\sin\beta}{\sin\alpha_T}C_{L_T} \tag{B.12}$$

$$C_L = \frac{\sin\alpha}{\sin\alpha_T}C_{L_T} \tag{B.13}$$

To solve Eq. (B.3), we use the fact that

$$\mathbf{y}'_\mathbf{B} = \mathbf{y}'_\mathbf{A} = \frac{\sin\alpha}{\sin\alpha_T}\mathbf{y_A} - \frac{\cos\alpha\sin\beta}{\sin\alpha_T}\mathbf{z_A} \tag{B.14}$$

and, getting expressions for $\mathbf{y_A}$ and $\mathbf{z_A}$ from the inverse transformation matrix Eq. (B.6), i.e.,

$$\mathbf{y_A} = -\cos\alpha\sin\beta\mathbf{x_B} + \cos\beta\mathbf{y_B} - \sin\alpha\sin\beta\mathbf{z_B} \tag{B.15}$$

$$\mathbf{z_A} = -\sin\alpha\mathbf{x_B} + \cos\alpha\mathbf{z_B} \tag{B.16}$$

leading to

$$\mathbf{y}'_{\mathbf{A}} = \frac{\sin\alpha\cos\beta}{\sin\alpha_T}\mathbf{y}_{\mathbf{B}} - \frac{\sin\beta}{\sin\alpha_T}\mathbf{z}_{\mathbf{B}} \tag{B.17}$$

Substituting the above results into Eq. (B.3) and solving for C_m and C_n gives

$$C_m = \frac{\sin\alpha\cos\beta}{\sin\alpha_T} C_{m_T} \tag{B.18}$$

$$C_n = -\frac{\sin\beta}{\sin\alpha_T} C_{m_T} \tag{B.19}$$

and, as a reminder, $C_l = 0$, unless C_{l_T} is defined to be non-zero.

Note that in the coefficient equations Eqs. (B.12), (B.13), (B.18) and (B.19), there exists a singularity for $\alpha_T = 0, \pi$. However, from the definition of α_T in Fig. B.2 we see that when $\alpha_T = 0$, this implies that both α and β must be zero as well. In that case the "regular" coefficients are equal to the "total" coefficients, and thus $C_S = C_n = 0$. ■

B.2 Winged Re-entry Vehicle: HORUS-2B

B.2.1 General

Initially, HORUS-2B was designed as a fully reusable second stage to the Ariane-5 launcher. Basically, the vehicle was unpowered although it had been equipped with a deorbitation engine and attitude-control thrusters. Later on, a rocket engine was added to the design and this adapted version became the manned, second stage of Sänger, the German Two-Stage-To-Orbit reference concept. Here, we use the original design being that of a winged, unpowered re-entry vehicle (see Fig. B.4). In Table B.5, the main characteristics of HORUS are listed.

Table B.5 Main characteristics of HORUS-2B

Total vehicle length	25.0 m
Maximum fuselage width	5.4 m
Maximum fuselage height	4.5 m
Wing span	13.0 m
Wing chord	23.0 m
Wing area	110.0 m^2
Maximum payload mass	7,000 kg
Re-entry mass (nominal)	19,029 kg
Re-entry mass (incl. payload)	26,029 kg

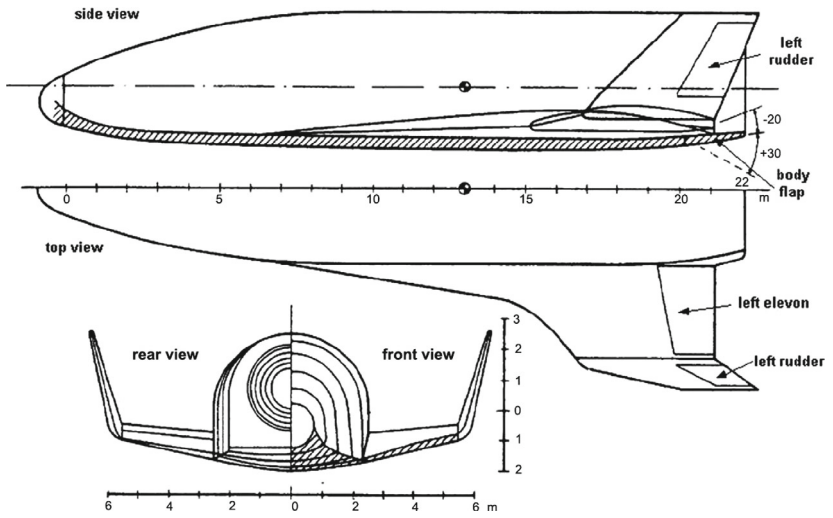


Fig. B.4 The HORUS-2B, adapted from MBB (1988)

The complete database of HORUS is described by MBB (1988), where the aerodynamic coefficients are given in graphical form; the corresponding numerical values have been measured from these graphs and published by Mooij (1995). We will briefly summarise the key elements of the vehicle. The gross mass of the vehicle at re-entry is $m = 26{,}029$ kg, which is the configuration when a 7,000 kg payload is returned to Earth. The c.o.m. is located in the vehicle's symmetry plane at 13 m from the nose (Fig. B.4). This configuration has the following principal moments of inertia:

$$\tilde{I}_{xx} = 119,000 \text{ kg m}^2 \qquad \tilde{I}_{yy} = 769,000 \text{ kg m}^2 \qquad \tilde{I}_{zz} = 806,000 \text{ kg m}^2$$

The inclination of the principal X-axis with respect to the X-axis of the body frame (negative for 'nose up') is $-1.7°$.

The HORUS-2B has a number of control surfaces, which have schematically been depicted in Fig. B.4: two rudders, two elevons and one body flap. The sign definitions for the deflection angles are as follows:

- Left and right rudder, deflection angles $\delta_{r,l}$ and $\delta_{r,r}$: positive outboard
- Left and right elevon, deflection angle $\delta_{e,l}$ and $\delta_{e,r}$: positive down
- Body flap, deflection angle δ_b: positive down.

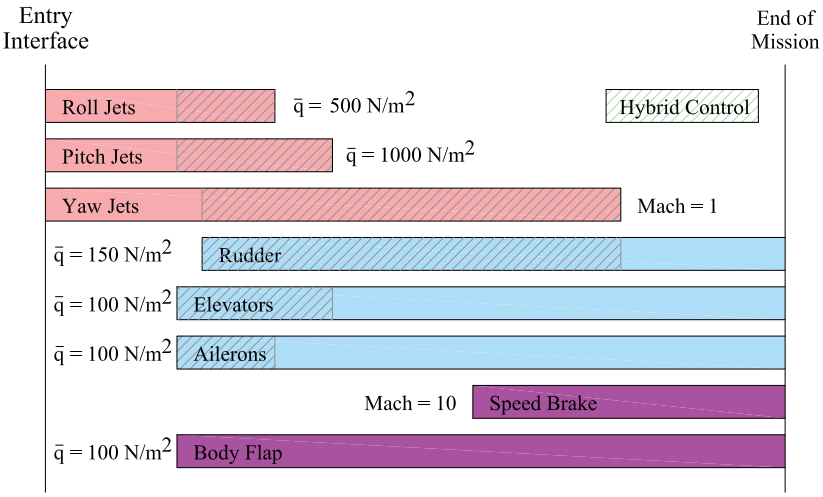


Fig. B.5 Entry control modes for HORUS

The rudders are outward movable only, which means that for yaw control only one rudder is active at a time. Deflecting both rudders at the same time would result in the so-called speed-brake function, but is not considered here.[8] The elevons combine both the elevator and aileron function. The deflections commanded by the attitude controller should be combined to give the corresponding left and right elevon deflection.

In Fig. B.5, the entry control modes for HORUS have been depicted. The reaction control thrusters are operated in the early phase of re-entry, when dynamic pressure is too low to allow for aerodynamic control. The aileron and elevator start operating once the dynamic pressure has reached a minimum value of 100 N/m^2. The rudder, which effectiveness is low, is being operated from a dynamic pressure of 150 N/m^2 and above. To support yaw control, the yaw thrusters continue to operate down to a Mach number of 1.

The control surfaces will only be effective above a certain dynamic pressure. For the upper layers of the atmosphere HORUS is equipped with reaction-control thrusters. These thrusters can also be used when large corrective manoeuvres are required and the control surfaces alone are not sufficient to deal with this. In Fig. B.5 the activation scheme of all control effectuators is shown. It is clear that the yaw thrusters are operational throughout the flight due to the low efficiency of the rudders.

The available thruster moments, as well as the control-surface moments per degree deflection (linearised aerodynamics), have been plotted in Fig. B.6. It can easily be

[8] Speed brakes can be used in the approach and landing phase to increase the drag and therefore decrease the velocity. Since the entry guidance of HORUS is based on energy control by regulating the drag through the angle-of-attack dependency, it might be worthwhile to study the use the speed brakes in this phase as well.

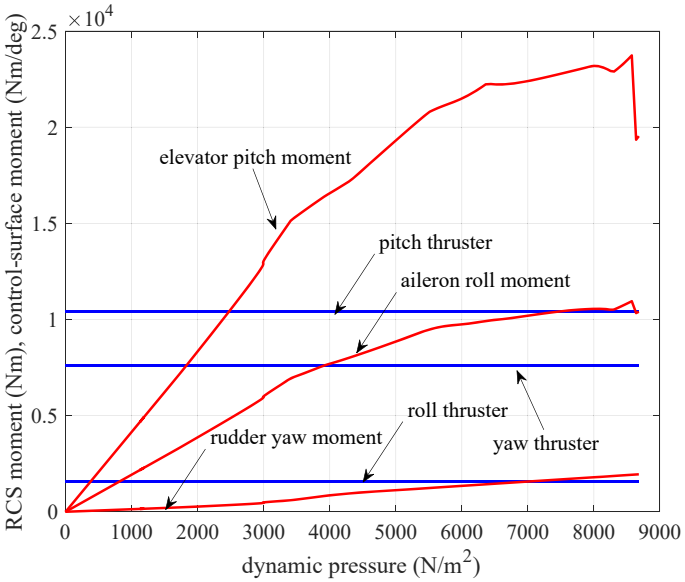


Fig. B.6 HORUS maximum available moments (RCS thrusters) and moments per degree deflection (control surfaces)

seen that for the elevons, there are three regions of a more or less linear increase with dynamic pressure. The rudder effectiveness seems to be completely linear with dynamic pressure. The dip in aileron effectiveness is, because the elevator-trim deflection suddenly changes, which has an effect on the point of linearisation.

B.2.2 Aerodynamic Database

The aerodynamic database of HORUS-2B is based on some simplifying assumptions. The coefficients are steady coefficients for the untrimmed super/hypersonic flight regime, neglecting the following aerodynamic effects:

- no aeroelastic effects (the vehicle is a non-elastic body);
- the influence of the Reynolds number on the skin-friction drag is not presented explicitly but included in the drag coefficient as averaged skin friction drag along a standard trajectory down to an altitude of 20 km. For trajectory calculations at lower altitudes, a drag altitude decrement is included in the data set. This decrement should be added to the total drag coefficient valid for 20 km;
- no special landing aerodynamics, like undercarriage drag, ground effects on pitch, lift and drag;
- no interference effects of the flaps;

- the influence of sideslip is simplified by linearisation over two degrees of angle of sideslip β. Note that β-derivatives or β-dependent coefficients can be extremely non-linear for $\beta > 2°$.

The definition of the aerodynamic coefficients is the following:

$$C_D = C_{D_0} + \Delta C_{D_{r,l}} + \Delta C_{D_{e,l}} + \Delta C_{D_b} + \Delta C_{D_{e,r}} + \Delta C_{D_{r,r}} - \Delta C_{D_h} \tag{B.20}$$

with

$$\begin{aligned}
C_{D_0} &= f(\alpha, M) \\
\Delta C_{D_{r,l}} &= f(\alpha, \delta_{r,l}, M) \\
\Delta C_{D_{e,l}} &= f(\alpha, \delta_{e,l}, M) \\
\Delta C_{D_b} &= f(\alpha, \delta_b, M) \\
\Delta C_{D_{e,r}} &= f(\alpha, \delta_{e,r}, M) \\
\Delta C_{D_{r,r}} &= f(\alpha, \delta_{r,r}, M) \\
C_{D_h} &= f(M, h)
\end{aligned}$$

$$C_S = \Delta C_{S_{r,l}} + \Delta C_{S_{e,l}} + \Delta C_{S_{e,r}} + \Delta C_{S_{r,r}} + \left[\left(\frac{\partial C_S}{\partial \beta} \right)_0 + \Delta \left(\frac{\partial C_S}{\partial \beta} \right)_{e,l} + \Delta \left(\frac{\partial C_S}{\partial \beta} \right)_{e,r} \right] \beta \tag{B.21}$$

with

$$\begin{aligned}
\Delta C_{S_{r,l}} &= f(\alpha, \delta_{r,l}, M) \\
\Delta C_{S_{e,l}} &= f(\alpha, \delta_{e,l}, M) \\
\Delta C_{S_{e,r}} &= f(\alpha, \delta_{e,r}, M) \\
\Delta C_{S_{r,r}} &= f(\alpha, \delta_{r,r}, M) \\
\left(\frac{\partial C_S}{\partial \beta} \right)_0 &= f(\alpha, M) \\
\Delta \left(\frac{\partial C_S}{\partial \beta} \right)_{e,l} &= f(\alpha, \delta_{e,l}, M) \\
\Delta \left(\frac{\partial C_S}{\partial \beta} \right)_{e,r} &= f(\alpha, \delta_{e,r}, M)
\end{aligned}$$

$$C_L = C_{L_0} + \Delta C_{L_{e,l}} + \Delta C_{L_b} + \Delta C_{L_{e,r}} \tag{B.22}$$

with

$$
\begin{aligned}
C_{L_0} &= f(\alpha, M) \\
\Delta C_{L_{e,l}} &= f(\alpha, \delta_{e,l}, M) \\
\Delta C_{L_b} &= f(\alpha, \delta_b, M) \\
\Delta C_{L_{e,r}} &= f(\alpha, \delta_{e,r}, M)
\end{aligned}
$$

$$
C_l = \Delta C_{l_{e,l}} + \Delta C_{l_{e,r}} + \left(\frac{\partial C_l}{\partial \beta}\right)_0 \cdot \beta \tag{B.23}
$$

with

$$
\begin{aligned}
\Delta C_{l_{e,l}} &= f(\alpha, \delta_{e,l}, M) \\
\Delta C_{l_{e,r}} &= f(\alpha, \delta_{e,r}, M) \\
\left(\frac{\partial C_l}{\partial \beta}\right)_0 &= f(\alpha, M)
\end{aligned}
$$

$$
C_m = C_{m_0} + \Delta C_{m_{e,l}} + \Delta C_{m_b} + \Delta C_{m_{e,r}} \tag{B.24}
$$

with

$$
\begin{aligned}
C_{m_0} &= f(\alpha, M) \\
\Delta C_{m_{e,l}} &= f(\alpha, \delta_{e,l}, M) \\
\Delta C_{m_b} &= f(\alpha, \delta_b, M) \\
\Delta C_{m_{e,r}} &= f(\alpha, \delta_{e,r}, M)
\end{aligned}
$$

$$
\begin{aligned}
C_n = \Delta C_{n_{r,l}} + \Delta C_{n_{e,l}} + \Delta C_{n_{e,r}} + \Delta C_{n_{r,r}} + \\
+ \left[\left(\tfrac{\partial C_n}{\partial \beta}\right)_0 + \Delta\left(\tfrac{\partial C_n}{\partial \beta}\right)_{r,l} + \Delta\left(\tfrac{\partial C_n}{\partial \beta}\right)_{e,l} + \Delta\left(\tfrac{\partial C_n}{\partial \beta}\right)_{e,r} + \Delta\left(\tfrac{\partial C_n}{\partial \beta}\right)_{r,r}\right]\beta
\end{aligned} \tag{B.25}
$$

with

$$
\begin{aligned}
\Delta C_{n_{r,l}} &= f(\alpha, \delta_{r,l}, M) \\
\Delta C_{n_{e,l}} &= f(\alpha, \delta_{e,l}, M) \\
\Delta C_{n_{e,r}} &= f(\alpha, \delta_{e,r}, M) \\
\Delta C_{n_{r,r}} &= f(\alpha, \delta_{r,r}, M) \\
\left(\frac{\partial C_n}{\partial \beta}\right)_0 &= f(\alpha, M) \\
\Delta\left(\frac{\partial C_n}{\partial \beta}\right)_{r,l} &= f(\alpha, \delta_{r,l}, M)
\end{aligned}
$$

$$\Delta\left(\frac{\partial C_n}{\partial \beta}\right)_{e,l} = f(\alpha, \delta_{e,l}, M)$$

$$\Delta\left(\frac{\partial C_n}{\partial \beta}\right)_{e,r} = f(\alpha, \delta_{e,r}, M)$$

$$\Delta\left(\frac{\partial C_n}{\partial \beta}\right)_{r,r} = f(\alpha, \delta_{r,r}, M)$$

Note that the current aerodynamic model of HORUS-2B does not comprise any damping coefficients, i.e.,

$$C_{l_p} = \frac{\partial C_l}{\partial\left(\frac{pb_{ref}}{V}\right)} = 0$$

$$C_{m_q} = \frac{\partial C_m}{\partial\left(\frac{qc_{ref}}{V}\right)} = 0$$

$$C_{n_r} = \frac{\partial C_n}{\partial\left(\frac{rb_{ref}}{V}\right)} = 0$$

The (supersonic) range of vehicle parameters and aerodynamic variables are given in Table B.6. Furthermore, there is a drag dependency on the altitude. This dependency is given as a tabulated function of M (1.20 and 1.50) and h (0, 5, 10, 15 and 20 km).

Summarised, the aerodynamic characteristics of HORUS are represented by tabulated functions of M, α and β, whereas the control-surface increments are also dependent on the related deflection angle. In total 34 tabulated coefficients are defined, i.e., seven for C_D, seven for C_S, four for C_L, three for C_l, four for C_m and nine for C_n.

Table B.6 Parameter ranges for the super/hypersonic aerodynamic coefficients

α	M	δ_r	δ_e	δ_b
(°)	(–)	(°)	(°)	(°)
0.0	1.2	0.0	−40.0	−20.0
5.0	1.5	10.0	−30.0	−10.0
10.0	2.0	20.0	−20.0	0.0
15.0	3.0	30.0	−10.0	10.0
20.0	5.0	40.0	0.0	20.0
25.0	10.0		10.0	30.0
30.0	20.0		20.0	
35.0			30.0	
40.0			40.0	
45.0				

Table B.7 HORUS drag coefficient (clean configuration) as a function of Mach number and angle of attack (MBB 1988; Mooij 1995)

Mach → α (°) ↓	1.2	1.5	2.0	3.0	5.0	10.0	20.0
0.0	0.10	0.09	0.08	0.07	0.07	0.07	0.05
5.0	0.12	0.10	0.09	0.08	0.07	0.07	0.05
10.0	0.20	0.14	0.12	0.10	0.09	0.08	0.06
15.0	0.33	0.23	0.18	0.15	0.14	0.12	0.09
20.0	0.52	0.36	0.29	0.23	0.20	0.18	0.14
25.0	0.77	0.54	0.42	0.35	0.31	0.29	0.24
30.0	1.08	0.77	0.60	0.51	0.45	0.42	0.36
35.0	1.08	1.03	0.82	0.69	0.63	0.60	0.52
40.0	1.08	1.32	1.07	0.90	0.82	0.79	0.70
45.0	1.08	1.32	1.33	1.21	1.05	1.02	0.92

Table B.8 HORUS lift coefficient (clean configuration) as a function of Mach number and angle of attack (MBB 1988; Mooij 1995)

Mach → α (°) ↓	1.2	1.5	2.0	3.0	5.0	10.0	20.0
0.0	−0.02	−0.02	−0.03	−0.04	−0.04	−0.05	−0.05
5.0	0.27	0.16	0.10	0.06	0.05	0.03	0.01
10.0	0.56	0.36	0.26	0.19	0.14	0.12	0.07
15.0	0.86	0.57	0.42	0.33	0.26	0.23	0.16
20.0	1.17	0.79	0.60	0.48	0.40	0.36	0.28
25.0	1.17	1.01	0.78	0.62	0.54	0.50	0.40
30.0	1.17	1.21	0.94	0.77	0.68	0.63	0.53
35.0	1.17	1.38	1.08	0.90	0.80	0.75	0.66
40.0	1.17	1.38	1.20	1.01	0.91	0.86	0.77
45.0	1.17	1.38	1.28	1.08	0.99	0.95	0.85

These coefficients include the static derivatives with respect to the angle of sideslip. No dynamic derivatives have been specified.

