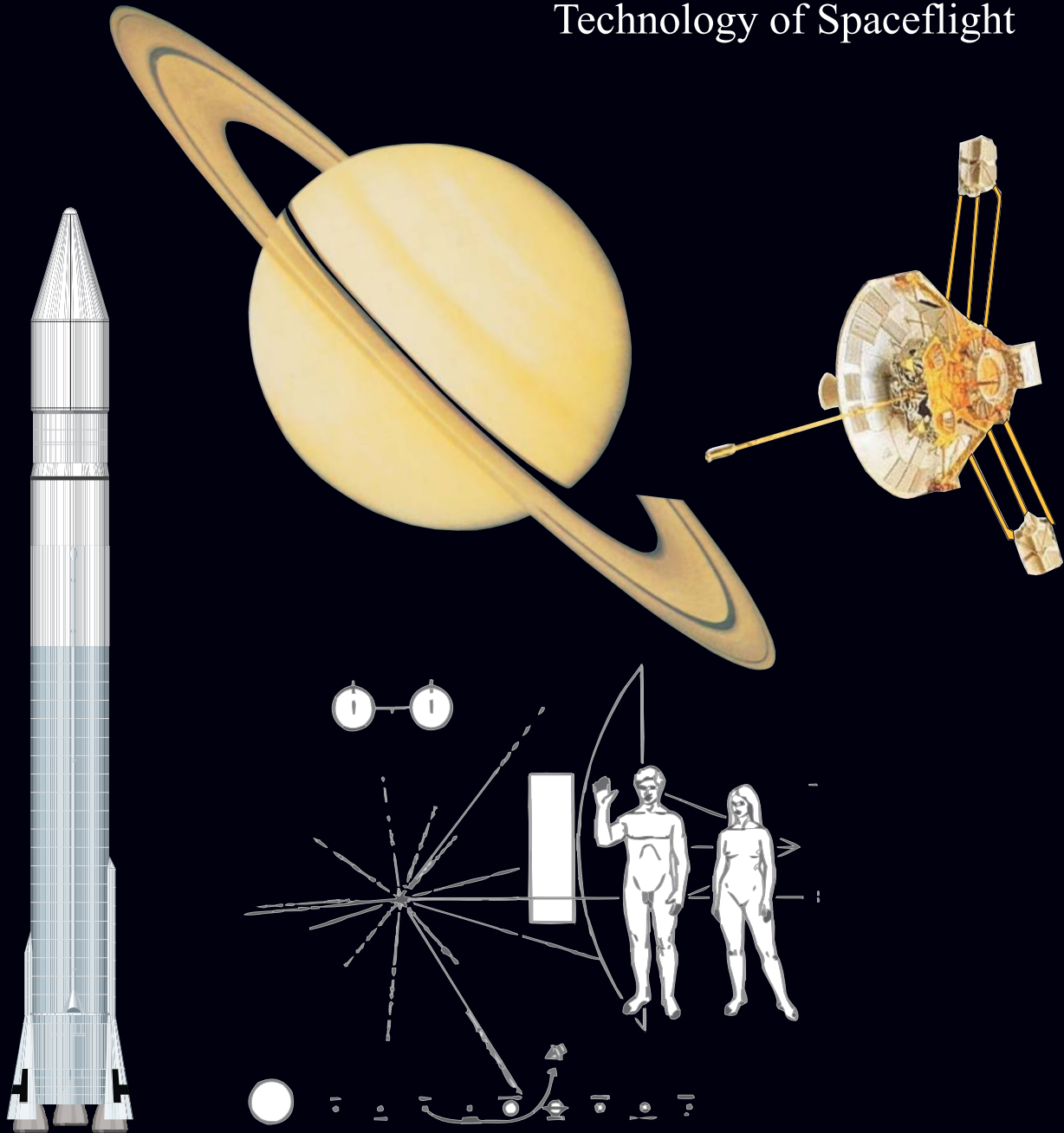


Hanfried Schlingloff

ASTRONAUTICAL ENGINEERING

An Introduction to the
Technology of Spaceflight



Including an interview with Prof.Dr.H.O.Ruppe about the past, the present and the future of astronautics

ASTRONAUTICAL ENGINEERING

An Introduction to the Technology of Spaceflight

by Hanfried Schlingloff

The cover page shows elements of the successful US-American Pioneer 11 project. The spacecraft, launched on April 6th 1973 by an Atlas-Centaur vehicle, performed the first Saturn flyby mission on September 1st 1979, after having passed Jupiter on December 4th 1974. The spacecraft left the solar system carrying on-board a metal plate that shows the images of a naked man and a woman, our solar system and the accurate position of our solar system with respect to some near stars (pulsars). There is, however, only an insignificant possibility that extraterrestrials might find and read the information: it is expected that the plate will corrode after a comparatively short time period, maybe after a few million years already...

ISBN 3-00-016289-5

©2005 Ingenieurbüro Dr.Schlingloff. This book was printed in Regensburg, Germany. All rights reserved. No part of this book may be reproduced in any way or by any means without explicit permission of the publisher.

Ingenieurbüro Dr.Schlingloff Publications, Kalkofenring 17, D-93077 Bad Abbach.

Today is Monday, the 21st March 2005, a pleasant evening in the conference room of the Institute Astronautical Engineering of the Technical University of Munich, and I have the privilege to talk to Prof.Dr.H.O.Ruppe, emeritus of the institute. Professor Ruppe is the author of more than 100 publications in the field of astronautics, including two comprehensive double-volume books; and, before he introduced astronautics to this university, he was head of the “Future Project Office” at NASA’s Marshall Space Flight center in Alabama. It is now 22 years ago that I finished my dissertation under his supervision at his institute. His professional help during the following years and many pleasant talks I had with him encouraged me to write this book. Now the book is finally finished, and at this occasion we want to talk about astronautics: the past, the present and the future.

Manned Space Flight - glorious past but gloomy prospects

Prof.Ruppe, astronautics is often used as a synonym for manned space flight. When I look at the American Space Shuttle System STS or at the International Space Station ISS, I get the feeling that presently manned space flight experiences a serious crisis. Do you agree with this point of view and, in case, can you explain to us how it came to the present situation?

Ruppe: *Till the end of world war II astronautics was essentially “manned astronautics”. At that time automated space flight was thought about but not really. It appeared necessary for all complicated functions to have man aboard. This may be compared to the development of aeronautics which had to do with human flight in machines. Of course there were models, usually from school children, but this was used to initiate the children to the science of flight. In the same way we thought about unmanned flights, for example to the moon, to bring us to human flight to the moon. In this way the Apollo mission followed the thoughts of the oldtimers: man to the moon. Such developed the competition between the two superpowers for a lunar race. Of course the total development was very complex. The then Soviet Union had wonderful successes in both, unmanned and manned astronautics. The United States experienced a disaster with respect to Cuba. As a consequence, the young president John F.Kennedy proposed the challenge of a lunar race. The logic was, that the advantage of the Soviet Union would not be decisive with respect to a manned lunar programme. We - that is the American oldtimers in matters of space - assured the president we could do! So project Apollo was born in 1961. We promised that the job would be accomplished within ten years. As we know now, we did it; the Soviet Union failed; and the United States preeminence in space was assured.*

Apollo was, of course, a manned mission. Every interested citizen could follow such a mission, since its duration was about ten days, a good time interval to capture the interest of a large population. A big question was: what comes after Apollo? There appeared no simple answer. A space station had already been realized both, by the United States and by the Soviet Union, those were small stations for small crews.

So as a follow-on project a large international space station appeared attractive and doable. As a tool for this international station a comfortable and large space shuttle seemed attractive. This tool became a goal of its own; and I think this started the trouble which continues till today. The space station cannot compare to Apollo, and problems with the Shuttle endanger the programme of the ISS.

But how about the manned Mars mission?

Ruppe: *President George W.Bush has suggested manned return to the moon and manned flight to Mars. He sees the lunar field as a thought of training ground for manned Mars flights. We also have quite a floury of unmanned Mars missions which followed the Viking missions. So Mars has been much in the news and it seems that we are ready to follow with man. I think this is very wrong. The manned Mars mission probably is the most difficult space mission we could do. Man to Mars is not a somewhat bigger Apollo, instead of weeks it takes years. It is not possible to fly with one vehicle from earth to the target. Instead we have to assemble a Mars fleet in earth orbit and fly from this orbit to Mars. A crew of three may not suffice. I do not think that the lunar training is of much help with the Mars job.*

Don't you think that a manned Mars mission would be a way out of the present crisis?

Ruppe: *Yes, I think so definitely. A manned Mars mission would again put an attractive goal for manned space missions. This would keep us busy for a long time - I think much more than ten years. It is a very attractive chance to help international cooperation.*

Recently we have seen the privately financed flight of Burt Rutan's Spaceship 1 vehicle. Isn't private astronautics another way out of the crisis?

Ruppe: *I do not think so. Rutan's vehicle showed us a jump about 100 kilometers high and a few hundred kilometers in distance. Even that is quite expensive; and you cannot develop real space flight from it. The minimum for space flight is in my opinion an earth satellite. To do this you need a step rocket with two or three stages. To return from orbit a capsule of sorts is necessary. We have just to look at American or Russian or Chinese beginnings. That is hundreds or thousands of millions of dollars away from Rutan.*

The shuttle must be replaced one day, maybe even soon if another accident occurs. Isn't a capsule the better solution for manned space flight?

Ruppe: *This is a very difficult question. At this time many investigations are under way both at NASA and at United States space industry. There is no clear answer yet. I think the main objective of the new vehicle will determine how the solution will look like. For a lunar landing you cannot use wings. The astronauts told me that they like the comfortable landing in a shuttle. From a reliability point of view I think there are no significant differences between both systems. Until now the shuttle has about 100*

flights with two significant accidents. I think that a future shuttle will have better rescue possibilities for the crew. Presently the capsule has better rescue capabilities - but we have had two fatal accidents in the Soviet programme. It appears to me that the struggle between capsule and shuttle is about even. It will be interesting to see what the future will hold. Obviously, the winged vehicle is limited to low earth orbit since in space the wings are just ballast and not really useful. Hence, for deep space or lunar environment capsules are the natural solution. Mars may be an exception. Wernher von Braun's first Mars study, circa 1950, utilized winged Mars landers. You should remember that, at this time, the Mars atmosphere was assumed to be more dense than we learned from Viking!

Launch Vehicles - tools for astronomical activities

Well, also for me it is evident that high-altitude airplanes are not the correct way to access space. Twenty years ago I believed that the shuttle was the best and safest option, but today the experience spoils that view. Today I think that conventional rockets are the best opportunity for space transportation. However, in my book I state that it was a bad mistake from the European space agency to take the commercially successful Ariane-4 launcher from the market; and I believe that the Ariane-5 launch vehicle in its present form is ill-conditioned for the transportation of commercial satellites, particularly because of the application of strong solid booster motors. You worked for more than 60 years on the subject of space launcher optimization. What do you think about the present space launcher situation?

Ruppe: *Well, the transport system "Ariane 4" could be adapted to mission requirements by relatively simple modifications of the first stage. This way is not useful for Ariane 5. Here the upper stage can be changed by using different propellents - hydrogen or storable. An advanced hydrogen version has been cancelled but is still being talked about. Nevertheless it remains basically the Ariane 5 with associated high transport costs. In my opinion the decision between solid or liquid propulsion is more or less even, at least as it concerns first stages; reasons of availability might be decisive for upper stages. Also there are rumors that in the case of the US-shuttle the solid propellant company "Thiokol" had to be involved. So you see that sometimes the problem of industrial participation might be decisive.*

European nationalistic reasons will keep the Ariane 5 in operation, for example for the transportation of military satellites. The space transport picture gets more difficult if the many available systems in US, Russia and China are considered. We can expect survival of only some of the many systems.

Application Satellites - a shrinking market

Well, then the space launch business is presently in a difficult situation, because the market for satellite services is actually shrinking. But why is this so? The telephone market is booming, television occupies more time in everyday life, and the new GPS navigation system in my car is really a useful device. I nearly never miss

the eight o'clock news, particularly because of the weather forecast for the next day. Considering all this, shouldn't we expect a booming satellite business?

Ruppe: *The most important part of application satellites has to do with communications. This market is fairly saturated at present. Only the failed satellites and of course extensions of the system have to be replaced. But I understand that glass fibers with optical communication give a stiff competition to the satellites. So this market may shrink compared to the past. Of course this is all undecided and only the future will tell. Weather satellites pose difficulties of their own. We are quite accustomed that weather surveys and predictions are provided by the governments. So private weather prediction is rare. Military weather prediction falls into the same class: it is not derived from private satellites. The military serve their own purposes.*

After Global Positioning System we are accustomed to non-pay services. Again the government gives us such satellites and the infrastructure which is required. It is not a private service! Recently there has been a big Tsunami in Asia. This led to the expectation of catastrophe warning satellites. Again, this is no field for private enterprise; I think that governments should provide this service. Observation of the surface of the earth is a typical task for spacecraft. The military require good resolution in the order of 10 centimeters, they have a close supervision of this field. They will do so in all foreseeable future. Because of this, private earth observation is rare.

Scientific Missions - brilliant jewels in astronautics

You often told me that you see the search for extraterrestrial life as the ultimate goal for astronautical activities. Well, up to now we haven't found any aliens in our solar system; but the scientific deep space probes were surely worth the money. How poor would be our understanding of the solar system without them. I am always fascinated when I look at the pictures which the deep space probes returned to the earth, many of them are prettier than a space artist could paint.

Ruppe: *I fully agree with what you say. There is a fascinating task to determine how life began. Did it come from space? Were comets important in this respect? How about life on other worlds? Do other suns have planets? I think of earthlike planets and not of giants like Jupiter. Some large planets of other suns have been found but not small bodies like our home planet. Observations from space promise to help; but these satellites have to be much larger than the Hubble space telescope. The Hubble telescope is the largest space observatory which has been launched so far. No doubt there will be more potent successors. In my opinion the never ending task of astronautics is the determination of the human role on earth and all the history of the universe.*

Prof.Ruppe, I want to thank you very much for this interesting interview.

Ruppe: *I am proud to accompany this fine book with this preface interview. I wish you all success.*

1.	Space Environment	1
1.1.	The Solar System	1
1.1.1	The Sun	2
1.1.2	The Inner Planets	3
1.1.3	Asteroids and Comets	5
1.1.4	The Gas Giants	6
1.1.5	Pluto and the Kuiper belt	7
1.2.	Gravity Fields	11
1.2.1	Gravitational Attraction of a Spherical Celestial Body	11
1.2.2	The Potential Field of Oblate Celestial Bodies	13
1.3.	The Structure of the Atmosphere	14
1.3.1	Barometric Scale Height	14
1.3.2	The Earth's Atmosphere	15
1.3.3	Atmospheres of other Celestial bodies	18
1.4.	The Structure of the Magnetosphere	20
1.4.1	Planetary Magnetic Fields	20
1.4.2	Magnetism and Particle Radiation	20
1.5.	Man in Space	22
1.5.1	Man on the way to space and back	22
1.5.2	Man Living in Space	23
1.5.3	Health Hazards in Space	25
1.6.	Progress in Technology and Future Space Flight	27
1.6.1	Technological Environment for Space Projects	27
1.6.2	Future Exploration of our Solar System	28
1.6.3	Interstellar Spaceflight	30
2.	Rocket Propulsion	31
2.1.	Theory of Rocket Motors	31
2.1.1	Expulsion of Propellant	31
2.1.2	Thrust and Impulse of a Rocket Stage	34
2.1.3	Multi Stage Rockets	36
2.2.	Rocket Propellant	41
2.2.1	Properties of Chemical Rocket Propellant	41
2.2.2	Cryogenic Propellant	42
2.2.3	Storable Liquid Propellant	43
2.2.4	Solid Propellant	44
2.3.	The Combustion Process	46
2.3.1	Combustion of Liquid Propellant	46
2.3.2	Combustion of Solid Propellant	48
2.3.3	Propellant Application and Performance	50
2.3.4	Computation of the Chamber Temperature	51
2.3.5	The Composition of the Chamber Gas	57
2.4.	Hypersonic Nozzles	63
2.4.1	Thermodynamic Phenomena inside the Rocket Nozzle	63
2.4.2	Velocity and Temperature in the Isentropic Nozzle	69
2.4.3	Overexpanding Nozzles	73
2.4.4	The Shifting Chemical Equilibrium	75
2.4.5	Nozzle Geometry	81
2.5.	Liquid Rocket Stages	88
2.5.1	Tanks and Structures	88
2.5.2	Thrust Chambers	90
2.5.3	Turbopump Feed Systems	92
2.5.4	Liquid Engine Sub-Masses	98

2.6.	Solid Rocket Motors	100
2.6.1	Solid Propulsion for Space Applications	100
2.6.2	The Burning Rate of Solid Propellant	102
2.6.3	Grain Configuration	103
2.6.4	Solid Motor Cases	104
2.6.5	Solid Motor Nozzles	105
2.6.6	Ignition and Control of Solid Rocket Motors	107
2.7.	Attitude and Trajectory Control Thrusters	108
2.7.1	Gas Pressure Feed Systems	108
2.7.2	Auxiliary Propulsion Systems	109
3.	Launch Services	111
3.1.	Applications for Space Launchers	111
3.1.1	Target Orbits	112
3.1.2	Commercial Market for Satellite Launch Services	113
3.1.3	Influences on the Commercial Launch Market	114
3.2.	General Space Launcher Design Aspects	115
3.2.1	Performance Optimization.	115
3.2.2	The Appropriate Propellant	117
3.2.3	Optimal Launcher Size	119
3.2.4	Flight Safety Aspects	120
3.3.	Evaluation of Commercial Launch Services	122
3.3.1	Launch Vehicles	122
3.3.2	Launcher Costs	136
3.3.3	Launcher Payload	140
3.3.4	Launcher Quality Comparison	142
3.4.	Optimization of Space Launchers	145
3.4.1	The Concept Ariane-5 Commercial	145
3.4.2	The Airborne Ariane-4 Launcher	149
3.5.	The Promise of Modern Technology	152
3.5.1	The Concept of Reusability	152
3.5.2	Single Stage to Orbit Vehicles	156
3.5.3	Aerospace Planes	158
3.5.4	The Future of Space Transportation	161
3.5.5	Infeasible Ways to Access Space	162
4.	Spacecraft Trajectories	163
4.1.	Vectors and Coordinate Systems	163
4.1.1	Moving Coordinate Systems	164
4.1.2	Coordinate Transformation	166
4.1.3	Rotation Velocity	169
4.1.4	Vector Differentiation in a Rotating Coordinate System	171
4.2.	Mechanics of Rocket-Powered Flight	173
4.2.1	The Gravitational Force	173
4.2.2	Aerodynamic Forces	175
4.2.3	Thrust Acceleration	178
4.2.4	Equations of Motion	181
4.2.5	Mechanical Energy of a Trajectory	182
4.2.6	The Angular Momentum Vector	183
4.3.	Orbital Flight	184
4.3.1	Conic Orbits	184
4.3.2	Flight Time on Conic Orbits	187
4.3.3	Orbital Elements	191
4.3.4	Perturbation of Low Earth Orbits	192
4.3.5	Destination Orbits	196

4.4.	Impulsive Transfers	199
4.4.1	Impulsive Thrusts	199
4.4.2	The Hohmann Transfer	202
4.4.3	Impulsive Escape Maneuvers	206
5.	Orbit Optimization	209
5.1.	Optimal Transfer Trajectories	209
5.1.1	Application of the Hamilton-Lagrange theory	210
5.1.2	Optimum Control of Thrust Arcs	216
5.1.3	Control Equations for Impulsive Transfers	221
5.1.4	Final Conditions for Optimal Transfer Trajectories	224
5.2.	Behaviour of Lagrange Multipliers and Switch Function	228
5.2.1	Vector Representation of the Optimal Trajectories	228
5.2.2	Plane Optimal Trajectories	231
5.2.3	Elimination of the Lagrange Multipliers	236
5.2.4	The Course of the Switch Function	239
6.	Attitude Stabilization	251
6.1.	The Angular Motion of Satellites	251
6.1.1	Newton's Law for an Object with Physical Size	252
6.1.2	The Angular Momentum Law for another Reference Point	254
6.1.3	Inertia of a Rigid Body	256
6.1.4	Kinetic Energy of a Rigid Body	262
6.1.5	Equations of Motion of a Rotating Rigid Body	262
6.2.	Spin Stabilization	265
6.2.1	Motion of a Torque-free Satellite	265
6.2.2	Attitude Maneuver of a Spinning Satellite	267
6.2.3	Stability of a Spinning Satellite	269
6.2.4	The Single-Spin Satellite	271
6.2.5	The Dual-Spin Satellite	274
6.3.	Three-Axis Stabilization	277
6.3.1	The Gyroscope Wheel	277
6.3.2	The Gyro as Sensor	280
6.3.3	The Gyro as Actuator	285
6.4.	Gravity Gradient Stabilization	289
6.4.1	The Motion of a Gravity Gradient Satellite	289
6.4.2	Stability of the Gravity Gradient Satellite	294
6.4.3	The Rotating Space Station	297
7.	Launcher Dynamics	299
7.1.	Thrust Vector Optimization	299
7.1.1	Control of the Thrust Angle	299
7.1.2	Control of the Thrust Magnitude	301
7.1.3	Control of the Exhaust Velocity	304
7.2.	Launcher Trajectory Optimization	310
7.2.1	Ascent of Conventional Space Launchers	310
7.2.2	Ascent Trajectories of Airborne Launchers	323
7.3.	Space Launcher Navigation	326
7.3.1	Guidance Systems	326
7.3.2	The Lambda-Matrix Method	330
7.4.	Space Launcher Attitude Stabilization	333
7.4.1	Control of the Angular Motion	333
7.4.2	The Flexible Rocket	341
7.4.3	Sloshing Propellents	351

8. Planetary Missions	353
8.1. Lunar Trajectories	354
8.1.1 Velocity Requirement for a Lunar Mission	354
8.1.2 Lunar Circumnavigation	358
8.1.3 Liberation Points of the Earth-Moon System	364
8.1.4 Escape from the Earth-Moon System	370
8.2. Planetary Flights	374
8.2.1 Earth Departure	374
8.2.2 Planetary Flyby	376
8.2.3 Heliocentric Transfer Trajectories	378
8.2.4 Return Orbits	385
8.3. Gravity Propelled Missions	387
8.3.1 Gravity Assist on Earth-Return Trajectories	387
8.3.2 Mars Swingby on Earth-Return Trajectories	391
8.3.3 Gravity Assist from Venus and Earth	395
8.3.4 Gravity Assist on the Way to the Outer Planets	398
8.4. Manned Mars Missions	405
8.4.1 Martian Trajectories	405
8.4.2 Flight Operations at Mars	410
8.4.3 Expenditures for a Manned Mars Mission	414
9. Low Thrust	419
9.1. Low Thrust Propulsion Systems	420
9.1.1 Thermo-electric Engines	420
9.1.2 Ion Engines	422
9.1.3 Magneto-plasma Dynamic Thrusters	426
9.1.4 Electric Energy Sources	430
9.1.5 Solar Sails	433
9.2. Electrically Propelled Missions	434
9.2.1 Optimization of the Propulsion System	434
9.2.2 Low Thrust Earth Escape	436
9.2.3 The Comet Nucleus Sample Return Mission	439
9.2.4 The Multi Asteroid Rendezvous Mission	443
9.3. Solar Sailing	449
9.3.1 Earth Escape with Solar Sails	449
9.3.2 Sailing in Interplanetary Space	453
9.3.3 The Sailcraft on an Earth-Return Trajectory	455
10. Reentry Maneuvers	457
10.1. High-Speed Gasdynamics	458
10.1.1 Flow Regimes	458
10.1.2 Aero-Thermodynamic Phenomena	461
10.1.3 Computation of Flow Fields	465
10.1.4 Aerodynamic Loads	467
10.2. Reentry Trajectories	470
10.2.1 Ballistic Reentry	470
10.2.2 Gliding Reentry	476
10.2.3 Control and Navigation	481
10.3. Thermal Protection Systems	484
10.3.1 High Temperature Materials for the Reusable Space Shuttle	484
10.3.2 Heat Shields	486

1. Space Environment

The word “space” determines an infinite dimension. Actually, every location which is higher than a certain altitude (say 90 km) above the surface of the earth is called space. Spaceflight or astronautics are synonyms for journeys to space. So, the word spaceflight covers a wide range of different missions to different destinations.

The first chapter of this book is concerned with the conditions required for spaceflight. In space we find different celestial bodies: all of them have a gravitational field, just that some of them have an atmosphere and some others have a magnetosphere. The vacuum of space is not perfectly empty: there we encounter particle radiation, x-rays, electro-magnetic waves, sunlight and meteorites. The astronauts on board of a spacecraft need a life-support system and protection from all the hazards in space. Spaceflight requires a high developed state of the technology and a powerful industry; sometimes also public money and the political will-power. Whenever a new planetary mission has been completed successfully, the books on the solar system are rewritten. For example, recommendable fundamental literature are the books of J.K.Beatty, B.O’Leary, A.Chaikin [“The New Solar System”, Cambridge University Press, 1982]; J.Gürtler,J.Dorschner [“Das Sonnensystem”, Johann Ambrosius Verlag Leipzig, 1993] and W.Engelhardt [“Planeten Monde Ringsysteme”, Birkhäuser Verlag, Basel, 1984].

1.1. The Solar System

The sun is the central celestial body of our solar system. Eight planets move on nearly circular orbits around it. The four inner planets (Mercury, Venus, Earth and Mars) consist mainly of solid material; the four outer planets (the gas giants Jupiter, Saturn, Uranus and Neptune) consist mainly of hydrogen and helium gas. Between the orbits of Mars and Jupiter there is the “asteroid main belt”, a ring around the sun which contains several thousand smaller objects. Outside the orbit of Neptune there is another ring of smaller objects, called “Kuiper belt”. The outermost planet Pluto is the most famous member of this Kuiper belt. It is assumed that comets come from a region far behind the Kuiper belt, called “Oort’s cloud” at a distance of about 1.7 light years. Except for the two innermost planets Mercury and Venus, all planets have at least one moon. The big gas planets have many moons. Saturn is famous for its ring system, but also the other big gas planets have rings. Objects which circle around moons have not been found in our solar system; “sub-moons” do not exist.

Celestial bodies of our solar system are the sun, planets, moons, asteroids and comets. As soon as mankind noticed that the earth is not the center of the whole universe, speculations on the existence of life on other planets came into fashion. Today we know that the earth is the only celestial body of our solar system that carries life. Our knowledge about the solar system has considerably improved with the beginning of the age of astronautics, but we still have to find answers to many questions. Astronautics is the only proper tool to explore the solar system.

1.1.1 The Sun

The biggest nuclear reactor of our solar system. The sun is just one of about 200 billion stars in our galaxy. It consists of hydrogen (73%) and helium (25%) and a remainder of heavier atoms. Our sun is not much different from many other stars, regarding its size, its brightness and its age, but for us the sun is the most important star: it provides the life-giving energy for all living beings. A nuclear furnace at its inner core heats up the sun. Temperatures, pressures and densities in the center of the sun are great enough to sustain nuclear fusion of hydrogen atoms. Every second, about 5 billion kg hydrogen are transformed into helium, releasing an enormous amount of energy (luminous power of the sun: $3.846 \cdot 10^{26}$ Watt.)