In Tables B.7 and B.8 the tabulated data of C_{D_0} and C_{L_0} can be found. These data can be used for simple point-mass calculations. For more accurate analyses, including guidance and control studies, obviously also the control-surface increments would be required. These data, in graphical form, can be found in the available documentation (MBB, 1988). However, all curves have been measured and the corresponding (numerical) data have been published in tabulated form by Mooij (1995). As the volume of data is very large, they cannot be included in this appendix.

B.2.3 Mission

The mission of HORUS starts in orbit with a deorbitation manoeuvre and successive coasting arc. The actual re-entry starts at the atmospheric entry interface ($h = 122$ km) with the following initial conditions:

$V = 7435.5$ m/s
$\gamma = -1.43°$
$\chi = 70.75°$
$h = 122$ km
$\tau = -106.58°$
$\delta = -22.3°$

The vehicle is assumed to be heading towards a runway in Kourou (French Guiana). Some 80 km from the landing site guidance switches to the Terminal Area Energy Management (TAEM) logic. However, here the main focus is on the hypersonic entry and descent.

The simulation has been executed for a trimmed condition (pitch equilibrium only). Originally, for the Mach range between 1 and 20, trim was realised with the body flap as much as possible, and the rest with the elevons (elevator function). (For $0 < M \leq 1$, although this flight regime is not of interest here, the HORUS-2B is trimmed with the elevators, while having the body flap maximum up). However, the upper Mach boundary had to be changed to a dynamic pressure boundary $\bar{q} \geq 100$ N/m^2, to prevent pitch instability and to be more in line with the body-flap activation of the Space Shuttle.

The nominal controls are shown in Fig. B.7, top. The nominal angle of attack is given as a function of flight time. HORUS enters the atmosphere with a high angle of attack of 40°. Because of this high angle, the maximum occurring heat flux can be kept sufficiently small. Further down the trajectory, the angle of attack is decreased in order to meet with the cross- and downrange requirements. The nominal bank-angle history (MBB, 1988) consists of absolute values only, guaranteeing a certain descent rate. The corresponding sign is determined by flying towards the targeting point near Kourou, while keeping the heading error within dead-band limits. Whenever the heading error exceeds the dead band, the sign of the bank angle is inverted. This manoeuvre, a so-called bank reversal, usually takes between 10 and 20 s. Since ideal control is applied, i.e., the actual attitude is equated to the commanded attitude, the reversals show as vertical lines. We see four reversals at $t \approx 724$ s, $t \approx 1076$ s, $t \approx 1184$ s and $t \approx 1240$ s. The simulation is finished when the Terminal Area, 0.75° ($\approx$ 83 km) from the runway, has been reached. This is about 1250 s after re-entry. In the lower plot of Fig. B.7, the altitude-velocity profile is plotted, whereas in Fig. B.8 the ground track is shown.

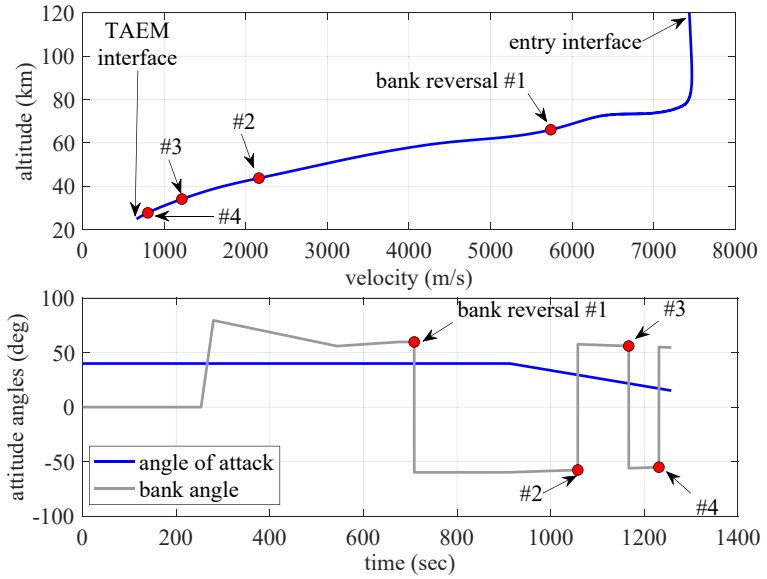


Fig. B.7 HORUS altitude-velocity profile (top) and nominal control history (bottom)

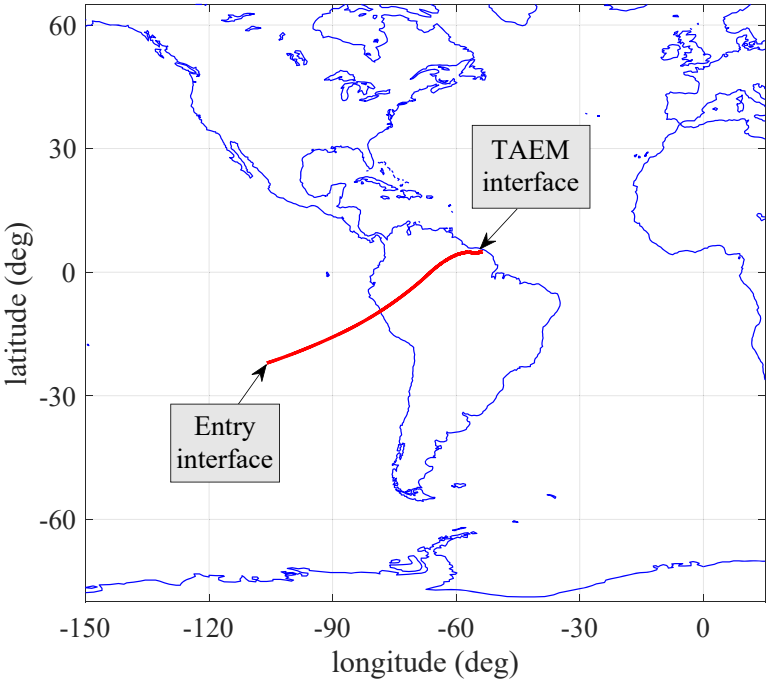


Fig. B.8 HORUS ground track for nominal mission

B.3 Entry Capsule: Huygens

ESA's Huygens missions concerned an entry into the atmosphere of Titan, Saturn's largest moon. It was launched together with NASA's Cassini mission on October 15, 1997, and was separated from Cassini on December 25, 2004, and landed on Titan on January 14, 2005 near the Adiri region. Until the date of writing (early 2020) it has been the only landing accomplished in the outer Solar System, and was also the first landing on a moon other than Earth's Moon. This part of the appendix will provide a summary of mass and geometry (Sect. B.3.1), aerodynamic data (Sect. B.3.2), and the parachute system (Sect. B.3.3) that were used during a period of shadow engineering, and provides a consistent database for preliminary design studies. The spin-vane configuration of the descent module has been introduced in the main text (Sect. 9.2.6) and will be summarised in Sect. B.3.4 for the sake of completeness.

It is stressed that these data do not describe the latest configuration, nor are they in any way corrected by or based on flight data. Source of the data is the mission-analysis report by Mooij (1992), where more details can be found, as well as the original data sources.

B.3.1 Mass and Geometry

Figure B.9 shows an artist impression of the entry and descent scenario. It shows that the probe enters the atmosphere with its aeroshell facing the flow. The parachute-deployment sequence is initiated when the axial deceleration has reached a value of 8.25 m/s^2 (corresponding with a Mach number of $M \approx 1.5$), and begins with the release of the aft cover. Thirty seconds after main parachute deployment, the heat shield is separated (the Mach number is about 0.6, which is considered to be a safe speed). The Huygens probe is therefore described by two distinctly different configurations, i.e., the entry configuration and the descent configuration under the parachute. The mass and geometry properties will therefore be provided as separate elements (Fig. B.10a): the descent module, aeroshell, and the back cover.

All coordinates are defined with respect to one and the same body-fixed reference frame (see Fig. B.10b). This frame has its origin in the nose of the heat shield, the X-axis is the longitudinal axis of symmetry, positive pointing aft, the Z-axis is positive in upward direction and the Y-axis completes the right-handed system. The mass properties for the entry configuration can be determined by combining the elements, summing the mass properties of the elements, and applying Steiner's parallel axis theorem to compute the global inertia tensor with respect to the global centre of mass.

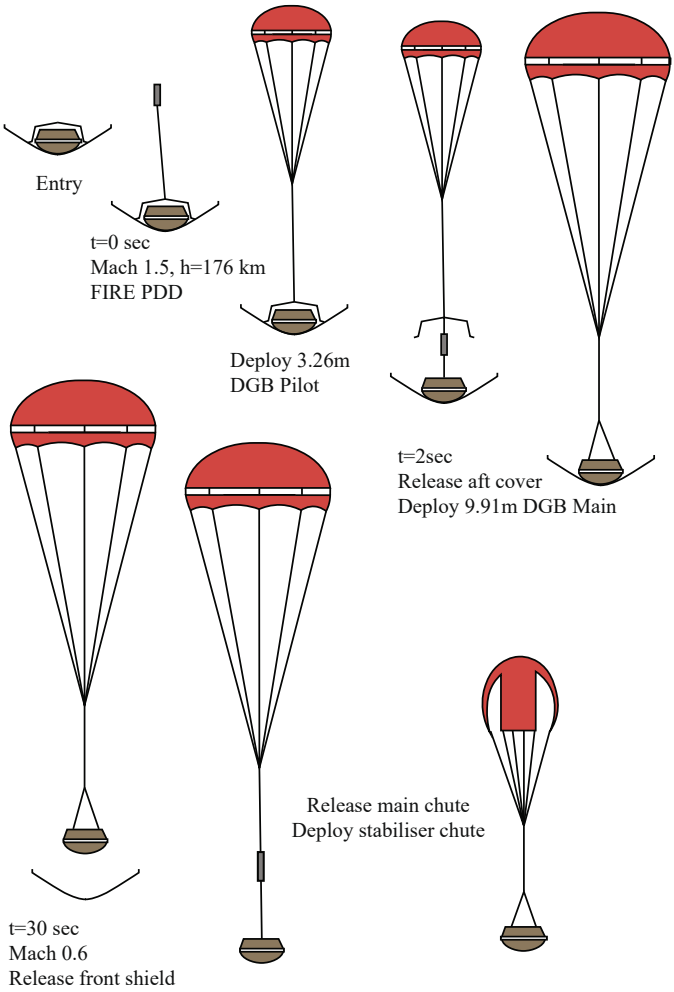
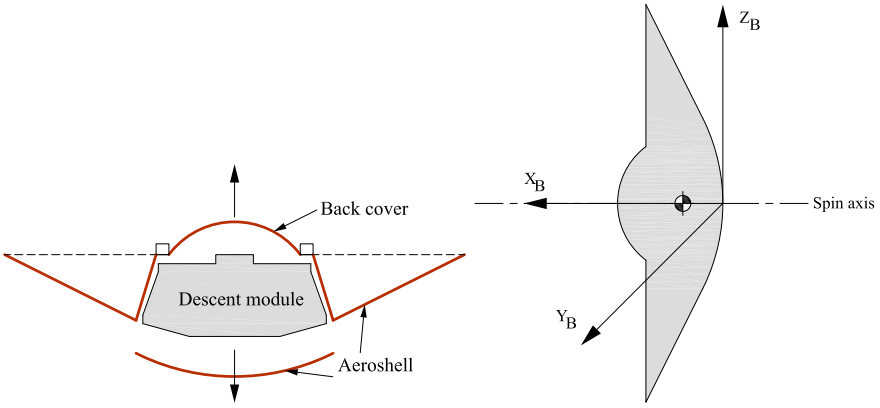


Fig. B.9 Huygens entry and descent scenario

Descent Module

$$
\begin{aligned}
m_{descent} &= 193.7 \text{ kg} \\
x_{cm,descent} &= 0.4049 \text{ m} \\
y_{cm,descent} &= 0.0015 \text{ m} \\
z_{cm,descent} &= 0.0006 \text{ m} \\
I_{xx,descent} &= 36.5 \text{ kg m}^2 \\
I_{yy,descent} &= 20.8 \text{ kg m}^2 \\
I_{zz,descent} &= 21.0 \text{ kg m}^2
\end{aligned}
$$



(a) Hardware elements (b) Body frame

Fig. B.10 Huygens hardware elements and body-frame definition

$I_{xy,descent} = I_{xz,descent} = I_{yz,descent}$ ≈ 0 kg m^2, i.e., the body is assumed rotational symmetric.

Aerodynamic decelerator

$m_{dec} =$ 77.7 kg
$x_{cm,dec} =$ 0.3988 m
$y_{cm,dec} = z_{cm,dec} =$ 0 m
$I_{xx,dec} =$ 92.5 kg m^2
$I_{yy,dec} = I_{zz,dec} =$ 52.4 kg m^2
$I_{xy,dec} = I_{xz,dec} = I_{yz,dec} =$ 0 kg m^2, i.e., the body is rotational symmetric.

Back cover

$m_{cover} =$ 15.3 kg
$x_{cm,cover} =$ 0.5810 m
$y_{cm,cover} = z_{cm,cover} =$ 0 m
$I_{xx,cover} =$ 6.7 kg m^2
$I_{yy,cover} = I_{zz,cover} =$ 3.8 kg m^2
$I_{xy,cover} = I_{xz,cover} = I_{yz,cover} =$ 0 kg m^2, i.e., the body is rotational symmetric.

Given the above mass elements, one can compute the global mass properties for the entry configuration, using the definition of centre of mass, and Steiner's parallel-axis theorem. This exercise is left to the reader, and we will simply state the results.

Entry configuration

$m_{tot} = 286.7$ kg
$x_{cm,tot} = 0.4126$ m
$y_{cm,tot} = 0.0015$ m
$z_{cm,tot} = 0.0006$ m
$I_{xx,tot} = 135.70$ kg m^2
$I_{yy,tot} = 77.46$ kg m^2
$I_{zz,tot} = 77.66$ kg m^2
$I_{xy,tot} = -0.0023$ kg m^2
$I_{yz,tot} = 0.00006$ kg m^2
$I_{xz,tot} = -0.0009$ kg m^2.

Note: in most flight simulation software, the body frame is defined opposite to the one shown in Fig. B.10b, i.e., the actual one being rotated 180° about the Y-axis. This gives an X-axis that points forward and allows for the "normal" definition of angle of attack ($-180° \leq \alpha < 180°$). In this situation the ballistic entry attitude is represented by $\alpha = 0°$, rather than $\alpha = 180°$, which was used for all Huygens' mission analyses done by industry. To remain consistent with the official documentation, the body-frame definition of Fig. B.10b has been adhered to for the definition of geometry and mass properties. To convert to the flight-simulation definition, one simply negates the X- and Z-coordinates, and thus also I_{xy} and I_{yz}.

B.3.2 Aerodynamic Database

For the Huygens probe, we have to consider two aerodynamic configurations. The first one is the descent module plus the jettisonable decelerator and the second one is the descent module by itself. For either configuration the moment reference centre is the stagnation point of the heat shield. The way of modelling is in both cases the same, and is briefly summarised below. Huygens is an axisymmetrical body, so in principle the aerodynamic coefficients could be defined as a function of the so-called total angle of attack, i.e., which can be seen as the arc summation of angle of attack and angle of sideslip. Only three so-called total aerodynamic coefficients are in that case needed: the total drag, lift and pitching moment coefficients. These coefficients can then be transformed to the actual drag, side and lift force coefficients and pitching and yawing moment coefficients (see also Intermezzo B.1 in Sect. B.1.2). Because most flight-dynamics software will have an interface with the latter set of coefficients, we will provide those and use the definition of $-180° \leq \alpha < 180°$.