The energy works out its way to the visible surface of the sun, the photosphere. The photosphere is a thin layer of gas, just 400 km thick. Here the diameter of the sun is 1392000 km, about 100 times as much as the diameter of the earth. At an effective temperature of 5785 K the energy is emitted into space. Above the photosphere, there is an invisible layer called chromosphere (a layer with lower radiation intensity and a temperature of about 4800 K). The sun has also an extended atmosphere, called corona. The inner corona is the location where magnetic fields hold so-called protuberances, glowing gas clouds with a life-span of 2 to 3 months. The corona is constantly changing, it spreads out far into space (several diameters of the sun). Plasma which flows in radial direction away from the sun is called "solar wind".

The sun as variable star. The photosphere shows two characteristic phenomena: sunspots and solar flares. Up to now the nature of sunspots has still not been completely explored, but it is well-known that they have something to do with magnetism. They are about 2000 K colder than the rest of the bubbling photosphere. It is remarkable that the quantity of sunspots is not constant in time. Sunspots appear with a typical frequency of about 11 years (strictly speaking between 8 and 15 years) when also the radiation from sun has a distinct maximum. During a time period of many sunspots, the sun shines brighter than normal (approximately 0.1%). The increased sun radiation has a noticeable influence on the climate of the earth.

Solar flares are eruptions which suddenly spray upward from groups of sunspots. These phenomena are accompanied by intensive emission of x-rays, they have a life-span of about 30 minutes. Even though the phenomenon of solar flares has been known for many decades, their origin is still not clear and a prediction is not possible.

Exploring the sun. Several spacecraft were designated to explore the complex physical phenomena of the sun: it started with the American probes Pioneer 5 to Pioneer 9 (interplanetary solar orbit, 1959-1968), continued with the European-American probes Helios-1 and Helios-2 (near sun orbit, 1974-1976) and Ulysses (the poles of the solar system, 1990). The American Solar-Max satellite explored the period of maximum sun activity (low earth orbit, 1990). The European probe SOHO explored the solar plasma flow from a position in space outside the geomagnetic field (it attained a halo orbit around libration point L1 of the sun-earth system, 1995).

1.1.2 The Inner Planets

Mercury. Since Mercury is always close to the sun, it can be observed either early in the morning before sunrise or shortly after sunset, otherwise the sun is too bright. During two minutes of a total solar eclipse you can have the best view to Mercury. In comparison with the earth, the innermost planet Mercury is much smaller (it has just about 5% of the mass of the earth). With its diameter of 4878 km, Mercury is intermediate in size to Moon and Mars. The images of Mercury's surface (taken by American probe Mariner 10, 1974-1975) show a lunarlike terrain, bespangled with craters. The orbit of the planet has a major semiaxis of just 0.388 AU. Consequently, the sunlight is 6.25 times more intensive on Mercury than on the earth. The surface on the planet's day side is quite hot (425°C), but Mercury is a good place to observe the sun. The night-side of Mercury (-170°C) is a good place to monitor the sun near space, for example to find smaller objects which can become dangerous for the earth.

Venus. Sometimes, at sunrise or at sunset, you can see an extremely bright star on the firmament. Probably it is Venus; the morning star and also the evening star. Sometimes Venus is so bright that people do not believe that what they see is a planet (they think it must be an airplane or something else). Observed through a good binocular telescope, Venus appears as a tiny disk with a moon-shaped silhouette. Venus is a "sister planet" of the earth when we regard size and inner composition; but at a sun distance of 0.723 AU, the solar radiation at Venus is about twice as severe as it is near the earth (1 AU). The conditions on the surface of Venus do not at all resemble the conditions on the surface of the earth. The environment on Venus is extremely unfriendly; the planet is without any life. Russian spacecraft landed on the surface of Venus and transmitted pictures (Venera 7 to Venera 14, 1970-1981). These images show sandy deserts with rocks and orange skies. The visibility on the surface is surprisingly good (several kilometers), but the skies are always overcast. Thick clouds never allow a look at the surface from orbit. Orbiting spacecraft mapped the entire surface of Venus, using radar imaging devices (the Russian probes Venera 15 and Venera 16, 1983; and the American probe Magellan, 1989). Nowadays Venus is often visited by spacecraft, but many of these spacecraft are actually on the way to another destination. Seen from the earth, Venus is the nearest planet. This is the reason why spacecraft can ideally use Venus for a gravity swingby maneuver.

Earth. Our home planet is without any doubt the "jewel" of the solar system. Its surface is covered with deep-blue oceans, colored continents, yellow-brown deserts and white clouds. The earth is the only planet in our solar system that carries life. Our planet is the most important space destination: just for a few exceptions, nearly all space missions are "missions to planet earth". Astronautics helps to make our life easier: satellites are used for broadcasting, weather forecast, navigation, surface monitoring, geography, ecology and so on. Space stations in low earth orbit allow the study of biology and material science. Astronauts report that the earth seen from space is extraordinarily beautiful, but it appears also as very vulnerable. Unfortunately, there is also a large number of payloads that have a military character.

Moon. The wonderful big moon of planet earth is actually an uninhabitable desert. In a perfect full-moon night, the moonlight is so bright that you can even read a book. You can see mountains, big impact craters, bright and dark regions (“mare”) on the moon by the naked eye; and with a good binocular telescope in your hand and some fantasy in your mind, you can walk on the moon through valleys and over mountains.

Sometimes the moon looks like a disk, sometimes like a crescent; but an observer on the earth sees always the same side of the moon: tidal forces have coupled the rotation period of the moon with the period of its orbit around the earth. An astronaut on the surface of the moon sees the daily rotation of the earth, but he sees the earth always at the same position in the sky (there is no “earthrise” or “earthset” on the moon).

For life on the earth, the moon and its phases are much more than a pretty spectacle: the moon influences our weather and is responsible for tides (together with the sun). Some people state that the existence of a big moon is necessary for terrestrial life. With its gravity the moon stabilizes the orientation of the rotation axis of the earth.

The manned lunar missions of the American Apollo programme explored the moon thoroughly (Apollo 11 - Apollo 17, 1969-1972). For a long time their success has prevented new lunar missions; but some small spacecraft have recently started the revival (US-Clementine, 1994; Lunar Prospector, 1998; and Japan’s probe Lunar-A, 1999).

Mars. When you want to find Mars in the night sky, you have to look for a bright star with reddish yellow (orange) color. Planet Mars is much brighter than some other (fixed) stars which are similar in color. With the aid of a small telescope you can see the disk of Mars, but you cannot see any surface details. Only the camera of a space probe makes surface details visible. Mars has the highest volcano of the solar system (called “Mons Olympus”); Mars has a deep canyon (called “Valles Marineris”). Like other planets, Mars has impact craters, basins and mountains. The geology of Mars, the history, the composition of surface material, the climate and meteorological phenomena are now subject of intensive studies.

In astronautics, Mars is probably the most attractive planetary space destination. Its desert surface has been visited several times by American probes (Viking-1 and Viking-2, 1975; Mars Pathfinder, 1996). In 1996, the probe Mars Global Surveyor mapped the surface of Mars with high accuracy (resolution better than 2 m per pixel). Planet Mars is the object of wild fantasies: some people dream about human settlements on Mars, with agricultural activities (“terraforming”). Unfortunately, up to now no-one really started with the concrete preparation of a manned mission; but at least, unmanned missions are often on the way or in the state of preparation.

Phobos and Deimos. Mars has two tiny moons: Phobos ($14 \times 11 \times 9$ km) and Deimos ($8 \times 6 \times 5$ km); fear and terror which accompany war. These bodies are probably captured asteroids (at least they are very similar to asteroids concerning size and composition). They surround the equator of Mars on nearly circular orbits: Phobos in 7.6 hours at a radius of 9370 km; Deimos in 30.3 hour at 23460 km.

1.1.3 Asteroids and Comets

Main belt asteroids. Denoted in the year 1772, a rule named “Bode’s law” gives a good estimate for the mean distances of the planets to the sun (consider the sequence of numbers 0, 3, 6, 12, 24, 48, 96, and 192; add 4 to each and divide the result by 10). Today it is known that this sequence cannot be deduced from laws of celestial mechanics (and does not hold for Neptune and Pluto), but at that time it surprised that obviously a planet was missing between Mars and Jupiter, and stimulated the search for one. The law seemed to be satisfied when actually in 1801 a planet with about only 1000 km diameter, named 1-Ceres, was found; but some time later also many others appeared on the scene. The brightest of them, 4-Vesta, is visible by the naked eye. Today, many thousand smaller celestial bodies are known, most of them being contained in a ring surrounding the sun between the orbits of Mars and Jupiter, called “main belt”. Initially it was assumed that these bodies were fragments of a broken planet, but today it is known that in this region material does not allocate to form a planet (due to gravity forces of sun and Jupiter). These bodies consist of the very original material, which had already been present at a time when our solar system was formed. Albedo and spectrum of the asteroids (or correct planetoids) are not uniform, making a classification into several types necessary. On its way to Jupiter, the US-American probe Galileo visited two S-type asteroids: 951-Gaspra (October 29, 1991) and 243-Ida (August 28, 1993). A discovery that surprised the scientific world very much was a moon of Ida, called Dactylus.

Near earth asteroids. From the astronomical point of view, so-called “near earth asteroids” are important for two reasons. First, a visit seems to be particularly easy. In 2000, the American probe NEAR performed a rendezvous maneuver with 433-Eros. Then, the ones which come very near to the earth might be dangerous. The orbits of these near earth asteroids are monitored anxiously, even if the actual danger of an impact is very small (these asteroids move on periodic orbits, as well as the earth). When an asteroid comes close to the earth, a radar image of its shape can be taken.

Comets. There are two types of comets: periodical comets which appear in certain, well-defined time intervals; and non-periodical comets which only appear once and never show up again. The origin of all comets is the outer solar system, which constantly supplies new comets. We can conclude this from the fact that all visible comets are quite young (normally, they are just several thousand years old). When comets come close to the gravitational field of a big planet (Jupiter, for example), they can change orbit. In this way a non-periodical comet can change to become a periodical comet, when it passes the inner solar system for the very first time. Comets are often called “dirty snowball”, referring to their material composition of water ice, stones and dust. A part of the water ice vaporizes when the comet comes near to the sun, and solar radiation forms the impressive tail; but the comet loses the material. Micrometeorites, meteors and dust accompany the orbit of a comet. After several perihelion passages, all the water ice is lost; just the rocky core of the comet is left over. Finally the comet continues the now more quiet phase of its life as an asteroid.

1.1.4 The Gas Giants

Jupiter. Gas giant Jupiter is the biggest planet of the solar system. Alone it contains about 70% of the mass of all planets. When Jupiter is visible as a bright star in the night sky, it is a worthwhile spectacle if you have a good binocular telescope at hand. You can see brilliant smaller dots near the disk of Jupiter: the Galilean moons. To observe the famous “Great Red Spot” (a gas cloud on Jupiter’s southern hemisphere) you need a very good telescope. Jupiter has also a ring system and several other moons, but these objects are too small to be visible by a terrestrial telescope.

American deep space probes unveiled some of the secrets of the Jupiter system. The research started with Pioneer 10 and Pioneer 11 (1972-1973) and continued with Voyager 1 and Voyager 2 (1979). While the first probes performed flyby maneuvers, the spacecraft Galileo (1995) attained orbit around Jupiter and studied the system for an extended time period. The images of Jupiter are fascinating: colored clouds form marvelous patterns on its gaseous surface, intensive storms with thunder and lightning occur in the atmosphere. Jupiter’s moons Io, Europa, Ganymed and Callisto are similar to each other just in size, but very different in composition and topography.

Saturn. In comparison with Jupiter, Saturn has about 30% of the mass, 80% of the diameter and twice the distance from the sun; and, because of its famous ring system, it is also a bright star on the firmament. When you use a binocular telescope you can see the elliptical silhouette of the ring planet. The rings are extremely thin compared with their other dimensions; they consist of ice, stones and dust. After a swingby maneuver at Jupiter, the American probes Pioneer 11 (1979), Voyager 1 (1980) and Voyager 2 (1981) visited Saturn. They examined the ring system, found several moons and they took pictures from the pale yellow cloud structures on Saturn’s gaseous surface. The Cassini mission established orbit around Saturn in summer 2004; later a sub-spacecraft of Cassini (called Huygens) landed on Saturn’s moon Titan. Titan is the only moon of the solar system with a thick non-transparent atmosphere.

Uranus. Planet Uranus, discovered in 1781, has about the double distance from the sun than Saturn. Since Uranus is much smaller than Jupiter and Saturn (it has just 5% of the mass of Jupiter), it is a faint object in the night sky. When you want to observe it with terrestrial telescope, you have to know its exact position. Uranus appears as a light-blue disk; five moons (Titania, Oberon, Ariel, Umbriel and Miranda) with approximately spherical shape are visible. Several other moons and a ring system were discovered when Voyager 2 passed Uranus on January, 24th 1986.

Neptune. Regarding size and composition, Neptune and Uranus are similar planets. The existence of an eighth planet was assumed soon after the discovery of Uranus, because the motion of Uranus deviated from the calculated orbit. In 1846, refined orbit calculation methods made the discovery of Neptune and its moon Triton possible. Triton has a retro-grade orbit; it indicates that the origin of this moon is probably not the Neptune system (Triton is maybe a captured asteroid). When Voyager 2 passed Neptune on August 25th, 1989, it discovered a ring and several other moons.

1.1.5 Pluto and the Kuiper belt

Asteroids of the Kuiper belt. Since 1930 we have known that there is one more small planet beyond the orbit of Neptune: it is called Pluto. In 1978 Pluto's moon Charon was discovered. Since 1992 we have known that the solar system has many smaller planetoids which are farther remote than Pluto. The biggest of them, object 2003 UB₃₁₃, is even bigger than Pluto. Some dozens with diameters of a few 100 kilometers are known, many others are expected. These objects are called the "Kuiper-belt", named for an astronomer who expected this primitive-body family already in 1951 (his expectation was based on characteristics of some long period comets). Pluto, Charon but also Neptune's moon Triton are probably the rest of a bigger population, captured by Neptune in their actual orbits. Pluto is probably the only planet in our solar system which maintained nearly all its original characteristics.

Pluto. Pluto is with a diameter of 2284 km a little smaller than our moon. The solar system's smallest and most remote planet has also the highest inclination (17.123°) and the highest eccentricity ($\varepsilon = 0.255$). All this indicates a non-standard origin. For a certain time interval Pluto is nearer to the sun than Neptune. Pluto's perihelion passage was on October 11th, 1989; aphelion passage will be 124.01 years later (perihelion radius 29.67 AU, aphelion radius 49.99 AU). Observed from Pluto's surface, the sun would appear as a tiny disk with a diameter of 1 to 0.6 arc-minutes. The sunlight on Pluto's surface, however, is still brighter than moonlight on the earth during a perfect full-moon night.

NASA's Hubble Space Telescope photographed nearly the entire surface, when Pluto rotated on its axis in late 1994. Several separate images were taken by ESA's faint object camera onboard the telescope and assembled by computer image processing software. The actual map has a resolution of approximately 466 km, it covers 85% of Pluto's surface. It confirms that Pluto is a complex object with a dark equatorial belt and bright polar caps. Except for the earth, Pluto has more large-scale contrast than any other planet or moon in our solar system. This includes Pluto's often cited twin, Neptune's large moon Triton. According to the pictures of the Hubble Space Telescope, Pluto is not at all a twin of Triton. The brightness variations on the map may be the result of topographical features like basins and impact craters, but we know that Pluto has a thin atmosphere, where methane is the main constituent. Most of the surface features are likely produced by the complex distribution of frost that migrates across Pluto's surface with its orbital and seasonal cycles.

Charon. Pluto's only known satellite must be an impressive bright spectacle for every observer on the surface of Pluto. Charon has a diameter of approximately 1197 km and a orbital radius of 19640 km; thus its virtual diameter appears on Pluto's surface as mighty 5° . This is about ten times as much as our moon seen from the earth. Topographical features on Charon's surface have not been discovered yet. The orbital period of the moon is synchronized with the planets's rotation: once in 6.39 days. The "two celestial body system" Pluto-Charon has an extraordinary mass ratio: Charon's mass amounts 10%, this is more than any other moon in our solar system.

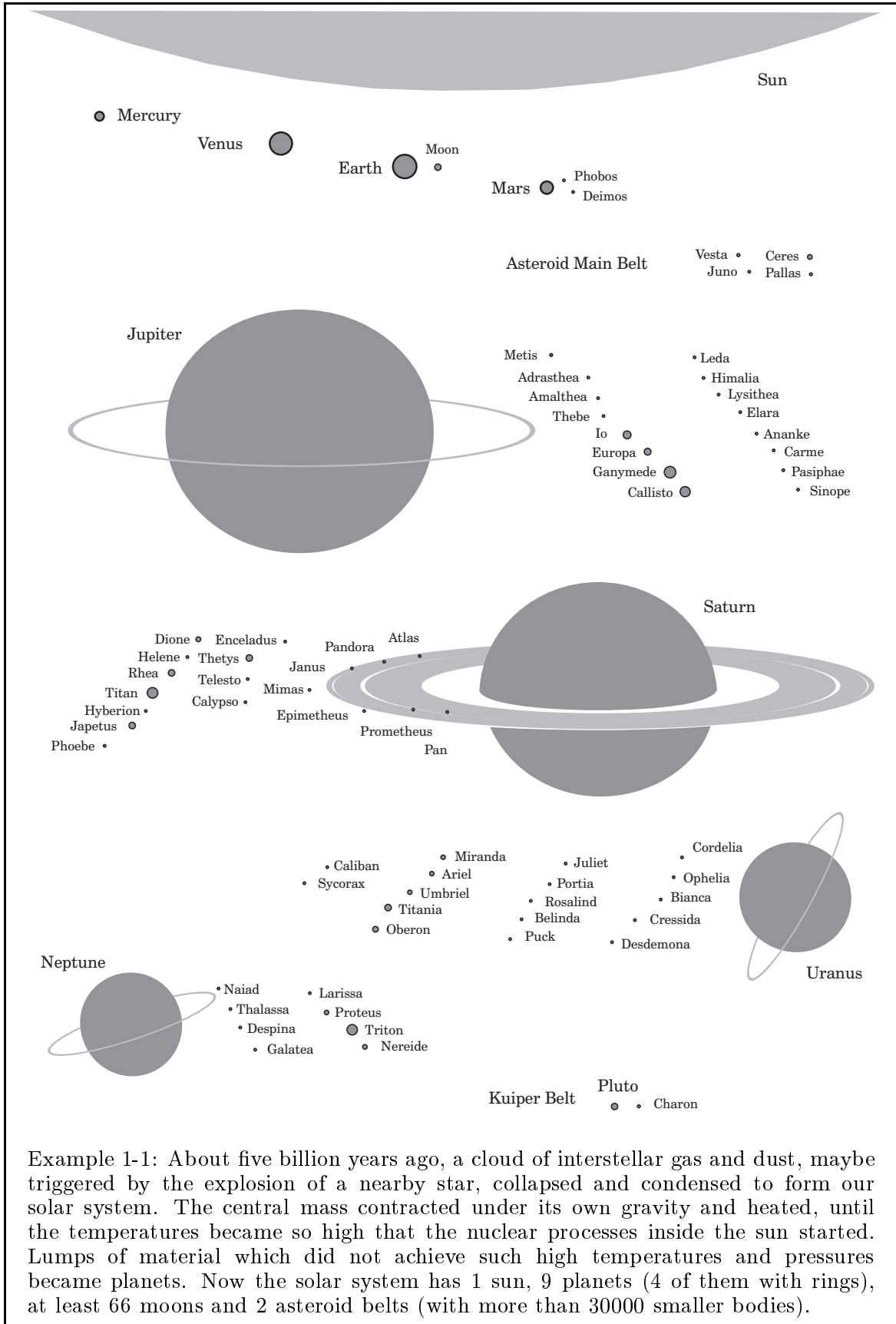
Objects of the solar system	total number	mass compared with earth
sun as central body	1	333000
all planets	9	446.6 (Jupiter 318)
all moons	at least 66	≈ 0.1
comets (Oort's cloud)	$\approx 10^{11}$	≈ 1
asteroids	at least 30000 but $\approx 10^5$	$\approx 5 \cdot 10^{-4}$
planetary ring systems	4	$\approx 10^{-5}$
meteorites, dust	-	$\approx 10^{-8}$

Planets	mass [earth = 1]	equatorial radius	rotational period [d]	app. diameter [arcseconds]	albedo	moons
Mercury	0.0558	0.382	58.646	4.6 - 12.9	0.106	-
Venus	0.8150	0.949	243.01	9.8 - 65.2	0.650	-
Earth	1.0000	1.000	0.9972	-	0.367	1
Mars	0.1074	0.532	1.0260	3.6 - 26.0	0.150	2
Jupiter	317.89	11.27	0.4135	30 - 49	0.52	16
Saturn	95.147	9.44	0.4375	15.5 - 20.5	0.47	18
Uranus	14.54	4.10	0.650	3.7 - 3.9	0.51	20
Neptune	17.23	3.88	0.768	2.2 - 2.5	0.41	8
Pluto	0.0022	0.180	6.387	0.08- 0.14	0.3	1

Moons	of planet	radius [km]	orbital period [d]	eccentricity	inclination (equ.) [deg]	major semi- axis [km]
Moon	Earth	1739	27.322	0.055	5.15 (ecl.)	384400
Phobos	Mars	14/11/9	0.319	0.015	1.02	9380
Deimos	"	8/6/5	1.263	0.0005	1.82	23460
Io	Jupiter	1815	1.769	0.004	0.04	421600
Europa	"	1569	3.551	0.009	0.47	670900
Ganymede	"	2631	7.155	0.002	0.20	1070000
Callisto	"	2400	16.69	0.007	0.28	1883000
Titan	Saturn	2575	15.95	0.029	0.33	1221850
Titania	Uranus	800	8.706	0.002	0.14	435800
Oberon	"	775	13.46	0.001	0.10	582600
Triton	Neptune	1353	5.891	0.000	157.	354800
Charon	Pluto	593	6.387	0?	98.8 (ecl.)	19600

Comets	perihelion radius [AU]	orbital period [y]	eccentricity	inclination [deg]	perihelion date
Encke	0.331	3.28	0.850	11.94	Feb 9.4 1994
Temple 2	1.484	5.48	0.522	11.98	Mar 16.8 1994
Halley	0.587	76	0.967	162.2	Feb 9.6 1986
1947 XII	0.110	∞	1.00003	138.5	Dec 2.6 1947

Asteroids	radius [km]	Spectral class	mass [10^{20} kg]	orbital period [y]	rotational period [h]	year of discovery
1-Ceres	480/453	C	11.7	4.60	9.08	1801
2-Pallas	287/263/250	U	2.79	4.61	7.81	1802
3-Juno	134	S		4.36	7.21	1804
4-Vesta	292/266/233	U	2.75	3.63	5.34	1807



Example 1-1: About five billion years ago, a cloud of interstellar gas and dust, maybe triggered by the explosion of a nearby star, collapsed and condensed to form our solar system. The central mass contracted under its own gravity and heated, until the temperatures became so high that the nuclear processes inside the sun started. Lumps of material which did not achieve such high temperatures and pressures became planets. Now the solar system has 1 sun, 9 planets (4 of them with rings), at least 66 moons and 2 asteroid belts (with more than 30000 smaller bodies).

planets epoch:3-24-1995	semiaxis a [AU]	eccentricity ε	asc. node ψ [°]	perihel φ_p [°]	inclination ϑ [°]	anomaly [°] $\theta - \varepsilon \sin \theta$
Mercury	0.387099	0.205637	48.3371	29.1124	7.0053	235.7137
Venus	0.723328	0.006747	76.6918	54.8474	3.3948	135.5220
Earth	1.000028	0.017112	354.1990	108.5165	0.0017	78.3096
Mars	1.523614	0.093443	49.5738	286.4927	1.8502	185.2087
Jupiter	5.202628	0.048371	100.4701	275.2222	1.3046	233.7331
Saturn	9.545509	0.052420	113.6511	338.0258	2.4856	259.8524
Uranus	19.282056	0.045563	74.0764	102.1806	0.7732	116.4084
Neptune	30.235456	0.007191	131.7686	235.1270	1.7707	287.4172
Pluto	39.831295	0.255112	110.3736	114.1766	17.1233	7.5334

asteroids epoch:9-9-1997	semiaxis a [AU]	eccentricity ε	asc. node ψ [°]	perihel φ_p [°]	inclination ϑ [°]	anomaly [°] $\theta - \varepsilon \sin \theta$
1-Ceres	2.7690927	.0769712	80.55944	73.25918	10.58116	185.87045
2-Pallas	2.7728107	.2328096	173.25555	309.80354	34.81354	173.50326
3-Juno	2.6693302	.2576633	170.18207	247.98232	12.96662	49.46662
4-Vesta	2.3608754	.0903554	103.96277	149.91330	7.13490	111.25959
5-Astraea	2.5772145	.1911376	141.71088	357.20124	5.36761	198.62862
6-Hebe	2.4250183	.2017319	138.89078	239.05890	14.77124	185.51936
7-Iris	2.3850563	.2305475	259.91860	145.22170	5.52355	187.29471
8-Flora	2.2015570	.1561537	111.05874	284.94617	5.88602	331.31918
9-Metis	2.3863410	.1210961	69.01416	5.70874	5.57818	325.62800
10-Hygiea	3.1357636	.1199964	283.68483	314.71559	3.84481	189.08876
11-Parthenope	2.4520184	.0997924	125.66838	194.17183	4.62259	53.44652
12-Victoria	2.3343696	.2192811	235.65538	69.29354	8.36918	110.38326
13-Egeria	2.5750848	.0870386	43.37934	80.56691	16.53068	104.80392
14-Irene	2.5878213	.1655902	86.81060	95.13860	9.10784	95.25810
26-Proserpina	2.6560657	.0881191	45.96084	194.46278	3.56143	105.44513
52-Europa	3.1005751	.1003653	129.05837	341.59927	7.46934	250.69796
177-Irma	2.7704222	.2342227	347.77359	37.92238	1.39662	6.85209
192-Nausikaa	2.4030293	.2467574	343.53297	29.78380	6.82445	68.51919
243-Ida	2.8619410	.0449065	324.42172	112.46606	1.13701	72.11222
433-Eros	1.4582573	.2229903	304.43503	178.64638	10.83095	305.01398
951-Gaspra	2.2093476	.1736485	253.36181	129.44412	4.10195	202.89332
1046-Edwin	2.9848203	.0627181	11.05380	53.39441	7.90637	53.81159
1249-Rutherfordia	2.2241198	.0766448	259.28688	222.68622	4.87295	141.97293
2703-Rodari	2.1933312	.0573875	49.59598	170.45353	6.03666	227.92885
3133-Sendai	2.1809904	.1601494	37.29982	357.96608	6.56824	330.85690
3840-Mimistrobell	2.2492033	.0830521	42.42390	167.48345	3.92165	164.37706

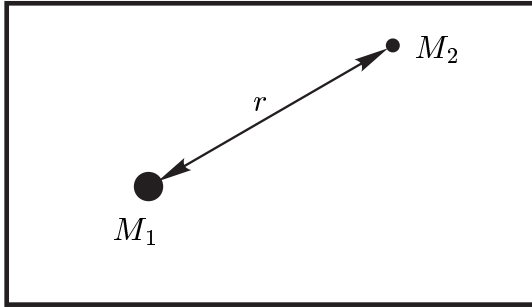
comets	semiaxis a [AU]	eccentricity ε	asc. node ψ [°]	perihel φ_p [°]	inclination ϑ [°]	perihel date month,day,year
Kopff	3.4646	0.544	120.3	162.8	4.72	7,2,1996
Wild2	3.4415	0.540	135.6	41.7	3.25	5,7,1997
Wirtanen	3.09850	0.65685	82.1948	356.328	11.724	3,14,1997
dVicoSwift	3.76845	0.43068	359.018	1.929	6.096	4, 9.466,1995
Finlay	3.57468	0.71031	42.048	323.540	3.674	5, 5.042,1995
Clark	3.11791	0.50207	59.721	208.845	9.504	5,31.245,1995
dArrest	3.48707	0.61404	138.987	178.050	19.523	7,27.362,1995
Reinmuth 1	3.76596	0.50249	119.741	13.288	8.129	9, 3.317,1995
Longmore	3.65173	0.34306	15.656	195.797	24.410	10, 9.320,1995

1.2. Gravity Fields

Gravity holds the solar system together. The existence of gravitational attraction between all material bodies was first recognized by Newton; later Einstein refined our understanding of gravitational fields. Even though we actually do not know what really causes the mutual attraction of material bodies, we can quite well compute it.