The force coefficients are defined by the drag, side-force and lift coefficients, i.e., C_D, C_S, and C_L. The numerical values for the entry configuration are listed in Tables B.9, B.10, and B.11, whereas those for the descent configuration are included in Tables B.16, B.17, and B.18. The nominal roll-moment coefficient C_l is considered to be zero, but there is a non-zero damping coefficient. Thus:

Table B.9 Huygens drag coefficient C_D (entry configuration) as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−90	−20	−10	−5	0	5	10	20	90
0.4	0.72	0.83	0.90	0.93	0.94	0.93	0.90	0.83	0.72
0.6	0.62	0.84	0.91	0.94	0.95	0.94	0.91	0.84	0.62
0.8	0.58	0.90	0.99	1.03	1.03	1.03	0.99	0.90	0.58
1.0	0.66	1.06	1.15	1.20	1.21	1.20	1.15	1.06	0.66
1.2	0.58	1.17	1.28	1.33	1.34	1.33	1.28	1.17	0.58
1.4	0.49	1.24	1.37	1.43	1.44	1.43	1.37	1.24	0.49
1.6	0.40	1.29	1.44	1.50	1.51	1.50	1.44	1.29	0.40
2.0	0.44	1.34	1.49	1.55	1.56	1.55	1.49	1.34	0.44
2.2	0.51	1.34	1.48	1.54	1.55	1.54	1.48	1.34	0.51
3.0	0.55	1.27	1.47	1.53	1.54	1.53	1.47	1.27	0.55

Table B.10 Huygens side-force coefficient C_S (entry configuration) as a function of Mach number and angle of sideslip

β (°) → Mach (–) ↓	−90	−20	−10	−5	0	5	10	20	90
0.4	0.020	0.245	0.179	0.051	0.000	−0.051	−0.179	−0.245	−0.020
0.6	0.030	0.265	0.190	0.056	0.000	−0.056	−0.190	−0.265	−0.030
0.8	0.040	0.307	0.214	0.066	0.000	−0.066	−0.214	−0.307	−0.040
1.0	0.000	0.367	0.252	0.078	0.000	−0.078	−0.252	−0.367	0.000
1.2	−0.020	0.428	0.292	0.092	0.000	−0.092	−0.292	−0.428	0.020
1.4	−0.007	0.481	0.325	0.105	0.000	−0.105	−0.325	−0.481	0.007
1.6	0.006	0.524	0.351	0.115	0.000	−0.115	−0.351	−0.524	−0.006
2.0	−0.014	0.537	0.360	0.117	0.000	−0.117	−0.360	−0.537	0.014
2.2	0.001	0.523	0.353	0.113	0.000	−0.113	−0.353	−0.523	−0.001
3.0	−0.097	0.477	0.343	0.110	0.000	−0.110	−0.343	−0.477	0.097

$$C_l = C_{l_p} \frac{pd}{V} \tag{B.26}$$

where p is the angular roll-rate with respect to $\mathcal{F}_R$ (rad/s), d is the aerodynamic reference length (m), and V is the relative flow velocity (m/s). The pitch-moment coefficient C_m has two components, one term independent of the rotational rate of the probe, C_{m_0}, and a damping term, C_{m_q}, which is dependent on the normalised rotational rate about the Y-axis. It is represented by

Table B.11 Huygens lift coefficient C_L (entry configuration) as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−90	−20	−10	−5	0	5	10	20	90
0.4	0.020	0.245	0.179	0.051	0.000	−0.051	−0.179	−0.245	−0.020
0.6	0.030	0.265	0.190	0.056	0.000	−0.056	−0.190	−0.265	−0.030
0.8	0.040	0.307	0.214	0.066	0.000	−0.066	−0.214	−0.307	−0.040
1.0	0.000	0.367	0.252	0.078	0.000	−0.078	−0.252	−0.367	0.000
1.2	−0.020	0.428	0.292	0.092	0.000	−0.092	−0.292	−0.428	0.020
1.4	−0.007	0.481	0.325	0.105	0.000	−0.105	−0.325	−0.481	0.007
1.6	0.006	0.524	0.351	0.115	0.000	−0.115	−0.351	−0.524	−0.006
2.0	−0.014	0.537	0.360	0.117	0.000	−0.117	−0.360	−0.537	0.014
2.2	0.001	0.523	0.353	0.113	0.000	−0.113	−0.353	−0.523	−0.001
3.0	−0.097	0.477	0.343	0.110	0.000	−0.110	−0.343	−0.477	0.097

$$C_m = C_{m_0} + C_{m_q} \frac{qd}{V} \tag{B.27}$$

where q is the angular pitch-rate with respect to $\mathcal{F}_R$ (rad/s). The yaw-moment coefficient C_n is defined similarly to the pitching-moment coefficient:

$$C_n = C_{n_0} + C_{n_r} \frac{rd}{V} \tag{B.28}$$

with r being the angular yaw-rate with respect to $\mathcal{F}_R$ (rad/s). Numerical values for the moment coefficients are listen in Tables B.12 through B.15 for the entry configuration, and in Tables B.9 through B.23 for the descent configuration[9]. The values of the reference length and area, used to compute the aerodynamic forces and moments, are for the entry configuration $d = 3$ m and $S_{ref} = 7.069$ m^2, and $d = 1.393$ m and $S_{ref} = 1.524$ m^2 for the descent configuration.

B.3.3 Parachute System

During the descent part of the mission, three parachute systems are used: a pilot chute, a main parachute, and a stabilising drogue chute. There were no public data available about the pilot chute, so we will restrict ourselves to the latter two, which

[9] It is noted that the symmetric coefficients, i.e., those dependent on the angle of attack, are not affected by a non-zero angle of sideslip, and *vice versa*, the lateral coefficients are not affected by a change in angle of attack.

Table B.12 Huygens pitch-moment coefficient C_{m_0} (entry configuration) as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−90	−20	−10	−5	0	5	10	20	90
0.4	0.401	0.093	0.046	0.026	0.000	−0.026	−0.046	−0.093	−0.401
0.6	0.327	0.075	0.037	0.020	0.000	−0.020	−0.037	−0.075	−0.327
0.8	0.292	0.068	0.033	0.019	0.000	−0.019	−0.033	−0.068	−0.292
1.0	0.336	0.077	0.038	0.021	0.000	−0.021	−0.038	−0.077	−0.336
1.2	0.318	0.073	0.037	0.020	0.000	−0.020	−0.037	−0.073	−0.318
1.4	0.220	0.052	0.026	0.014	0.000	−0.014	−0.026	−0.052	−0.220
1.6	0.166	0.040	0.020	0.011	0.000	−0.011	−0.020	−0.040	−0.166
2.0	0.182	0.042	0.021	0.012	0.000	−0.012	−0.021	−0.042	−0.182
2.2	0.193	0.046	0.023	0.013	0.000	−0.013	−0.023	−0.046	−0.193
3.0	0.256	0.060	0.026	0.013	0.000	−0.013	−0.026	−0.060	−0.256

Table B.13 Huygens pitch-damping-moment coefficient C_{m_q} (entry configuration) as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−25	−10	−7	−5	0	5	7	10	25
0.0	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08
0.4	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08
0.6	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12
0.8	−0.16	−0.16	−0.16	−0.16	−0.16	−0.16	−0.16	−0.16	−0.16
0.9	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20
1.0	0.10	0.05	0.05	0.00	−0.10	0.00	0.05	0.05	0.10
1.1	3.50	1.40	0.75	0.40	−0.20	0.40	0.75	1.40	3.50
1.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.6	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20
100.0	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20

are both Disk Gap Band (DGB) parachutes. Below, the related parameters for the two chutes are listed (see also Fig. B.11). It is noted that during the design phase two stabilising−drogue configurations were studied, linked to the descent time under the main parachute. For the first configuration, a larger drogue parachute would be deployed after 600 s of main-parachute descent. The second configuration would have a main-parachute descent of 2,550 s, at which moment a minimum-sized, stabilising drogue would be deployed. The mass and geometry data for both configurations are provided.

Table B.14 Huygens yaw-moment coefficient C_{n_0} (entry configuration) as a function of Mach number and angle of sideslip

β (°) → Mach (–) ↓	−90	−20	−10	−5	0	5	10	20	90
0.4	0.401	0.093	0.046	0.026	0.000	−0.026	−0.046	−0.093	−0.401
0.6	0.327	0.075	0.037	0.020	0.000	−0.020	−0.037	−0.075	−0.327
0.8	0.292	0.068	0.033	0.019	0.000	−0.019	−0.033	−0.068	−0.292
1.0	0.336	0.077	0.038	0.021	0.000	−0.021	−0.038	−0.077	−0.336
1.2	0.318	0.073	0.037	0.020	0.000	−0.020	−0.037	−0.073	−0.318
1.4	0.220	0.052	0.026	0.014	0.000	−0.014	−0.026	−0.052	−0.220
1.6	0.166	0.040	0.020	0.011	0.000	−0.011	−0.020	−0.040	−0.166
2.0	0.182	0.042	0.021	0.012	0.000	−0.012	−0.021	−0.042	−0.182
2.2	0.193	0.046	0.023	0.013	0.000	−0.013	−0.023	−0.046	−0.193
3.0	0.256	0.060	0.026	0.013	0.000	−0.013	−0.026	−0.060	−0.256

Table B.15 Huygens yaw-damping-moment coefficient C_{n_r} (entry configuration) as a function of Mach number and angle of sideslip

β (°) → Mach (–) ↓	−25	−10	−7	−5	0	5	7	10	25
0.0	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08
0.4	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08
0.6	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12
0.8	−0.16	−0.16	−0.16	−0.16	−0.16	−0.16	−0.16	−0.16	−0.16
0.9	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20
1.0	0.10	0.05	0.05	0.00	−0.10	0.00	0.05	0.05	0.10
1.1	3.50	1.40	0.75	0.40	−0.20	0.40	0.75	1.40	3.50
1.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.6	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20
100.0	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20	−0.20

Main parachute

n_p = number of parachutes = 1
m_p = mass of parachute = 3.897 kg
m_1 = mass of suspension lines = 1.038 kg
d_0 = nominal diameter = 9.91 m
l_s = suspension line length = 20.43 m
l_r = riser line length = 10.18 m

Table B.16 Huygens drag coefficient C_D (descent configuration) as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−90	−25	−15	−5	0	5	15	25	90
0.0	0.28	0.95	1.04	1.05	1.04	1.05	1.04	0.95	0.28
0.5	0.28	0.95	1.04	1.05	1.04	1.05	1.04	0.95	0.28
0.8	0.79	1.03	1.10	1.11	1.08	1.11	1.10	1.03	0.79
0.9	0.72	1.15	1.23	1.27	1.25	1.27	1.23	1.15	0.72
1.4	0.95	1.32	1.39	1.42	1.43	1.42	1.39	1.32	0.95
2.1	0.85	1.29	1.43	1.47	1.48	1.47	1.43	1.29	0.85

Table B.17 Huygens side-force coefficient C_S (descent configuration) as a function of Mach number and angle of sideslip

β (°) → Mach (–) ↓	−90	−25	−15	−5	0	5	15	25	90
0.0	0.190	0.445	0.291	0.126	0.000	−0.126	−0.291	−0.445	−0.190
0.5	0.190	0.445	0.291	0.126	0.000	−0.126	−0.291	−0.445	−0.190
0.8	0.125	0.392	0.276	0.123	0.000	−0.123	−0.276	−0.392	−0.125
0.9	0.057	0.415	0.304	0.133	0.000	−0.133	−0.304	−0.415	−0.057
1.4	0.508	0.469	0.334	0.120	0.000	−0.120	−0.334	−0.469	−0.508
2.1	0.360	0.453	0.329	0.121	0.000	−0.121	−0.329	−0.453	−0.360

Table B.18 Huygens lift coefficient C_L (descent configuration) as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−90	−25	−15	−5	0	5	15	25	90
0.0	0.190	0.445	0.291	0.126	0.000	−0.126	−0.291	−0.445	−0.190
0.5	0.190	0.445	0.291	0.126	0.000	−0.126	−0.291	−0.445	−0.190
0.8	0.125	0.392	0.276	0.123	0.000	−0.123	−0.276	−0.392	−0.125
0.9	0.057	0.415	0.304	0.133	0.000	−0.133	−0.304	−0.415	−0.057
1.4	0.508	0.469	0.334	0.120	0.000	−0.120	−0.334	−0.469	−0.508
2.1	0.360	0.453	0.329	0.121	0.000	−0.121	−0.329	−0.453	−0.360

Table B.19 Huygens roll-damping-moment coefficient C_{l_p} (descent configuration) as a function of Mach number

Mach (–)	
0.04	−0.0021
0.10	−0.0038
0.20	−0.0053
0.27	−0.0058
0.40	−0.0530

Table B.20 Huygens pitch-moment coefficient C_{m_0} (descent configuration) as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−90	−25	−15	−5	0	5	15	25	90
0.0	0.068	0.058	0.039	0.016	0.000	−0.016	−0.039	−0.058	−0.068
0.5	0.068	0.058	0.039	0.016	0.000	−0.016	−0.039	−0.058	−0.068
0.8	0.068	0.058	0.039	0.017	0.000	−0.017	−0.039	−0.058	−0.068
0.9	0.068	0.056	0.036	0.017	0.000	−0.017	−0.036	−0.056	−0.068
1.4	0.060	0.049	0.030	0.012	0.000	−0.012	−0.030	−0.049	−0.060
2.1	0.053	0.046	0.029	0.011	0.000	−0.011	−0.029	−0.046	−0.053

Table B.21 Huygens pitch-damping-moment coefficient C_{m_q} (descent configuration) as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−5	−3	−2	−1	0	1	2	3	5
0.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.8	−0.857	−0.915	−0.974	−1.032	−1.149	−1.032	−0.974	−0.915	−0.857
0.9	−0.585	−0.674	−0.763	−0.853	−1.031	−0.853	−0.763	−0.674	−0.585
1.4	0.140	−0.031	−0.202	−0.374	−0.716	−0.374	−0.202	−0.031	0.140
2.1	−0.500	−0.522	−0.544	−0.566	−0.610	−0.566	−0.544	−0.522	−0.500

Table B.22 Huygens yaw-moment coefficient C_{n_0} (descent configuration) as a function of Mach number and angle of sideslip

β (°) → Mach (–) ↓	−90	−25	−15	−5	0	5	15	25	90
0.0	0.068	0.058	0.039	0.016	0.000	−0.016	−0.039	−0.058	−0.068
0.5	0.068	0.058	0.039	0.016	0.000	−0.016	−0.039	−0.058	−0.068
0.8	0.068	0.058	0.039	0.017	0.000	−0.017	−0.039	−0.058	−0.068
0.9	0.068	0.056	0.036	0.017	0.000	−0.017	−0.036	−0.056	−0.068
1.4	0.060	0.049	0.030	0.012	0.000	−0.012	−0.030	−0.049	−0.060
2.1	0.053	0.046	0.029	0.011	0.000	−0.011	−0.029	−0.046	−0.053

Table B.23 Huygens yaw-damping-moment coefficient C_{n_r} (descent configuration) as a function of Mach number and angle of sideslip

β (°) → Mach (–) ↓	−5	−3	−2	−1	0	1	2	3	5
0.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.8	−0.857	−0.915	−0.974	−1.032	−1.149	−1.032	−0.974	−0.915	−0.857
0.9	−0.585	−0.674	−0.763	−0.853	−1.031	−0.853	−0.763	−0.674	−0.585
1.4	0.140	−0.031	−0.202	−0.374	−0.716	−0.374	−0.202	−0.031	0.140
2.1	−0.500	−0.522	−0.544	−0.566	−0.610	−0.566	−0.544	−0.522	−0.500

f = swivel friction coefficient = 0.000013 m
x_{att} = X-coordinate attachment = 0.73 m (with respect to the nose of the heat shield).

The original aerodynamic coefficients are given as axial-force coefficient, C_A, normal-force coefficient, C_N, pitch-moment coefficient C_{m_0}, each one a function of angle of attack (Table B.24), and a (constant) pitch-damping-moment coefficient, C_{m_q}. Furthermore, C_A is depending on the Mach number as well, according to the following formula:

$$C_A = C_A\,[1.0 - 0.3268(M - 1.3)] \tag{B.29}$$

for Mach numbers larger than 1.3.

To be consistent with the defined coefficients for the vehicle, C_A and C_N, which are given as a function of angle of attack, can easily be converted to C_D and C_L, and for rotational-symmetry reasons, C_S is the same as C_L, but then as a function of angle of sideslip. Because of the Mach number dependency of C_A, both C_D, C_S and C_L are dependent on M. For the same reason of rotational symmetry, C_{n_0} and C_{n_r}

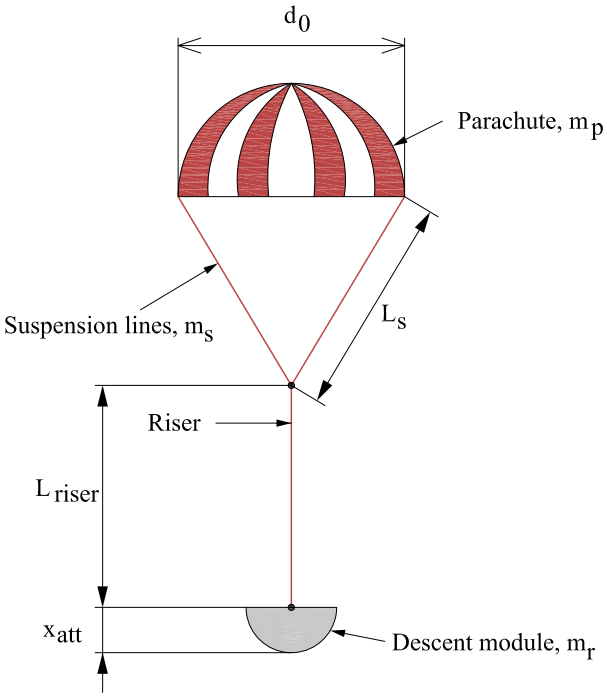


Fig. B.11 Parachute geometry

Table B.24 Available coefficients for the DGB parachute as a function of total angle of attack

α_T (°)	0.0	2.5	5.0	7.5	12.5
C_A	0.54	0.54	0.54	0.54	0.54
C_N	0.000	0.010	0.025	0.040	0.090
C_m	0.00000	−0.01364	−0.02955	−0.05454	−0.11364

are identical to C_{m_0} and C_{m_q}, but also they will be a function of β, rather than α. The aerodynamic coefficients are given in Tables B.25 through B.29; as mentioned, the damping-moment coefficients are constant and have a value of $C_{m_q} = C_{n_r} = -0.20$.

The reference point for which the coefficients are defined, is one parachute diameter d_0 below the skirt plane (Fig. B.12).

Stabilising Drogue Parachute (t_{deploy} = 600 s)

n_p = number of parachutes = 1
m_p = mass of parachute = 0.23 kg
m_s = mass of suspension lines = 0.084 kg

Table B.25 Drag coefficient C_D for the DGB parachute as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−12.5	−7.5	−5	−2.5	0	2.5	5	7.5	12.5
1.20	0.527	0.582	0.589	0.594	0.600	0.594	0.589	0.582	0.527
1.25	0.520	0.574	0.581	0.586	0.591	0.586	0.581	0.574	0.520
1.30	0.513	0.566	0.573	0.578	0.583	0.578	0.573	0.566	0.513
1.35	0.506	0.558	0.565	0.570	0.575	0.570	0.565	0.558	0.506
1.40	0.499	0.550	0.557	0.561	0.566	0.561	0.557	0.550	0.499
1.45	0.491	0.542	0.549	0.553	0.558	0.553	0.549	0.542	0.491
1.50	0.484	0.534	0.541	0.545	0.550	0.545	0.541	0.534	0.484

Table B.26 Side-force coefficient C_S for the DGB parachute as a function of Mach number and angle of sideslip

β (°) → Mach (–) ↓	−12.5	−7.5	−5	−2.5	0	2.5	5	7.5	12.5
1.20	0.024	0.036	0.026	0.015	0.000	−0.015	−0.026	−0.036	−0.024
1.25	0.023	0.035	0.025	0.015	0.000	−0.015	−0.025	−0.035	−0.023
1.30	0.021	0.034	0.025	0.015	0.000	−0.015	−0.025	−0.034	−0.021
1.35	0.020	0.033	0.024	0.014	0.000	−0.014	−0.024	−0.033	−0.020
1.40	0.018	0.032	0.023	0.014	0.000	−0.014	−0.023	−0.032	−0.018
1.45	0.016	0.031	0.022	0.014	0.000	−0.014	−0.022	−0.031	−0.016
1.50	0.015	0.029	0.022	0.013	0.000	−0.013	−0.022	−0.029	−0.015

Table B.27 Lift coefficient C_L for the DGB parachute as a function of Mach number and angle of attack

α (°) → Mach (–) ↓	−12.5	−7.5	−5	−2.5	0	2.5	5	7.5	12.5
1.20	0.024	0.036	0.026	0.015	0.000	−0.015	−0.026	−0.036	−0.024
1.25	0.023	0.035	0.025	0.015	0.000	−0.015	−0.025	−0.035	−0.023
1.30	0.021	0.034	0.025	0.015	0.000	−0.015	−0.025	−0.034	−0.021
1.35	0.020	0.033	0.04	0.014	0.000	−0.014	−0.024	−0.033	−0.020
1.40	0.018	0.032	0.023	0.014	0.000	−0.014	−0.023	−0.032	−0.018
1.45	0.016	0.031	0.022	0.014	0.000	−0.014	−0.022	−0.031	−0.016
1.50	0.015	0.029	0.022	0.013	0.000	−0.013	−0.022	−0.029	−0.015

Table B.28 Pitch-moment coefficient C_{m_0} for the DGB parachute as a function of angle of attack

α (°)	C_{m_0}
−12.5	0.11364
−7.5	0.05454
−5.0	0.02955
−2.5	0.01364
0.0	0.00000
2.5	−0.01364
5.0	−0.02955
7.5	−0.05454
12.5	−0.11364

Table B.29 Yaw-moment coefficient C_{n_0} for the DGB parachute as a function of angle of sideslip

β (°)	C_{n_0}
−12.5	0.11364
−7.5	0.05454
−5.0	0.02955
−2.5	0.01364
0.0	0.00000
2.5	−0.01364
5.0	−0.02955
7.5	−0.05454
12.5	−0.11364

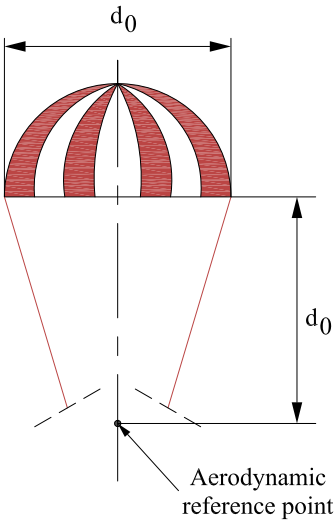


Fig. B.12 Reference-point definition for the aerodynamic coefficients

d_0 = nominal diameter = 2.09 m
L_s = suspension line length = 3.505 m
L_{riser} = riser line length = 9.13 m
f = swivel friction coefficient = 0.000013 m
x_{att} = X-coordinate attachment = 0.73 m (with respect to the nose of the heat shield).