1.2.1 Gravitational Attraction of a Spherical Celestial Body

Gravitational acceleration. Regarded are two particles, for example a bigger mass M_1 and a smaller mass M_2 . The attractive force is proportional to the product of both masses and inversely proportional to the square of the distance r . The gravitational acceleration vector of mass M_2 is the gradient vector of the potential field of mass M_1 :



$$\text{attractive force} \sim \frac{M_1 \cdot M_2}{r^2} \quad (1-1)$$

$$\text{potential field} \sim -\frac{M_1}{r} \quad (1-2)$$

Using the universal gravity constant $\gamma_o = 6.672 \cdot 10^{-11} \text{ m}^3/(\text{kg s}^2)$, the gravitational acceleration (exerted by the bigger mass M_1 on the smaller mass M_2) becomes:

$$g = \frac{\gamma_o M_1}{r^2} \quad (1-3)$$

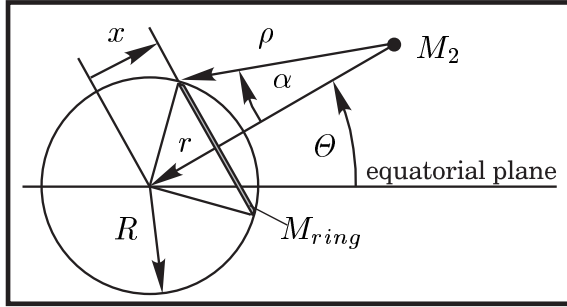
Then the potential field of mass M_1 is given by the following equation:

$$e_{\text{potential}} = -\frac{\gamma_o M_1}{r} \quad (1-4)$$

For every location inside the gravitational field of M_1 , we can find the acceleration vector of gravity from the potential field simply by coordinate differentiation. The gravitational acceleration aligned with the distance is $\partial e_{\text{potential}}/\partial r = \gamma_o M_1/r^2$; the gravitational acceleration vector has no component in any direction orthogonal to the straight line of distance r . The vector aims at mass M_1 .

Obviously, the gravitational attraction is infinitely strong when infinitely small particles touch each other ($r \rightarrow 0$), but real masses are never without size. At the moment when the small mass M_2 gets closer to the bigger body M_1 , the mass distribution of big body M_1 becomes important. Then every element of mass M_1 attracts mass M_2 .

Gravitational acceleration of a spherical celestial body. In case the mass M_1 is a celestial body, we may assume a more or less spherical mass distribution. Now we calculate as a first step the gravitational attraction g_{ring} between mass M_2 and a ring-shaped mass M_{ring} on the surface of the spherical celestial body:



$$\rho^2 = r^2 + R^2 - 2rx$$

$$\cos \alpha = (r - x)/\rho$$

Distance r is the straight line that connects particle M_2 with the center of the celestial body; distance ρ is the distance between mass M_2 and all the elements of the ring (ρ is determined if we know the coordinate x of the ring, the radius R of the celestial body and the distance r). When we integrate all the infinitesimal small attractions of the ring elements, just components count which are parallel with the straight line r . We have to use the factor $\cos \alpha$, since components orthogonal to r cancel out. Thus:

$$g_{ring} = \gamma_o \frac{M_{ring}}{\rho^2} \cdot \cos \alpha = \gamma_o M_{ring} \frac{(r - x)}{\sqrt{(r^2 + R^2 - 2rx)}^3} \quad (1 - 5)$$

The next step is to find the gravitational attraction of the entire surface of the spherical celestial body. The surface area is $4\pi R^2$, the area of the ring is $2\pi R dx$ (the ring radius is $\sqrt{R^2 - x^2}$, the arc length at this location is $\sqrt{R^2/(R^2 - x^2)} dx$). We can follow that $M_{ring} = M_{surface} dx/2R$, assuming a homogeneous mass distribution. Then we insert the relationship into the equation above and integrate:

$$g_{surface} = \frac{\gamma_o M_{surface}}{2R} \int_{-R}^{+R} \frac{(r - x) dx}{\sqrt{(r^2 + R^2 - 2rx)}^3} = \gamma_o \frac{M_{surface}}{r^2} \quad (1 - 6)$$

Regarding its gravitational attraction, the spherical surface of the celestial body behaves as if it were collapsed on the center of the celestial body. Without further calculations we may conclude that then the whole celestial body behaves the same. The only requirement is that the celestial body must consist of concentric spheres with homogeneous mass distribution. Right in the midst there is the center of gravity. The body attracts as if its entire mass were concentrated on this center of gravity:

$$g_{celestial\ body} = \frac{\gamma_o \cdot M_{celestial\ body}}{r^2} \quad (1 - 7)$$

The gravitational acceleration is a linear function of the mass of the celestial body. We can simplify the notation for the gravitational acceleration when we introduce individual gravitational constants for every celestial body: $\gamma_{planet} = \gamma_o \cdot M_{planet}$.

1.2.2 The Potential Field of Oblate Celestial Bodies

Oblateness of celestial bodies. Celestial bodies are never perfectly spherical. Particularly if a planet rotates rapidly, it has a pronounced lentiform deviation from the spherical shape. The radius of the planet is larger at the equator than at the poles. The oblateness is distinct for the gas planets (Jupiter, Saturn, Uranus and Neptune). Since the planets Mercury and Venus rotate slowly, their oblateness is very small; but the earth completes one full rotation in 24 hours and Mars in 24.6 hours, consequently these planets bulge at the equator (the earth \approx 21 km; Mars \approx 17 km).

Accurate representation of the field of gravity. We return to equation (1-4). It describes the potential field $e_{potential}$ of a spherical celestial body as function of the distance r from the center of attraction; but equation (1-4) is just an approximation for oblate celestial bodies. A more accurate approach considers distance r and latitude angle θ of every location inside the field. Then the potential is given by:

$$e_{potential} = -\frac{\gamma_{\circ} M_{planet}}{r} \cdot \left[1 - J_2 \cdot \left(\frac{R_{planet}}{r}\right)^2 \cdot \left\{ \frac{3}{2} \sin^2 \theta - \frac{1}{2} \right\} \right. \\ \left. - J_3 \cdot \left(\frac{R_{planet}}{r}\right)^3 \cdot \left\{ \frac{5}{2} \sin^3 \theta - \frac{3}{2} \sin \theta \right\} \right. \\ \left. - J_4 \cdot \left(\frac{R_{planet}}{r}\right)^4 \cdot \left\{ \frac{35}{8} \sin^4 \theta - \frac{30}{8} \sin^2 \theta + \frac{3}{8} \right\} \dots \right] \quad (1-8)$$

Planet	M_{planet} [kg]	R_{planet} [km]	J_2	J_3	J_4
Mercury	$3.3022 \cdot 10^{23}$	2439.7	0.00008	-	-
Venus	$4.8690 \cdot 10^{24}$	6051.9	0.000006	+0.0000078	-0.0000016
Earth	$5.9742 \cdot 10^{24}$	6378.140	0.0010826268	-0.0000025356	-0.0000016234
(Moon	$7.35 \cdot 10^{22}$	1738	0.0002024	+0.0000089	-0.0000117)
Mars	$6.4191 \cdot 10^{23}$	3397	0.0019592	+0.0000296	-0.0000102
Jupiter	$1.8988 \cdot 10^{27}$	71492	0.014697	+0.0000014	-0.000584
Saturn	$5.6850 \cdot 10^{26}$	60268	0.016331	-	-0.000914
Uranus	$8.6625 \cdot 10^{25}$	25559	0.003516	-	-0.000032
Neptune	$1.0278 \cdot 10^{26}$	24764	0.003708	-	-
Pluto	$\approx 1.5 \cdot 10^{22}$	1151	-	-	-

In the equation above, term R_{planet} is the equatorial radius of the considered celestial body. The expressions in curly braces are called “Legendre functions” of argument $\sin \theta$ and degree 2,3,4... (without degree 1). Factor J_2 is the most important factor, it describes the flattening of the celestial body (the body is rather “lentiform” than exactly “ball-shaped”). Factor J_3 is required if the radius of the north pole is different from the radius of the south pole (then the body is also a little bit “pear-shaped”). If the celestial body bulges at the equator and at the poles, term J_4 is required. The series can be continued, but J_5 is already not significant any more. Usually, also the mass distribution along the equatorial line is not exactly homogeneous. Then a similar series can be established for the potential field as function of the longitude.

1.3. The Structure of the Atmosphere

The earth has a thin layer of gas above its surface; we call it the “atmosphere”. The atmosphere protects life on the earth; it provides oxygen for the animals and carbon dioxide for the vegetation. Accurate knowledge about the structure of the atmosphere is very important to space engineers: launchers experience atmospheric drag on their way to space; and reentry vehicles use the atmosphere for braking.

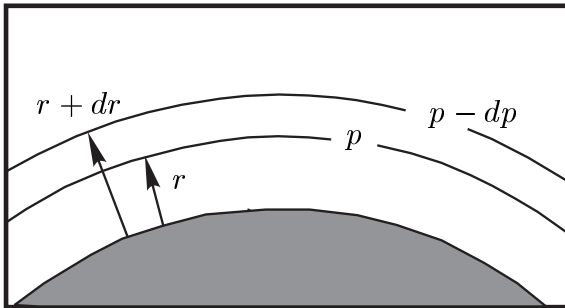
1.3.1 Barometric Scale Height

Atmospheric pressure. The motion of gas molecules is a function of the gas temperature T . The hotter the gas is, the faster its molecules move. Now we consider gas inside a gas container. When many molecules hit the wall of the gas container, these impacts are noticed as a gas pressure p . The pressure is a function of the temperature, since it depends on the impact velocity of the gas molecules. When we increase the density ϱ of the gas by a reduction of the volume of the container, the number of molecules that hit the container wall increases correspondingly: the gas pressure p is also a function of the gas density ϱ . This relationship can be expressed as:

$$p = \Re \cdot T \cdot \varrho \quad (1 - 9)$$

Term \Re is the gas constant (it is the Boltzmann-constant $\Re_0 = 8.3143 \text{ J/molK}$ divided by the mean molecular mass \mathcal{M} of the gas). If a gas fulfills condition (1-9) exactly, it is called a “perfect gas”. For example, the terrestrial atmosphere is almost such a perfect gas ($\mathcal{M} = 29 \text{ gram/mol}$ and $\Re = 287.05287 \text{ J/kgK}$).

Barometric representation. The atmosphere is tied up to the celestial body by the gravitational attraction. Now we consider an infinitesimal thin layer of the atmosphere as a gas container. If we move upwards and change the radius from r to $r + dr$, the pressure diminishes by dp , since the gas in the layer has the specific weight $g \cdot \varrho$:



$$dp = - g \cdot \varrho \cdot dr \quad (1 - 10)$$

Using the perfect gas law (1-9), we find:

$$\frac{dp}{p} = - \frac{g}{\Re T} \cdot dr \quad (1 - 11)$$

The gravitational acceleration g is a function of the distance r from the center of gravity. However, in comparison with the whole celestial body the atmosphere is very thin, usually; and thus we make nearly no mistake when we approximate g by the constant value g_0 (in case of the earth $g_0 = 9.81066 \text{ m/s}^2$).

Scale height. It follows from equation (1-11) that the atmospheric pressure is an exponentially declining function of the altitude. The scale height h_{scale} is introduced as:

$$h_{scale} = \Re T/g \quad (1-12)$$

The perfect gas law (1-9) proves a geometrical interpretation of the scale height: h_{scale} is the thickness of a hypothetical atmosphere, where the “sea-level” conditions (density and temperature) do not change with the altitude. For the earth we can calculate a scale height of $h_{scale} = 8.431$ km (assuming 288.15 K ground temperature); but pressure, density and temperature are functions of the altitude. When we assume that at least the temperature remains approximately constant, we can readily integrate equation (1-11). We find a formula which defines the aerostatic pressure p as function of the altitude (expression $r - r_o$). Then a similar relationship must apply to the air density ϱ , since the density is a linear function of the pressure for a perfect gas with constant temperature. Therefore we can denote:

$$p(r) = p_o \cdot e^{-(r - r_o)/h_{scale}} \quad (1-13)$$

$$\varrho(r) = \varrho_o \cdot e^{-(r - r_o)/h_{scale}} \quad (1-14)$$

On the surface of the earth (sea-level radius $r_o = 6378$ km) we have an aerostatic pressure of $p_o = 101325$ N/m² and an air density of $\varrho_o = 1.225$ kg/m³, approximately. Equation (1-13) is called “barometric representation” of the atmospheric pressure, but actually the real pressure deviates pretty much from the value given by the barometric representation. Particularly the assumption of a constant atmospheric temperature is not correct. The atmosphere consists of several layers with individual temperature profile, where the radiation from sun and earth is differently absorbed. The atmosphere is continuously in motion and changes its content of dissolved water (there are wind, clouds, rain, hail, snow and many other meteorological phenomena). The climate of a particular geographical location has a considerable influence on the properties of the atmosphere. Pressure, temperature and density vary continuously.

1.3.2 The Earth’s Atmosphere

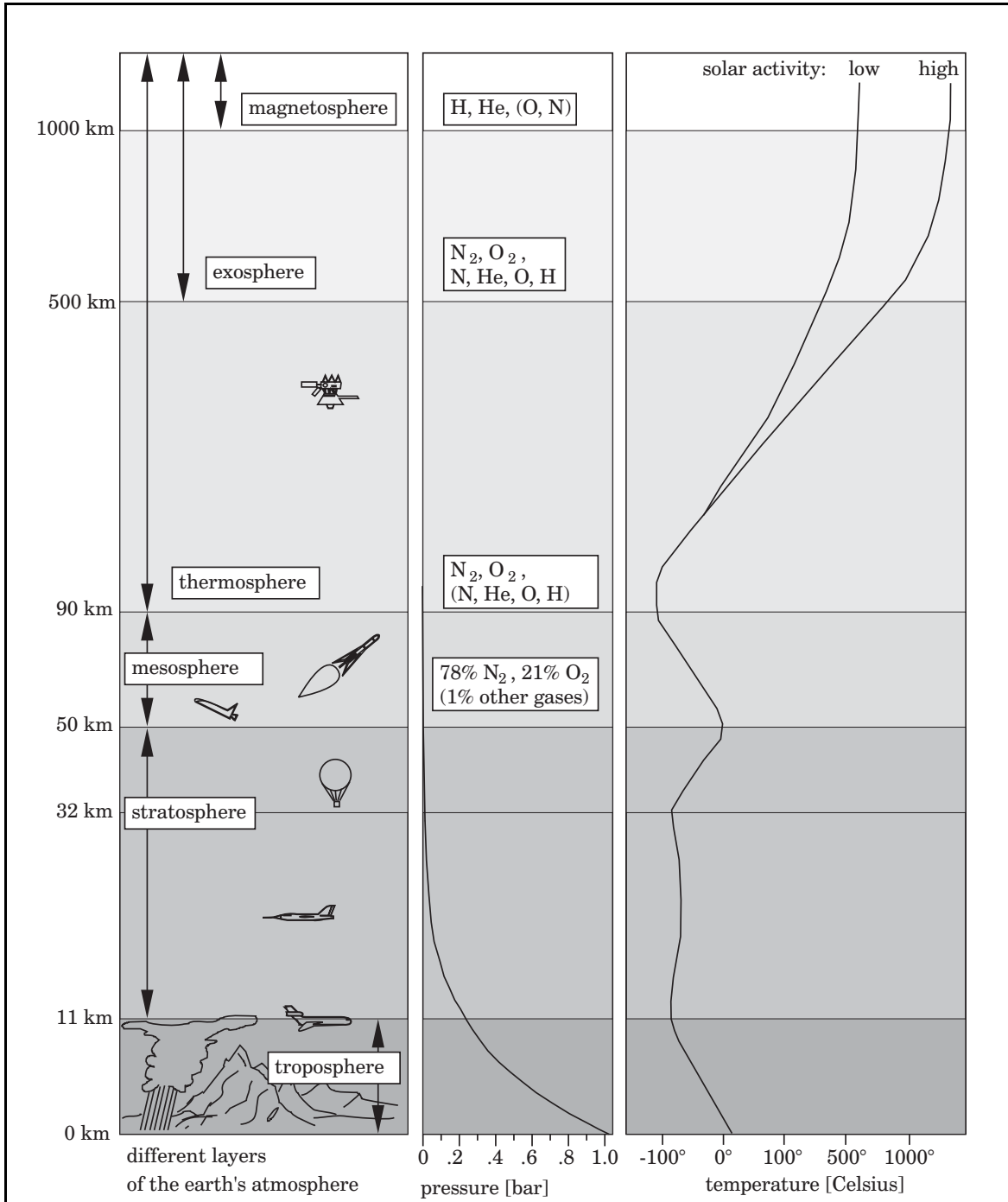
The troposphere. The lowest layer of the atmosphere is called troposphere; it is the zone of weather phenomena. It ranges from sea-level up to an altitude of about 11 km (the highest altitude of the troposphere depends on the geographical latitude, it is higher on the equator and lower on the poles). The troposphere is characterized by a declining temperature profile: when we climb, it gets colder. The temperature gradient is not exactly constant, it ranges from 5° up to 10° for every kilometer. A typical profile for a summer day in Europe is a linear drop of the temperature from 15° Celsius at sea-level to −56.5° at 11 km altitude. The pressure declines exponentially from 1 bar at ground level to 0.226 bar at 11 km altitude. The molecular composition of the air is almost invariant: the atmosphere consists of 78% nitrogen, 21% oxygen and 1% other gases (water steam, carbon dioxide, some inert gases). The air composition nearly does not change up to an altitude of 90 km.

The stratosphere. The next two layers of the earth's atmosphere are both together called stratosphere. The stratosphere ranges from the top of the troposphere (about 11 km) up to an altitude of about 50 km. Its first layer is characterized by a nearly constant temperature profile (approximately -56.5° Celsius from 11 km up to 20 km). This layer absorbs the infrared radiation of the earth. The second layer absorbs the ultraviolet radiation of the sun, and the temperature increases slightly with the altitude (approximately 1° Celsius for every kilometer). For example, at an altitude of 32 km there is the aerostatic pressure of just 0.009 bar and the air density of 0.0132 kg/m^3 , at a temperature of about -44.5° Celsius. Jet airplanes and balloons can access the stratosphere. Sometimes airplanes encounter the stratospheric "jetstream", a strong horizontal wind which blows constantly in one direction. From the meteorological point of view the stratosphere is stable: without vertical heat convection there are no clouds in the stratosphere (in contrast to the troposphere).

The mesosphere. The mesosphere ranges from top of the stratosphere ($\approx 50 \text{ km}$) up to an altitude of about 90 km. Altitudes higher than 90 km are called "space". Thus, the mesosphere contains all layers of the atmosphere between stratosphere and space. The mesosphere is characterized by a slightly decreasing temperature profile. Aircraft and balloons can just exceptionally enter this airspace; space launchers and reentry vehicles can remain inside the mesosphere just for a short time period.

The thermosphere. In space, molecular dissociation and other photochemical reaction processes are dominant, and the composition of the rest-atmosphere changes with the altitude (the mean molecular weight becomes smaller with the altitude). The rest-atmosphere reacts sensitively on solar radiation. Consequently, the atmospheric properties (density, temperature, composition) are functions of the intensity of the sunlight: they depend on the hour of the day, on the geographical latitude, on the season, and on time-variant activity of the sun (sun spots and solar flares). The density can be several times higher on the day-side than on the night-side. Particularly uncertainties in the prediction of solar flares complicate seriously the calculation of the drag force for a satellite on orbit. The thermosphere begins in space (90 km) and ends at an altitude of about 500 km. Referring to the predominance of positive ions and free electrons, the thermosphere is also called ionosphere.

The exosphere. Collisions between particles become rare above 500 km altitude, where the exosphere begins. The motion of charged particles is dominated by the magnetic field of the earth; and neutral particles move on conic orbits. Above 1000 km we find nearly no oxygen or nitrogen molecules any more (just atoms). The higher we climb, the less we can encounter heavy atoms. Above 10000 km, hydrogen is practically the only constituent of the atmosphere (protons and hydrogen atoms). In the exosphere, these hydrogen particles can reach a velocity faster than the earth escape velocity (about 11 km/s): the hydrogen gas disappears from the earth's atmosphere. Like in the thermosphere, the concentration of particles in the exosphere depends on the solar activity. The exosphere ends at about 50000 km, where we practically cannot encounter particles from the atmosphere any more.



Example 1-2: The earth's atmosphere consists of several layers with different temperature profile. The troposphere, the lowest layer, is the region of weather phenomena. The stratosphere, on top of the troposphere, is a meteorological stable region which can be accessed by high altitude airplanes and balloons. The transition region between stratosphere and near-earth space is called mesosphere. In near earth space we encounter hot ionized gas of extremely low density, called thermosphere. The uppermost layer of the atmosphere is called exosphere; it is a region where still light weight gas particles can be encountered. Always some particles are faster than the escape velocity of earth, thus gas from the exosphere is constantly lost into space.

1.3.3 Atmospheres of other Celestial bodies

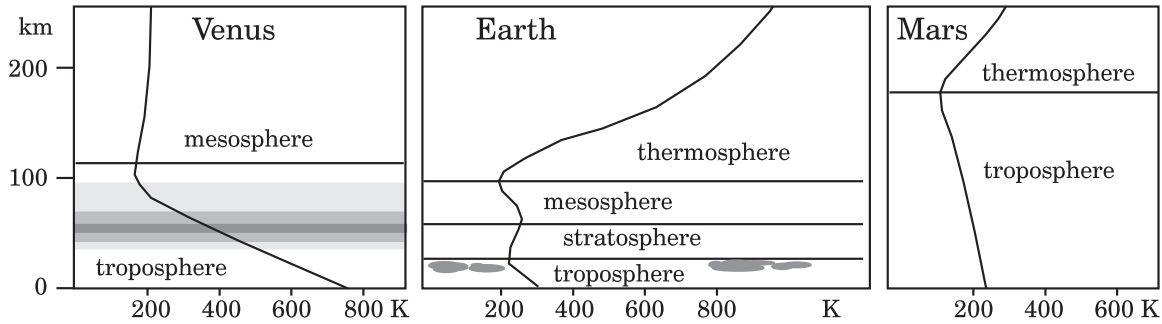
Conditions for having an atmosphere. What we call exosphere is a gas with extremely low density. Its temperature is a function of the average velocity of the gas particles, where collisions between the particles are rare. Some particles move slower, some others move faster. Always, some particles are faster than the escape velocity of the celestial body: these particles leave the gravitational field on a hyperbolic orbit.

As a consequence, the exosphere will rarefy until all the gas has escaped into space (detailed analysis shows that the permanence of the exosphere is proportional to the expression $\exp(v^2/(2\mathfrak{R}T))$, where v is the parabolic escape velocity, \mathfrak{R} the gas constant and T the gas temperature). Thus three factors increase the probability that a celestial body has an atmosphere: the mass of the planet (velocity v); the molecular weight of the atmospheric gas (higher molecular weight means a lower gas constant \mathfrak{R}) and the degree of coldness of the gas (the temperature T inside the exosphere). In case of the earth, hydrogen is the dominant constituent of the exosphere, but hydrogen disappears from the exosphere into space. Continuously the lost hydrogen is replaced (by photochemical splitting of water molecules, by volcanic eruptions, solar wind, and other processes). Planet Venus has about the same mass as planet earth; but Venus can keep a much denser atmosphere because the temperatures inside its exosphere are much colder. It is well-known that our moon has no atmosphere. With a diameter of 2284 km Pluto is smaller (approximately 2/3) than our moon, but the outermost planet Pluto is so cold that it can even keep a methane atmosphere.

Origin of atmospheres. The atmospheres of the gas giants Jupiter, Saturn, Uranus and Neptune are called “primary atmospheres”, because they consist of original gas. This gas was already present when the planets were born. Probably, these big planets have no defined surface (the density of the gas is more or less a continuous function of the depth). Main constituents of the atmospheres are hydrogen (more than 90% of the volume) and helium; methane and traces of other constituents are insignificant. The existence of aerosols (ammonia) is responsible for clouds. Particularly in case of Jupiter, strong winds form impressive patterns of clouds (strongest winds blow in equatorial regions, with 150 m/s in case of Jupiter and 500 m/s in case of Saturn).

Since the gravitational fields of Venus, Earth and Mars are considerably smaller, these planets have lost their primary atmospheres (original hydrogen and helium gas); but volcanic eruptions replaced the lost gas and formed secondary atmospheres. Chemical, physical and geological processes (and in case of the earth also biological processes) have changed the original composition of these atmospheres substantially. Until now the atmospheres of Venus, Earth and Mars are changing continuously.

The gravity of planet Mercury is too small to keep gases. Also moons of our solar system are too small for having real atmospheres (except for Saturn's moon Titan). When the volcanic processes had stopped on these celestial bodies, they started losing their secondary atmospheres; but ionized gas atoms are permanently liberated by radioactivity, and now there are extremely thin atmospheres above their surfaces.



The Venus atmosphere. The troposphere of the atmosphere of Venus is much hotter than the troposphere of the earth. At the surface we find temperatures around 464° Celsius. The surface temperature varies slightly with the latitude and with the day-night cycle (a Venus day lasts 117 earth-days). The aerostatic pressure is 95 bar; there is nearly no wind and no other meteorological phenomenon. The main constituents of the atmosphere are carbon dioxide (96.5% volume) and nitrogen (3.5%). Thick clouds cover the sky at altitudes between 48 km and 70 km. They are formed by droplets of concentrated sulphuric acid. The clouds surround Venus once in a time period of four earth-days: the cloud cover rotates faster than the rest of the planet. The mesosphere begins at an altitude of about 110 km, where temperature conditions are similar to the conditions inside the mesosphere of the earth; but when we move upwards towards the exosphere of Venus, the temperature does not increase much. Carbon dioxide and clouds prevent heat radiation from the surface of Venus and thus the exosphere of Venus is much colder than the exosphere of the earth.