Stabilising Drogue Parachute (t_{deploy} = 2,550 s)

n_p = number of parachutes = 1
m_p = mass of parachute = 0.11 kg
m_s = mass of suspension lines = 0.054 kg
d_0 = nominal diameter = 1.36 m
L_s = suspension line length = 1.23 m
L_{riser} = riser line length = 10.34 m
f = swivel friction coefficient = 0.000013 m
x_{att} = X-coordinate attachment = 0.73 m (with respect to the nose of the heat shield)

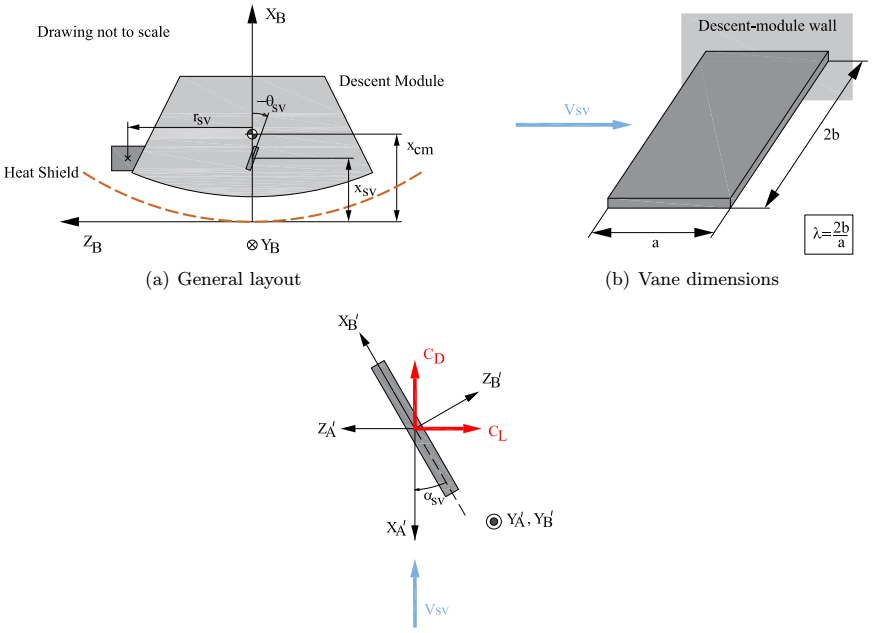
The aerodynamic coefficients for both stabilising drogues are the same as for the main parachute.

B.3.4 SpinVane Model

The conceptual spin-vane model for the descent module, to enforce a certain minimum spin rate under the parachute, consists of a set of vanes, considered to be rectangular flat plates (Fig. B.13).

The nominal spin-vane configuration is determined by the following parameters:

n_{sv} = number of spin vanes = 24
λ = $2b/a$ = aspect ratio of spin vane = 2
a, b = chord length and vane span = 0.0315 m
θ_{sv} = (constant) offset angle = $-2°$
r_{sv} = mean local radius descent module = 0.624 m
x_{sv} = X-coordinate spin vane with respect to the nose of the heat shield = 0.27 m
η_{sv} = efficiency (scale) factor = 1.0



(a) General layout (b) Vane dimensions

(c) Aerodynamic forces

Fig. B.13 Spin-vane model

References

MBB Space Communication and Propulsion Systems Division (1988) Study on re-entry guidance and control, ESA report reference: ESA CR (P) 2652

Mooij E (1992) ESA Huygens probe entry and descent analysis. ESTEC working paper EWP 1679, ESTEC, Noordwijk

Mooij E (1995) The Horus-2B reference vehicle, Memorandum M-692, Delft University of Technology

Moseley Jr, WC, Moore Jr RH, Hughes JE (1967) Stability characteristics of the Apollo command module, NASA TN D-3890

North American Aviation Inc (1965) Aerodynamic data manual for project Apollo, NASA-CR-82907

Appendix C
Mathematical Tools and Techniques

In this appendix several mathematical tools and techniques that come in handy when numerically simulating entry and descent problems are discussed. Section C.1 kicks off with a brief summary of spherical trigonometry. Next, in Sect. C.2, the concept of Taylor Series is introduced, which is then applied in Sect. C.3 to discuss the process of linearisation. Section C.4 presents several ways to interpolate in tabulated data, followed by Sect. C.5 with a short overview of numerical integration. Section C.6, finally, concludes this appendix with an introduction to root finding.

C.1 Spherical Trigonometry

Navigating on a flat surface is simply dictated by straight-line geometry. When the surface is curved, most notably the surface of a sphere, the closest analogue to straight lines are the so-called great and small circles. Great circles are circles of which the centre coincides with the centre of the sphere. For example, when we assume that the Earth is a perfect sphere, the meridians and the equator are great circles on its surface. On the other hand, if the plane does not contain the centre of the sphere then the circle is called a small circle. For a spheroidal Earth, the non-equatorial lines of latitude are small circles. On a plane, a line segment represents the shortest distance between two points. Similarly, an arc of a great circle on a sphere between two endpoints, is the shortest path on this sphere, and is covering less than 180°. An area on the sphere bounded by arcs of great circles is called a spherical polygon or spherical triangle, see also Fig. C.1a.

The sides of these polygons are specified by the angles at the sphere's centre spanned between the two endpoints of the sides. In the figure, these sides are indicated by a, b, and c (which may have different values for each of the indicated "spherical triangles"). Of course, these arc angles (in radians) can be transformed to an arc length (in m) when multiplied by the sphere's radius. The corresponding corner angles are indicated by A, B, and C. Spherical trigonometry allows us to do

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1

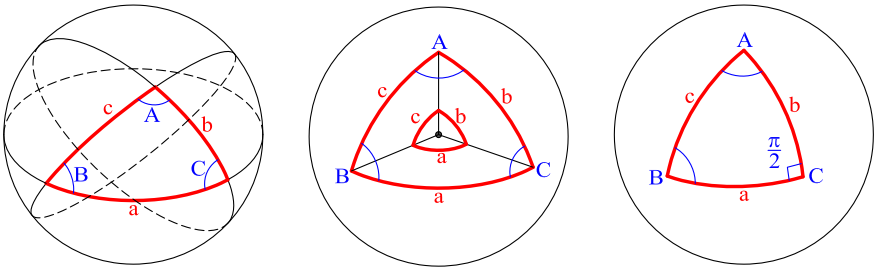


Fig. C.1 Spherical trigonometry

algebraic manipulations with these arc and corner angles, and is of great importance for calculations in astronomy, geodesy and navigation. Brummelen (2012) gives an extensive treatment of spherical trigonometry, starting from a historical perspective, providing the mathematical framework, and relating this to issues in astronomy and navigation. In this section, we will give a summary of the most important relations, without any derivation or providing context. The interested reader is referred to the earlier mentioned reference.

Spherical triangles satisfy a spherical law of cosines (note the cyclic behaviour):

$$\cos a = \cos b \cos c + \sin b \sin c \cos A \tag{C.1}$$

$$\cos b = \cos c \cos a + \sin c \sin a \cos B \tag{C.2}$$

$$\cos c = \cos a \cos b + \sin a \sin b \cos C \tag{C.3}$$

In Fig. C.1b, the geometry of the spherical triangle is indicated, including a scaled-down version of the triangle, which does not affect the angles involved.

A variation on the law of cosines, the second spherical law of cosines, or the cosine rule for angles, states that

$$\cos A = -\cos B \cos C + \sin B \sin C \cos a \tag{C.4}$$

$$\cos B = -\cos C \cos A + \sin C \sin A \cos b \tag{C.5}$$

$$\cos C = -\cos A \cos B + \sin A \sin B \cos c \tag{C.6}$$

Spherical triangles also satisfy an analogue of the law of sines:

$$\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c} \tag{C.7}$$

In case the angle C is a right angle, Fig. C.1c, then Eq. (C.3) changes to

$$\cos c = \cos a \cos b \tag{C.8}$$

In total, there are 12 relations for a right-handed triangle, details of which can be found in Todhunter (1886). They will not be stated here, as without context they are not well understood.

Finally, spherical triangles satisfy the half-side formulae:

$$\sin\left(\frac{A}{2}\right) = \sqrt{\frac{\sin(s-b)\sin(s-c)}{\sin b \sin c}} \tag{C.9}$$

$$\cos\left(\frac{A}{2}\right) = \sqrt{\frac{\sin s \sin(s-a)}{\sin b \sin c}} \tag{C.10}$$

$$\tan\left(\frac{A}{2}\right) = \sqrt{\frac{\sin(s-b)\sin(s-c)}{\sin s \sin(s-a)}} \tag{C.11}$$

$$\sin\left(\frac{a}{2}\right) = \sqrt{\frac{-\cos S \cos(S-A)}{\sin B \sin C}} \tag{C.12}$$

$$\cos\left(\frac{a}{2}\right) = \sqrt{\frac{-\cos(S-B)\cos(S-C)}{\sin B \sin C}} \tag{C.13}$$

$$\tan\left(\frac{a}{2}\right) = \sqrt{\frac{-\cos S \cos(S-A)}{\cos(S-B)\cos(S-C)}} \tag{C.14}$$

where $s = (a+b+c)/2$ and $S = (A+B+C)/2$. The 12 other sides/angles combinations follow by cyclic permutation.

One application of spherical trigonometry is the calculation of ground-track range, i.e., the ground-projected distance that a vehicle is covering over a spherical central body. The total ground-track range, d_{total}, is calculated with the so-called Haversine function (Brummelen 2012), which is historically used in navigation (the *haversine* is half the versed sine, i.e., a flipped sine function with half the period). In this particular case the distance along the trajectory is integrated using spherical trigonometry. Given the position (τ,δ) at two time points t_i and t_{i+1}, the distance covered in the time interval $\Delta t = t_{i+1} - t_i$ is calculated with:

$$a_{1,i} = \sin^2\left(\frac{\Delta\delta}{2}\right) + \cos\delta_i \cos\delta_{i+1} \sin^2\left(\frac{\Delta\tau}{2}\right) \tag{C.15}$$

$$a_{2,i} = 2\text{atan}\left(\frac{\sqrt{a_{1,i}}}{\sqrt{1-a_{1,i}}}\right) \tag{C.16}$$

$$d_{total} = R_e \sum_{i=1}^{n_{total}} a_{2,i} \tag{C.17}$$

with $\Delta\tau = \tau_{i+1} - \tau_i$ and $\Delta\delta = \delta_{i+1} - \delta_i$, $n_{total} =$ total number of arc segments, and $R_e =$ equatorial radius of the central body.

C.2 Taylor Series

In mathematics, a Taylor series is a representation of a function as an infinite sum of terms that are calculated from the values of the function's derivatives at a single point. The Taylor series is named after the English mathematician Brook Taylor, who formally introduced the concept in 1715. If the Taylor series is centred at zero, then that series is also called a Maclaurin series, named after the Scottish mathematician Colin Maclaurin. He made extensive use of this special case of Taylor series in the 18th century.

The Taylor series of a real or complex function $f(x)$ that is infinitely differentiable in a neighbourhood of a real or complex number a is the power series

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \ldots \tag{C.18}$$

which can be written in a more compact form as

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{C.19}$$

In the above equation, $n!$ is the factorial of n and $f^{(n)}$ is the nth derivative of f evaluated at the point a. The zeroth derivative of f is defined to be f itself and $(x-a)^0$ and 0! are both defined to be 1. As mentioned above, in the case that $a = 0$, the series is also called a Maclaurin series.

It is common practice to approximate a function by using a finite number of terms of its Taylor series instead of Eq. (C.19). Taylor's theorem gives quantitative estimates on the error in this approximation. Any finite number of initial terms of the Taylor series of a function is called a Taylor polynomial.

Example C.1 (*MacLaurin series of* $\cos x$) As an example of Taylor series we will derive the Mclaurin series of $f(x) = \cos x$. Taking the first four derivatives of $f(x) = \cos x$ gives $f^{(0)}(x) = \cos x$, $f^{(1)}(x) = -\sin x$, $f^{(2)}(x) = -\cos x$, $f^{(3)}(x) = \sin x$, $f^{(4)}(x) = \cos x$, after which the cycle repeats itself. Evaluating the derivatives for $x = a = 0$ yields $f^{(0)}(0) = 1$, $f^{(1)}(0) = 0$, $f^{(2)}(0) = -1$, $f^{(3)}(0) = 0$, $f^{(4)}(0) = 1$, etc.

Substituting these values in Eq. (C.19) we get:

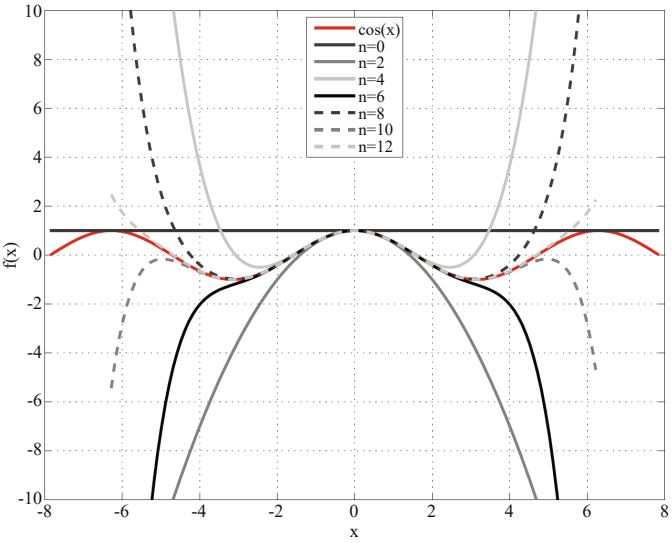


Fig. C.2 As the degree of the Taylor polynomial rises, it approaches the correct function. Here, $\cos x$ and Taylor approximations (polynomials of degree 0, 2, 4, 6, 8, 10 and 12) are shown

$$\cos x = \frac{1}{0!}x^0 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \ldots = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 + \ldots \tag{C.20}$$

In Fig. C.2, the Maclaurin series for $\cos x$ has been plotted for different degrees of the Taylor polynomials. It is clear that the approximation improves with increasing n. ■

C.3 Linearisation

In mathematics and its applications, linearisation refers to finding the linear approximation to a function at a given point. In the study of dynamical systems, linearisation is a method for assessing the local stability of an equilibrium point of a system of nonlinear differential equations or discrete dynamical systems. One can use the eigenvalues of the Jacobian matrix evaluated at an equilibrium point to determine the nature of that equilibrium. If any of the eigenvalues is positive, the equilibrium is unstable; if they are all negative the equilibrium is stable; and if the values are of mixed signs, the equilibrium is possibly a saddle point. Any complex eigenvalues will appear in complex conjugate pairs and indicate an oscillatory motion. Further, application of certain classical control design and analysis techniques requires a system to be linear. During the linearisation process, the Taylor series is mainly used for approximating functions when one can identify a small parameter.

Linearisation is an effective method for approximating the output of a function $y = f(x)$ at any $x = x_0$ based on the value and slope of the function at $x = x_0 \pm \Delta x$ given that $f(x)$ is continuous over the interval Δx and that Δx is sufficiently small. In, short, linearisation approximates the output of any differentiable function near $x = x_0$.

The differentiable function $y = f(x)$—for the sake of the discussion f is one dimensional and real—can be written in a Taylor-expansion form as:

$$\begin{aligned} y =& f(x_0) + f'(x_0)(x - x_0) + \frac{f'(x_0)}{2!}(x - x_0)^2 \\ &+ \frac{f^{(3)}(x_0)}{3!}(x - x_0)^3 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \ldots \end{aligned} \tag{C.21}$$

where f' denotes the first derivative of function f with respect to x, and f'' denotes the second derivative, etc. If the function would be two dimensional, i.e., the state $\mathbf{x}$ is given by $\mathbf{x} = (\mathbf{x}, \mathbf{u})$, the Taylor expansion about state (x_0, u_0) of a real function can be written as:

$$\begin{aligned} y =& f(x_0, u_0) + [f_x(x_0, u_0)\Delta x + f_u(x_0, u_0)\Delta u] \\ &+ \frac{1}{2!}\left[f_{xx}(x_0, u_0)\Delta x^2 + 2f_{xu}(x_0, u_0)\Delta x \Delta u + f_{uu}(x_0, u_0)\Delta u^2\right] \\ &+ \frac{1}{3!}\left[f_{xxx}(x_0, u_0)\Delta x^3 + 3f_{xxu}(x_0, u_0)\Delta x^2 \Delta u\right. \\ &+ \left. 3f_{xuu}(x_0, u_0)\Delta x \Delta u^2 + f_{uuu}(x_0, u_0)\Delta u^3\right] \\ &+ \ldots \end{aligned} \tag{C.22}$$

where f_x denotes the partial derivative of f with respect to x and f_u denotes the partial derivative of f with respect to u, etc. The Taylor expansion can become rather large when the dimension of the state grows, although the general principle of obtaining the derivatives does not change.

Linearisation of a function about a point is defined as expanding the function in a Taylor series and taking only the first term and the terms with a first (partial) derivative from that expansion. This means that terms, which contain (partial) derivatives of order higher than 1 are discarded. So, for a function with an n-dimensional state the linearised function about $\mathbf{x} = \mathbf{x_0}$ becomes:

$$\begin{aligned} y &= f(\mathbf{x_0}) + f_{x_1}(\mathbf{x_0})\Delta x_1 + f_{x_2}(\mathbf{x_0})\Delta x_2 + \ldots + f_{x_n}(\mathbf{x_0})\Delta x_n \\ &= f(\mathbf{x_0}) + \sum_{i=1}^{n} f_{x_i}(\mathbf{x_0})\Delta x_i \end{aligned} \tag{C.23}$$

Equation (C.23) can also be written in matrix-vector form as $y = f(\mathbf{x_0}) + \mathbf{A}\Delta\mathbf{x}$, with $\mathbf{A} = \begin{bmatrix} f_{x_1} & f_{x_2} & \ldots & f_{x_n} \end{bmatrix}$ being a row vector and $\Delta\mathbf{x} = [\Delta x_1 \ \Delta x_2 \ \ldots \ \Delta x_n]^T$ being a column vector.

Example C.2 (*Kinematic equation*) The change in altitude h of an aircraft or re-entry vehicle, or equivalently the radial distance R to the origin of a planet-fixed reference frame, is given by the following kinematic equation of motion:

$$\dot{R} = \dot{h} = V \sin\gamma \tag{C.24}$$

with V being the velocity and γ being the flight-path angle.

Following Eq. (C.23) gives for the partial derivatives

$$\left.\frac{\mathrm{d}\dot{h}}{\mathrm{d}V}\right|_{\mathbf{x}=\mathbf{x_0}} = \sin\gamma_0 \quad \text{and} \quad \left.\frac{\mathrm{d}\dot{h}}{d\gamma}\right|_{\mathbf{x}=\mathbf{x_0}} = V_0 \cos\gamma_0$$

These derivatives substituted in Eq. (C.23) results in the following linearised kinematic equation of motion:

$$\Delta\dot{h} = \sin\gamma_0 \Delta V + V_0 \cos\gamma_0 \Delta\gamma \tag{C.25}$$

■

Equation (C.23) can be generalised for any number of equations. One particular application could be the application of classic control theory to design a linear state feedback controller. In that case it is necessary that the system of equations of motion is Linear Time Invariant (LTI). Since the general equations of motion are time varying and highly non-linear, some assumptions are usually made. The nominal trajectory is considered to be an equilibrium path that the vehicle follows under ideal conditions. Looking only at small deviations from the nominal trajectory, the equations can be linearised. Furthermore, by discretising the trajectory into a number of sufficiently small time intervals, in each of the intervals the linearised state of the vehicle can be assumed to be constant. As a result, the overall system has been divided into a sequence of LTI systems.

With the general form of the equations of motion given by $\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}, \mathbf{u})$ with $\mathbf{x}$ being the nx1 state vector and $\mathbf{u}$ the mx1 control vector, linearisation of these n differential equations $\mathbf{f}$ leads to n equations of the form of Eq. (C.23). For a time point k, the LTI system can be written in matrix or *state-space* form as

$$\Delta\dot{\mathbf{x}} = \mathbf{A}(t_k)\Delta\mathbf{x} + \mathbf{B}(t_k)\Delta\mathbf{u} \tag{C.26}$$

In this equation, $\mathbf{A}$ and $\mathbf{B}$ are the system and control matrix of appropriate dimensions, respectively. Note that the above model is sometimes also referred to as the local stability model. For each time point t_k, $\mathbf{A}$ and $\mathbf{B}$ are constant, and can be obtained from n equations of motion f_i by

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_{n_p}}{\partial x_n} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{bmatrix} \tag{C.27}$$

evaluated for the nominal state and control, $\mathbf{x} = \mathbf{x_0}$ and $\mathbf{u} = \mathbf{u_0}$.

To look at linearisation in a slightly different manner, let us again consider the ordinary differential equation

$$\dot{x} = f(t, x) \tag{C.28}$$

To linearise this equation we may approximate the state x as the sum of a nominal (or *equilibrium*) value, x_0, and a small perturbation, Δx, yielding:

$$\dot{x}_0 + \Delta\dot{x} = f(t, x_0 + \Delta x) \tag{C.29}$$

Here, $\dot{x}_0 = f(t, x_0)$.

Subtracting $\dot{x}_0$ from both sides in Eq. (C.29) gives us the differential equation for the behaviour of the perturbations, or, in other words, the linearised system:

$$\Delta\dot{x} = f(t, x_0 + \Delta x) - f(t, x_0) \tag{C.30}$$

To illustrate this consider the following differential equation:

$$\dot{x} = x^2 \tag{C.31}$$

Applying Eq. (C.29) gives

$$\dot{x}_0 + \Delta\dot{x} = (x_0 + \Delta x)^2 = x_0^2 + 2x_0\Delta x + (\Delta x)^2 \approx x_0^2 + 2x_0\Delta x \tag{C.32}$$

The term $(\Delta x)^2$ can be neglected with respect to the other two terms, since Δx is small, and therefore $(\Delta x)^2$ is even smaller. We say that this term is of *second order*, and therefore not important for the linearisation, where we only consider terms of the *first order*.

Subtracting the nominal differential equations, i.e., $\dot{x}_0$ from the left side and x_0^2 from the right, results in

$$\Delta\dot{x} = 2x_0\Delta x \tag{C.33}$$

which is the linearised equation.

Example C.3 The differential equation that we consider for linearisation is

$$\dot{x} = \frac{1}{x} \tag{C.34}$$

Similarly as before, we get:

$$\dot{x}_0 + \Delta\dot{x} = \frac{1}{x_0 + \Delta x} \tag{C.35}$$

To subtract the nominal states from both sides we rewrite the right-hand side of Eq. (C.35) as

$$\frac{1}{x_0 + \Delta x} = \frac{1}{x_0}\frac{1}{(1 + \frac{\Delta x}{x_0})} \tag{C.36}$$

To further process the above equation, the term $\frac{1}{(1+\frac{\Delta x}{x_0})}$ is approximated by a Taylor series:

$$\frac{1}{(1 + \frac{\Delta x}{x_0})} = 1 - \frac{\Delta x}{x_0} + \left(\frac{\Delta x}{x_0}\right)^2 + \ldots \tag{C.37}$$

Since $\frac{\Delta x}{x_0}$ is small, the second- and higher-order terms can be ignored, so Eq. (C.35) becomes

$$\frac{1}{x_0 + \Delta x} = \frac{1}{x_0}\left(1 - \frac{\Delta x}{x_0}\right) \tag{C.38}$$

Subtracting the nominal state from both sides finally gives

$$\Delta\dot{x} = \frac{1}{x_0 + \Delta x} - \frac{1}{x_0} = \frac{1}{x_0}\left(1 - \frac{\Delta x}{x_0}\right) - \frac{1}{x_0} = -\frac{1}{x_0^2} \tag{C.39}$$

■

Example C.4 As a continuation of the previous example we consider the following differential equation:

$$\dot{x} = \frac{1}{\cos x} \tag{C.40}$$

Substituting $x = x_0 + \Delta x$ gives us:

$$\dot{x}_0 + \Delta\dot{x} = \frac{1}{\cos(x_0 + \Delta x)} \tag{C.41}$$

For $\cos(x_0 + \Delta x)$ we can write, using the angle sum rule for the cosine function:

$$\cos(x_0 + \Delta x) = \cos x_0 \cos \Delta x - \sin x_0 \sin \Delta x \approx \cos x_0 - \sin x_0 \Delta x \tag{C.42}$$

where we have used the fact that Δx is small, so $\cos \Delta x \approx 1$ and $\sin \Delta x \approx \Delta x$. Equation (C.42) substituted in Eq. (C.41) yields

$$\dot{x}_0 + \Delta\dot{x} = \frac{1}{\cos x_0 - \sin x_0 \Delta x} \tag{C.43}$$

This equation can be rewritten as

$$\dot{x}_0 + \Delta\dot{x} = \frac{1}{\cos x_0}\left(\frac{1}{1 - \tan x_0 \Delta x}\right) \tag{C.44}$$

As in the previous example, realising that $\tan x_0$ is a constant and Δx is small, we can approximate the second term on the right-hand side with a Taylor series, retaining only the first-order term:

$$\frac{1}{1 - \tan x_0 \Delta x} \approx 1 + \tan x_0 \Delta x$$

which gives us for Eq. (C.44)

$$\dot{x}_0 + \Delta\dot{x} = \frac{1}{\cos x_0}(1 + \tan x_0 \Delta x) \tag{C.45}$$

Finally, subtracting the nominal function from both sides results in

$$\Delta\dot{x} = \frac{1}{\cos x_0}(1 + \tan x_0 \Delta x) - \frac{1}{\cos x_0} = \frac{\tan x_0}{\cos x_0}\Delta x = \frac{\sin x_0}{\cos^2 x_0}\Delta x \tag{C.46}$$

■

From this last example, it is clear that this way of linearising can be quite tedious. It is therefore advised to use the direct way of linearising according to Eq. (C.23). For instance: with $\Delta\dot{x} = A\Delta x$ and $A = \frac{\mathrm{d}\dot{x}}{\mathrm{d}x}\big|_0$ it easily follows for the function of Example C.4 that:

$$\left.\frac{\mathrm{d}\dot{x}}{\mathrm{d}x}\right|_0 = \frac{\mathrm{d}\left(\frac{1}{\cos x}\right)}{\mathrm{d}x} = -\frac{1}{\cos^2 x_0}(-\sin x_0) = \frac{\sin x_0}{\cos^2 x_0} \tag{C.47}$$

To conclude this section we will have another go at Example C.2 to review the steps for the more elaborate manner.

Example C.5 As before, the kinematic equation for altitude motion is given by

$$\dot{R} = V \sin\gamma \tag{C.48}$$

Substituting $V = V_0 + \Delta V$, $\gamma = \gamma_0 + \Delta\gamma$ and $R = R_0 + \Delta R$ we obtain:

$$\dot{R}_0 + \Delta\dot{R} = (V_0 + \Delta V)\sin(\gamma_0 + \Delta\gamma) \tag{C.49}$$

Using the double-angle formula, $\sin(\gamma_0 + \Delta\gamma)$ expands to

$$\sin(\gamma_0 + \Delta\gamma) = \sin\gamma_0 \cos\Delta\gamma + \cos\gamma_0 \sin\Delta\gamma \tag{C.50}$$

Since $\Delta\gamma$ is small, $\sin\Delta\gamma \approx \Delta\gamma$ and $\cos\Delta\gamma \approx 1$, and the above equation reduces to:

$$\sin(\gamma_0 + \Delta\gamma) = \sin\gamma_0 + \cos\gamma_0 \Delta\gamma \tag{C.51}$$

With this result, Eq. (C.49) can be written as:

$$\dot{R}_0 + \Delta\dot{R} = V_0 \sin\gamma_0 + V_0 \cos\gamma_0 \Delta\gamma + \sin\gamma_0 \Delta V + \cos\gamma_0 \Delta V \Delta\gamma \tag{C.52}$$

The last term on the right-hand side, i.e., $\cos\gamma_0 \Delta V \Delta\gamma$, is a second-order term and can be neglected compared to the others, so we obtain:

$$\dot{R}_0 + \Delta\dot{R} = V_0 \sin\gamma_0 + V_0 \cos\gamma_0 \Delta\gamma + \sin\gamma_0 \Delta V \tag{C.53}$$

Finally, subtracting the nominal state, i.e., $\dot{R}_0 = V_0 \sin\gamma_0$, from both sides of Eq. (C.53) yields:

$$\Delta\dot{R} = V_0 \cos\gamma_0 \Delta\gamma + \sin\gamma_0 \Delta V \tag{C.54}$$

which is the result we have been looking for, but with a lot more effort compared to the method shown in Example C.2. ■

C.4 Interpolation

When building simulation models to study atmospheric flight, input data are quite often given in tabular form. Also atmosphere models are sometimes only available as data tables. During simulation the actual value of the input or model parameters has to be obtained through interpolation. When the data are sparse, e.g., in case of aerodynamic coefficients, it is common practice to use linear interpolation (Sect. C.4.1), even though this may give a rather ragged-looking result. When there are sufficient data available one would expect that linear interpolation would get more accurate, though. In general this is the case, but there are situations that smoother interpolation is required to avoid discontinuities in the first derivative at the nodal points.

One such case is the atmospheric density obtained from an altitude-dependent data table. Even a small discontinuity in the first derivative will give a noticeable "jump" in the acceleration, as the density discontinuity translates into a discontinuity in the aerodynamic force, and thus the corresponding acceleration. Therefore, a different interpolation method is required. From the available methods, cubic-spline interpolation can be chosen as a fast and easy-to-implement method. This method will be discussed in Sect. C.4.2, because it is quite straightforward and works well for one-dimensional data sets with many data points. Multi-variate cubic-spline interpolation exists as well, but is a lot more complicated and forms a special topic, well beyond the scope of this appendix.

In the case the input data set is sparse, (cubic-)spline interpolation does not always produce correct results. One could try and fit another type of function through the data, for instance, with least-squares fitting, but fortunately there is also a simpler alternative: Hermite polynomials, which will be discussed in Sect. C.4.3.

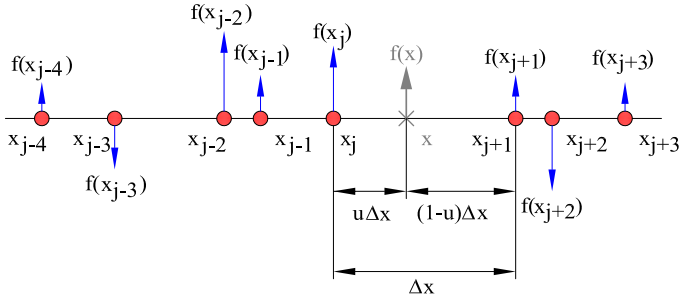


Fig. C.3 Linear interpolation in one dimension

C.4.1 Linear Interpolation

Suppose we know the values of a function $f(x)$ at a set of n data points $x_1, x_2, \ldots, x_n$ (with $x_1 < x_2 < \ldots < x_n$). If we want to know the value of $f(x)$ at an arbitrary point x and we do not have an analytic expression of $f(x)$ at hand, then we can use the coordinate pairs $[x_i, f(x_i)]$ to approximate $f(x)$ at x. If x lies within the interval $[x_1,x_n]$, this approximation is called interpolation and if it is outside this interval we speak of extrapolation.

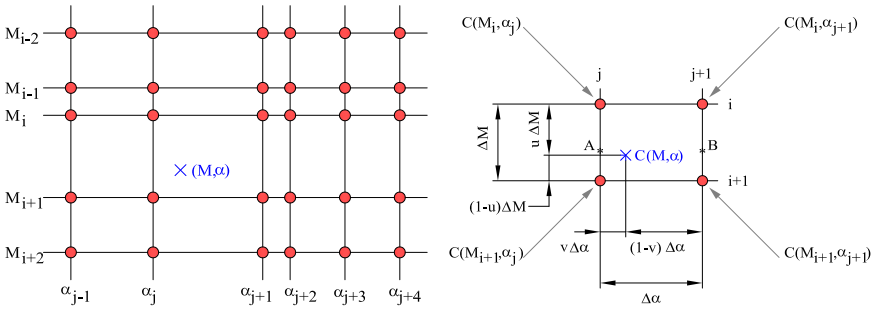
If we use k data points to calculate $f(x)$ we have a $(k-1)$th-order interpolation. In our case we will only use two data points and is therefore called linear interpolation. Lagrange's formula gives for linear interpolation in one dimension (Fig. C.3):

$$f(x) = \frac{x - x_{j+1}}{x_j - x_{j+1}} f(x_j) + \frac{x - x_j}{x_{j+1} - x_j} f(x_{j+1}) = (1 - v) f(x_j) + v f(x_{j+1}) \tag{C.55}$$

To further the discussion, as an example we will consider an aerodynamic coefficient C depending on two variables, e.g., the Mach number M and the angle of attack α. This dependency is given in the form of tabulated data: for every point (M_i,α_j) the value of C is given. The independent variables have to be ordered (but do not have to be equally spaced), so that the interpolation process can be executed. Thus, $M_1 < M_2 < \ldots < M_m$ and $\alpha_1 < \alpha_2 < \ldots < \alpha_n$. In Fig. C.4a the tabulated data are visualised as the points in a rectangular grid.

The problem to solve is: how do we calculate the value of C in an arbitrary point (M, α), with $M_i \le M \le M_{i+1}$ and $\alpha_j \le \alpha \le \alpha_{j+1}$. If we use linear interpolation in both directions (bilinear interpolation), we get for the M-direction (see also Fig. C.4b):

$$\begin{aligned} C_A &= \tfrac{M-M_{i+1}}{M_i-M_{i+1}} C(M_i, \alpha_j) + \tfrac{M-M_i}{M_{i+1}-M_i} C(M_{i+1}, \alpha_j) \\ &= (1-u) C(M_i, \alpha_j) + u C(M_{i+1}, \alpha_j) \end{aligned} \tag{C.56}$$



(a) The grid. (b) Related parameters.

Fig. C.4 Bilinear interpolation

$$\begin{aligned} C_B &= \tfrac{M-M_{i+1}}{M_i-M_{i+1}} C(M_i, \alpha_{j+1}) + \tfrac{M-M_i}{M_{i+1}-M_i} C(M_{i+1}, \alpha_{j+1}) \\ &= (1-u)C(M_i, \alpha_{j+1}) + uC(M_{i+1}, \alpha_{j+1}) \end{aligned} \tag{C.57}$$

And, if we apply Eq. (C.55) for the α-direction using Eqs. (C.56)–(C.57), we get:

$$\begin{aligned} C(M, \alpha) &= \tfrac{\alpha-\alpha_{j+1}}{\alpha_j-\alpha_{j+1}} C_A + \tfrac{\alpha-\alpha_j}{\alpha_{j+1}-\alpha_j} C_B \\ &= (1-v)C_A + vC_B \\ &= (1-u)(1-v)C(M_i, \alpha_j) + u(1-v)C(M_{i+1}, \alpha_j) + \\ &+ (1-u)vC(M_i, \alpha_{j+1}) + uvC(M_{i+1}, \alpha_{j+1}) \end{aligned} \tag{C.58}$$

With Eq. (C.58) we can calculate C for every value of M and α as defined on the grid. This also includes the boundary of the grid. If one (or both) of the values of M or α is outside the grid, i.e., one of them is larger or smaller than any of the tabulated data, there are three possibilities:

1. C is not defined; if the interpolation algorithm is used in a computer program, an error flag has to be set in order to warn the user that computations cannot proceed.
2. The boundary value of M (or α) is used, so C can be calculated by means of interpolation if necessary.
3. C will have to be computed by means of extrapolation. Eq. (3.3.34) also applies in this case. For the values of M_i and M_{i+1} (or the corresponding values of α) we use the boundary value and the one just preceding (succeeding) it.

Equation (C.58) easily extends to higher dimensions by following the same steps as shown for the 2-dimensional case.

C.4.2 Cubic Spline Interpolation

Spline functions yield smooth interpolating curves, which are less likely to exhibit the large oscillations, characteristic of high degree polynomials (Stoer and Bulirsch, 1980). To interpolate in one dimensional functions, e.g., atmospheric data tables that require continuity in the first derivatives, cubic spline interpolation can be used. It is necessary that we have, apart from the function values f, also the second derivatives f'' at our disposal.

Suppose that the atmospheric variables temperature, T, pressure, p, and density, ρ, are available in tabular form, as a function of a discrete number of altitude values. Based on these function values, it is possible to compute the second derivatives if the continuity of the first derivative is required (Press et al. 2007). The resulting tables (of T, p and ρ and their second derivatives) form the tabulated atmosphere models discussed before.

To compute T, p, and ρ at an arbitrary altitude h, we proceed as follows. Consider each of the atmosphere variables as a tabulated function $f_i = f(x_i)$, with $i = 1 \ldots n$. For a particular interval between x_j and x_{j+1} we get the following interpolation formula, if we apply linear interpolation (see also Sect. C.4.1):

$$f(x) = \frac{x_{j+1} - x}{x_{j+1} - x_j} f(x_j) + \frac{x - x_j}{x_{j+1} - x_j} f(x_{j+1}) = Af(x_j) + Bf(x_{j+1}) \quad \text{(C.59)}$$

with

$$A = \frac{x_{j+1} - x}{x_{j+1} - x_j} \quad \text{(C.60)}$$

$$B = 1 - A = \frac{x - x_j}{x_{j+1} - x_j} \quad \text{(C.61)}$$

The cubic spline interpolation formula is, using the same notation, given by (Press et al. 2007):

$$f(x) = Af(x_j) + Bf(x_{j+1}) + Cf''(x_j) + Df''(x_{j+1}) \quad \text{(C.62)}$$

with

$$C = \frac{1}{6}(A^3 - A)(x_{j+1} - x_j)^2 \quad \text{(C.63)}$$

$$D = \frac{1}{6}(B^3 - B)(x_{j+1} - x_j)^2 \quad \text{(C.64)}$$

From the definition of C and D we see the term "cubic" arise: A^3 (or B^3) $\sim x^3$.