The Mars atmosphere. Mars has a thin cloudless atmosphere which consists of carbon dioxide (95.3 volume %), of nitrogen (2.7%), argon (1.6%) and oxygen (0.1%). Unfortunately, there is nearly no water on Mars. Between 0.007 bar and 0.010 bar, the surface pressure is comparable with the aerostatic pressure at an altitude between 30 km and 33 km above earth. This is well sufficient for an aerodynamic braking maneuver of an entry vehicle. The mean surface temperature is -53° Celsius, but the temperature varies substantially with the day-night cycle (about 50° Celsius) and with the latitude (about 90° Celsius). Even positive degrees Celsius can be encountered on a summer day near the equator of Mars. The poles are covered with carbon dioxide ice (maybe there is also some water ice). The temperature variations are responsible for strong winds on Mars, which cause dust clouds and change the topography of Mars.

The atmosphere of Saturn's moon Titan. The only moon in our solar system with a thick non-transparent atmosphere is Saturn's Titan. The pressure at Titan's surface amounts 1.6 bar at cold temperatures (about -180° Celsius). The atmosphere consists of nitrogen gas with small fractions of methane and other substances which form haze and clouds. Hydrogen gas from Titan's exosphere gets lost into space.

Pluto's atmosphere. Planet Pluto has also an atmosphere above its ice cover. The atmosphere is extremely cold, thin and consists mainly of methane. Surface pressure and temperature vary between $200 \cdot 10^{-6}$ bar at -213° Celsius (near sun) and $0.2 \cdot 10^{-6}$ at -233° Celsius (distant sun); planet Pluto moves on an eccentric orbit. The colder it is, the more methane is frozen and the less dense is Pluto's atmosphere.

1.4. The Structure of the Magnetosphere

Some celestial bodies are surrounded by a magnetic field. The reasons for planetary magnetism are not completely explored (responsible are probably electrical currents inside the planet). The magnetic fields are strongly influenced by particle radiation from the sun. There is a region around a planet where its magnetic field dominates the motion of charged particles. We call this region the magnetosphere of the planet.

1.4.1 Planetary Magnetic Fields

Magnetic dipole moment. The earth has such a magnetosphere. The strength of the magnetic field of the earth is determined by its dipole moment, $8 \cdot 10^{25}$ Gauss cm^3 . Usually, the magnetic dipole inside a planet is not coincident with the rotation axis. In case of the earth, the inclination is about 11° ; the dipole is displaced about 500 km from the center of the earth. Astonishingly, the moon and the other inner planets Mercury, Venus and Mars have nearly no magnetic field (in every case, the magnetic field of the earth is more than 1000 times stronger); but all the big gas planets are mighty magnetic dipoles. We find the strongest planetary magnetism near Jupiter, $1.6 \cdot 10^{30}$ Gauss cm^3 . The magnetism is responsible for intensive particle radiation.

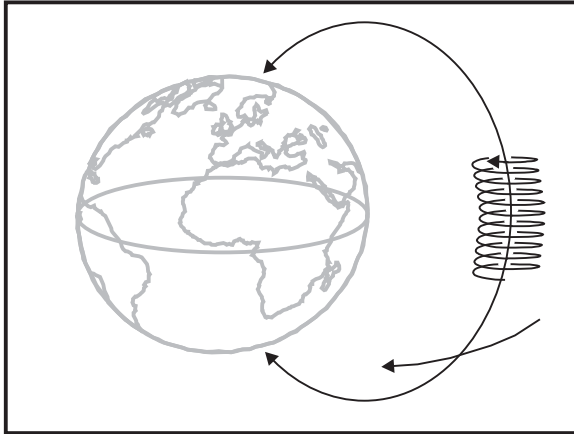
Orientation of the magnetic dipole. Planetary magnetism is not time invariant. Often, aircraft make use of the so-called “magnetic heading angle” for navigation. Geographical information on navigation charts has to be changed in a typical time interval of 10 years, because the location of the magnetic north pole moves slowly. The earth reverses the polarity of its magnetic dipole in about 200000 years.

Magnetic field lines. Outside the planet, the magnetic dipole is noticed as a magnetic field. The field lines are an infinite number of imaginary curves which connect the magnetic poles from the south to the north. The magnetic field is strong in the vicinity of the planet; more distant the field is weaker. Far away from the planet the magnetic field lines are deformed by radiation from the sun: the solar wind consists of moving protons which, on their part, have also a magnetic field.

1.4.2 Magnetism and Particle Radiation

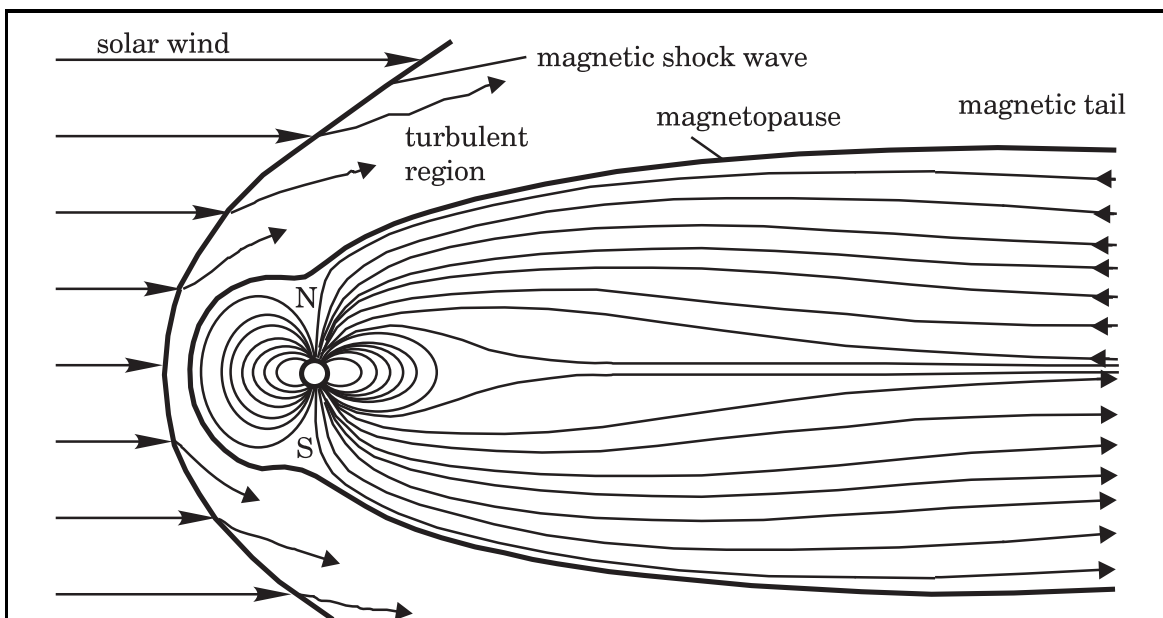
The radiation belts. The hot ionized gas inside the magnetosphere (called plasma) consists mainly of protons and electrons. The particles move rapidly, but the plasma is sort of trapped inside the magnetosphere. In case of the earth, the hot plasma forms two radiation belts (so-called “van-Allen belts”): there is an inner belt (between $r = 7600$ km and $r = 16000$ km) which consists of high-energetic protons; and there is an outer belt of high energetic electrons (maximum intensity at about $r = 22300$ km). Nearer to earth there is the comparatively cold plasma of the inner magnetosphere.

The motion of charged particles. The trajectory of a charged particle inside the magnetic field is the result of the superposition of three individual movements. First, there is the so-called Lorentz force which causes the particle to execute a spiral movement around magnetic field lines (rotation period 0.001-0.00001 s).



Second, there is an oscillation (up and down between the poles) parallel with the field lines of the magnetic field (oscillation period about 0.1s to 1 s). Third, there is a drift of the plasma in direction parallel with the equator (rotation period about 30 minutes). Ions drift in another direction than electrons (in case of the earth protons move westward, electrons eastward).

Interaction of solar wind and magnetic field. Proton radiation from the sun perturbs the outer magnetosphere substantially. When the solar wind flows around the planet, the magnetosphere causes similar phenomena like an obstacle in a supersonic wind channel. There is a hypersonic shock wave, the magnetosphere is compressed at the day side and stretched at the night side. The “magnetic tail” on the night side ranges deeply into space (for example in case of planet earth about 1000 times the radius, in case of Jupiter even behind the orbit of Saturn). The outer magnetosphere shows a very turbulent behaviour (“magnetic storms”); all phenomena are subjected to the constantly changing strength of the solar wind.



Example 1-3: The magnetosphere of the earth is substantially influenced by solar wind, it causes a similar flow characteristic like an obstacle in a hypersonic.

1.5. Man in Space

According to our knowledge, humans are the only beings who ever managed to travel to space. However, space is a dangerous environment for man: in space humans are weak, pretentious and vulnerable. They need complicated technical systems to maintain life in space. An extensive training programme of several years is necessary to prepare astronauts for the hazards of space flight.

1.5.1 Man on the way to space and back

Accelerations. Before launch the crew feels normal weight ($1 g_o$ or 9.81066 m/s^2). Lift-off takes place with a low acceleration (for example $1.1 g_o$). Then the launcher speeds up and the crew experiences an increase of weight. When the operative stage is nearly empty, the launcher accelerates rapidly and weight seems to be several times higher (for example $3 g_o$, sometimes up to $7 g_o$). The stress lasts for a few minutes. Finally, after burnout of the upper stage, the astronauts feel weightlessness (zero-g). The crew is exposed to similar accelerations during the re-entry flight phase: for a time interval of a few minutes a deceleration of $3 g_o$ can be measured inside a space shuttle (or $6 g_o$ inside a capsule). Acceleration stress is uncritical if the astronauts sit in an upright position on a special arm-chair, where the acceleration vector points in direction of their eyesight. In such a position, a man can stand $10 g_o$ for some minutes (for shorter time intervals up to $25 g_o$). Special centrifuges have been developed for the training of astronauts, in which launch accelerations can be simulated.

Noise and vibrations. Typically, 0.3% to 0.8% of the power of rocket engines is released in form of noise. In case of big space launchers the sound pressure level is several hundred times higher than the sound pressure level of civil jet airplanes. Unprotected exposure of a space crew would lead immediately to permanent deafness. The structure of the cabin and the helmet reduces the noise to an acceptable level; but still the noise is distracting for the crew. Inside the helmet, the launch is as loud as for example a train which passes nearby. Astronauts reported that the firing of an attitude control thruster appears to be as loud as a shotgun blast. Also low frequency vibrations can cause discomfort to the astronauts, particularly if some organs of the body resonate in response (the body as a whole resonates with a frequency between 3-12 Hz; the head between 20-30 Hz, the eyeballs between 40-80 Hz).

Return from space. During the reentry flight phase, ionized air flows around the vehicle with extremely high temperature. Thermal protection is required to keep the temperature inside the cabin in the allowed limits (for example below 23° Celsius). Communication with the ground station is not possible in this flight phase: the plasma air flow that surrounds the vehicle causes a radio transmission “black out”.

Space shuttles land like conventional airplanes; capsules use parachutes for landing. In case of abnormal landing, the crew needs life-saving equipment. An emergency sender must ensure that the rescue party arrives soon after a crash landing.

1.5.2 Man Living in Space

Size of the cabin. If the cabin is too small for the crew, it will cause psychological stress, claustrophobia and aggressiveness between crew members, particularly in case of long space missions. For short time intervals (maximum 1-2 days) a volume of $1 \text{ m}^3/\text{man}$ may be sufficient. If mission duration is longer (for example about a week), the crew needs at least a volume $2\text{-}3 \text{ m}^3/\text{man}$. “Comfortable” are $7\text{-}10 \text{ m}^3/\text{man}$ for long duration space missions, comparable with submarines and polar expeditions.

Atmosphere for breathing. If the astronauts carry no space suits, their cabin must be pressurized. Without pressure, humans would become unconscious after about 15 seconds and die after 2 minutes (June 29th 1971, three Russian cosmonauts died when their Sojus 11 spacecraft suddenly lost pressure). Already an atmospheric pressure higher than 0.07 bar prevents that the blood boils, but in a low pressure atmosphere a man needs pure oxygen for breathing. Man can just survive in normal air (with 21% oxygen and 78% nitrogen) if the cabin pressure is at least 0.3 bar. For structural reasons a low cabin pressure is advantageous. Conventional jet airplanes use a cabin pressure which corresponds to a flight altitude of 4 km (0.6 bar), passengers consider a low pressure as more comfortable. The American space station SKYLAB used an atmospheric pressure of 0.35 bar, with 72% oxygen and 28% nitrogen. Space cabins are never perfectly closed: a small amount of gas disappears through the seals of the hatches and cables (“outgassing”). A low atmospheric pressure in a space station has also the advantage that the effect of outgassing is automatically reduced. The cabin air must be ventilated, since it does not circulate automatically in zero-gravity.

When astronauts leave their cabin for extra vehicular activities, a space suit must carry their life support system. Sometimes, light weight and movableness of the space suits require that the pressure inside the space suit is lower than the cabin pressure. Then the normal cabin pressure must be reduced slowly, or the spacecraft must provide a special lock chamber for a slow decompression of the astronauts.

Cabin temperature. The temperature control system of the human body regulates the inner core temperature of the body always to 37° Celsius. With comfortable clothes, an ambient temperature of 20° Celsius is considered as optimal for working. Inside space cabins, the temperature ranges from 19° to 22° Celsius, typically.

If the ambient temperature deviates just a few degrees from these optimal values, man soon notices discomfort. Temperatures below 10° Celsius cause the fingers and limbs to become stiff and numb. Working is physically fatiguing at temperatures higher than 25° Celsius; the heat causes the body to sweat and slow down.

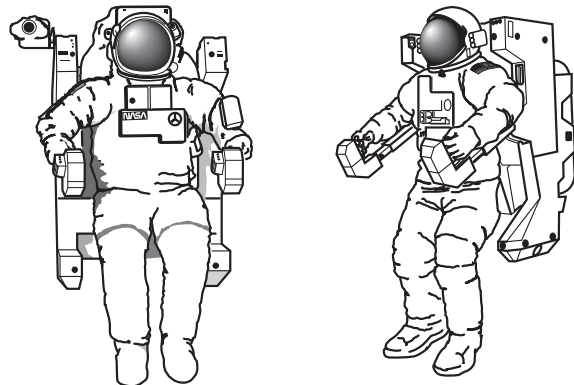
Every day the human metabolism requires food equivalent to 3.3 kWh (or 2840 kcal). The air condition system of the cabin must consider that the astronauts themselves are heaters. A normally working man produces 0.2 kW (168 kcal/h) heat energy; sleeping he produces still 0.075 kW (65 kcal/h). The heat is lost through the skin (87%), through the exhaled air in the lungs (11%), and by drinking cold water (2%).

Food, water, oxygen and hygiene. During an average working day, the human metabolism consumes about 0.630 kg of dry food, 3.344 kg of water and 0.826 kg of oxygen. The body expels 3.630 kg water (2.300 kg by perspiration and exhaled breath, 1.330 kg urine) and exhales 1.015 kg carbon dioxide. Inside a space cabin, carbon dioxide and other waste gas must be removed from the atmosphere. Filtering requires about 1.3 kg chemicals per man and per day (for example lithium hydroxide). Thus, the life of an astronaut must be supplied with about 6 kg materials every day. A comfortable bath in a shower cubicle requires about 5 kg water (a sponge is more economical). In case of a long space trip, also a laundry facility should be available.

The consumption of water is expensive, particularly in case of a long space mission. It is comparatively easy to recycle the water for washing. Sometimes also the recycling of drinking water is considered, but astronauts prefer to drink fresh water anyway (and water processing equipment is also heavy). Waste water as well as fresh water can be used in an ideal manner for radiation shielding, therefore flat water tanks should be integrated in the outer cabin walls. Drinking water can be produced from hydrogen and oxygen in fuel cells, while electrical energy is generated additionally.

Weightlessness. The feeling of weightlessness can be experienced on board of an aircraft when it flies a special “zero-g” trajectory. The feeling is strange and tensing, however, the feeling just lasts for half a minute on board of the aircraft. On board of an orbiting spacecraft, weightlessness is the usual gravity environment for the crew. The short time effect of “zero-g” is sea-sickness. The bad feeling is so strong that astronauts are incapable of working for about one or two days. Later the weightlessness is perceived as comfortable, but the serious long term effect of weightlessness is muscular atrophy and a softening of the bones. Physical training and sports can diminish this effect just partially. Particularly the reduction of calcium in the bones is an unwanted effect, decalcification can limit the duration of long time space missions. It is possible to simulate gravity inside a large space station: the station may have the form of a ring, and rotate about its main axis; but it is necessary that the radius of the station is sufficiently large (several 100 meters). If the radius is too small, other gyroscopic effects are strong and the rotation is more disturbing than helpful.

Example 1-4: Working in space outside the space cabin is extremely exhausting. Astronauts have to carry a space suit when they leave their cabin for such extra-vehicular activities (servicing, repair or erection of a space station). For the purpose of extra-vehicular activities the USA developed a special space chair (MMU, “manned maneuvering unit”). The space chair is an element of the space shuttle system; it is equipped with an autonomous life support system and a propulsion unit (control thrusters).



1.5.3 Health Hazards in Space

Radiation. A nuclear process inside the sun generates energy; this energy is emitted from the surface of the sun by radiation. Solar radiation contains electro-magnetic waves (visible light, infrared and ultraviolet light, microwaves in the radio-frequency spectra), X-rays and gamma rays, electrons, protons (solar wind), neutrons and alpha particles. Particle emission from the sun increases considerably during time periods of solar flares. The solar radiation interacts with the magnetic field of the earth and forms the so-called “van Allen radiation belt” (90% high energetic protons, 10% alpha particles). While the radiation inside the van Allen belt is high, charged particles are trapped, and thus the van Allen belt is also a protection for the earth. This protection does not exist outside the van Allen belt, for example on moon or on Mars, or on other planets without a magnetic field.

Solar radiation is not the only radiation which can be encountered in space: other high energetic rays come from processes inside or outside the galaxy, for example from the birth or death of stars or from the genesis of the universe (the “big bang”).

We distinguish between ionizing and non-ionizing radiation. The non-ionizing radiation consists of light and microwaves which can effectively be blocked by appropriate shielding. Ionizing radiation produces ions when it passes through matter. It has a damaging effect on living beings since it destroys molecules inside the cells directly. The “dose” is decisive for the grade of the destruction: radioactivity has no negative effect on man if the dose is lower than a certain (not really well-known) level. Also after having been exposed to a harmful dose, man can recuperate quite well; but if the dose exceeds a certain limit, man will witness nausea immediately and die later. The damaging effect of radioactivity seems to be more serious for animals with a higher evolutionary state; for example insects survive when the dose is hundred times higher than the dose which would be sufficient to kill a man.

On earth, life is protected from cosmic radiation by the atmosphere and the magnetic field of the earth. In space radiation can be seen as a serious health hazard for astronauts. Particularly during time periods of solar flares, the radiation from the sun can be so strong that it would kill an unprotected crew.

The radiation protection of the atmosphere and the magnetic field on the surface of earth corresponds to a layer of about 10 meters of water or 1 meter of lead (1 kg/cm^2 water). Unfortunately, such a good protection is impossible in space: normal space cabins have protection material that corresponds to a thickness of just $2\text{-}3 \text{ g/cm}^2$. This is not sufficient to protect the crew of an interplanetary spacecraft in case of solar flares: then a special protection room must be available, with about 10 times as much surface material ($10\text{-}30 \text{ g/cm}^2$, where for example drinking water containers as well as waste water containers can be used). The Apollo capsule had a protection capability equivalent to 7.5 g/cm^2 , the Lunar Module just 1.5 g/cm^2 : the crew would not have survived the strike of a strong solar flare during the moon landing mission.

Sunlight. Also the non-ionizing radiation from the sun is harmful to a space crew. Sunlight consists of 51% infrared light, 40% visible light and 9% ultraviolet light. The light is so bright that even a short glance at the sun would burn the retina of unprotected eyes. While the human eye sees light in visible spectra and senses the infrared wavelengths, the eye does not detect the most harmful ultraviolet radiation. Windows of space cabins and visors of space helmets must attenuate the brightness of sunlight and filter out or reflect the hazardous ultraviolet light. Seen from outside, windows of space cabins and visors of space helmets look like silver or golden mirrors.

Meteorites and space debris. Fortunately, there is just a small probability that a meteorite or the piece of broken satellite hits an orbiting spacecraft. If cabin pressure is lost in consequence of an impact, the crew has to slip on space suit immediately and repair the leak (assuming the leak is accessible). So-called “meteoric showers” consist of dust from comets, microscopically small particles with a mass smaller than a microgram (the occurrence of meteoric showers can be predicted in some limits). Meteoric showers can destroy sensitive instruments and optics, or even disable a deep space probe completely; but for the space crew the cabin of the spacecraft or even the space suit provides sufficient protection against the impact of micrometeorites.

The human health factor. Even the healthiest astronaut can get sick during a long space travel. Smaller health problems include sea-sickness, headache, respiratory diseases, inflammation of the eyes, diarrhea and other digestion problems. To provide help in case of a smaller disease, medicine has to be available inside the space cabin. The astronaut has also a psyche which can get sick, especially during long space trips. The drug kit must supply stimulants, sleeping tablets, pain killers and also narcotica. Also the soul needs attention: astronauts want to talk regularly to their relatives over the telephone; audio and video entertainment is necessary on long space journeys.

If a space traveling astronaut gets seriously ill he has to return to earth immediately. Since a fast return to earth is not possible during a long interplanetary space mission, the crew members should be able to perform smaller surgical operations themselves.

Example 1-5: “Star-Trek” is the name of a famous TV series. However, just in science-fiction films it is possible to travel to remote galaxies. We know that physical laws prohibit that a spacecraft can fly faster than the speed of light. Will we ever be able to explore remote star systems and find other civilizations, or will extraterrestrial creatures come to visit our planet? Surely, we cannot expect that the technology for interstellar flight will be available on earth one day.



1.6. Progress in Technology and Future Space Flight

Satellites, space stations and interplanetary probes are complicated systems; “high-tech” is the well-known basis for spaceflight. Particularly computers, but also other electronic equipment (telemetry, telecommunication and power supply devices) experience a rapid development in our present time. The advance in propulsion technology is slow (compared with electronics), but also rocket propulsion systems can still be improved if better materials and refined construction methods are used. When astronautics is considered, many proposals for future space projects appear. These proposals are considered and discussed everywhere in public (for example at congresses, in magazines and on television). You can observe that people discuss future space projects assuming a certain “basis of technology”. Often the discussion is conflicting because everyone takes a different technological environment as a basis. Certainly, some ideas are not feasible today because the technology is missing. It can be expected that more advanced projects will become possible in the future; but like always, the prediction of the future is a difficult undertaking. Nearly always the reality turns out to be different from what we expected it to be.

1.6.1 Technological Environment for Space Projects

Basis of available technology. No-one really doubts that a project is feasible if there is a very similar project that has already been executed; but also if it is predictable that the technology required for a certain project can be acquired in a short period of time, we may say that the project is based on “available technology”. For example, projects based on “available technology” do not depend on a space transport capacity which is 1000 times cheaper than today (space hotels for everyone), or new computers which are 1000 times faster than today.

Basis of anticipated technology. At least in some areas it is reasonable to expect great technological progress. Seen from today, computers will become faster, cheaper and smaller; but what are the limits? Is it possible to equip computers with human-like intelligence one day? Today we use the chemical energy of the propellant for rocket propulsion, but much more energy can be released by a nuclear process. The technology for nuclear spaceflight is not available today, but it could be available one day. Maybe the procurement of this non-available technology is impossible or will make every effort. If a spaceflight project requires such a technology, it is just fair when we state in a discussion that the project is based on “anticipated technology”. Depending on the point of view, the procurement of the technology is more or less easy; but in any case, the availability of the technology in the future is not proved.

Basis of science fiction. If a project assumes a technology that surely violates well-known physical laws, we leave the basis of “anticipated technology”. Nearly all science fiction movies fall into the category of infeasible space missions. Often, astronautical engineers study these unrealistic space projects as “mental exercises”.

1.6.2 Future Exploration of our Solar System

Manned missions to planet earth. Some space scientists dream about future cities in space: people could live there, or at least spend their vacations there. Such a space-trip would be an impressive experience; the earth is very beautiful when visitors observe it from space. Particularly the moon is an exciting place. A comfortable hotel on the surface of the moon would offer visitors incomparable holidays. To make such a travel possible for everybody, space-trips should be cheap.

Today, the feasibility of manned missions to earth orbit is sufficiently proved. The American Apollo project proved the feasibility of manned moon missions; but immense costs, high risk and serious health hazards are always involved; today they prohibit “space tourism” and “space cities”. Drastic cost reductions for space transportation (much more than factor 10) are just possible when “anticipated technology” is taken as a basis for discussion. When we take “available technology” as a basis for discussion, the ideas of cities in space or hotels on moon are unrealistic.

Manned missions to other planets. Planet Mars is the next candidate for a manned exploration. Several detailed studies exist; these studies examine how such a Mars project could be realized. All these studies state that a future manned Mars mission will be expensive (probably more expensive than the Apollo project) and dangerous for the life of the crew; but all studies also state clearly that a manned Mars mission is possible with available technology. A manned spacecraft could land on the surface of Mars approximately ten years after project initialization. Within the mission, visiting the moons of Mars (Phobos and Deimos) is possible.

Venus is the planet that moves on an orbit around sun with the nearest distance to the earth. In the past some American space probes established orbit around Venus, Russian probes (Venera 13 and Venera 14) landed on the surface and transmitted some pictures. They demonstrated that it is possible to reach the surface of Venus; but a return from the surface of Venus back to earth seems to be impossible: first, Venus has about the same gravity as earth, a large space launcher is necessary to reach orbit from surface. Then, there is the extreme environment on the surface of Venus (pressure, temperature and chemistry): the Russian probes died shortly after landing. Even when we discuss an unmanned Venus sample return mission, we must admit that such a mission is not possible if we assume the basis of available technology. Certainly, a manned Venus mission requires anticipated technology.

The American probe Mariner 10 demonstrated the feasibility of an unmanned Mercury flyby mission. Planet Mercury has a much smaller gravity than planet Venus (Mercury is bigger than our moon and smaller than Mars); but Mercury has no atmosphere which could be used for a braking maneuver. Thus, landing on Mercury would require an immense amount of rocket propulsion. For a spacecraft that comes from earth, even the establishment of an orbit around Mercury is extremely difficult. Thus, also a manned Mercury mission requires non-existing propulsion technology.

Between the orbits of Mars and Jupiter there is the “asteroid main belt”. This main belt contains most of the about 30000 known asteroids and comets; smaller celestial bodies with more or less well-known orbits. If we compare their orbits with the orbit of Mars, most of the asteroids and comets are farther distant from earth; but asteroids and comets have nearly no gravitation; and, once orbit has been established, landing on the surface seems to be easy. Considering this, a mission to a well-located asteroid or comet seems to be even easier than a mission to planet Mars; but the large majority of asteroids and comets are no targets for a manned mission.