C.4.3 Cubic Hermite Splines

One of the main problems of using cubic-spline interpolation is that when applied to relatively sparse data sets, the resulting polynomial may oscillate in between the nodes. For one dimensional problems, such as the interpolation in a vector with guidance or control nodes, or aerodynamic coefficients as a function of a single variable, a good alternative is found in (cubic) Hermite splines or polynomials. Not an interpolation scheme per se,[10] cubic Hermite polynomials can be constructed for a data interval, consisting of two nodes x_i and x_{i+1}, if besides the function values also the first derivative in the nodes are available. In case these are not readily available, they can be constructed using two additional nodes, x_{i-1} and x_{i+2}. Typically, a central difference scheme is used, apart from the first and last node of the data set, where either forward (right derivative) or backward (left derivative) differencing.

The polynomial thus constructed is called an *interpolating polynomial*, and joining all interval polynomials together is an example of a piecewise-continuous polynomial. The resulting process is known as piecewise cubic Hermite interpolation, which is continuous over the full data set, because of the use of the first derivatives. Formally, this is called C_0 and C_1 continuity.

For the Hermite interpolation that satisfies:

$$p(x_j) = y_j \qquad p'(x_j) = d_j \qquad p(x_{j+1}) = y_{j+1} \qquad p'(x_{j+1}) = d_{j+1} \tag{C.65}$$

the polynomial is defined by (Moler 2008):

$$\begin{aligned} p(x) = \; & y_j \left[1 + \frac{2}{x_j - x_{j+1}}(x_j - x)\right]\left(\frac{x - x_{j+1}}{x_j - x_{j+1}}\right)^2 + \\ & + y_{j+1}\left[1 + \frac{2}{x_{j+1} - x_j}(x_{j+1} - x)\right]\left(\frac{x - x_j}{x_{j+1} - x_j}\right)^2 \\ & + d_j(x - x_j)\left(\frac{x - x_{j+1}}{x_j - x_{j+1}}\right)^2 + d_{j+1}(x - x_{j+1})\left(\frac{x - x_j}{x_{j+1} - x_j}\right)^2 \end{aligned} \tag{C.66}$$

The shape of the polynomial relies on the estimates of the derivatives, d_j and d_{j+1}, at the endpoints x_j and x_{j+1}. Unfortunately, these are unknown. Suitable derivative values, such that the Hermite interpolant has a desirable shape, could be determined by trial and error or finite differences. An alternative method to estimate the derivatives is provided by (Moler 2008), which is used in Matlab, giving shape-preserving "visually pleasing" interpolants with no overshoots.

[10] Hermite interpolation does exist, but its discussion is beyond the scope of this text. This form of interpolation can be univariate, bivariate, and trivariate, and is often used in computer graphics.

Table C.1 Sparse data set

x	y
0.00	0.00
0.08	0.30
0.25	0.57
0.45	0.02
1.00	0.10

Let d_k be the slope at point k, if the slope left of k and right of k, δ_{k-1} and δ_k respectively, have opposite signs, or one of them is equal to zero then $d_k = 0$. If δ_{k-1} and δ_k have similar sign then d_k is given by the weighted harmonic-mean of the slopes:

$$\frac{w_1 + w_2}{d_k} = \frac{w_1}{\delta_{k-1}} + \frac{w_2}{\delta_k} \tag{C.67}$$

with w_1 and w_2 being the weights. These are required when the two intervals have different lengths and computed by:

$$w_1 = 2h_k + h_{k-1} \tag{C.68}$$

$$w_2 = h_k + 2h_{k-1} \tag{C.69}$$

with h being the length of the interval.

Example C.6 (*Interpolation*) Consider a sparse data set, listed in Table C.1, which could be one (of many) guidance profile(s), used in a trajectory-optimisation problem. Suppose that the x-coordinate is normalised time, and during the simulation an output as a function of time has to be provided. One can choose from several options, i.e.,

1. Per interval, keep the output constant, and allow for discrete changes at the interval boundaries,
2. Linear interpolation,
3. Cubic-spline interpolation, and
4. Cubic-Hermite polynomials.

In Fig. C.5, the four different methods have been plotted. Even though linear interpolation seems to give a reasonable result, it is clear that there are large changes in the first derivatives that may adversely affect the stability of the guidance response. The cubic-spline solution is smooth, but the sparsity of the input grid results in a very large overshoot. In a real-life situation this might even lead to infeasible solutions. The best option to use is the Hermite polynomial: continuous and smooth everywhere, without any overshoots. Finally, in Fig. C.6 the effect of using either central differencing to calculate the derivatives, or the weighted harmonic-mean of the slopes, Eq. (C.67). Using a central-difference scheme introduces an unwanted oscillation that is gone when using the weighted harmonic-mean of the slopes. ■

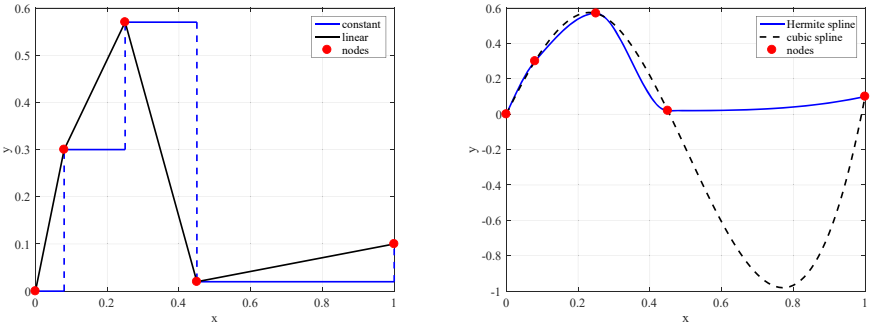
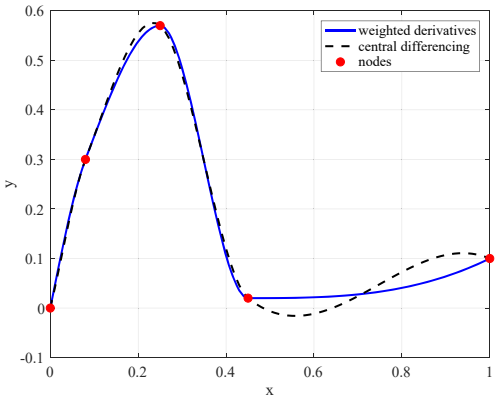


Fig. C.5 Different interpolation schemes

Fig. C.6 Hermite polynomials: derivative computation with central-difference scheme and a weighted harmonic-mean of the slopes



C.5 Numerical Integration

An ordinary differential equation of first order is defined as

$$y' = f(x, y(x)) \tag{C.70}$$

with $y = y(x)$ is a differentiable function, whose derivative $y'(x)$ satisfies some function f, and x is the independent variable.

In general, there are infinitely many different functions y, which are solutions of Eq. (C.70). However, if we impose an additional requirement in the form of an initial condition, we can single out one solution. Thus, an initial value problem can be defined by Eq. (C.70), satisfying an initial condition of the form

$$y(x_0) = y_0 \tag{C.71}$$

In the case the problem at hand is defined by more than one differential equation we can generalise for a system of n ordinary differential equations of first order:

$$\begin{aligned} y_1' &= f_1(x, y_1(x), \ldots, y_n(x)) \\ y_2' &= f_2(x, y_1(x), \ldots, y_n(x)) \\ &\vdots \\ y_n' &= f_n(x, y_1(x), \ldots, y_n(x)) \end{aligned} \tag{C.72}$$

If an initial value problem is considered, we also need n initial conditions, i.e., $y_{1_0}, y_{2_0}, \ldots, y_{n_0}$. Written in vector form we get:

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}) \quad \text{with} \quad \mathbf{y}(x_0) = \mathbf{y}_0 \tag{C.73}$$

where

$$\mathbf{y}' = \left(\mathbf{y_1'}, \ldots, \mathbf{y_n'}\right)^{\mathrm{T}} \quad \mathbf{f}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} f_1(x, y_1, \cdots, y_n) \\ \vdots \\ f_n(x, y_1, \cdots, y_n) \end{pmatrix} \quad \mathbf{y_0} = \left(\mathbf{y_{1_0}}, \ldots, \mathbf{y_{n_0}}\right)^{\mathrm{T}}$$

Newton's second law relates force to acceleration. With acceleration being the second (time) derivative of the position—velocity is the first derivative—the corresponding differential equation is of second order. In case of three-dimensional motion, we would have a system of three second-order equations. However, such a system can always be transformed into an equivalent system of (in this case six) first-order differential equations.

Considering the general form of an ordinary differential equation of mth order,

$$z^{(m)}(x) = f(x, z(x), z'(x), \ldots, z^{(m-1)}(x)) \tag{C.74}$$

we introduce the following auxiliary functions

$$\begin{aligned} y_1(x) &= z(x) \\ y_2(x) &= z'(x) \\ &\vdots \\ y_m(x) &= z^{(m-1)}(x) \end{aligned} \tag{C.75}$$

which gives the following system of first order differential equations:

$$\left(\mathbf{y}'\right) = \begin{pmatrix} y'_1 \\ \vdots \\ y'_{m-1} \\ y'_m \end{pmatrix} = \begin{pmatrix} y_2 \\ \vdots \\ y_m \\ f(x, y_1, y_2, \cdots, y_m) \end{pmatrix} \tag{C.76}$$

Given these equations, how can we solve these? The methods and results for initial-value problems for systems of ordinary differential equations of first order are essentially independent of the number n of unknown functions (Stoer and Bulirsch 1980). In the following sections, we will discuss the methods of numerical integration limited to the case of *one* ordinary differential equation of first order for only *one* unknown function. The results are applicable to systems of n ordinary differential equations of first order, with every scalar replaced by a vector in that case.

Both the equations of translational motion and the equations of rotational motion are ordinary first order differential equations. Doing a three-degrees-of-freedom (d.o.f.) simulation, a system of six coupled differential equations has to be solved, and for a six-d.o.f. simulation we have to find a solution for a system of (at least) twelve coupled equations. Due to the complexity of the equations considered here, no analytical solution exists. It is necessary to use a numerical integration method to do a trajectory simulation.

Some of the most important characteristics of the set of problems are:

- large accelerations (entry of ballistic vehicles),
- large initial velocities (atmospheric capture of vehicles coming from parabolic or hyperbolic orbits, e.g., Apollo),
- large flight time (descent in a dense atmosphere (Titan), parachute descent (Huygens probe)),
- fast change of gas-dynamic forces and moments because of the change in atmospheric density.
- high frequency oscillations of uncontrolled re-entry vehicles,
- physical sensitivity to initial conditions, and
- large variation in flow regime.

According to these problems, some of the required characteristics of the integration method must be pointed out.

Stability

The stability can be characterised in terms of the rate of error throughout the computation. So a selected integration method has to be stable enough with respect to the characteristics mentioned above.

Accuracy

The accuracy of algorithms can be identified by the error introduced in computing one point of the solution, including:

- local truncation error, which decreases with the order of the algorithm and increases with the step size.
- round-off errors, which increase with the order and complexity of the algorithms.

Speed

The speed of an algorithm can be defined by the time required to compute one point of the solution relative to real time. The speed requirement must be directly linked with the accuracy requirement, since it is always possible to speed up a simulation by increasing the step size, which may result in far less accurate results.

For the numerical integration of ordinary differential equations, a broad variety of integrators is available. However, not all of them are suitable to solve the problem stated above. Here, we will only discuss a two integration methods, i.e., the fixed-step, fourth-order Runge-Kutta method (RK4) and the variable-step, fourth-order Runge-Kutta-Fehlberg with fifth-order step size control (RKF45), to illustrate the process of numerically solving first-order ordinary differential equations. Higher-order variable step-size methods, developed by Fehlberg (1968), are extensions of the lower-order methods and need not be discussed. However, depending on the specific application these methods are worthwhile considering.

The RK4 integration method is probably the most well-known for engineering problems and has proven to be very reliable. To maintain a reasonable high accuracy, a very small step size is required in regions where the solution fluctuates. This is a drawback of any fixed-step method, since in regions where the solution is a very smooth function, one could do with a larger step size, so the method is here unnecessarily slow.

We consider the initial value problem of Eq. (C.70), but replacing x with time t being the independent variable, because our equations of motion are formulated as such. To solve this equation for the interval $[a, b]$ we divide this interval into n sub-intervals with fixed size h, so that $h = (b-a)/n = t_{i+1} - t_i$, with $i = 0, 1, \ldots, n$, $t_0 = a$ and $t_n = b$. t_{i+1} and t_i are the boundaries of the $(i+1)$th sub-interval. For the numerical solution $\hat{y}_{i+1}$ of $y(t_{i+1})$ we can now write, using the fourth-order method of Runge-Kutta (Stoer and Bulirsch 1980; Press et al. 2007):

$$\begin{aligned} k_1 &= h \cdot f(t_i, \hat{y}_i) \\ k_2 &= h \cdot f(t_i + \frac{1}{2}h, \hat{y}_i + \frac{1}{2}k_1) \\ k_3 &= h \cdot f(t_i + \frac{1}{2}h, \hat{y}_i + \frac{1}{2}k_2) \\ k_4 &= h \cdot f(t_i + h, \hat{y}_i + k_3) \\ \hat{y}_{i+1} &= \hat{y}_i + (k_1 + 2k_2 + 2k_3 + k_4)/6 \end{aligned} \tag{C.77}$$

The coefficients k_j $(j = 1, \cdots, 4)$ are the so-called predictor values of $\hat{y}$.

The method of Runge-Kutta is an explicit one step method, requiring four function evaluations per step. Explicit means in this case, that for step $i+1$ (to solve $\hat{y}_{i+1}$) only $\hat{y}_i$ appears on the right-hand side of Eq. (C.77). Implicit schemes, on the contrary, also include $\hat{y}_{i+1}$, so an iteration process is needed to evaluate $\hat{y}_{i+1}$. The local truncation error is found to be of fifth order relative to the time step.

Table C.2 Butcher tableau for integrator coefficients (Butcher 2008)

c_1					
c_2	a_{21}				
c_3	a_{31}	a_{32}			
$\vdots$	$\vdots$	$\vdots$	$\ddots$		
c_s	a_{s1}	a_{s2}	$\dots$	$a_{s,s-1}$	
	b_1	b_2	$\dots$	b_{s-1}	b_s

Table C.3 Butcher tableau for RK4

0				
1/2	1/2			
1/2	0	1/2		
1	0	0	1	
	1/6	1/3	1/3	1/6

It is common practise to summarise the related coefficients of the integration method in so-called Butcher tableaus (Butcher 2008), see Table C.2 for the generic layout. To give the Butcher tableau for RK4, Eq. (C.77) is rewritten in the proper format:

$$\begin{aligned} k_1 &= h \cdot f(t_i, \hat{y}_i) \\ k_2 &= h \cdot f(t_i + c_2 h, \hat{y}_i + a_{21} k_1) \\ k_3 &= h \cdot f(t_i + c_3 h, \hat{y}_i + a_{32} k_2) \\ k_4 &= h \cdot f(t_i + c_4 h, \hat{y}_i + a_{43} k_3) \\ \hat{y}_{i+1} &= \hat{y}_i + b_1 k_1 + b_2 k_2 + b_3 k_3 + b_4 k_4 \end{aligned} \tag{C.78}$$

which gives the Butcher tableau of Table C.3.

Using a fixed-step method can be quite constraining, as the dynamics of the system is not considered. The advantage of an adaptive-step-size algorithm is that it does take the shape of the derivative function into account. It can take smaller steps when required (increasing accuracy), but also larger steps whenever the dynamics allows it (increasing speed). The algorithm determines the value of Δt such that the truncation error is kept below a predefined value for each step; to do this, the truncation error must be estimated at each time step.

Adaptive-step-size Runge-Kutta algorithms are based on the embedded Runge-Kutta formulae, originally developed by Erwin Fehlberg. He discovered a fifth-order Runge-Kutta method requiring six function evaluations where another combination of the six functions results in a fourth-order method; this method is commonly referred to as the RKF45 method (Fehlberg 1969). At each step, it determines whether the proper Δt is being used by comparing two different approximations of the solution—obtained using the fifth-order method and the embedded fourth-order formula. The difference is used as an estimate of the truncation error. If the

two approximations are sufficiently close, the step size is accepted. If the two values do not agree to a specified accuracy, Δt will be reduced. On the other hand, when the solution agrees to more significant digits than necessary, Δt is increased (Moler 2008).

The following intermediate values are defined for each RKF45 step:

$$\begin{aligned}
k_1 &= h \cdot f(t_i, \hat{y}_i) \\
k_2 &= h \cdot f(t_i + c_2 h, \hat{y}_i + a_{21} k_1) \\
k_3 &= h \cdot f(t_i + c_3 h, \hat{y}_i + a_{31} k_1 + a_{32} k_2) \\
k_4 &= h \cdot f(t_i + c_4 h, \hat{y}_i + a_{41} k_1 + a_{42} k_2 + a_{43} k_3) \\
k_5 &= h \cdot f(t_i + c_5 h, \hat{y}_i + a_{51} k_1 + a_{52} k_2 + a_{53} k_3 + a_{54} k_4) \\
k_6 &= h \cdot f(t_i + c_6 h, \hat{y}_i + a_{61} k_1 + a_{62} k_2 + a_{63} k_3 + a_{64} k_4 + a_{65} k_5)
\end{aligned} \tag{C.79}$$

The fifth-order solution is given by

$$\hat{y}_{i+1} = \hat{y}_i + b_1 k_1 + b_2 k_2 + b_3 k_3 + b_4 k_4 + b_5 k_5 + b_6 k_6 + O(\Delta t^6) \tag{C.80}$$

and the embedded fourth-order formula is

$$\hat{y}^*_{i+1} = \hat{y}_i + b^*_1 k_1 + b^*_2 k_2 + b^*_3 k_3 + b^*_4 k_4 + b^*_5 k_5 + O(\Delta t^5) \tag{C.81}$$

The sets of constants a, c, b and b^* can have several values, depending on the assumptions that were made during the derivation of the formulae. The values, as derived by (Fehlberg 1969), are given in Table C.4.

Table C.4 Butcher tableau for RKF45 (Fehlberg 1969)

0							
1/4	1/4						
3/8	3/32	9/32					
12/13	1932/2197	-7200/2197	7296/21970				
1	439/216	-8	3680/513	-845/4104			
1/2	-8/27	2	-3544/2565	1859/4104	-11/40		
b	16/135	0	6656/12825	28561/56430	-9/50	2/55	
b^*	25/216	0	1408/2565	2197/4104	-1/5	0	

The truncation error estimate is defined as:

$$\varepsilon \equiv \hat{y}_{i+1} - \hat{y}^*_{i+1} = \sum_{i=1}^{6} \left(b_i - b_i^*\right) k_i \tag{C.82}$$

which scales with Δt^5. If a given step Δt produces an error ε, the step Δt^* that would have provided the desired accuracy ε^* can be estimated:

$$\Delta t^* = \Delta t \left| \frac{\varepsilon^*}{\varepsilon} \right|^{\frac{1}{5}} \tag{C.83}$$

The change in Δt to achieve the desired accuracy naturally follows.

C.6 Root Finding

Root finding is the generic process of finding the solution to

$$f(x) = 0 \tag{C.84}$$

in scalar form or $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ for an N-dimensional vector-valued function. For the sake of this appendix, a scalar function is considered.

Many methods exist for root finding (Press et al. 2007), with one of the main characteristics whether the derivative of the function is available or not. In this section, two methods will be described, one for each category. The preferred root-finding algorithm for functions without derivative information is recursive bisection, which is a robust method that for all practical purposes does not lead to divergence, which is not necessarily true for fixed-point iteration.

The concept behind bisection is simple, yet robust and reliable. Given an interval $[a, b]$, the bisection method can be used as shown in Algorithm 1. This method is based on Bolzano's theorem, which states that if $f(a)$ and $f(b)$ have opposite signs, there exists a $c \in [a, b]$ such that $f(c) = 0$. Note that $f(x)$ has to be continuous in all of the interval $[a, b]$.

Data: function $f(x)$, interval $[a, b]$ and accuracy ε
Result: root of $f(x)$ inside interval $[a, b]$

```
if f(a)f(b) ⩽ 0 then
    while abs(b − a) ⩾ ε do
        compute mid point: c = (a+b)/2;
        if f(a)f(c) ⩽ 0 then
            root is in [a, c];
            b ← c;
        else
            root is in [c, b];
            a ← c;
        end
    end
else
    no single root exists in [a, b]
end
```

Algorithm 1: Bisection method.

Given a prescribed accuracy ε,

$$n = \frac{\ln\left(\frac{\Delta}{\varepsilon}\right)}{\ln 2} \tag{C.85}$$

will give the number of iterations, n, that is required for convergence. Here, $\Delta = b - a$ is the size of the original interval.

A popular method for root finding of a (one-dimensional) function when derivative information is available is the method of Newton-Raphson. Given a function $f(x)$, its derivative $f'(x)$, and a starting value $x = x_0$, successive estimates of x can be found with

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{C.86}$$

This process can be continued until $\text{abs}(x_{n+1} - x_n) < \varepsilon$, a pre-defined tolerance, or when the maximum number of iterations has been reached. Putting a limit to the number of iterations will prevent infinite loops, as the Newton-Raphson method does not always converge.

In case no analytical formula is at hand for the derivative, one can still use the Newton-Raphson formula, but now with a numerical difference to approximate the true local derivative:

$$f'(x) \approx \frac{f(x + dx) - f(x)}{dx} \tag{C.87}$$

although this is not recommended. The derivative calculation requires an additional function evaluation, and the derivative is also sensitive to the selected dx, and will thus affect the algorithm's performance.

References

Butcher JC (2008) Numerical methods for ordinary differential equations, 2nd edn. Wiley

Fehlberg E (1968) Classical fifth-, sixth-, seventh-, and eighth-order Runge-Kutta formulas with stepsize control, NASA-TR-R-287, Washington, DC

Fehlberg E (1969) Low-order classical Runge-Kutta formulas with stepsize control and their application to some heat transfer problems, NASA-TR-R-315, Washington, DC

Moler C (2008) Numerical computing with MATLAB. The mathworks

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes: the art of scientific computing, 3rd edn. Cambridge University Press

Stoer J, Bulirsch R (1980) Introduction to numerical analysis. Springer, New York

Todhunter I (2006) Spherical trigonometry. For the use of colleges and schools, 1886. The Project Gutenberg

Van Brummelen G (2012) Heavenly mathematics: the forgotten art of spherical trigonometry. Princeton University Press

Index

A
θ-β-M equation, 211
θ-β-M plot, 211–212
Ablation modelling, 1181–1183
Ablation resistance, 1135, 1182, 1190–1191, 1247–1248
Ablative materials
  Carbon phenolic, 1191, 1192–1193, 1219–1220, 1248
  Phenolic Impregnated Carbon Ablator (PICA), 1105–1106, 1109–1111, 1137, 1181, 1183, 1191, 1220, 1248
  Phenolic-Nylon, 1107, 1190–1191, 1218–1219, 1247
  SLA-561V, 1105–1106, 1138, 1182
Ablator finite difference model, 1206–1213
Ablator model, 1203–1205
Absolute acceleration, 1470
Acceleration due to gravity, 1400
Active cooling, 1093, 1169, 1172–1175
  Enhanced radiation cooling, 1174–1175
  Nucleate boiling, 1173–1174
  Radiation shielding, 1172–1173
Active systems, 1168–1175
  Convective cooling, 1171–1172
  Film cooling, 1170–1171
  Transpiration cooling, 1168–1170
  Variants, 1172–1175
Active-cooling modeling, 1195–1199
Actuators
  Control surfaces, 636, 684–688
  Moving-mass systems, 636, 695
  Parachutes, 694
  Reaction control thrusters, 636, 688–690
  Retro rockets, 690–693
  Spin vanes, 696–697
Advanced alloys, 1091
Aero-assisted manoeuvres, 1259
  Aerobraking, 1261
  Aerocapture, 1284
  Aerogravity assist, 1324
Aerobraking guidance
  Atmospheric phase, 1282–1284
  Periapsis lowering manoeuvre, 1277–1279
  Periapsis corridor estimation, 1279–1281
  Walk-in, 1276
  Walk-out, 1276
Aerobraking, 1261–1284
  Concept, 1260–1261
  Guidance, 1275–1284
  Mechanical loads, 1269
  Missions, 1262–1268
  Theory, 1268–1275
  Thermal loads, 1269–1270
Aerocapture verification
  First-order method, 498–502
  Second-order method, 519–526
Aerocapture, 1284–1323
  Characteristic trajectories, 1288–1293
  Concepts, 1284–1285
  Entry corridor, 1312–1323
  Missions, 1285–1288
  Predictor-corrector guidance, 1293
Aerodynamic force, 170, 635
Aerodynamic heating, 283–286
Aerodynamic moment, 170
Aerodynamic reference frame, 326, 347
Aerodynamics, 158–159, 162
  Hypersonic, 226–230, 235

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

E. Mooij, *Re-entry Systems*, Springer Aerospace Technology,
https://doi.org/10.1007/978-3-031-62174-1

Rarefied flow, 211
Subsonic, 192–195
Supersonic, 158
Transonic, 194–195
Aerogravity assist Mars, 1375
Aerogravity assist, 1260, 1458–1459, 1463–1465
Atmospheric bending angle, 1347–1349
Concept, 1345–1346
Drag losses, 1356–1360
Equilibrium altitude, 1361–1362
Equivalent radius, 1355
Maximum bending angle, 1354
Total bending angle, 1349–1352
Aerogravity-assist guidance, 1365–1366, 1368
Air-data systems
Air-data probes, 651
Flush Air-Data Sensing (FADS), 651
Airbags
Landing phases, 1022
Operating principle, 1020–1022
Systems, 906, 1018–1025
Altitude at maximum deceleration, 1414
Altitude at maximum heat flux, 1416
Altitude-targeting entry corridor, 1313–1316
Altitude-velocity diagram, 481–484
Aluminium alloys, 1149–1150
2024-T4, 1261, 1230, 1232, 1235
Aluminium-Beryllium, 1160–1161, 1230, 1233, 1235
America Bomber project, 27
Angle of attack at maximum heat flux, 437
Angular momentum, 1475, 1477–1479
Angular rates, 334
Angular velocity, 329, 360
Aperiodic motion, 545
Apoapsis altitude, 1261–1262, 1268, 1271–1272, 1274
Apollo 10 Lunar Module, 1395
Apollo capsule, 1483–1489
Aerodynamics, 1484–1485
Axisymmetric body, 1485–1489
Mass and geometry, 1483
Apollo Lunar Module, 950–951, 1028–1034
Approach-face guidance, 1032
Breaking-face guidance, 1031
Powered descent, 1030
Soft landing, 1034
Surface scanning and guidance, 1032–1033
Velocity-nulling guidance, 1033
Apparent centripetal acceleration, 1470
Apparent forces, 323, 1473
Apparent moment, 1476
Apparent tangental acceleration, 1470
Apple-tree myth, 15
Asteroids,96–97
Asymmetrical motion
Coupling with symmetrical motion, 594–595
Atmosphere
Earth, 123–137
Mars, 137–145
Titan, 145–148
Venus, 148–151
Atmospheric density, 123, 1396
Atmospheric properties, 121
Atmospheric skip, 466
Attitude-control moments, 710
Attitude, 331–334

B

Ballistic entry, 1412–1413
Ballistic flight outside atmosphere, 1422
Ballistic flight, 383–403
Apollo entry, 393–395, 399–401
Altitude-velocity profile, 386–387
Ballistic parameter, 389
Deceleration, 390
Huygens entry, 403–411
Mars entry analysis, 411–422
Mars entry capsule, 412–413
Maximum heat flux, 396
Mechanical load, 389–393
Model verification, 393–395, 399–401, 403–411
Thermal load, 396–399
Trajectory shape, 385–386
Ballistic parameter, 389
Ballute, 958–962
Gemini escape system, 961
Low-density Supersonic-decelerator (LDSD), 960–962
Trailing-edge Isotensoid, 960
Bank angle, 332, 347, 1411, 1442–1444
Bank-angle control, 446–447
Influence on phugoid mode, 595
Bank reversal, 462–463, 751, 771, 1444
Battle of Kai-Fung-Fu, 4
Battle of Pollilur, 6
Bisection method, 1539
Blanchard, Jean-Pierre, 864

Body frame, 325, 333–334, 1477, 1479
Body reference frame, 325, 349–351
Boundary layer, 217–226
Boundary-layer effects, 1402
Boundary-layer transition, 166, 176, 1116–1117
Brahe, Tycho 12
Braun, Wernher von, 28–30
Bridging function, 276, 279, 282
Brillouin sphere, 106
Buran terminal-area guidance, 826–829

C
Carbon-Carbon, 1142, 1161, 1166, 1180, 1237–1238
Cartesian coordinates, 665, 677
Çelebi, Hezârfen Ahmed, 2
Çelebi, Lagari Hasan, 3
Central-field model, 1401
Ceramic materials,
  UHTC, 1217–1218, 1222, 1226
  ZrB.2SiC, 1217–1218, 1236
Chapman aerodynamic heating, 396
Characteristic motion, 533, 582, 613, 625
Charring ablator, 1185–1186
Chinese Rockets, 4–5
Circular velocity, 381, 399, 425, 1413
Classical mechanics, 322
Coatings, 1126, 1135–1137
  Ceramic, 1136
  High Efficiency Tantalum-based Composite (HETC), 1137
  Reaction Cured Glass (RCG), 1136
  Self-healing, 1136
  Toughened Unipiece Fibrous Insulation (TUFI), 1136
Cold-wall model, 1116
Complex eigenvalues, 540, 543–545
Composite materials, 1128, 1137–1138
Compressible flow, 288
Computational fluid dynamics (CFD), 176–182
Concorde, 199
Conduction, 1123–1124
Congreve Rockets, 6–7
Continuum flow, 167, 171–172
Contours of constant drag, 817–818
Control errors, 752–754
Control systems
  Adaptive control, 738–740, 1434
  Closed-loop control, 704–705
  Incremental Non-linear Dynamic Inversion (INDI), 735–738
  Optimal control, 731–735
  Output feedback, 706
  Performance metrics, 710–711
  PID control, 708–710, 1447
  State feedback, 705
Controllability, 705
Convection, 1118–1119
Convective cooling, 1158, 1171, 1181
Convective heat transfer, 292–295, 1401, 1422, 1457
Coordinate transformations, 353–358
Coordinated turn, 819,
  Equilibrium conditions, 818, 1450
  Horizontal centrifugal force, 819
  Load factor, 819
  Turn radius, 820, 1450–1451
Coriolis acceleration, 1470
Coriolis effect, 381–383
Coriolis force, 1406
Coriolis moment, 1476
Corrective burn, 1261
Corrective control, 755
Corrosion resistance, 1129–1130
Coupled mass-spring system, 549
Critical damping, 547
Cross-range, 451–453, 456–458
Crushable material
  Crushable impact ring, 1007
  Deceleration loads, 1008–1009
  Genesis mission, 1006
  Impact duration, 1007
Cubic-Hermite polynomials, 1529–1530
Cubic-spline interpolation, 1528

D
Daedalus, 1
Damping coefficient
  Damping ratio, 547, 1433
  Negative damping, 595, 617
Damping ratio, 546–547, 1433
Deceleration, 390, 433–434, 490–492, 1425–1427
Deceleration due to drag, 1402
Deployment dynamic pressure, 892–894, 963, 1037
Derivative approximation, 1539
Derivative control, 712–713
Descent angle, 427–428
Descent systems
  Airbag descent, 1018–1020
  Hazard detection, 972–979
  Parachute systems, 873–877

Powered descent, 950–957, 1453–1454
Shock attenuation, 1001–1025
Digital elevation model (DEM), 88–96
Direct-entry corridor, 1317–1319
Disk-gap-band parachutes
Mars Exploration Rovers (MER), 1052–1053
Mars Science Laboratory (MSL), 1056–1057
Stardust, 1046
Viking lander, 1036
Displacement thickness, 220, 223, 231
Drag coefficient, 168, 255–258
Drag force, 169
Dynamic pressure, 169
Dynamic pressure modulation, 824

E

E-guidance, 941–940
Closed-form solution, 943
ESA Lunar Lander, 948–950
Surveyor, 946–947
Time-to-go, 946
Early rocketry, 3–6
Earth
Atmosphere, 123–137
Gravity field, 105, 108
Shape, 87
Earth re-entry heat loads, 1414
Eigenfrequency, 542–543, 547
Eigenmotion, 1430
Aperiodic, 545
Damped, 547
Eigenvalue influence, *see* characteristic motion
Oscillatory, 545, 547
Eigenvalues, 538–540, 1427, 1430–1431
Analytical, 602, 604
Emissivity, 1113, 1119, 1172–1173, 1212, 1234
Coatings, 1142, 1144, 1146
Energy dissipation, 811, 814, 850
Energy height
Definition, 804–805
Rate of energy dissipation, 814
Energy thickness, 226
Energy-state approximation, 814
Energy dissipation, 814
Maximum-range glide path, 814–815
Entry corridor, 1311–1313
Altitude targeting, 1313–1316
Direct entry, 1317–1319
Heat-flux corridor, 1270
Periapsis heat-flux, 1268
Thermal constraints, 1269–1270
Entry corridor, 440–444, 765
Entry mechanisms, 376
Entry velocity profile, 174–175
Equations of planar motion, 381
Equations of rotational motion, 1476
Equations of translational motion, 1474
Equatorial radius, 86
Equilibrium conditions, 555
Equilibrium temperature, 301
Equilibrium trajectory, 555
Equilibrium-glide constraint, 442–443
Equivalent airspeed, 812
ESA Mars Climate Database, 415–422
Escape velocity, 466
Euler angles, 331, 340
Euler axis, 338, 340
Euler equations, 1481
Euler transformation, 332
Euler's Rotation Theorem, 337–340, 1406
Expansion wave, 212–216
Exponential atmosphere model, 122–123, 1396
External force, 1471
External forces, 321, 322

F

Fibre materials, 1134–1135, 1243–1247
Fibrous insulators, 1097–1098, 1132, 1136–1137
Filtering techniques
Extended Kalman Filter (EKF), 657–659
Linear Kalman Filter, 652–656
Recursive filters, 669
Finite element method, 177–182
Flat earth approximation, 805–807
Flight duration, 431–432
Flight dynamics, 322, 327
Flight kinematics, 322
Flight modes
Lateral oscillation, 594–595
Longitudinal oscillation, 594–595
Phugoid, 595
Flight range, 427
Flight-path angle, 330–331, 355–356
Flight-path angle error, 773–774
Flow complications, 234–235
Flow regimes, 172–174
Footprint, 446–448
Forebody wake, 902–903

Fourth-order Runge-Kutta, 1534–1535
Frame transformation, 335, 349, 1407–1408

Free-molecular flow, 271–283
French space program, 28
Friction coefficient, 547
Friction drag, 1403

G
Gagarin, Yuri, 36–37
Gain scheduling, 744–745, 749–750
Gain tuning, 710
Galileo mission, 1039–1044
Galileo, 11, 12
Gas chemistry, 248
Geoid, 106
Gimbal lock, 1410
Glide range, 428–430, 1425
Gliding flight, 422–464
  Cross-range, 451–453, 456–458
  Descent angle, 427–428
  Downrange, 453–454, 456–458
  Entry corridor, 440–444
  Equilibrium glide, 424–426
  Flight duration, 431–432
  Footprint, 446–448
  Glide range, 428–430
  Lift parameter, 425
  Lift-to-drag ratio, 430
  Maximum deceleration, 433–434
  Maximum heat flux, 434–438
  Trajectory constraints, 442–444
Global Positioning System (GPS)
  Carrier face difference, 647–648
  Clock bias, 646, 668, 678
  Observables, 645–648
  Pseudo range, 645
  Range rate, 646
  Signal errors, 644–645
Goddard, Robert, 22–24
GRAM atmosphere model
  Earth, 125–126
  Mars, 138, 141, 144
  Venus, 149
Graphite epoxy fabric, 1234–1135, 1137–1138
Gravitational characteristics, 101
Gravitational force, 100, 322–323
Gravity assist geometry, 1328
Gravity assist versus aerogravity assist, 1384
Gravity assist, 1462–1463
  Concept, 1324–1325
  Conservation of momentum, 1328
  Energy balance, 1332
  Maximum deflection angle, 1331
  Periapsis radius, 1331
  Periapsis velocity, 1130
  Velocity-vector diagram, 1329
Gravity field
  Earth, 105, 108
  Mars, 106
  Moon, 107
  Venus, 106
Gravity field, 1399
Gravity potential, 101–102
Gravity turn, 927–941
  Equation of motion, 928–930
  Flat moon approximation, 933
  Normalized thrust deceleration, 932
  Planar thrusted motion, 928
  Spherical body, 934–935, 940
Guidance systems
  Closed loop, 1434
  Commands, 637
  Lateral guidance, 768–773, 1434
  Lift and drag control, 635
  Steering commands, 633–634, 780
  Space Shuttle guidance, 759–766
  Tracking, 1445–1447
  Trajectory tracking, 766–768
Gunpowder rockets, 3–4

H
HAC turn, 849–851
  Energy dissipation during turn, 850
  Energy-tube concept, 846–849
  HORUS terminal-area flight, 829–853
  Optimal maximum-dive trajectory, 844–846
  Optimal trajectory, 850
  Radius influence, 850, 852
Hale, William, 7
Harmony coefficients, 102–103
Hazard Detection and Avoidance (HDA), 986–990
Hazard detection systems, 990–991
Heading, 330, 355
Heading Alignment Cylinder (HAC), 768–769, 801, 823, 827–828, 830
Heading-error dead band, 769–770
Heat balance, 307