Obviously, landing on one of the gas giants Jupiter, Saturn, Uranus and Neptune is impossible because none of these planets has a defined solid surface. Without doubt, their moons are more interesting targets for a manned planetary exploration. When we compare a manned mission to one of these moons with a mission to planet Mars, we have to realize that all these targets are extremely remote, extremely cold and located inside a strong gravity field. When we plan such a manned mission on the basis of available technology, we realize that these missions require unrealistic large spacecraft and unrealistic long travel times. To get spacecraft with a realistic size, we have to anticipate technology which does not exist today. Beyond Saturn there are Uranus, Neptune and Pluto, obviously too remote for a manned mission.

Thus, on the basis of available technology, there are just a few celestial bodies which are possible targets for manned missions: Moon, Mars and its moons and some well-located asteroids or comets. Missions to other targets require non-existing technology.

Unmanned planetary missions. The situation is different when we consider missions of unmanned space-probes: flyby missions open the entire solar system for unmanned exploration. Landing on the surface is extremely difficult if the planet or moon has no atmosphere. Landing on the surface of an asteroid or comet is easy (once orbit has been established). Sample return missions are possible in a few special situations only (again: moon, Mars and some asteroids and comets).

Terraforming. Some scientists study the possibility of exporting terrestrial life to other planets, for example to Mars or to Venus. The procedure should work as follows: genetically changed micro-organisms survive in the thin atmosphere of Mars. Then, as time goes by, these micro-organisms change the atmosphere of Mars and generate better conditions for other life on Mars. The atmosphere gets warmer and water ice (inside Mars) melts. As a next step plants are brought from earth, which can now exist there. When sufficient vegetation grows on Mars, animal life is exported. Finally, human Mars settlements are possible. Terraforming is a perfect example for a project based on anticipated technology. First, we do not know if the project is possible at all. Then, we do not know how difficult it is and how long it might take. Finally, if it is possible, we do not know whether we should do it. Today, the project is nothing else than an exciting idea. Space scientists should study such a project; but when they talk about terraforming they must clearly state that the technology for the project is not available. Otherwise the public will believe that space engineers really intend to realize the project “terraforming” soon.

1.6.3 Interstellar Spaceflight

The search for extraterrestrial life. Our knowledge about other remote solar systems is rather poor. On the basis of indirect measurements we may conclude that probably also other suns in our galaxy have planets. Obviously, some of these planets could carry extraterrestrial life. Maybe even intelligent beings could exist there. A mission to a distant star would be extraordinarily exciting.

Manned interstellar missions. It is not necessary to explain in detail why a manned interstellar mission is impossible when we take available technology as a basis for the discussion. The technology we would have to anticipate for a manned interstellar mission is far away from the technology which we have today. Certainly nuclear propulsion is required for an interstellar mission, and the technology of nuclear propulsion is not at all predictable. According to our present knowledge, flight with a velocity faster than light is impossible. Anyway, mission duration would become very long in comparison with the lifetime of humans (probably more than one generation). We hate to agree, but interstellar flight just works fine in science fiction films.

Unmanned interstellar missions. Also unmanned interstellar missions are unrealistic today, but the technology we have to anticipate is much more realistic than the technology we would need for a manned mission. A very near flyby at sun with large solar sails could give an extremely light spacecraft a velocity of up to 10% of the velocity of light, maybe. Equipped with an extremely small computer with human-like artificial intelligence and a long lifetime, such a spacecraft could fly autonomously to a near star. It could explore the distant solar system and return data to the earth. Today there is no reason why such a mission should be impossible in the future.

Visits of extra-terrestrial creatures on earth. In comparison to humans, other intelligent life can have a very advanced state, if it really exists in deep space. Some people state that earth is regularly visited by extraterrestrials. However, in spite of many reports, there is no single proof for this and there is no official statement of an acknowledged space agency about the existence of extraterrestrials. At least it is not evident that they have ever visited earth; and it is not plausible that governments or space agencies hide the proofs for their existence. Four plausible answers can be given to the question why extraterrestrials obviously do not frequently contact us:

- Intelligent life is extremely rare in space, and – if it exists – far remote.
- Interstellar flight is too difficult, even with any advanced technology.
- Every “intelligent” society extinguishes itself before interstellar flight is possible.
- Spacecraft fly frequently through our solar system, without being noticed by us.

The search for extraterrestrial life is one of the most challenging tasks of astronautics. On our planet, up to now humans are the only beings which have ever developed a technical civilization. This includes the capability of space-flight, the capability of manipulating genetic material and also the construction of atomic bombs. Finding other intelligent life in space could prove that it is possible for mankind to survive.

2. Rocket Propulsion

A vehicle which expels exhaust fumes to generate a thrust force is called "a rocket". Rocket motors are used to propel space launchers, satellites and missiles; because of immense noise and high specific fuel consumption, rocket motors are not used to propel ships, aircraft or cars. Astronautical activity, however, requires rocket motors. These motors work also in vacuum of space, in contrast to other engines; and in practice rockets are the only engines which work in space. Refueling on the way to orbit or in orbit is not possible, usually. Thus, extremely high performance is necessary to save any small amount of valuable propellant: rocket motors are high-technology power-drives. Many publications are concerned with rocket motors because of their importance to space flight. Fundamental books were written by G.P.Sutton ["Rocket Propulsion Elements", John Wiley & Sons Inc., New York, 1992 sixth edition], and M.Barrère, A.Jaumotte, B.Fraeijs de Veubeke, J.Vandenkerckhove ["Rocket Propulsion", Elsevier Publishing Company, Amsterdam, 1960].

The second chapter of this book is concerned with the technology of rocket propulsion. In the first section we want to examine the theory: the amount of propellant that a rocket can carry is limited, and this has as consequence that we need a vehicle with more than one stage to achieve a velocity fast enough to reach space. A large variety of chemical substances is theoretically usable as rocket propellant: the second and the third section of this chapter discuss performance and combustion properties of the few practically important chemicals. The expansion in a hypersonic nozzle, a complicated process that involves many physical and chemical phenomena, is studied in the fourth section. Finally we will have a look at the construction of liquid engines and solid motors, the only rocket propulsion systems which are really in use today.

2.1. Theory of Rocket Motors

One property enables rocket motors to operate also in space: rockets carry all working fluid for expulsion inside themselves. This feature makes the distinction between rocket motors and other jet propulsion systems: so-called duct engines "breathe" the gases of the ambient atmosphere, but the working fluid for rocket motors is exclusively propellant. Obviously, the amount of propellant a rocket can carry inside is limited. The theory of rocket motors examines the consequences of this property.

2.1.1 Expulsion of Propellant

Ejection principle. Consider the famous example of the hunter who is sitting in a boat on a lake. The boat is supposed to be in rest with respect to the water. Suddenly the hunter fires a shot. The bullet flies away with high velocity. After the shot has been fired, boat and hunter will not remain in rest: the gun recoil will drive the boat in the other direction, opposite to the direction of the shot.

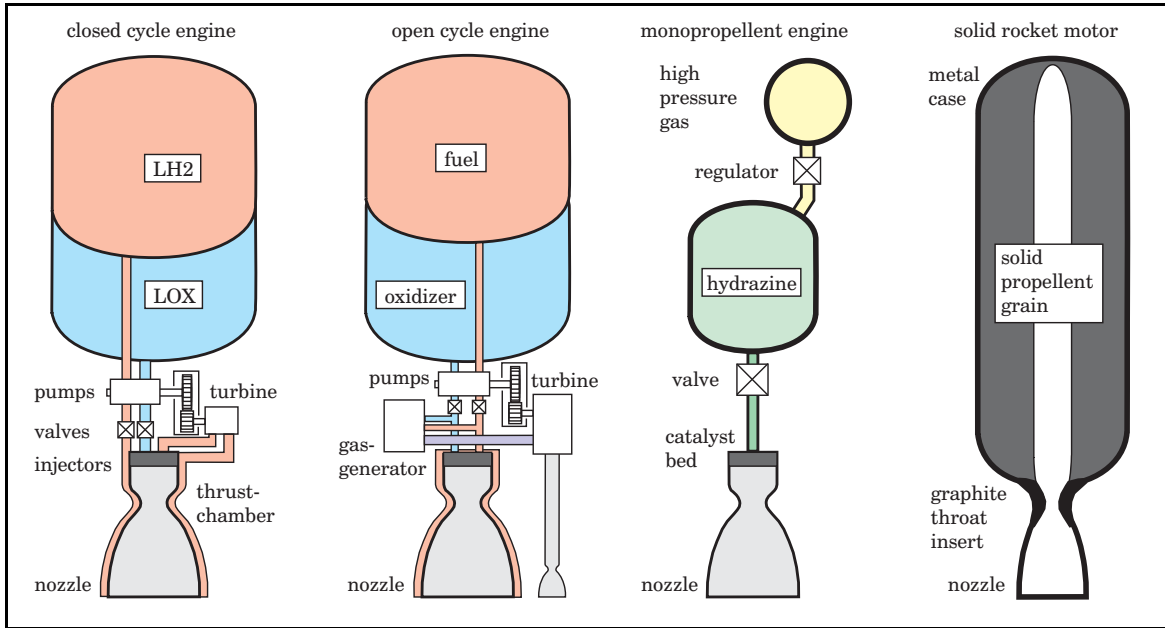
The linear momentum of a mechanical system (product of mass M and velocity \vec{v}) can just be changed by application of external forces, according to Newton's law. Before shooting the linear momentum of the whole system (boat, hunter and bullet) was certainly zero, because everything was in rest. Now, after shooting, everything is moving. The "new" linear momentum of the system must still be zero, as result of the absence of external forces. The linear momentum is a vector: the vector of the linear momentum of the bullet has the same length as the vector of linear momentum of boat and hunter. The two vectors aim exactly in opposite directions and cancel out.

In this example the bullet is the "working fluid"; and the propulsion energy is provided in the chemical form of gun powder. Note that it is possible to change the mechanical energy of the system without application of external forces. In contrast to the linear momentum ($\vec{V} = M \cdot \vec{v}$), the kinetic energy ($E = 1/2 \cdot M v^2$) is not a vector. The shot transformed chemical energy to mechanical energy (via thermal energy). After shooting the mechanical energy of the whole system is not zero any more.



The "hunter in a boat" is not a good example for a rocket, since shooting does not give a continuous thrust force, but when we exchange the gun by a fire extinguisher, we really have a rocket. We can observe that this propulsion system would also work in space. The thrust force is independent of the velocity of the vehicle. By using rocket propulsion the vehicle can achieve a velocity higher than exhaust velocity (the velocity of the expelled working fluid relative to the vehicle). We can also recognize that efficient rocket propulsion requires a high relative velocity of the working fluid.

Classification of rocket propulsion systems. A rocket motor requires working fluid (always stored inside the vehicle) and energy (not necessarily but often stored inside the vehicle). Energy is needed to accelerate the working fluid before exhaustion. The usual way is to heat up the working fluid in a thrust chamber and expand the gases in a nozzle. The energy can be provided by the working fluid itself (chemical reaction of the propellant), by solar cells (electric energy to supply an arc or resistor), by the sun directly (parabolic mirror), by a nuclear reactor (direct heating or using thermal-electric converters). Ion-engines accelerate the charged working fluid in an electric field; this avoids heating of the propellant and expansion in a nozzle. When we consider also combinations, we realize that many options exist. Of all these theoretical possibilities, in practice just two options are really used: chemical (bipropellant or monopropellant) liquid rocket engines and solid rocket engines. Despite of good potential for the future, electric propulsion systems still live a shadowy existence. Astronautics is done utilizing chemical motors. The energy is stored inside the working fluid (the propellant). A chemical reaction of the propellant releases this energy to generate heat for the acceleration of the exhaust fumes in a hypersonic nozzle.



Liquid propellant rocket motors. The combination of liquid hydrogen LH₂ as fuel and liquid oxygen LOX as oxidizer is often used as propellant for so-called “liquid rocket engines”. The stage carries fuel and oxidizer in separate tanks (or in one tank separated by a partition). For structural stability reasons a tandem arrangement of cylindrical tanks is favorable, and the tanks have to be pressurized. The engine needs pumps to feed fuel and oxidizer into the thrust chamber. Just turbines are light and powerful enough to drive these pumps. Two different ways exist to provide the thermal energy for the turbines: some more modern constructions use the heat of chamber wall cooling exclusively (closed “expander” cycle engines); more conventional constructions use a separate gas generator (open cycle engines). Closed cycle engines are more efficient but also more complicated (more expensive), and in practice this principle just works if LOX/LH₂ is used as propellant. It is necessary to cool the wall of the thrust chamber, otherwise the chamber would not stand temperatures of over 3000 K. Cooling is done via propellant that runs through small pipes at the outer surface of the chamber. Small attitude or trajectory control thrusters use a simple pressurized feed system, the chemical reaction is a catalytic decomposition of monopropellant (hydrazine). After the reaction the gases are expanded in a hypersonic nozzle.

Solid propellant rocket motors. A firework rocket is a perfect example for a “solid rocket engine”. Now the thrust chamber contains the total amount of propellant from the beginning, and pumps or turbines are not required. The solid propellant contains an appropriate mixture of fuel and oxidizer in one substance. After ignition the surface of the propellant burns, gases are generated and expanded in a hypersonic nozzle. To avoid that the structure is exposed to high temperatures for an extended time interval, the propellant burns usually from inside to outside. The burn velocity must be limited, otherwise the pressure would grow without limits and the motor would explode like a bomb. Now tank and chamber are just one unit, and the chamber pressure is consequently much smaller than in case of liquid engines (up to 60 bar).

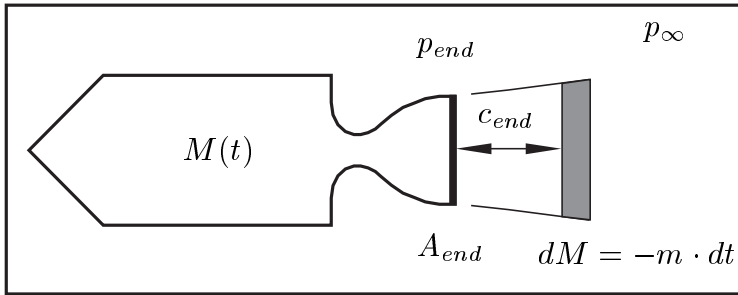
2.1.2 Thrust and Impulse of a Rocket Stage

Thrust acceleration. We consider a rocket propulsion vehicle: while the motor is operative the vehicle expels propellant, and the rocket mass M is a function of time. A snapshot is taken at the instant t . During the infinitesimal short time interval dt the infinitesimal small amount of propellant mass dM is ejected. We introduce the mass flow rate as $m = -dM/dt$ (m is also called “rate of expenditure of propellant”, or “fuel consumption rate”). Often, but not always, the mass flow rate is constant:

$$dM = -m \cdot dt \quad (2-1)$$

$$M(t) = M_0 - m \cdot t \quad (2-2)$$

If m is constant in time, we can readily integrate equation (2-1) to find equation (2-2). Term M_0 is the initial rocket mass. The exhaust fumes leave the final cross-sectional area of the rocket motor A_{end} with the average relative velocity c_{end} . It can be assumed that the final static pressure p_{end} (in area A_{end}) is different from the ambient static pressure p_∞ . The thrust force S is composed of two parts: the ejected mass dM imparts to the rocket the momentum $c_{end} \cdot dM$ (the change of the linear momentum during time interval dt is equivalent to the first part of the thrust force); and then the difference of the static pressure $p_{end} - p_\infty$ multiplied by the final cross-sectional area A_{end} is equivalent to the second part of the thrust force. We can write:



$$S = c_{end} \cdot m + A_{end} (p_{end} - p_\infty)$$

$$S = c \cdot m \quad (2-3)$$

The first part (momentum thrust) contributes much more to the thrust of a rocket engine than the second part (pressure thrust). Note that the “pressure part” can also be a negative thrust component, in case the nozzle is over-expanded ($p_{end} < p_\infty$). The thrust does not depend on the velocity of the vehicle, but on its flight altitude. In space the thrust is higher, since p_∞ equals zero in space. The notation can be simplified when we introduce the effective exhaust velocity c , a quantity that depends on the flight altitude: c is nearly a linear function of the ambient aerostatic pressure; but the effective exhaust velocity c is constant for the flight in the vacuum of space (because the available power for the expulsion of propellant is usually not a variable). We obtain the thrust acceleration s as the ratio of thrust force and rocket mass:

$$s = \frac{c \cdot m}{M} \quad (2-4)$$

Even when the thrust force is constant, the thrust acceleration is a function of time. Exhausting propellant the rocket mass diminishes, and the thrust acceleration grows.

Specific Impulse. The thrust of the rocket stage can vary during the time interval of operation. The integral of the total thrust is termed “impulse of the stage”. If we measure the impulse of a stage and divide it by the consumed propellant mass, we obtain the effective exhaust velocity (if we form $c = \int S dt / \int m dt$). Thus, the effective exhaust velocity c can be regarded as “mass specific impulse”. More in common use to characterize the performance of a rocket stage is famous so-called $I_{specific}$, the “weight specific impulse”:

$$I_{specific} = \frac{c}{g_o} \quad (\text{with: } g_o = 9.81066 \text{ m/s}^2) \quad (2-5)$$

The specific impulse $I_{specific}$ of a rocket stage has the dimension of seconds, it is an important performance parameter which can be measured without difficulties. The effective exhaust velocity c of the stage follows from multiplication of this value with the gravity acceleration g_o on the surface of the earth. We have to resist the temptation of using factor 10 instead of 9.81066: a change of the exhaust velocity by just two percent makes a lot of difference to the performance of a rocket stage.

There exists a transparent physical interpretation for the specific impulse: to lift-off from the surface of earth, the rocket stage of mass M_0 needs at least a mass flow rate of $m = M_0/I_{specific}$. Then the initial thrust acceleration s of the stage is exactly g_o . When we assume that this hypothetical stage may consume its total mass M_0 in form of usable propellant, then the specific impulse is exactly the burn time of the stage.

Ciolkovskij’s equation. The time integral of thrust acceleration s of a rocket stage (in zero-gravity vacuum flight) corresponds to a change of the velocity Δv . We form:

$$\Delta v = v_1 - v_0 = \int_0^1 s dt = -c \int_0^1 \frac{dM}{M} \quad (2-6)$$

The result of integral (2-6) is the famous Ciolkovskij equation:

$$\Delta v = -c \cdot \ln \frac{M_1}{M_0} \quad (2-7)$$

or:

$$M_1 = M_0 \cdot e^{-\Delta v/c} \quad (2-8)$$

Equation (2-7) is of fundamental importance to astronautics. Usually many forces act on a moving rocket, but Ciolkovskij’s equation was established neglecting all forces except for thrust. The result is a relationship between the mass ratio of the rocket stage (ratio of burnout mass M_1 and ignition mass M_0), the performance of its engine (effective exhaust velocity c), and the “ideal velocity” (or “delta-v” capability Δv) of the rocket stage. The Δv requirement of a maneuver can be found separately by trajectory computation, disregarding in an approximation the actual construction of the rocket. It is remarkable that the Δv capacity does not depend on the value of the thrust S itself, but on the exhaust velocity c and on the mass ratio M_1/M_0 only. If we make the thrust of a rocket stage lower, the stage just burns longer. Exclusively accelerated by thrust force, it will finally achieve the same characteristic velocity Δv .

2.1.3 Multi Stage Rockets

Payload mass ratio. We make use of a simple mass model. Essentially, the rocket stage consists of four sub-masses: payload, engine, structure (tanks) and propellant:

$$M_{ignition} = M_{payload} + M_{engine} + M_{structure} + M_{propellant} \quad (2-9)$$

$$M_{burnout} = M_{payload} + M_{engine} + M_{structure} \quad (2-10)$$

Difference between ignition mass and burnout mass is the amount of usable propellant (the simplified model considers any residual propellant as if it were structure mass). The stage will execute a flight mission which requires a particular Δv capacity. Using Ciolcovskij's equation (2-8) we find the relationship between Δv (divided by effective exhaust velocity c) and the required mass of usable propellant as:

$$M_{propellant} = M_{ignition} - M_{burnout} = M_{ignition} \cdot (1 - e^{-\Delta v/c}) \quad (2-11)$$

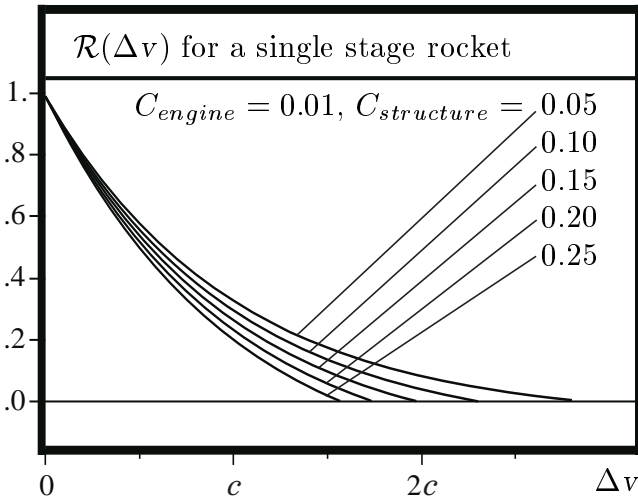
In particular, the model will analyze size effects. It will indicate what happens if we change the size of a rocket stage. Compared are stages of different size which have to perform approximately the same type of mission (for example the launching of a satellite). This premission justifies the following assumptions: approximately the engine mass is a linear function of the ignition mass; and approximately the structure mass is a linear function of the propellant mass. We may write in a simplification:

$$M_{engine} = C_{engine} \cdot M_{ignition} \quad (2-12)$$

$$M_{structure} = C_{structure} \cdot M_{propellant} \quad (2-13)$$

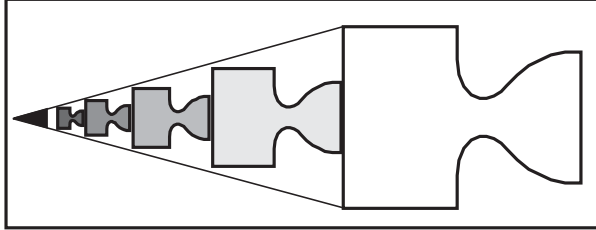
Compared to their weight liquid rocket engines develop high thrust, $C_{engine} \approx 1-5\%$. Light weight tanks may contain propellant of about 10 times their weight, $C_{structure} \approx 10-20\%$. Now we can relate all terms in equation (2-9) to the initial mass and write:

$$\mathcal{R} = \frac{M_{payload}}{M_{ignition}} = (1 + C_{structure}) \cdot e^{-\Delta v/c} - C_{engine} - C_{structure} \quad (2-14)$$



Term \mathcal{R} is called “mass ratio” of the rocket stage (the inverse value is called “growth factor” of the stage). The plot shows the ratio \mathcal{R} as exponentially declining function of the requirement Δv of the mission. For realistic values $C_{engine} \approx 0.01$ and $C_{structure} \approx 0.15$, the Δv capacity of the stage is much higher than the exhaust velocity c of its engine; but in any case the final velocity which can be reached is limited, even if the stage carries no payload.

Mass ratio of multi-stage rockets. Structure mass and engine mass limit the Δv capability of a rocket stage. If the empty stage had no weight, the rocket could accelerate to unlimited final velocity, but if a realistic rocket which carries a real payload should become fast, empty tank mass must be discharged. For high Δv requirements a multi-stage rocket has to be considered. In a “tandem arrangement” the payload mass of a stage is exactly the ignition mass of the next stage. Therefore:



$$\begin{aligned} M_{ignition} &= [M_{ignition}]_{stage1} \\ [M_{payload}]_{stage1} &= [M_{ignition}]_{stage2} \\ [M_{payload}]_{stage2} &= [M_{ignition}]_{stage3} \\ &\dots \\ [M_{payload}]_{stageN} &= M_{payload} \end{aligned} \quad (2-15)$$

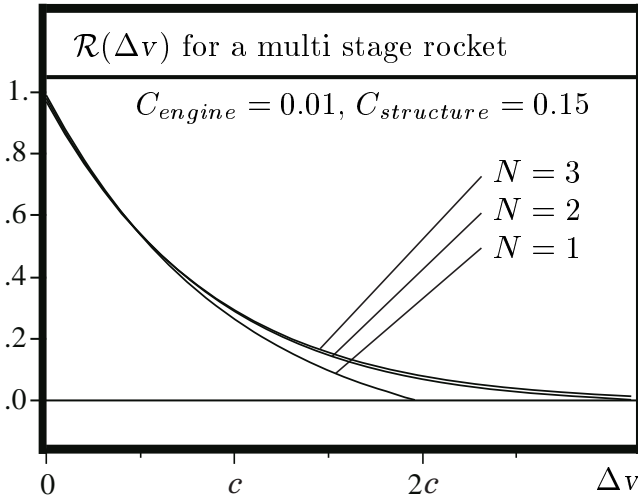
Equation (2-14) is valid for every stage, but we have to distribute the overall mission Δv requirement among all stages and write for the mass ratio \mathcal{R} of a N -stage rocket:

$$\mathcal{R} = \mathcal{R}_{stage1}(\Delta v_1) \cdot \mathcal{R}_{stage2}(\Delta v_2) \cdot \dots \cdot \mathcal{R}_{stageN}(\Delta v_N) \quad (2-16)$$

$$\text{with: } \Delta v = \Delta v_1 + \Delta v_2 + \dots + \Delta v_N \quad (2-17)$$

If the overall velocity requirement Δv is distributed in equal portions among all stages, and if the stages consist of the same technology (if for all stages exhaust velocity c and weight factors C_{engine} and $C_{structure}$ are the same), we may conclude:

$$\mathcal{R} = \frac{M_{payload}}{M_{ignition}} = [(1 + C_{structure}) \cdot \sqrt[N]{e^{-\Delta v/c}} - C_{engine} - C_{structure}]^N \quad (2-18)$$



The figure shows mass ratio \mathcal{R} of N -stage rockets as function of Δv , with stage number N as parameter. Quantity $M_{payload}$ means the payload mass of the multi-stage rocket (and the payload mass of its last stage as well). Term $M_{ignition}$ means the total mass of all stages together at the beginning of flight, including payload mass $M_{payload}$. Also for multi-stage rockets, \mathcal{R} is an exponentially decreasing function of the mission requirement Δv .