Heat flux indicator, 298–299, 1269
Heat flux tracking, 774–790
Heat flux, 292, 298, 300, 1112–1116, 1119
Heat transfer, 291
Heat-flux constraint, 443–444
Height mode, 594–595, 614, 615–618
Heliocentric model, 10–11
Hermite-spline interpolation, 1398
High gate, 1031–1032
Hiten aerobraking, 1262
Hohmann transfer, 1337–1444, 1459–1462
  Concept, 1338
  ΔV, 1338, 1343
  Interplanetary flight, 1341
  Transfer time, 1339, 1343
Hooke, Robert, 16–17
Horizontal centrifugal force, 819
Horizontal range, 484
HORUS attitude controller, 741–756
  Control law, 743, 746
  Entry control modes, 742
  Lateral control, 745–750
  Longitudinal control, 741–745
HORUS navigation system, 664–683, 1433
  Blackout phase, 681
  Estimated position velocity, 676–677
  GPS and IMU coupling, 663–664, 669
  Integrated navigation, 662
  Loose coupling, 662–663
  Measurement equation, 641–642, 645–647, 1435
  Tight coupling, 663–664
  Visible GPS satellites, 679
HORUS-2B, 1489–1498
  Aerodynamics, 1492–1496
  Entry control mode, 1491–1492
  Mass and geometry, 1489
  Mission, 1497–1498
Hot atomic oxygen, 1148–1149
Hu, Wan, 4
Huygens entry, 403–411
Huygens mission analysis, 1060–1074
  Descent-time analysis, 1067–1069
  Mission, 1061–1062
  Nominal descent, 1065–1067
  Parachute system, 1062–1063
  Spin-rate analysis, 1069–1072
  Wind analysis, 1073–1074
Huygens probe, 1499–1513
  Aerodynamics, 1502–1504
  Mass and geometry, 1499–1502
  Parachute system, 1504–1513
Hydrostatic equation, 1396
Hyperbolic trajectory, 1326–1327
Hyperion-1, 776–778
Hypersonic flow chemistry, 1146–1147
Hypersonic flow, 226–230, 235
Hypersonic vehicle, 174

I
Icarus, 1
Ideal gas law, 121, 159
Impulse response, 545
Inertia, 560–563
  Asymmetric mass, 560–563
  Moment of inertia, 561
Inertial frame, 1468
Inertial Measurement Unit (IMU), 639–642
  Aiding GPS, 642, 677
  Instrument errors, 640–641
  Measurement equation, 641
Inflatable Aerodynamic Decelerator (IAD), 962–971
  Ballute, 958–962
  Deployment range, 963
  Demonstrator, 965–969
  Technological challenges, 970–971
  Types, 964
Insulation, 1094, 1097–1099, 1109, 1131
  Fibre insulation, 1135
  Materials, 1133, 1236, 1240–1247
  Multi-Layer Insulation (MLI), 1109, 1131

Integral control, 713–714
Integrated heat load, 300
Interpolation in atmosphere models, 1398
Inviscid approximation, 249
Inviscid flow, 193
Irregular bodies, 96–99
Itokawa, 99

K
Kalman Filter
  Gain calculation, 656, 659, 661
  State estimation, 652
Kepler's Laws, 13
Kepler, Johannes, 12–13
Kleopatra, 98
Knudsen number, 167

L
Laminar flow, 217–219
Landing bags, *see* airbags
Landing ellipse, 973–974, 1055

Landing legs
Landing-gear configuration, 1010–1011
Energy absorber, 1012
Touchdown stability, 1013–1017
Lateral aerodynamic stiffness, 604
Lateral oscillation, 582–583, 595, 597, 614–617
Legendre polynomials, 101, 103–105
Lift coefficient at maximum heat flux, 437
Lift coefficient, 168
Lift force, 169
Lift parameter, 425
Lift-down flight, 503–504, 510
Lift-to-drag ratio, 254, 283
Lift-up, lift-down trajectory, 1303–1306
Linear interpolation, 1526–1527
Linear momentum, 1471
Linearisation of equations, 559, 1428–1432
Linearisation, 1519–1520
Linearised models, 652, 657–658
Liquid-fueled rockets, 22–25
Load Factor, 819
Local inclination methods, 249–251
Base pressure, 259
Gabeaud base pressure, 260
Modified Newtonian, 252
Shock expansion, 261–262
Tangent cone, 258–259
Low gate, 1033
Luna-9, 871
Lunar return, 502–507

M
MA956 alloy, 1129, 1230, 1234
Mach number, 164
Mach-number independence principle, 229
Mach-wave patterns, 180, 197, 240
Magellan spacecraft, 1263
Magnetometer, 651
Malmesbury, Eilmer of, 1
Mars
Atmosphere, 137–145
Gravity field, 106
Shape, 87, 93–94
Mars climate database, 139–141
Mars entry vehicle, 412–413, 411–422, 403–411
Mars Exploration Rovers (MER), 1049–1054
Aeroshell design, 1050
Airbag landing, 1020, 1053
Mission profile, 1053
Parachute deployment, 1053
Mars Odyssey aerobraking, 1265, 1270
Mars Pathfinder, 1049
Mars Reconnaissance Orbiter aerobraking, 1266
Mars Science Laboratory (MSL)
Guided entry, 1054–1055
Mission profile, 1056–1059
Parachute design, 1016
Parachute performance, 1055
Parachute testing, 1057
Sky crane, 1059–1060
Mass distribution, 1472
Mass Estimation Relationships (MER), 1180–1183
Afterbody, 1182
Forebody, 1181
Materials selection, 1096, 1149
Maximum deceleration, 1414
Maximum heat flux, 396, 434–438, 495–496, 1416
Maximum-range glide, 814–815
Measurement equation, 641–642, 645–647, 1435
Mechanical energy, 804
Mechanical load, 389–393, 433–434, 490–492, 1411, 1414–1415
Mechanical-load constraint, 444
Metallic TPS, 1093, 1102, 1115, 1145, 1158, 1198, 1213
Metals
Inconel 617, 1102, 1234–1135, 1144–1145, 1151, 1153, 1155, 1161, 1230
PM-1000, 1102, 1129–1130, 1145, 1150, 1156, 1162
Stainless Steel AISI 310, 1161, 1229–1231, 1235
Mid-air recovery, 35, 868–869
Moment of inertia, 1480
Momentum thickness, 220, 223
Monte Carlo analysis, 411–422, 1414–1419

Montgolfier brothers, 864
Motion, 322
Multi-Layer Insulation (MLI), 1109, 1131

N
Natural frequency, 547
Navier-Stokes equations, 307–312
Navigation errors, 637, 1435–1436
Newton, Isaac, 13–15
Newton's Laws of Motion, 13–15, 322–323, 1467

Newton-Raphson method, 1539
Newton's Second Law, 1406
Nomex felt, 1133–1134, 1236, 1240, 1246
Non-inertial frame, 1468
Normal acceleration, 1411
Normal shock, 199–206
Numerical integration, 1531–1534

O
Oberth effect, 27
Oberth, Hermann 25–26
Oblateness, 86
Oblique shock, 206–212
Observability, 705
Opening loads, 892–896
Optimal aerocapture, 1298–1311
  Minimum ΔV, 1300–1302
  Minimum radiative heat load, 1307–1311
  Minimum thermo-mechanical loads, 1302–1307
Optimal control
  Linear Quadratic Regulator (LQR), 703, 731–735, 1441–1445
  Pontryagin's minimum principle, 714–730
Optimal glide path, 811
Orbital decay, 1260
Orbital elements, 327–329
Orbital motion, 324, 329
Orbital-elements corrections, 1278–1279
Ottoman aviators, 2
Overshoot boundary, 440–441, 465
Oxidation resistance, 1096, 1129, 1140, 1148
Oxidation stability, 1129, 1150

P
Parachute clustering, 899–902
Parachute deployment, 1412
  Drogue gun, 877
  Inflation process, 889–892
  Opening loads, 892–896
Parachute drag area, 1412–1413
Parachute dynamics simulation
  Added mass, 913–914
  Equilibrium descent, 907–910
  Mass properties, 914–917
  Multi-body simulation, 907–908
  Pendulum motion, 905
  Scissors motion, 905
  Simulation tools, 904
  Swivel moment, 917–919
  Terminal descent, 909
  Reefing, 919–920
  Spin vanes, 921–922
Parachute reefing, 896–899, 919–920, 1451, 1453
Parachute systems
  Canopy, 877
  Drag versus stability space, 878–880
  Drogue chutes, 874, 877
  Hardware elements, 876–877
  Harness, 877
  Materials, 885–889
  Parachute selection, 881
  Performance analysis, 881
  Pilot chutes, 874, 877
  Requirements, 873
  Riser, 877
  Sequence of events, 875
  Stability, 679
  Suspension lines, 877
Parachute types
  Conical ribbon parachutes, 881, 883
  Cruciform parachutes, 877
  Disk-gap-band parachutes, 882, 884
  Guide surface parachutes, 881–882
  Ring-sail parachutes, 883, 885
Parafoil, 870
Passive actuators, 694
Passive systems, 1159–1164
  Heat-sync structure, 1159–1161
  Hot structures, 1161–1162
  Insulated structures, 1163–1164
Pendulum motion, 1436–1441
  Damped, 546–539
  Undamped, 542
Performance curve, 808
Periapsis adjustment, 1268, 1272, 1274, 1276, 1278–1279
Periapsis corridor, 1280–1281
Phase plane analysis, 541, 544
Phugoid motion, 466
Phugoid, 582–583, 594, 617
Pitch angle, 331–332
Planar motion equations, 1411
Planetary shape
  Earth, 87, 90–91
  Mars, 87, 93–94
  Moon, 87, 91
  Venus, 87, 95
Planetary shape parameters, 87
Planetocentric reference frame, 1473

PM-2000 alloy, 1129, 1144–1145, 1150–1151, 1153, 1156, 1162, 1230, 1235
Polyhedron model, 97–99, 113
Pontryagin's minimum principle, 714–730
Position vectors, 1470
Powered Descent Initiation (PED), 1029
Powered descent, 950–957, 1453–1454
Prandtl number, 166
Precision landing, 941, 972, 980–981
Predictor-corrector guidance
  Equilibrium-glide entry phase, 1294–1296
  Exit phase, 1296–1298
Pressure coefficient, 180
Pressure distribution, 245, 249, 251–252, 1402
Pressure gradient, 231
Principia Mathematica, 15–16
Proportional control, 711–712
Pyrolysis zone, 1185–1186

Q
Quadratic guidance law, *see* gravity turn, E-guidance
Quaternion, 333

R
Radial acceleration, 1400
Radiation shielding, 1172–1173
Radiation, 1119–1123
  Absorption, 1120–1122
  Emissivity, 1119
  Net radiative heat flux, 1123
  Radiative surface, 1119–1121
  View vector, 1119–1121
Radiative heat transfer, 300–306, 1420–1422
Radiative peak, 306
Ramjet, 27–28
Rarefied flow, *see* free-molecular flow
Rate of descent, 809
Real gas effects, 243–245
Recursive estimation, 656
Reference frames
  Aerodynamic reference frame (airspeed), 326, 347
  Aerodynamic reference frame (groundspeed), 325, 346
  Body reference frame, 325, 349
  Inertial reference frame, 324, 340
  Local orbital reference frame, 325, 342
  Orbital reference frame, 324, 341–342
  Planetocentric reference frame, 324, 345
  Rotating planetocentric reference frame, 325, 343
  Trajectory reference frame, 326, 346
  Vertical reference frame, 325, 345
  Wind reference frame, 326, 352
Reinforced Carbon-Carbon (RCC), 1096–1098, 1142
Relative acceleration, 1470
Relative moment, 1476
Relative velocity, 1470
Response analysis, 706–707
Retro rockets, 993, 1051, 1055
Reusability, 1094, 1096, 1226, 1130
Reynolds analogy, 291
Reynolds number, 165, 1402
Rigid body, 1477–1478
Rigid body dynamics, 1406
Robust control, 702–703
Roll angle, 331–332
Roll mode, 595
Root finding in interpolation, 1422
Root finding, 1538–1539
Rosetta aerogravity assist, 1385–1387
Rotating frame, 1468
Rotation vector, 1472
Rotational dynamics, 560–563
Rotational motion, 358, 360–361
Runge-Kutta-Fehlberg, 1536–1538

S
Sänger, Eugen 27–28
Saffil insulator, 1236, 1238, 1243–1247
Scale height, 1396
Second-order approximation, 508–519
Semi-major axis, 328
Semi-passive systems, 1164–1168
  Ablative systems, 1166–1168
  Heat pipes, 1164–1166
Sensors
  Accelerometers, 641
  Flush Air-Data Sensing (FADS), 651
  Global Positioning System (GPS), 643
  Gyroscopes, 642
  Horizon sensors, 636
  Lidar, 648, 649
  Magnetometer, 651
  Pressure sensor, 651
  Star trackers, 650
  Terminal descent sensor, 650
Service hazards
  Craters, 976

Illumination, 977
Roughness, 976
Slope, 975–976
Surface changes, 978–979
Shock interaction, 240–242
Shock stand-off distance, 237
Shock waves, 196, 198, 1402
Shock-attenuation systems
Aeroshell penetrator, 1002–1003
Crushable material, 1003–1008
Landing bags, 1018–1021
Landing legs, 1009–1013
Shockwave formation, 198
Side-slip angle, 332, 349
Siemienowicz, Kazimierz, 5
Silverbird Bomber, 28
Skin friction, 222–223
Skipping flight, 464–507
Altitude-velocity diagram, 481–484
Apollo skip entry, 502–507
Entry corridor, 441
Lift-to-drag ratio, 492
Lowest point, 1426
Maximum deceleration, 490–492, 1425–1427
Maximum heat flux, 495–496
Phugoid motion, 466
Range modulation, 465
Silverbird Bomber, 465–466
Skipping legs, 482
Velocity reduction, 1426
Skipping legs, 482
Small perturbation method, 558
Solar System gravity assist, 1334–1335
Solidification Principle, 1472
Space exploration, 20–21
Space Shuttle terminal-area guidance, 822–826
Space Shuttle TPS, 1098–1101
Tiles, 1099–1101
Space Shuttle, 175, 187, 245–246, 263, 267–273
Specific heat capacity
Ceramics, 1236–1240
Insulators, 1240–1247
Metals, 1229–1235
Specific heats, 1397
Speed brake, 813, 821–824, 831, 834
Speed of sound, 1397
Sphere of influence, 1473
Spherical coordinates, 321, 330
Spherical-harmonics model, 104–105, 1396
Spherical trigonometry, 1515–1518
Spheroid model, 86, 88
Spin-rate analysis, 625–629
Spin-vane model, 1513–1514
Sputnik, 20, 34
Stability analysis, 601–605
Entry capsules, 606, 625
Winged vehicle, 582
Stability derivatives, 605
Stagnation point, 163, 252, 292
Stagnation temperature, *see* equilibrium temperature
State definitions
Angular rates, 334
Attitude, 331–334
Cartesian coordinates, 329, 354
Orbital elements, 327–328
Position and velocity, 327–331, 355
Spherical coordinates, 330, 355–356
State estimation
A-posteriori estimation, 653, 658
A-priori estimation, 653, 658
State feedback, 705
State variables, 1406
State-space form, 556, 566, 574, 603, 611–613, 704, 1433
State-space system, 1521–1522
Steady glide, 807–812, 1448–1449
Equilibrium glide angle, 808
Maximum endurance, 811
Maximum range, 811
Steady-glide condition, 808–809
Steady-state approximation, 807–812
Šteins, 96–97
Strength and stiffness, 1149–1154
Stress and deformation, 1150–1152
Subsonic flow, 192–195
Supersonic flow, 195–226
Supersonic pockets, 1402
Surface energy balance, 1183–1184, 1190
Surface models, 88
Surface stability, 1135, 1148
Surface temperature effects, 1401
Surveyor, 1025–1028
Symmetrical motion, 595

T

Tangential velocity, 1396
Targeting point, 770

Tauber and Sutton model, 1310–1311
Taylor series, 1518
Temperature gradient, 231
Terminal Area Energy Management (TAEM), 799–801
Terminal descent, 950–957
  Burn time, 954
  Engine-on altitude, 955
  Equations of motion, 952
  Layout, 953
  Mars, 955–957
Terminal velocity, 1412–1413
Terminal-area guidance, 820–829
  Acquisition phase, 822
  Energy control, 823–824
  Heading-alignment phase, 823
  Pre-final phase, 823
  S-turn phase, 822
Terrain Absolute Navigation (TAN), 981–984
Terrain Relative Navigation (TRN), 985–986
Thermal analysis, 1180, 1199–1203
Thermal conductivity, 166, 310–311, 1113
  Ceramics, 1236–1240
  Insulators, 1240–1247
  Metals, 1229–1235
Thermal expansion, 1157
Thermal load, 396–399, 434–438, 495–496
Thermal properties, 1229–1247
  Specific heat coefficient
  Thermal conductivity
Thermal protection system, 286
Thermal strain, 1155–1156
Thermal stress, 1154–1157
Thermo-mechanical loads, 389–393, 396–399
Thrust force, 327, 350
Time derivative of a vector, 1468
Time-dependent atmosphere effects, 1396
Tipu Sultan, 6
Titanium
  6Al-4V, 1230, 1232, 1235
  TA2 (Unalloyed), 1233
TPS characterisation, 1158–1159
  Active systems, 1168–1175
  Passive systems, 1159–1164
  Semi-passive systems, 1164–1168
TPS design approach, 1176–1180
TPS design, 1455–1458
TPS examples, 1087, 1105
  Capsules, 1087–1089
  Low cross-range orbiter, 1095–1096
  Huygens, 1109
  Mars Science Laboratory, 1108–1110
  Orion, 1111
  Space Shuttle, 1096, 1101
  SR-71 *Blackbird*, 1091–1092
  Stardust, 1105–1107
  Viking, 1104–1105
  X-15, 1089–1091
  X-20, 1092–1094
  X-33, 1101–1102
TPS finite difference model, 1199–1203
TPS requirements, 1112–1114
Trajectory constraints, 442–444
Translational motion, 321, 329, 570, 596, 601, 614–618, 625–626
Transonic flow, 194–195
True anomaly, 328–329
Tsiolkovsky, Konstantin, 18, 19
Turbulent flow, 217–219

U
Undershoot boundary, 440–441, 465
Universal gravitation, 15

V
V-2 rocket, 33–34
Variable mass systems, 1472, 1474, 1476
Vehicle design, 246, 184
Vehicle trim, 751, 754, 756
Velocity components, 329, 355
Velocity loss, 1350–1351
Velocity profiles, 1402
Venera program, 871–872
Venus
  Atmosphere, 148–151
  Gravity field, 106
  Shape, 87, 95
Venus entry simulations, 1414–1419
Venus Express aerobraking, 1267
Verne, Jules, 17, 18
Vernier engines, 928, 1026–1027
Vertical equilibrium, 1412–1413
Vertical plane dive, 812–813
Viking mission, 1034–1039
  Entry, descent, landing, 1038
  Parachute deployment, 1038
Vinci, Leonardo da, 8, 870
Viscosity, 217, 293, 310
Viscous drag, 1402
Viscous interaction, 236–237
Vostok-1 orbit, 1395
Vostok-1, 869

W
WAC Corporal rocket, 1395
Wall catalycity, 1146–1147
Waverider aerodynamics, 1377
Waverider geometry, 1376
Werner and Scheeres method, 96, 113
Wind tunnel
  Subsonic, 184–186
  Supersonic, 188–189
  Transonic, 187–188
  Hypersonic, 189–192

Y
Yaw angle, 331–332
Yaw-pitch-roll sequence, 1406
Yield strength, 1127, 1150–1151
Young's modulus, 1152–1153

Z
Ziegler-Nichols tuning rules, 709
Zirconia felt, 1127, 1131, 1135