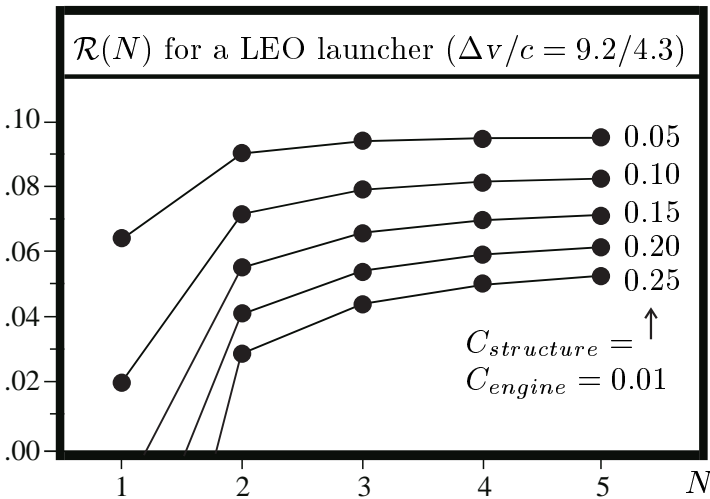
As expected, for a given ratio \mathcal{R} the Δv capacity increases with the number of stages; but the benefit is negligible small for missions with low Δv requirement. If the mission requirement Δv is lower than the value of exhaust velocity c , the mass ratio \mathcal{R} of a single stage rocket is nearly as good as the ratio \mathcal{R} of a multi stage rocket; but if Δv is higher than $2c$, a multi-stage rocket is necessary. Then ratio \mathcal{R} becomes very small, indicating that either the payload is very small or the rocket is very big.

Optimal number of stages. The next point in question is the optimal stage number N of a multi-stage rocket. We regard N as parameter open for optimization. As we have seen, of strong influence on the optimal number N are the technology of the stages (characterized by the exhaust velocity c) and the Δv requirement of the mission. The more stages are used, the more complex and expensive the rocket system becomes. Also for safety reasons a configuration with too many stages should be avoided. On the other hand, the rocket shall have a sufficient payload capacity.

Ruppe's formula serves for finding the optimal stage number N . We must divide the Δv requirement of the mission by the average exhaust velocity c , and round the result to a natural number. The result is approximately the optimal number of stages:

$$N_{optimal} \approx \text{integer}\left(\frac{\Delta v}{c}\right) \quad (2-19)$$

Ruppe's formula is an empirical law. It is considered that costs and complexity grow with the number of stages. Formula (2-19) can be clarified using an example:

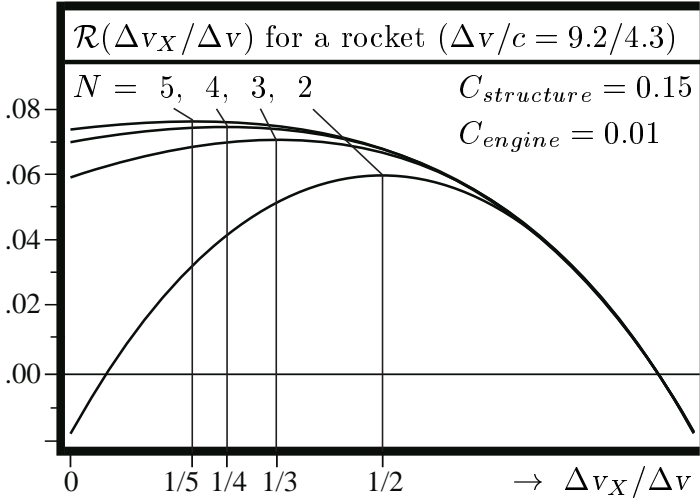


A launcher is developed for satellite transportation to low earth orbit ($\Delta v \approx 9.2$ km/s); the space launcher will use high energetic liquid propellant LOX/LH2 ($c \approx 4.3$ km/s). The optimal number follows as $N = 2$. For a single-stage system ($N = 1$) the payload is too small (or even vanishing); for many stages ($N = 3, 4, 5\dots$) the payload is not much better than for a two-stage system.

Optimal size of the stages. Equation (2-16) shows that all stages contribute in the same way to the performance of a multi-stage launcher (total mass ratio \mathcal{R} is the product of all the individual mass ratios of the stages). For equal mass ratios we may expect a payload maximum. If stage X of a N -stage rocket has a different mass ratio (stage X can be any one of the stages), equation (2-18) has to be rewritten as:

$$\mathcal{R} = \frac{M_{payload}}{M_{ignition}} = [(1 + C_{structure}) \cdot e^{-\Delta v_X/c} - C_{engine} - C_{structure}] \cdot [(1 + C_{structure}) \cdot \sqrt[N-1]{e^{-(\Delta v - \Delta v_X)/c} - C_{engine} - C_{structure}}]^{N-1} \quad (2-20)$$

In the equation above it is not necessary that stage X shares the same (equal) portion of the total Δv requirement of the mission. The differential quotient $\partial \mathcal{R} / \partial \Delta v_X$ vanishes for $\Delta v_X = \Delta v / N$, indicating a maximum. The stage is of appropriate size, if it makes approximately the same contribution to mission Δv as all the other stages.



Multi-stage rocket systems are optimized if every stage contributes approximately with the “same portion” Δv_X to the total flight performance (if $\Delta v_X = \Delta v/N$). The plot shows mass ratio \mathcal{R} of a multi-stage rocket as function of ratio $\Delta v_X/\Delta v$; stage X may be anyone of the N stages (for $\Delta v_X = 0$ we actually have $N - 1$ stages; for $\Delta v_X = \Delta v$ we have just one single stage).

Stages of equal size. Usually, the first stage of a rocket is big and the upper stage is small. Considered is now a multi-stage rocket where all stages are exactly of the same size as the payload mass. Then we can express the mass ratio \mathcal{R} of the stages (equation 2-14) in form of a simple series, beginning with N (the last stage) and counting down to 1 (the first stage). For a different payload mass the sequence of \mathcal{R}_X shows a more complicated but similar behaviour. We can use equation (2-14) to calculate the velocity capability Δv_X of stage X as a function of ratio \mathcal{R}_X :

$$\mathcal{R}_X = \frac{N - X + 1}{N - X + 2} \quad \left(= \dots \frac{4}{5}, \frac{3}{4}, \frac{2}{3}, \frac{1}{2} \right) \quad (2-21)$$

$$\Delta v_X = -c \cdot \ln \left(\frac{\mathcal{R}_X + C_{engine} + C_{structure}}{1 + C_{structure}} \right)$$

$$\Delta v = \Delta v_1 + \dots + \Delta v_X + \dots + \Delta v_N \quad (2-22)$$

Upper stages share bigger parts of the total flight performance than other stages, usually. Equation (2-22) explains why: the improvement of a small upper stage is cheaper and easier than the improvement of a big booster stage (raising the exhaust velocity c or reducing the factor $C_{structure}$). Particularly if the upper stage is not too small, improvements have a strong influence on the total flight performance Δv .

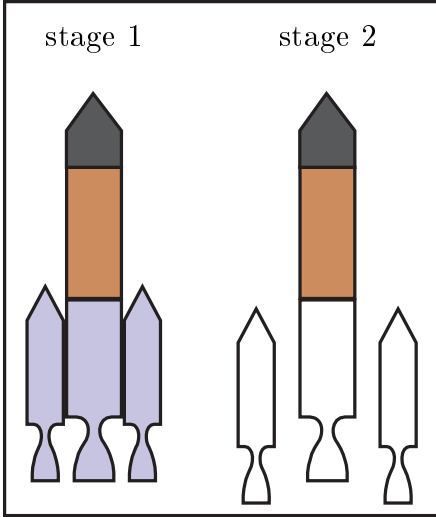
Another interesting result can be obtained when we consider the empty stages as weightless. Now the model corresponds to a single stage rocket, where structure mass and engine mass are treated in the same way as if they were a part of the payload mass. The total propellant mass is divided into equal portions (equal to the payload mass). With $C_{engine} = 0$ and $C_{structure} = 0$, equation (2-22) takes the simple form of:

$$\Delta v = -c_1 \cdot \ln(\mathcal{R}_1) \dots - c_X \cdot \ln(\mathcal{R}_X) \dots - c_N \cdot \ln(\mathcal{R}_N) \quad (5-23)$$

The “stages” can have different exhaust velocities c , when a single stage system which uses more than one sort of propellant is under consideration. Since the “upper stage” has the lowest mass ratio ($\mathcal{R}_N = 1/2$), we can conclude that the propellant with the best exhaust velocity c_N should be burned last. At the beginning of the flight (when $\mathcal{R}_1 < \approx 1$), the rocket should expend propellant with the lowest exhaust velocity c_1 .

Parallel staging. The situation is similar for multi-stage rockets in “parallel arrangement”. Many launchers lift-off while core stage and strap-on booster motors are burning simultaneously (it improves flight safety greatly if no engines have to be started during the ascent mission). If we arrange the sub-masses of such a launcher appropriately, we may regard it as if it were a tandem-staged vehicle.

Considered is a rocket where core stage and several booster motors burn parallel. During flight phase 1 of stage 1 the vehicle increases its velocity by Δv_1 . Flight phase



2 of stage 2 begins when the booster motors are empty and dropped. The core stage continues the mission alone and increases its velocity by Δv_2 . Stage 1 consists of the following sub-masses: the booster motors, the amount of propellant which the core stage expends while the boosters are burning, and the structure necessary to store this part of propellant. Sub-masses of stage 2 are the complete structure of the core stage, the rest of propellant in the core stage and the engines of the core stage. Thus, the structure of the core stage is thought as intersected into two sub-structures. For mass ratio \mathcal{R} we can write the same relationship as if the launcher were a tandem-staged vehicle:

$$\mathcal{R} = \frac{M_{payload}}{M_{ignition}} = [(1 + C_{structure1}) \cdot e^{-\Delta v_1/c} - C_{engine1} - C_{structure1}] \cdot [(1 + C_{structure2}) \cdot e^{-\Delta v_2/c} - C_{engine2} - C_{structure2}] \quad (2 - 24)$$

Again, overall velocity requirement is $\Delta v = \Delta v_1 + \Delta v_2$. Term c indicates the average exhaust velocity of a stage; but factors C_{engine} and $C_{structure}$ are different now:

$$C_{engine1} = \frac{M_{engines,boosters}}{M_{ignition,phase1}}, \quad C_{structure1} = \frac{M_{structure,boosters} + M_{structure,phase1}}{M_{propellant,boosters} + M_{propellant,phase1}}$$

$$C_{engine2} = \frac{M_{engines,corestage}}{M_{ignition,phase2}}, \quad C_{structure2} = \frac{M_{structure,corestage}}{M_{propellant,phase2}}$$

Compared to tandem-staged vehicles, factor $C_{engine1}$ of the first stage is a little better now: without counting, the engines of the core stage push additionally. The flight phase 2 starts when the core stage is partially empty; and the factor $C_{structure2}$ of the second stage is considerably deteriorated. Since engine weight is subordinated in comparison to structure weight, from the energetic point of view the concept of parallel staging is worse than the concept of tandem staging. The core stage wastes energy to accelerate the booster motors. The concept requires that the core stage burns much longer than the booster motors. As a consequence, parallel-staged launchers need strong booster motors supporting a comparatively weak core stage.

2.2. Rocket Propellant

As we have seen, rocket propulsion systems require not only working fluid but also energy to accelerate the working fluid before the exhaustion. In case of so-called chemical engines the propellant itself is the container for this energy. The combustion process transforms the chemical energy of the propellant into thermal energy (heat); the exhaustion process transforms the thermal energy into kinetic energy. The content of energy available for propulsion is the most important characteristic factor for chemical propellant. Theoretically, a large variety of chemical substances can be used as rocket propellant. Practically, just a few options are really used today.

2.2.1 Properties of Chemical Rocket Propellant

Propellant performance. Obviously, a high content of chemical energy is desirable because this implies a high relative velocity of the exhausted gas. However, if large and heavy tanks are necessary for storing the propellant, the efficiency of the entire propulsion system is considerably deteriorated. Ciolcovskij's equation (2-7) shows that two factors have a main influence on the performance of a rocket stage: the exhaust velocity c and the ratio M_1/M_0 (the ratio burnout mass to ignition mass). For low density rocket propellant more structure weight is needed: not only the mass-specific energy content is important, but also the volume-specific energy content. Solid propellant has a high density, solid booster motors are comparatively compact. So-called cryogenic propellant (liquefied gases) has a comparatively high energy content per unit of mass. Unfortunately, this propellant has also a low density.

In-flight handling. Liquid propellant has to be stored in tanks and conducted via pipes to the engines. Therefore special materials have to be used if the propellant is a chemical aggressive substance. Corrosion must be avoided. Cryogenic liquids require that tanks and pipes are equipped with special insulating material for heat protection. Usually, the propellant is used for thrust chamber cooling before it is combusted (it runs through small pipes at the outer surface of the thrust chamber). Attention has to be paid to the specific heat of the substances, their heat transfer and pumping properties, and to the variation of these properties with the temperature (vapor pressure, viscosity, freezing point, chemical decomposition and so on). When the propellant enters the thrust chamber, oxidizer and fuel are injected through small orifices and mix in form of thin sprays. It is an advantage if the propellant ignites spontaneously: for so-called "hypergolic propellents" no ignition system is required.

Storability and stability. Long duration space missions require storable propellant. For this liquid propellant should have good chemical stability and a low vapor pressure even at elevated temperature. Cryogenic liquid propellant cannot be stored for an extended time period, it vaporizes and disappears through seals and even through tank walls. On the other hand, solid propellant can be kept stable for many years without chemical decomposition and minimal reaction with the atmosphere.

Propellant production and ground handling. The production of propellant should be a simple chemical process which requires just ordinary industrial equipment. The source substances and materials for propellant production should be cheap and available in sufficient quantity. If the substances are dangerous, ground handling involves safety concerns. Personnel that handles dangerous substances must be appropriately protected: if the propellant is toxic or chemically aggressive, special clothes against accidental spills are necessary. Use of gas masks and a limited exposure time is prescribed for personal that gets into contact with toxic gases or vaporized propellant. Non-toxicity is a highly desirable property of propellant, as well as non-toxicity of its chemical reaction products. Toxic exhaust gases can impose limits to the flight rate of surface launched vehicles. Chemical rocket propellants have the property that their fuel component can burn when it gets mixed with atmospheric oxygen. If the fuel is an explosive substance this means severe fire hazard; and special precautions are necessary during the entire production process.

2.2.2 Cryogenic Propellant

Liquefied gases. A gas changes to its liquid phase when it is cooled down below a certain temperature. Liquid substances which are under normal ambient conditions gaseous are called “cryogenic liquids”. These liquids are produced from gas by a thermodynamic process, subsequently compressing, cooling and expanding the gas. It is difficult to keep liquefied gas for an extended period of time in rocket tanks: cryogenic propellant cannot be used for long duration space missions; but because of its high specific impulse, cryogenic propellant is often used for space launchers.

Liquid hydrogen LH₂ as fuel. Hydrogen when burned with oxygen produces water vapor, releasing comparatively much thermal energy per unit of mass. Another positive property is that LH₂ is an excellent regenerative coolant for rocket engines, but liquefied hydrogen is also a substance with extremely low boiling point (20 K) and very low density (70 kg/m³). Large-volume tanks of cold-resistant material are required to store the liquid. Tanks may not lose their strength if hydrogen atoms enter the wall material. Heat insulating jackets are necessary to prevent that air moisture condenses at the outer surface of tanks and forms unwanted ice layers. The fueling procedure of a hydrogen tank is not unproblematic: since other gases and liquids solidify in liquid hydrogen, improper fueling can create solid particles which can later plug valves, orifices or small pipes. Sometimes liquid hydrogen can be replaced by liquid methane, a denser cryogenic substance which is lower in performance.

Liquid oxygen LOX as oxidizer. Liquid oxygen can easily be manufactured by purification of liquefied air. Compared with liquid hydrogen, the oxidizer is much denser and less volatile (at normal pressure conditions liquid oxygen has a boiling point of 90 K and a density of 1140 kg/m³). LOX is non-toxic, non-corrosive and safely storable in clean tanks. Like LH₂-tanks, LOX-tanks require special heat insulation. Liquid fluorine as oxidizer would give a higher specific impulse than LOX, but fluorine is an extremely aggressive and toxic substance, improper for space applications.

Kerosene as fuel. Kerosene in combination with liquid oxygen can be seen as cryogenic propellant, since the oxidizer is a liquefied gas. Compared with liquid hydrogen, kerosene is denser and thus the fuel requires smaller tanks. As a disadvantage, the energy content per unit of propellant mass is much smaller. Kerosene is a colorless oily liquid, it is non-toxic, everywhere available and easy to handle. Burning with oxygen it forms a yellow-white brilliantly radiating flame. The physical properties of kerosene fuel can vary slightly, depending on the source substances (mineral oil) and on the production process. Typically, kerosene has a density of 810 kg/m^3 (289 K) and a freezing point of 230 K. Kerosene can be replaced by other hydrocarbon fuels with similar properties (petroleum derivatives like jet fuel, diesel or aviation gasoline). So-called rocket propellant RP-1 is a specifically refined petroleum product which is particularly suitable for the application in rocket motors (RP-1 has a low content of aromatic substances which could cause deposits inside cooling passages).

substance	freezing point	boiling point	density [kg/m ³]	viscosity [10^{-6} Ns/m ²]	vapor press. [bar]
liquid hydrogen (H ₂)	14.0 K	20.4 K	76 (14 K) 71 (20 K)	24 (14 K) 13 (20 K)	2.03 (23 K) 8.70 (30 K)
liquid methane (CH ₄)	90.6 K	111.6 K	445 (93 K) 226 (293 K)	2000 (93 K)	0.33 (100 K) 26.8 (173 K)
liquid oxygen (O ₂)	54.4 K	90.1 K	1230 (78 K) 1144 (90 K)	870 (54 K) 190 (90 K)	0.05 (89 K)
liquid fluorine (F ₂)	53.5 K	85.0 K	1505 (85 K) 1440 (93 K)	300 (77 K) 412 (69 K)	1.58 (422 K) 0.065 (66 K)
hydrocarbon fuel RP-1 (CH _{1.97})	225-239 K	336 K	807 (289 K) 580 (422 K)	750 (289 K) 210 (366 K)	0.02 (344 K) 0.23 (422 K)

2.2.3 Storable Liquid Propellant

Hydrazine. The chemical substance hydrazine N₂H₄ can either be used as liquid monopropellant (for trajectory or attitude control thrusters) or as fuel in bipropellant rocket stages (in combination with the oxidizer nitrogen tetroxide N₂O₄). Hydrazine can be stored in clean, sealed tanks for many years (decades). It is a colorless liquid, which spontaneously decomposes when it gets into contact with a catalyst, or it ignites when it is mixed with the oxidizer. The ground handling of hydrazine is not easy: for personnel the substance is a toxic, cancer-causing chemical. Its vapor may explode when mixed with air, accidentally spilled on cloth it can ignite spontaneously. The density of hydrazine is with 1023 kg/m^3 (293 K) comparatively high.

Unsymmetrical dimethylhydrazine and monomethylhydrazine. Both fuels UDMH ((CH₃)₂NNH₂) and MMH (CH₃NHNH₂) are derivatives of hydrazine with similar properties. These fuels are extensively used in bipropellant rocket propulsion systems (sometimes in form of a mixture with pure hydrazine). Usually, nitrogen tetroxide is the oxidizer. Compared to pure hydrazine, UDMH and MMH give slightly lower performance to rocket engines, but these substances are used because they are more stable liquids with better shock resistance and better heat transfer properties.

Nitrogen tetroxide as oxidizer. The yellow-brown liquid nitrogen tetroxide (N_2O_4) is the most common oxidizer for the fuel hydrazine and its derivatives. Pure N_2O_4 is already quite corrosive, but particularly when it absorbs air moisture (or gets otherwise mixed with water) N_2O_4 forms an aggressive chemical substance. Inflammable material can ignite spontaneously when it gets into contact with N_2O_4 . The exhaust fumes are extremely toxic. Liquid nitrogen tetroxide has a high density, but the high vapor pressure causes comparatively heavy tanks. The substance is in its liquid phase just between 262 K (freezing point) and 294.3 K (boiling point). In space applications attention must be taken to avoid freezing this liquid. N_2O_4 can be stored for decades, if clean sealed tanks made of appropriate material are used.

substance	freezing point	boiling point	density [kg/m ³]	viscosity [10^{-6}Ns/m^2]	vapor press. [bar]
hydrazine (N_2H_4)	274.7 K	386.6 K	1023 (293 K) 951 (350 K)	970 (293 K) 580 (330 K)	0.014 (293 K) 0.50 (366 K)
unsymmetrical dimethyl- hydrazine ((CH_3) ₂ NNH ₂)	216 K	336 K	611 (228 K) 850 (244 K)	4400 (220 K) 490 (300 K)	0.13 (289 K) 8.0 (339 K)
monomethylhydrazine (CH_3NHNH_2)	221 K	361 K	879 (293 K) 863 (311 K)	855 (293 K) 400 (344 K)	0.075 (300 K) 6.9 (428 K)
nitrogen tetroxide (N_2O_4)	262 K	294 K	1447 (293 K) 1370 (322 K)	423 (293 K) 320 (315 K)	0.96 (293 K) 4.13 (328 K)

2.2.4 Solid Propellant

Solid propellant for space applications. Solid propellant is used in large booster stages, in smaller launch-assist motors, in spin-stabilized kick-stages for satellites, in tactical missiles and in gas-generators. Each of these different applications needs a different optimal propellant. High performance and comparatively low thrust is very important to upper stages. Launch-assist motors put emphasis on another property: non-toxicity of the exhaust gases. Minimum smoke is important to some missile applications, but unimportant for space applications. Solid propellant in rocket motors is used to generate thrust (the combustion temperature is higher than 2400 K). If solid propellant is burned in gas generators, the purpose is to generate power (then the combustion temperature is much lower, usually between 800 and 1200 K).

The chemical composition of solid propellant influences its burn velocity, the combustion temperature and the combustion pressure. Basically, solid propellant consists of fuel, oxidizer and binder (all in one substance); but some properties of solid propellant can be influenced by adding other ingredients to the composition. Up to ten additives are sometimes used to get a “tailored” solid propellant. The production of solid propellant is a complicated process which requires a lot of empirical knowledge (gained from experiments). As a consequence, the exact composition of a particular solid propellant varies not only among manufacturers, but is also different for different motor applications. Practically, the industry develops a new propellant composition for every new motor. Particularly, the production of solid propellant is a dangerous and expensive procedure when large booster motors are manufactured.

Composite propellant. The most common solid propellant for space applications consists of a heterogeneous mixture of aluminum powder as fuel (5% to 22% Al), crystals of ammonium perchlorate as oxidizer (65% to 70% NH_4ClO_4) and polybutadiene as binder (8% to 14%). The inflammable binder keeps fuel and oxidizer close together and forms a rubbery (or plastic-like) substance. The size, the shape and the distribution of the particles inside the grain influence the burn characteristics. Usually, a blend of particle sizes is used to maximize the volume specific weight of the propellant. Up to 3% curing agent is added to solidify the binder (to form long molecular chains from the prepolymer molecules). Its performance can be increased by adding high energetic radicals (for example acrylic acid) or explosives (crystalline nitramines) to the binder. In comparison with other solid propellants, the production of composite propellant is easy, the propellant is reliable and its combustion properties are well-known. It has a high density (between 1700 kg/m^3 and 1800 kg/m^3) and a good performance. The flame temperature is about 3300 K, and the exhaust fumes are toxic (HCl) and smoky (burned aluminum powder).

Double-base propellant. Lower in performance and more dangerous in production is so-called double-base propellant. Essentially, double-base propellant consists of nitrocellulose dissolved in nitroglycerin, with minor percentages of plasticizers and other ingredients. The propellant grain is a homogeneous substance, where both the main components are explosives. The performance of double-base propellant increases if more nitroglycerin is used (40% to 60%), but physical properties get poor if the percentage of nitroglycerin is high. In comparison with composite propellant, double-base propellant has a lower density ($\approx 1600 \text{ kg/m}^3$) and lower flame temperature (for example 2550 K). Burning the propellant gives non-toxic, clean exhaust fumes. A mixture of composite propellant and double-base propellant in one substance is possible (so-called composite double-base propellant).

Solid propellant ingredients. Besides fuel, oxidizer and binder, several other ingredients are used to optimize the properties of solid propellant. Burn rate modifiers are used to accelerate or decelerate the combustion velocity. Plasticizers are used to improve the mechanical properties of the solid propellant during the production process. Curing agents (or cross-linkers) cause the binder to solidify, where a long processing time is desirable. Stabilizers and bonding agents improve the chemical properties of the propellant, these substances prohibit deterioration if the propellant is stored for a long time interval. Energetic binders or energetic plasticizers contribute by an exothermal chemical reaction to the combustion process.

solid propellant processing method	composite cast (STS-SRM)	double base extruded	composite/double-base extruded
typical solid propellant composition in percent of weight	69.93% amm.perchl. 16.00% aluminum 12.04% polymer binder 1.96% epoxy curative 0.07% catalyst (iron oxide)	51.5% nitrocellulose 43.0% nitroglycerin 3.2% diethyl phthalate 1.2% potassium sulfate 1.0% ethyl centralite 0.1% additives (wax)	20.4% amm.perchl. 21.1% aluminum 21.9% nitrocellulose 29.0% nitroglycerin 5.1% triacetin 2.5% stabilizers

2.3. The Combustion Process

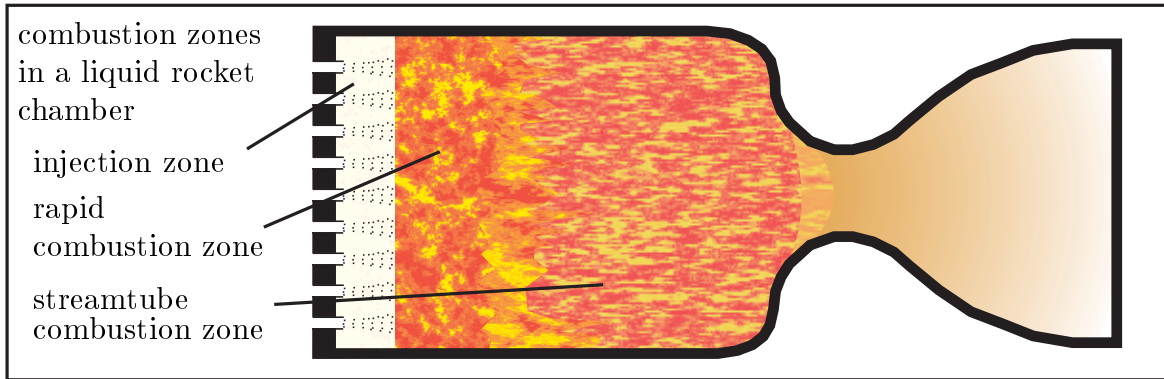
The combustion of liquid or solid propellant is a quite complex process, where many different chemical and physical phenomena are involved (including thermodynamics, reaction kinetics, fluid dynamics, heat transfer and so on). The combustion inside the thrust chamber of a rocket engine releases the chemical energy of the propellant. The process should be efficient and stable, it should generate high-temperature uniform gas. Our knowledge about combustion is based on experiments rather than on analytical understanding. We know the composition of the substances which enter the thrust chamber and we know the composition of the exhaust fumes, but we have hardly more than an idea what happens inside the thrust chamber. Even though mathematical models and computer programs do exist and are in use, engineering depends on empirical techniques and experiments. Some phenomena do still resist analysis of computer programs (computational fluid dynamics): usually neglected or simplified are reaction kinetics, turbulent effects, uncertainties in the size and distribution of droplets, time variations in local temperature and gas composition. Only experienced engineers can read the results of computer simulations correctly.

2.3.1 Combustion of Liquid Propellant

Injection zone. Fuel and oxidizer are injected through a matrix of many small orifices, they enter the thrust chamber with high velocity (between 7 and 60 m/s) and mix in form of thin sprays. Heat radiation from the combustion zone lets the small droplets of fuel and oxidizer evaporate. The region inside the thrust chamber near to the injection orifices contains a heterogeneous mixture of liquid and vaporized propellant, as well as already some hot-burning gas. The gas flow is unsteady (turbulent), it includes first chemical reactions, local explosions and shock waves.

Rapid combustion zone. The gas flow enters the rapid combustion zone where its temperature and its specific volume increase substantially. Now nearly all droplets are vaporized. Chemical reactions occur intensively in form of many small local explosions, here the combustion is not a steady flow process. The reaction front is carried along mainly via shock waves, but also via radiative and convective heat transfer. Temperature, pressure, and gas composition change rapidly in time for every individual small region within this zone.

Stream-tube zone. Then the flow enters the stream-tube zone where its axial velocity increases considerably. Now turbulent mixing across streamline boundaries and local radial velocities become smaller. Chemical reactions continue at a lower rate; the gas composition is driven to its chemical equilibrium. High chamber temperatures cause the chemical reactions to proceed rapidly: a few milliseconds after having been injected, the propellant leaves the chamber again and enters the convergent part of the nozzle. The propellant is burned in the combustion chamber at high temperatures and high pressures, therefore the combustion rate in rocket motors is very high. Combustion efficiency ranges from 95% to 99.5% in well designed thrust chambers.



Instabilities. Combustion instabilities can cause extensive oscillations and destroy the rocket motor. Therefore it is the aim of a proper design to prevent the occurrence of instabilities and keep the combustion always stable. Instabilities are excited if the combustion is not “smooth” but “rough”, and interacts with the natural frequencies of the feed system or with the structure of the vehicle.

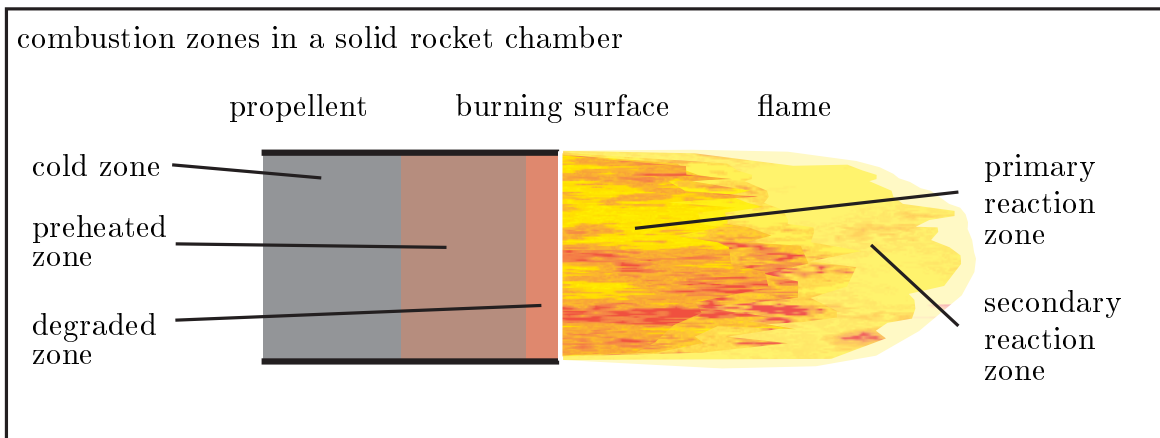
type of instability	frequency	excited by interactions between
longitudinal “pogo” vibrations (propellant flow rate disturbances)	10-50 Hz	thrust force and propellant feed system
“chugging” (feed system instability)	10-400 Hz	chamber pressure and propellant feed system
“buzzing” (acoustic resonance)	400-1000 Hz	structure, injectors and propellant feed system
“screaming” or “screeching” (thrust chamber pressure waves)	above 1000 Hz	chamber resonance frequency and combustion process

The occurrence of combustion instabilities can hardly be predicted by means of computer programs; and the avoidance of instabilities is more an empirical than an analytical process. The natural frequencies of the vehicle and its propellant feed systems can readily be calculated or measured. Often it is easy to alter these natural frequencies by minor structural modifications. Low frequency “pogo” oscillations can be damped by including energy absorbing devices into the fluid flow lines and into the tanks. “Buzzing” is usually more annoying (noisy) than really dangerous. More dangerous is high-frequency “screaming” (1000-4000 Hz), accompanied by more vibratory energy. This phenomenon takes place exclusively inside the thrust chamber, but it can destroy an engine in less than one second. Shock waves inside the chamber are reflected by the injector face, the converging part of the nozzle and the chamber wall. Unfortunately, high-frequency instabilities are most common in the development of new engines. It is tried to make the combustion stable at all flight conditions by the incorporation of “fixes”: injector face baffles, acoustic energy absorbers or energy absorption cavities, perforated combustion chamber liners are added to the design (often by using an experimental approach “trial and error”). The design of the injector has a major influence on instabilities; and seemingly minor modifications can have a substantial effect on the appearance of instabilities.

2.3.2 Combustion of Solid Propellant

Motor ignition. The solid rocket motor is ignited by an electric signal which triggers a hot gas generator. Within milliseconds the complete surface of the propellant grain burns. The free volume of the chamber fills with hot gas and the chamber pressure grows until equilibrium flow has been established.

Burning solid propellant grain. Like the combustion of liquid propellant, the combustion of solid propellant involves complex chemical and physical phenomena. At the burning surface of the solid propellant grain, the constituents of the propellant transform into their liquid or into their gaseous phase. Chemical reactions take place in a heterogeneous mixture of solid material, liquid droplets and flowing gas. The physical and chemical processes which occur during the combustion are not completely understood, and mathematical models are oversimplified and unreliable. The processes depend sensitively on many quite unknown variables: accumulation of solid particles, size of liquid droplets, intermediate chemical reaction products, three-dimensional flow behaviour and velocity of chemical reactions.



Experimental analysis with test burners give insight into the combustion process. It can be observed that most of the chemical reactions occur inside the flame above the burning surface. With increasing chamber pressure the visible length of this flame becomes shorter, and the volume specific energy release increases. In case of composite propellant the flame is irregular (flickering) and attached to the burning surface; and in case of double-base propellant the flame is more regular and seemingly detached from the surface. More distant from the surface the flame is less brilliant, but still reactions take place until the gas has reached equilibrium composition.

Essentially, the burning surface can be divided into two main zones: the outermost zone (“degradation zone”) accepts the heat radiated by the flame. This zone is very thin, here the solid propellant melts and vaporizes, and the first chemical reactions take place. The combustion proceeds via heat conduction from the degradation zone to the “preheated zone” inside the propellant grain. The thickness of the preheated zone is strongly influenced by the actual burn velocity of the solid propellant.

The burn velocity must be a stable function of the chamber pressure: the motor explodes if chamber pressure and burn velocity grow simultaneously without limits. The burn velocity can be influenced by adding catalysts to the solid propellant. Catalysts can influence the reactions inside the flame; the heat transfer rate from the flame to the propellant grain and the heat conduction properties inside the grain (catalysts influence just the velocity but not the chemical equilibrium of a reaction).

The combustion process will stop when the chamber pressure is rapidly diminished. Sudden depressurization occurs when most of the propellant has been consumed, or when a solid motor is stopped intentionally by opening additional gas escape areas.

Combustion instabilities. During the time interval of normal operation, solid rocket motors experience often substantial variations of the combustion pressure. The burn velocity increases if the pressure grows and decreases if the pressure diminishes. When instabilities occur, the pressure oscillates by sometimes more than 30% of its nominal value. Even when they do not destroy the motor, these pressure oscillations must be avoided because they lead to an unacceptable growing of the thrust.

Initially, the volume inside the thrust chamber of the solid motor is loaded with about 90% of solid propellant; and the motor is more or less empty at burnout. During the time interval of burning, the geometry of the gas cavity inside the motor is very complex and alters its size substantially. The remaining solid propellant is an energy-absorbing viscoelastic material, but the combustion process is a strong energy source for vibrations. Low frequency pressure variations essentially uniform throughout the volume can be excited (“bulk mode” oscillations, where the gas velocity in the nozzle oscillates with a frequency below 150 Hz). The wall of the thrust chamber can also reflect noise and establish acoustic waves (in radial and axial direction). Usually these waves are at lower frequency (because of the size of the thrust chamber), but also harmonic frequencies of the basic vibration modes can be excited. Normally these oscillations disappear after a few seconds, as the burning proceeds and the resonance behaviour of the chamber alters.

Avoidance of instabilities in the design phase of solid rocket motors is a complicated procedure. The physical and chemical phenomena are influenced by many complex factors which are simply too difficult for a mathematical simulation. Minor changes in the grain composition and grain geometry affect substantially the occurrence of instabilities. Consequently, the avoidance of instability is mainly an empirical process. Not always in the development of a solid motor occurs unstable burn behaviour, but making experiments with full-scale motors is unfortunately very expensive.

The vibrational properties of the burning solid propellant are analyzed with sub-scale models (“T-burners”), but the conclusions for the full-scale motor drawn from these experiments are dubious. Experiments with T-burners are appropriate to prove that a change in the propellant composition has no effect on oscillations, rather that they can prove the stability of a new motor design. Particularly if the solid rocket motor is big, the design of an appropriate grain configuration becomes an expensive procedure.

2.3.3 Propellant Application and Performance

Liquid rocket propellant. The performance of a rocket stage is characterized by the value of its specific impulse. The following table gives an overview over some example applications for propellant in liquid rocket engines:

fuel mix.(50%/50%)	LH2	LH2	RP-1	kerosene	UDMH	MMH
oxidizer	LOX	LOX	LOX	LOX	N ₂ O ₄	UDMH
ignition	electric spark		hypergolic runon		hypergolic	
engine name	SSME	RL-10	F1	RD-107	Viking-V	TR-201
application	USA	USA	USA	Russia	Europe	USA
vehicle	Shuttle	Centaur	Saturn-5	Vostok	Ariane-4	Delta
flow rate:						
oxidizer [kg/s]	408	14.0	1585	226	173.3	8.92
fuel [kg/s]	68.0	2.79	790	91	101.9	5.62
chamber:						
pressure [bar]	206	32.6	69	60	58	7.1
temperature [°C]	3300	3250	3280	3520	3000	3300
performance:						
$I_{specific}$ (v) [s]	452.9	444.4	301	314	278.4	303
thrust (v) [kN]	2091	73.4	8044	1000	752	44
thrust (sl) [kN]	1668	-	6770	821	678	-

Propellant for control thrusters. Usually, small control thrusters work with the catalytic decomposition of monopropellant hydrazine over a heated catalyst test bed; but also bipropellant systems are used, particularly if better performance is required:

attitude control thrusters	chamber press.[bar]	thrust [N]	$I_{specific}$ (v) [s]
hydrazine catalytic decomposition:			
CHT 0.5, Germany satellites	5.5-22.0	0.20-0.75	216-227
CHT 400, Germany, upper stages	5.5-22.0	110-450	218-235
MRE-1, USA, deep space probes	5.4-38.0	0.9-5	210-220
bipropellant (N ₂ O ₄ (+NO) / MMH at 1.65):			
Leros 20, UK, satellites	7.5	22	294
R-40A, USA, space shuttle	10.5	3114-5338	281-306

Composite solid rocket propellant. Solid rocket propellant has many disadvantages for space applications (for example, it is comparatively bad in performance). Nevertheless it is used for big and small boosters, and also for upper stages:

composite propellant	total mass [t]	propellant [t]	thrust [kN]	$I_{specific}$ [s]
SRM, big shuttle booster	570.0	503.5	11521 (avg.)	265(v), 246 (sl)
Castor 4A, small booster	11.6	10.1	436.7 (avg.)	238(sl)
Star 27, upper stage	0.362	0.334	27 (avg.)	288(v)

2.3.4 Computation of the Chamber Temperature

Propellant performance analysis. The combustion process inside the chamber liberates the energy which is stored in a chemical form inside the propellant. Then this energy heats up the products of the chemical reaction (usually exhaust gases). A flow process in the nozzle transforms the heat of the gases into kinetic energy.

The chamber enthalpy. To analyze the combustion process and the subsequent flow process we make use of the concept of the “enthalpy”. In thermodynamics, the enthalpy H of a gas is the quantity useful to characterize the capability of the gas to generate mechanical work. By definition, the enthalpy H of a gas is the sum of its inner energy E (the potential energy and the kinetic energy of all gas molecules) and the product of pressure and volume $p \cdot v$ (the work done on or by the gas; v is the chamber volume). For the conditions inside the thrust chamber we may write:

$$H_{chamber} = E_{chamber} + p_{chamber} \cdot v_{chamber} \quad (2 - 25)$$

$$= E_{chamber} + M \cdot \mathfrak{R} \cdot T_{chamber} \quad (2 - 26)$$

We get equation (2-26) from equation (2-25) by using the perfect gas law $p v = M \mathfrak{R} T$. Term \mathfrak{R} (the gas constant) is the universal gas constant $\mathfrak{R}_o = 8.3145 \text{ J/mol K}$ divided by the mean molecular mass \mathcal{M} (in gram/mol). Then term M (in grams) is the mass of the entire gas which is being contained at the moment inside the combustion chamber; and term $H_{chamber}$ is the absolute enthalpy of the chamber gas (in Joule or Nm). To find the specific enthalpy h we have to divide this value by mass M of the gas.

The pressure inside the chamber. We consider the burning gas (mass M) inside the combustion chamber as a mixture of perfect gases; a constituent has the mass M_k . The molecular mass \mathcal{M}_k of the constituent is nearly identical with the number of protons and neutrons of the gas molecule. For example, hydrogen H_2 has the molecular mass 2 gram/mol, oxygen O_2 32 gram/mol, water steam H_2O 18 gram/mol and so on. We use the term n_k to specify the amount of substance of a constituent of the gas; n_k is the mass of a constituent divided by the molecular mass of this constituent:

$$n_k = \frac{M_k}{\mathcal{M}_k} \quad (2 - 27)$$

Term n_k has the dimension of mols. The gas mixture has the mean molecular mass:

$$\mathcal{M} = \frac{\sum n_k \mathcal{M}_k}{\sum n_k} = \frac{M}{\sum n_k} \quad (2 - 28)$$

Every constituent of the gas inside the chamber contributes to the total pressure with its partial pressure p_k . Since all constituents are treated as perfect gases, we have:

$$p_{chamber} = \sum p_k = \sum n_k \frac{\mathfrak{R}_o \cdot T_{chamber}}{v_{chamber}} \quad (2 - 29)$$

The sum of all partial pressures p_k is exactly the total chamber pressure $p_{chamber}$. Equation (2-29) is the perfect gas law for a mixture of gases.

The chamber gas is a mixture of perfect gases; therefore we can divide the partial pressure p_k of a constituent by the total pressure $p_{chamber}$ to verify that this ratio is equal to the ratio $n_k / \sum n_k$. The amount of a substance divided by the total amount of all substances is called molar fraction of the constituent ($p_k / p_{chamber} = n_k / \sum n_k$).

Specific heat. Since we consider the gas in the combustion chamber as perfect gas, its inner energy E and its enthalpy H are just functions of the gas temperature T . The specific heat at constant pressure c_p , the specific heat at constant volume c_v and their ratio κ are important factors (specific means “per mol of the gas”, to get the mol-specific values of enthalpy and inner energy we have to multiply the absolute enthalpy H and the absolute inner energy E with the factor \mathcal{M}/M). Thus:

$$c_p = \frac{\partial(H_{chamber} \cdot \mathcal{M}/M)}{\partial T}, \quad c_v = \frac{\partial(E_{chamber} \cdot \mathcal{M}/M)}{\partial T} \quad (2-30)$$

The factors c_p and c_v have the dimension J/mol K. Using the perfect gas law we find:

$$c_p = \frac{\kappa}{\kappa - 1} \mathfrak{R}_o = c_v + \mathfrak{R}_o, \quad c_v = \frac{1}{\kappa - 1} \mathfrak{R}_o = c_p - \mathfrak{R}_o, \quad \kappa = \frac{c_p}{c_v} \quad (2-31)$$

The relationships above determine the values for c_v and κ when the value for c_p is well-known. The ratio κ (without dimension) of the specific heats is always greater than 1, typical values for κ in rocket thrust chambers range between $\kappa = 1.1$ and $\kappa = 1.25$.

Every constituent has its individual specific heat value $c_{p,k}$. We have to find a mean value for the entire chamber gas, using the molar fractions of all constituents:

$$c_p = \frac{\sum n_k \cdot c_{p,k}}{\sum n_k} = \frac{1}{p_{chamber}} \sum p_k \cdot c_{p,k} \quad (2-32)$$

The specific heats $c_{p,k}$ allow us to define the chamber enthalpy as a function of the chamber temperature. Then the change of the chamber enthalpy in relation to a reference temperature is:

$$\Delta H_{chamber} = \sum [n_k \cdot \int_{T(reference)}^{T(chamber)} c_{p,k} dT] \quad (2-33)$$

Unfortunately, the values for $c_{p,k}$ are not constant, they increase slightly with T (except for gases where the molecules consist of just one atom, for example inert gases). The combustion temperatures inside a rocket combustion chamber are so high that it would be inadequate to assume constant values $c_{p,k}$ for the gases. Specific heats are functions of the temperature, tabulated in chemistry handbooks. More important for us are the integral values (so-called “mean-specific heat values”):

$$c_{p,mean\ value}(T) = \frac{1}{T - 273.15 [K]} \cdot \int_{273.15 [K]}^T c_p dT \quad (2-34)$$

Reference temperature for these mean integral values of c_p is 0° Celsius, usually. Using such tables avoids the integration of the specific heat temperature functions.

The following tables show values of c_p (dimension J/mol K) for a selection of gases; rocket chambers contain a mixture of some of these gases after the chemical reaction. (the values for atomic hydrogen, nitrogen and oxygen are constant, $c_p = \frac{5}{2} \cdot \mathfrak{R}_o$).

T [°C]	H ₂	O ₂	H ₂ O	OH	CO ₂	CO	N ₂	NO	NH ₃
0	28.617	29.274	33.499	29.990	35.860	29.123	29.115	29.977	35.002
500	29.559	33.549	38.406	29.844	50.815	31.707	31.250	32.590	49.949
1000	31.284	35.914	44.723	31.966	56.773	34.470	34.043	35.165	62.300
1500	33.394	37.242	49.639	34.030	59.411	35.856	35.546	36.371	69.794
2000	35.056	38.406	52.930	35.508	60.654	36.597	36.367	37.024	74.148
2500	36.316	39.502	55.161	36.580	61.085	37.053	36.869	37.422	76.744
3000	37.388	40.277	56.522	37.476	61.378	37.388	37.263	37.744	78.377

The concerning integral values of c_p are (J/mol K, reference temperature 0° Celsius):

T [°C]	H ₂	O ₂	H ₂ O	OH	CO ₂	CO	N ₂	NO	NH ₃
0	28.617	29.274	33.499	29.990	35.860	29.123	29.115	29.977	35.002
500	29.249	31.334	35.630	29.605	44.573	30.099	29.864	30.869	42.241
1000	29.789	33.118	38.619	30.220	49.392	31.665	31.313	32.456	49.367
1500	30.647	34.282	41.525	31.179	52.348	32.858	32.502	33.582	54.927
2000	31.548	35.169	43.995	32.075	54.290	33.708	33.373	34.369	59.188
2500	32.385	35.927	46.017	32.887	55.617	34.336	34.022	34.943	62.440
3000	33.159	36.676	47.696	33.578	56.606	34.834	34.583	35.383	64.960

When the chamber temperature is approximately well-known, the temperature functions of the specific heats of the chamber gases can be interpolated from the tables (for example, we can approximate these functions by parabolic curves).

Exhaust velocity and chamber enthalpy. We use the principle of “conservation of energy” to calculate the propellant performance from the chamber enthalpy. The expansion in the nozzle is considered as an isentropic flow process, without friction, without heat transfer through the walls and without heat input. Since the gas does not do any mechanical work inside the nozzle, the sum of enthalpy and kinetic energy must be the same for every location inside the nozzle. For the conditions at the entrance and exit of the nozzle we can write:

$$H_{chamber} + \frac{M}{2}c_{chamber}^2 = H_{end} + \frac{M}{2}c_{end}^2 \quad (2 - 35)$$

Usually, the velocity c inside the chamber is comparatively small, it can be neglected; but the velocity c_{end} at nozzle exit should be as high as possible. We may conclude that two factors are important for the best performance: a good thrust chamber has to work with a high mass-specific chamber enthalpy and with a low exit enthalpy. A high specific chamber enthalpy is achieved if high energetic propellant is used and if the combustion efficiency is good. The rocket motor has a low exit enthalpy if the pressure inside its chamber is high and the gas has a low specific heat ratio κ .

The chemical reaction process. A chemical reaction is actually a complicated procedure, it involves several intermediate reactions where high energetic atoms and free radicals play a part. For the calculation of the energy that is released by a chemical reaction, it is usually sufficient to know the composition of the reactants before and after the reaction. The overall amount of converted energy is independent of the intermediate states and the velocity of the reaction (reaction kinetics).

We can calculate the heat of the reaction $\Delta H_{reaction}$ of a rocket propellant when we know the gas composition before (reactants n_j) and after (products n_k) the reaction. For an exothermic (-) or endothermic (+) reaction, the following formula can be used:

$$\Delta H_{reaction} = \sum_{k=1}^{products} [n_k \cdot \Delta H_{formation,k}] - \sum_{j=1}^{reactants} [n_j \cdot \Delta H_{formation,j}] \quad (2 - 36)$$

substance	phase	molecular mass \mathcal{M} [gram/mol]	$\Delta H_{formation}$ [kJ/mol]	$\Delta \mathcal{S}_{standard}$ [J/mol K]	$c_p(25^\circ\text{C})$ [J/mol K]
O	gaseous	15.9994	249.18	161.059	21.913
O ₂	gaseous	31.9988	0	205.152	29.376
H	gaseous	1.00794	217.998	114.717	20.786
H ₂	gaseous	2.01588	0	130.680	28.836
OH	gaseous	17.0073	47.52	189.395	29.892
H ₂ O	gaseous	18.0153	-241.83	188.835	33.590
C	gaseous	12.0107	716.7	158.10	20.838
CO	gaseous	28.0101	-110.53	197.660	29.142
CO ₂	gaseous	44.010	-393.51	213.785	37.129
CH ₄	gaseous	16.0476	-74.873	186.25	35.639
C ₂ H ₆	gaseous	30.069	-84.73	229.60	50.019
C ₃ H ₈	gaseous	44.096	-103.85	270.02	73.508
N	gaseous	14.00674	472.7	153.301	20.786
N ₂	gaseous	28.0135	0	191.609	29.125
NO	gaseous	30.0061	90.25	210.76	29.901
NO ₂	gaseous	46.0055	33.10	240.04	36.974
NH ₃	gaseous	17.0306	-45.898	192.77	35.652
HNO ₃	gaseous	63.0128	-134.306	266.4	53.326
N ₂ O ₄	gaseous	92.011	9.079	304.376	77.256
N ₂ H ₄	gaseous	32.0451	95.353	238.719	50.813
H ₂ O	liquid	18.0153	-285.830	69.93	75.351
C ₆ H ₁₄	liquid	86.175	-198.8	292.5	195.01
C ₇ H ₁₆	liquid	100.20	-224.4	328.0	224.65
C ₈ H ₁₈	liquid	114.23	-250.0	361.2	254.05
N ₂ O ₄	liquid	92.011	-19.564	209.198	142.509
N ₂ H ₄	liquid	32.0451	50.626	121.544	98.840
Al ₂ O ₃	liquid	101.9612	-1620.567	67.23	79.015
Al	crystal	29.9815	0	28.275	24.204
C	graphite	12.0107	0	5.74	8.517
NH ₄ ClO ₄	solid	117.485	-295.767	184.18	128.07

In tables on thermochemical data we find the “heat of formation” of substances, so-called $\Delta H_{formation}$. The heat of formation of a chemical substance is the enthalpy released (-) or absorbed (+) when one mol of a substance is formed from its constituent atoms at standard conditions (1 bar and 25° Celsius or 298.15 K). The heat of formation is a relative value. By definition, the values for elementary substances (H_2 , O_2 , N_2 , Al, graphite-C and so on) equal zero.

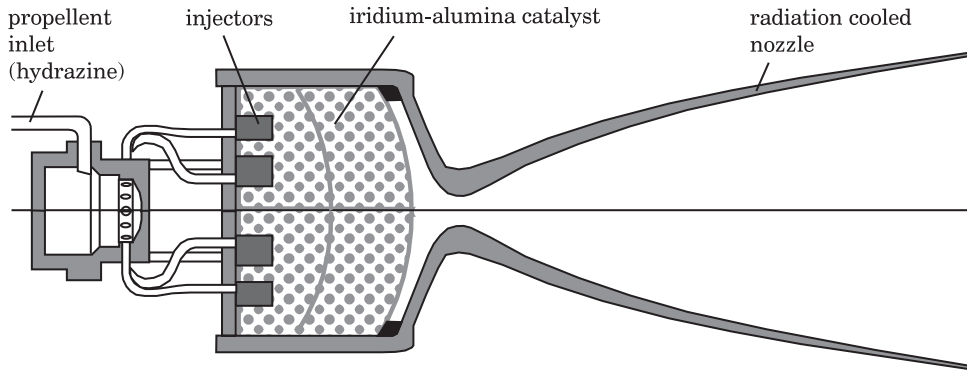
For example, 1 mol H_2 together with 1/2 mol O_2 produces 1 mol of H_2O . The reactants (hydrogen and oxygen gas) are supplied at standard conditions, after the reaction the product (water or water steam) is transformed back again to standard conditions. The enthalpy release amounts 285.84 kJ when water is produced and 241.83 kJ when the product is steam. The standard heat of formations of both reactants is zero by definition. Burning of 1 mol OH gas with 1/2 mol hydrogen gas H_2 produces water steam H_2O and an enthalpy of $241.83-47.52=194.31$ kJ. Gas that consists of free hydrogen atoms H can be produced by splitting H_2 molecules, the reaction consumes 217998 J/mol at standard conditions (25° Celsius).

The heat of formation is not a function of the pressure in case of perfect gases, but the reaction is performed at a temperature which is different from 25° Celsius. Then the conversion of energy is the same, but the energy which is produced (or absorbed) by the reaction differs slightly from the standard heat of reaction. Reactants and reaction products have different specific heat constants (equation 2-30), usually. It requires a slightly different amount of enthalpy to bring the reactants from 25° Celsius to the reaction temperature than to bring the reaction products from reaction temperature back to 25° Celsius. The standard heat of formation must be corrected, using the reaction temperature and the specific heats of reactants and products.

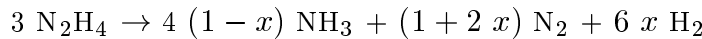
Heat of reaction and chamber enthalpy. The reactants which are present in the chamber before the reaction are not at standard conditions, usually. They may have entered the chamber colder or hotter than 25° Celsius. When the fuel has been used for chamber wall cooling, it enters the combustion chamber already at elevated temperature (if staged combustion is used, the propellant can be quite hot already). First, we have to calculate the enthalpy which is required to bring the reactants from their actual temperature $T_{injection}$ to the standard temperature 298.15 K. Then we have to calculate the enthalpy which is required to bring the reaction products from standard temperature to combustion temperature $T_{chamber}$. The reaction at standard temperature changes the gas composition and produces the standard heat of reaction. The heat of reaction is used to bring the reactants from $T_{injection}$ to 298.15 K and to bring the reaction products from 298.15 K to $T_{chamber}$:

$$\Delta H_{reaction} = - \sum_{j=1}^{reactants} [n_j \cdot \int_{T(injection)}^{298.15 \text{ K}} c_{p,j} dT] - \sum_{k=1}^{products} [n_k \cdot \int_{298.15 \text{ K}}^{T(chamber)} c_{p,k} dT] \quad (2-37)$$

Term n_j is the amount of substance of a reactant; term n_k of a reaction product. When we use the equation above together with law (2-36) we are able calculate the combustion temperature, provided that the composition of the chamber gas is known.



Example 2-1: Hydrazine is often used as monopropellant in small control thrusters. It decomposes over a heated catalyst bed according to the equation:



Term x is the degree of ammonia dissociation, x can be influenced by the size, geometry and temperature of the catalyst bed, and by the stay time of the propellant in the catalyst bed (anyway $0 \leq x \leq 1$). The reaction of 1 mol hydrazine involves:

$$n_{\text{N}_2\text{H}_4} = 1 \text{ mol} \rightarrow n_{\text{NH}_3} = 4(1-x)/3 \text{ mol}, n_{\text{N}_2} = (1+2x)/3 \text{ mol}, n_{\text{H}_2} = 2x \text{ mol}$$

Using the molecular masses of ammonia (≈ 17), nitrogen (≈ 28) and hydrogen (≈ 2), we can find the mean molecular mass of the reaction products as a function of x . Then we estimate the combustion temperature (approximately 1500 K), take the specific heats of the products from a table and calculate the mean specific heat value:

$$\mathcal{M}_{\text{products}} = \frac{n_{\text{NH}_3} \mathcal{M}_{\text{NH}_3} + n_{\text{N}_2} \mathcal{M}_{\text{N}_2} + n_{\text{H}_2} \mathcal{M}_{\text{H}_2}}{n_{\text{NH}_3} + n_{\text{N}_2} + n_{\text{H}_2}} \quad \left(\approx \frac{96}{5+4x} \frac{\text{gram}}{\text{mol}} \right)$$

$$c_{P,\text{products}} = \frac{n_{\text{NH}_3} c_{P,\text{NH}_3} + n_{\text{N}_2} c_{P,\text{N}_2} + n_{\text{H}_2} c_{P,\text{H}_2}}{n_{\text{NH}_3} + n_{\text{N}_2} + n_{\text{H}_2}} \quad \left(\approx \frac{315-10x}{5+4x} \frac{\text{J}}{\text{molK}} \right)$$

We can observe that the average molecular weight and the average specific heat decrease slightly with the value of x . The reaction enthalpy can be calculated knowing the heat of formation of the gaseous reaction products and the liquid reactant. By definition, the values of heat of formation of hydrogen H_2 and nitrogen N_2 are zero:

$$\begin{aligned} \Delta H_{\text{reaction}} &= n_{\text{NH}_3} \Delta H_{\text{NH}_3} + n_{\text{N}_2} \Delta H_{\text{N}_2} + n_{\text{H}_2} \Delta H_{\text{H}_2} - n_{\text{N}_2\text{H}_4} \Delta H_{\text{N}_2\text{H}_4} \\ &= (4/3)(1-x)(-45898) + 0 + 0 - 1 \cdot 50626 \text{ [J]} \end{aligned}$$

If no ammonia is dissociated ($x = 0$), the reaction delivers the highest amount of heat. We may expect the best performance of the thruster at this point. We can calculate the chamber temperature approximately as:

$$\begin{aligned} T_{\text{chamber}} &= T_{\text{injection}} + \frac{-\Delta H_{\text{reaction}}}{n_{\text{NH}_3} c_{P,\text{NH}_3} + n_{\text{N}_2} c_{P,\text{N}_2} + n_{\text{H}_2} c_{P,\text{H}_2}} \\ &(\approx T_{\text{injection}} + (111823 - 61197x)/(105 - 3x) \text{ [K]}) \end{aligned}$$

The chamber temperature is nearly a linear function of x . If $x = 0$, the reaction products NH_3 , N_2 and H_2 are about 1200 K hotter than the liquid reactant, but the temperature increases just by 600 K if the ammonia is completely dissociated.

2.3.5 The Composition of the Chamber Gas

The principle of conservation of matter. It is well-known that a chemical reaction changes the molecules of the reactants, but not their atoms. After the reaction the products are composed of exactly the same atoms as the reactants before the reaction. The amount of substance remains unchanged for every chemical element. To find the amount of a chemical element (or atom), we have to count the number of atoms in a molecule, then multiply this number by the amount of substance of the concerning molecular species and add up the result for all molecules. This amount is the same before and after the reaction for any one atomic species that participates. Term $atoms(n)$ is the number of atoms of a particular atomic species in substance n :

$$\text{for each atomic species :} \quad \sum_{j=1}^{\text{reactants}} atoms(n_j) \cdot n_j = \sum_{k=1}^{\text{products}} atoms(n_k) \cdot n_k \quad (2-38)$$

When mixed with oxygen and ignited, hydrogen burns with a nearly colorless flame. The product of the reaction is water or steam. The complete combustion of hydrogen in a stoichiometric mixture ($H_2 + 1/2 O_2 \rightarrow H_2O$) transforms 1 mol H_2 and 1/2 mol O_2 into 1 mol H_2O (for example the complete combustion of 2 kg hydrogen with 16 kg oxygen produces 18 kg water vapor, with -241826 kJ heat of formation). LOX/LH2 rocket engines operate “fuel rich”, it means they do not burn the propellant combination LOX/LH2 at (or even near) the stoichiometric proportion of 8 (much more hydrogen is injected, a typical mixture ratio is between 4.5 and 6.0). Additionally, the high combustion temperature causes an incomplete reaction. Apart from water vapor H_2O , we find in the reaction products also H_2 and O_2 , atomic gases (hydrogen H and oxygen O) and hydrogen monoxide gas (OH). Applied to the combustion of LOX/LH2 in a rocket chamber, the atom mass balances (2-38) are:

$$\begin{aligned} \text{O:} \quad & 2 \cdot n_{O_2, oxidizer} = 1 \cdot n_{H_2O} + 2 \cdot n_{O_2} + 1 \cdot n_O + 1 \cdot n_{OH} \\ \text{H:} \quad & 2 \cdot n_{H_2, fuel} = 2 \cdot n_{H_2O} + 2 \cdot n_{H_2} + 1 \cdot n_H + 1 \cdot n_{OH} \end{aligned} \quad (2-39)$$

The fuel hydrazine (N_2H_4) combined with the oxidizer nitrogen tetroxide (N_2O_4) can be used as storable rocket propellant. Also rocket engines which work with such a combination do not operate with the stoichiometric proportion of 1.4375, usually. Hydrazine burns ($2 N_2H_4 + N_2O_4 \rightarrow 4 H_2O + 3 N_2$) and decomposes simultaneously ($3 N_2H_4 \rightarrow 4 NH_3 + N_2$). The atom mass balances according to equation (2-38) are:

$$\begin{aligned} \text{N:} \quad & 2 \cdot n_{N_2H_4, fuel} + 2 \cdot n_{N_2O_4, oxidizer} = 2n_{N_2H_4} + 2n_{N_2O_4} + 2n_{N_2} + n_{NH_3} + n_{NO} + n_N \\ \text{O:} \quad & 4 \cdot n_{N_2O_4, oxidizer} = 4 \cdot n_{N_2O_4} + n_{H_2O} + 2n_{O_2} + n_O + n_{OH} + n_{NO} \\ \text{H:} \quad & 4 \cdot n_{N_2H_4, fuel} = 4n_{N_2H_4} + 2n_{H_2O} + 3n_{NH_3} + 2n_{H_2} + n_H + n_{OH} \end{aligned} \quad (2-40)$$

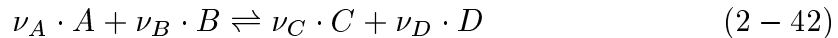
The main reaction products water steam (H_2O), ammonia (NH_3) and nitrogen (N_2) dissociate and form several gas components which consist of elementary molecules, atoms, and free radicals. A very small part of fuel and oxidizer remains unreacted. Other molecules which could be encountered theoretically (for example hydrogen peroxide H_2O_2) are neglected, because the probability of their existence is very small.

The chemical equilibrium. When we want to calculate the composition of the combustion products, we notice that the law of the conservation of matter (2-38) is not sufficient. For reversible chemical reactions there exist the so-called state of equilibrium, where the rate of forming products is exactly equal to the rate of the reverse reaction. Once equilibrium has been reached, the concentration of reactants and reaction products is constant in time. The equilibrium of a chemical reaction depends on the chemical aggressiveness of the substances (the chemical potential or “Gibbs free energy”); it is always a function of the temperature and the mixture of the reactants, sometimes also a function of the pressure. Using the conditions of the chemical equilibrium, we can calculate the composition of the gas in a rocket chamber (assuming that the residence time of the gas is long enough to reach this equilibrium).

When a reversible chemical reaction of a mixture of perfect gases is in its equilibrium, the so-called “equilibrium constant” relates the partial pressures of the participants. The equation of condition for the equilibrium constant $K(T)$ can be written as:

$$K(T) = \frac{1}{(p_{standard})^\nu} \cdot \prod p_k^{\nu_k}, \quad (\nu = \sum \nu_k) \quad (2-41)$$

In this equation, term $p_{standard}$ is the standard pressure, usually 1.01325 bar. Term p_k is the partial pressure of gas k that participates in the chemical reaction. Term ν_k is the stoichiometric molar concentration coefficient. The sign of ν_k is negative if the gas k is a reactant and positive if the gas k is a reaction product. For example, when we consider a reversible reaction with the reactants A and B and the reaction products C and D , we can write the equation of the reaction as:

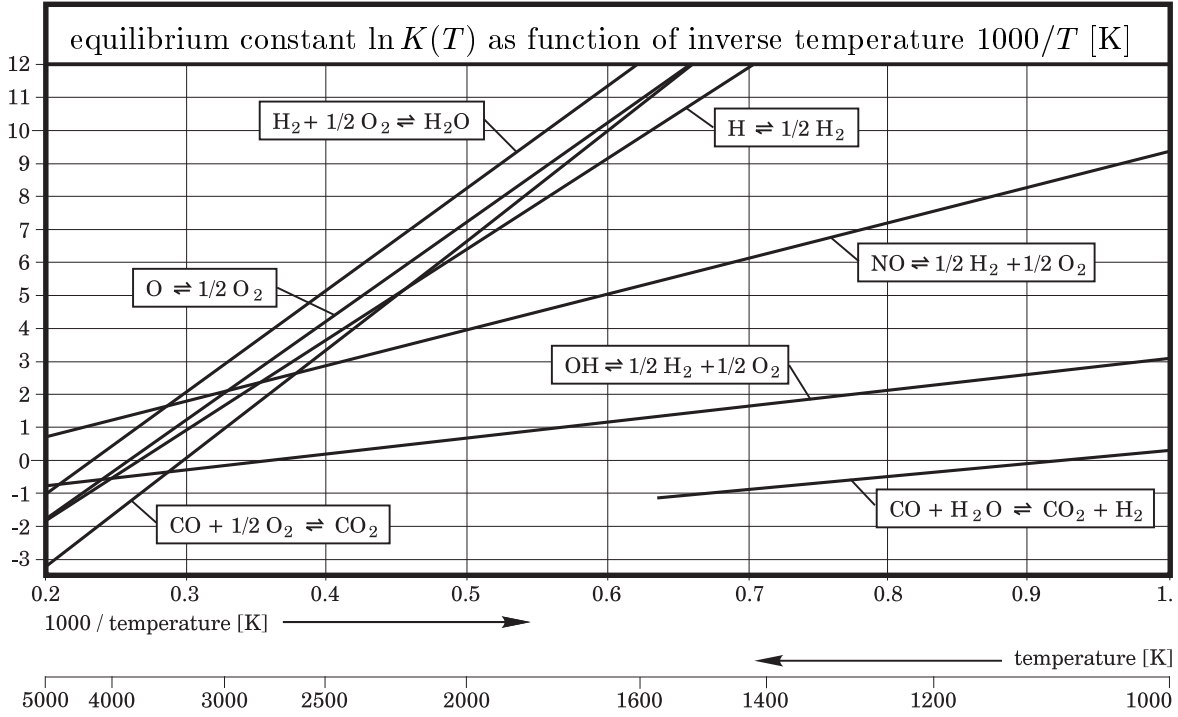


According to formula (2-41), the equilibrium constant $K(T)$ for the reaction above is:

$$K(T) = (p_{standard})^{(\nu_A + \nu_B - \nu_C - \nu_D)} \cdot \frac{p_C^{\nu_C} \cdot p_D^{\nu_D}}{p_A^{\nu_A} \cdot p_B^{\nu_B}} \quad (2-43)$$

Equation (2-43) can be transformed into a representation which uses expressions of n_k (amount of substance) instead of partial pressures p_k (using the perfect gas law 2-29). It can be observed that the chemical equilibrium for a reversible reaction is just independent of the pressure if the condition $\nu = \nu_A + \nu_B - \nu_C - \nu_D = 0$ is valid. For example, this is the case for the “water gas” reaction ($\text{CO} + \text{H}_2\text{O} \rightleftharpoons \text{CO}_2 + \text{H}_2$); but this is not the case for the “oxyhydrogen gas” reaction ($\text{H}_2 + 1/2 \text{O}_2 \rightleftharpoons \text{H}_2\text{O}$). A higher pressure drives the equilibrium to the side of a more complete combustion (the oxyhydrogen reaction reduces the mol-number, the water gas reaction does not).

The equilibrium constant is a function of the temperature. We can calculate $K(T)$ when we know the standard reaction enthalpy ΔH , the standard reaction entropy ΔS and the specific heats $c_{p,k}(T)$ of the constituents of the mixture of perfect gases (these values are tabulated in chemistry handbooks). For the combustion of gases, the values of $K(T)$ are extremely large if the temperature is low (indicating a complete combustion) and small if the temperature is high (indicating dissociating reactants).



$$\ln(K(T)) = \frac{1}{\mathfrak{R}_o} \sum \nu_k \left[\frac{\Delta H_k}{T} - \Delta S_k \right]; \quad \frac{d \ln(K(T))}{dT} = \frac{\sum \nu_k [\Delta H_k]}{\mathfrak{R}_o T^2} \quad (2-44)$$

$$\text{with: } \Delta H_k = \Delta H_{\text{standard},k} + \int_{T(\text{standard})}^{T(\text{chamber})} c_{p,k} dT$$

$$\text{and: } \Delta S_k = \Delta S_{\text{standard},k} + \int_{T(\text{standard})}^{T(\text{chamber})} \frac{c_{p,k}}{T} dT$$

Reaction enthalpy and entropy can be obtained from their standard conditions (usually $p_{\text{standard}}=1.01325$ bar, $T_{\text{standard}}=298.15$ K) by integration of the temperature functions of the specific heats $c_{p,k}(T)$ of the gases. Approximately we get:

$$K(T)_{\text{H}_2 + \frac{1}{2} \text{O}_2 \rightleftharpoons \text{H}_2\text{O}} = \sqrt{p_{\text{standard}}} p_{\text{H}_2\text{O}} / (p_{\text{H}_2} \sqrt{p_{\text{O}_2}}) \quad (\approx e^{(29500/T - 6.740)}) \quad (2-45)$$

$$K(T)_{\text{H} \rightleftharpoons \frac{1}{2} \text{H}_2} = \sqrt{p_{\text{standard}}} \sqrt{p_{\text{H}_2}} / p_{\text{H}} \quad (\approx e^{(27230/T - 7.288)}) \quad (2-46)$$

$$K(T)_{\text{O} \rightleftharpoons \frac{1}{2} \text{O}_2} = \sqrt{p_{\text{standard}}} \sqrt{p_{\text{O}_2}} / p_{\text{O}} \quad (\approx e^{(30220/T - 7.886)}) \quad (2-47)$$

$$K(T)_{\text{OH} \rightleftharpoons \frac{1}{2} \text{H}_2 + \frac{1}{2} \text{O}_2} = \sqrt{p_{\text{H}_2}} \sqrt{p_{\text{O}_2}} / p_{\text{OH}} \quad (\approx e^{(4720/T - 1.680)}) \quad (2-48)$$

$$K(T)_{\text{CO} + \frac{1}{2} \text{O}_2 \rightleftharpoons \text{CO}_2} = \sqrt{p_{\text{standard}}} p_{\text{CO}_2} / (p_{\text{CO}} \sqrt{p_{\text{O}_2}}) \quad (\approx e^{(33477/T - 9.993)}) \quad (2-49)$$

$$K(T)_{\text{NO} \rightleftharpoons \frac{1}{2} \text{N}_2 + \frac{1}{2} \text{O}_2} = \sqrt{p_{\text{N}_2}} \sqrt{p_{\text{O}_2}} / p_{\text{NO}} \quad (\approx e^{(10880/T - 1.485)}) \quad (2-50)$$

The exponential approximations cover a wide range of different temperatures T ; they are usually accurate enough when the combustion temperature is not exactly known (but they are not really accurate for a certain well-defined temperature).

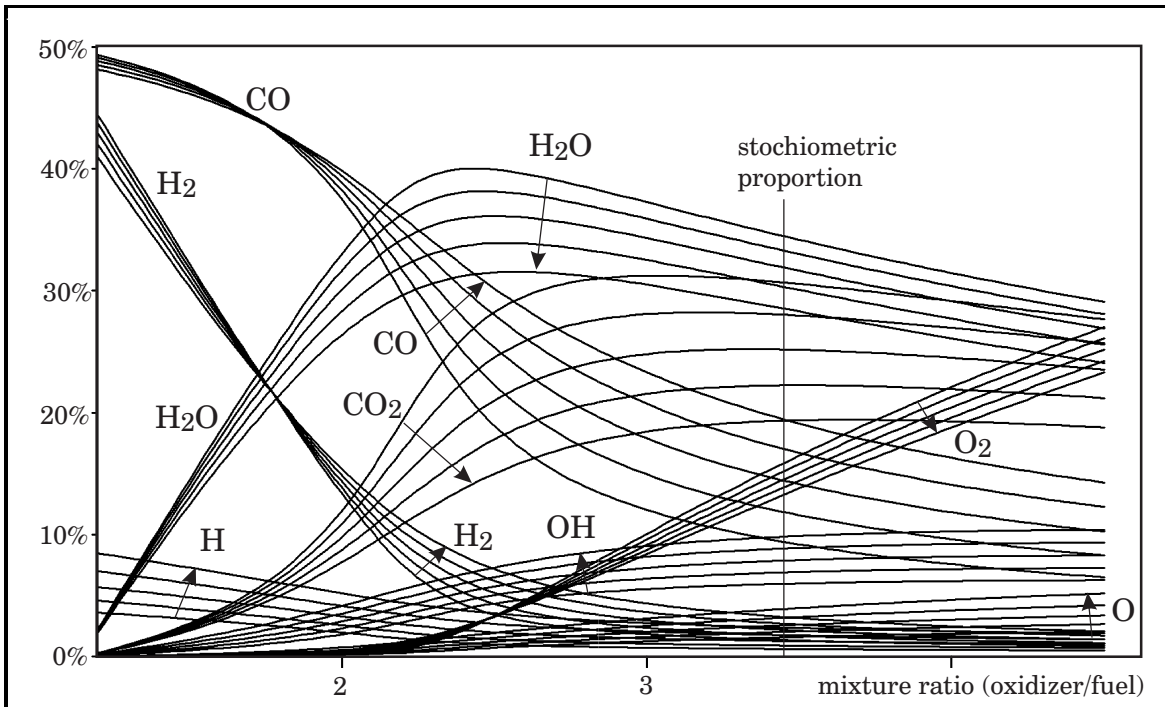
When hydrocarbon fuel is burned in a rocket chamber, the reaction products include water steam as well as carbon dioxide and carbon monoxide, and the water gas reaction takes place ($\text{CO} + \text{H}_2\text{O} \rightleftharpoons \text{CO}_2 + \text{H}_2$). We can find the equilibrium constant for the water gas reaction when we divide equation (2-49) by equation (2-45). Thus:

$$K(T)_{\text{CO}+\text{H}_2\text{O} \rightleftharpoons \text{CO}_2+\text{H}_2} = \frac{p_{\text{CO}_2} p_{\text{H}_2}}{p_{\text{CO}} p_{\text{H}_2\text{O}}} = \frac{K(T)_{\text{CO}+\frac{1}{2}\text{O}_2 \rightleftharpoons \text{CO}_2}}{K(T)_{\text{H}_2+\frac{1}{2}\text{O}_2 \rightleftharpoons \text{H}_2\text{O}}} \approx e^{(3977/T - 3.253)}$$

When water vapor dissociates it forms hydrogen gas and hydrogen monoxide gas $\frac{1}{2} \text{H}_2 + \text{OH} \rightleftharpoons \text{H}_2\text{O}$. This reaction is equivalent to the reaction $\text{H}_2 + \frac{1}{2} \text{O}_2 \rightleftharpoons \text{H}_2\text{O}$ running simultaneously with the reaction $\text{OH} \rightleftharpoons \frac{1}{2} \text{H}_2 + \frac{1}{2} \text{O}_2$. We may write for the concerning equilibrium constant:

$$K(T)_{\frac{1}{2}\text{H}_2+\text{OH} \rightleftharpoons \text{H}_2\text{O}} = \frac{\sqrt{p_{\text{standard}} p_{\text{H}_2\text{O}}}}{\sqrt{p_{\text{H}_2}} p_{\text{OH}}} = K(T)_{\text{H}_2+\frac{1}{2}\text{O}_2 \rightleftharpoons \text{H}_2\text{O}} \cdot K(T)_{\text{OH} \rightleftharpoons \frac{1}{2}\text{H}_2+\frac{1}{2}\text{O}_2}$$

We can observe that it is not necessary to consider a reaction in the calculation of the equilibrium conditions when a set of equivalent reactions has already been considered. When the oxyhydrogen gas reaction and the reaction of dissociating hydrogen monoxide have already been included in the equations, the calculation of the equilibrium for the reaction $\frac{1}{2} \text{H}_2 + \text{OH} \rightleftharpoons \text{H}_2\text{O}$ would provide no additional information anymore.



Example 2-2: The composition of the fire gases inside a rocket chamber is a function of the mixture ratio, the temperature and pressure. The figure shows the composition (in mole per cent) for the combustion of a hydrocarbon fuel RP-1 ($\text{CH}_{1.953}$), burned at a pressure of 50 bar with the oxidizer LOX, as a function of the mixture ratio for temperatures from 3300 K to 3700 K (the arrows indicate increasing temperatures).

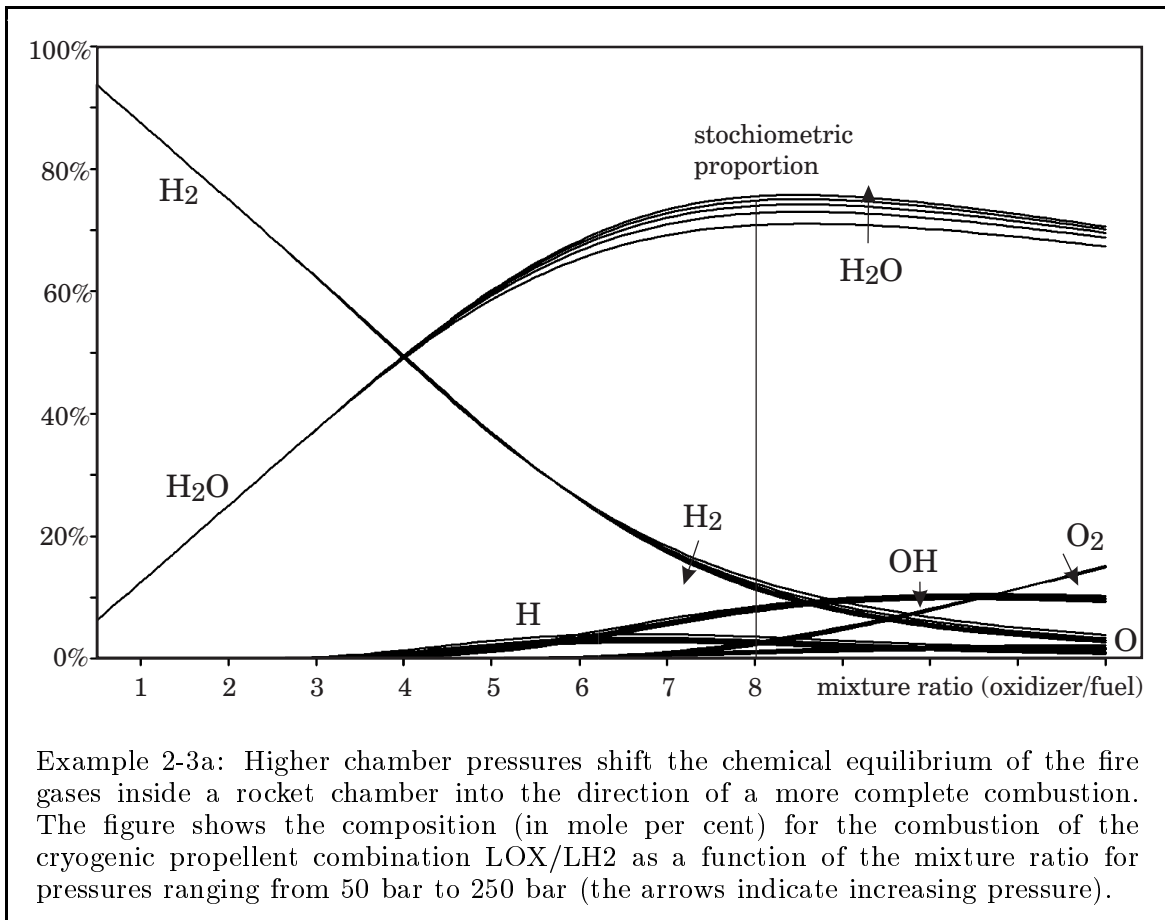
Let us now consider that the propellant combination LOX/LH2 is burned in a rocket chamber. The combustion gas is a mixture of H_2O , H_2 , O_2 , H , O and OH . The combustion chamber works with a certain (well-known) pressure $p_{chamber}$; and this pressure is sum of all the partial pressures of the constituents of the gas:

$$p_{chamber} = p_{H_2O} + p_{H_2} + p_{O_2} + p_H + p_O + p_{OH} \quad (2-51)$$

The mass balances (2-39) for a certain mixture ratio x (oxidizer mass divided by fuel mass) make one more equation available for the calculation of the partial pressures:

$$x \cdot \frac{2}{32} = \frac{n_{O_2}}{n_{H_2}} = \frac{p_{H_2O} + 2p_{O_2} + p_O + p_{OH}}{2p_{H_2O} + 2p_{H_2} + p_H + p_{OH}} \quad (2-52)$$

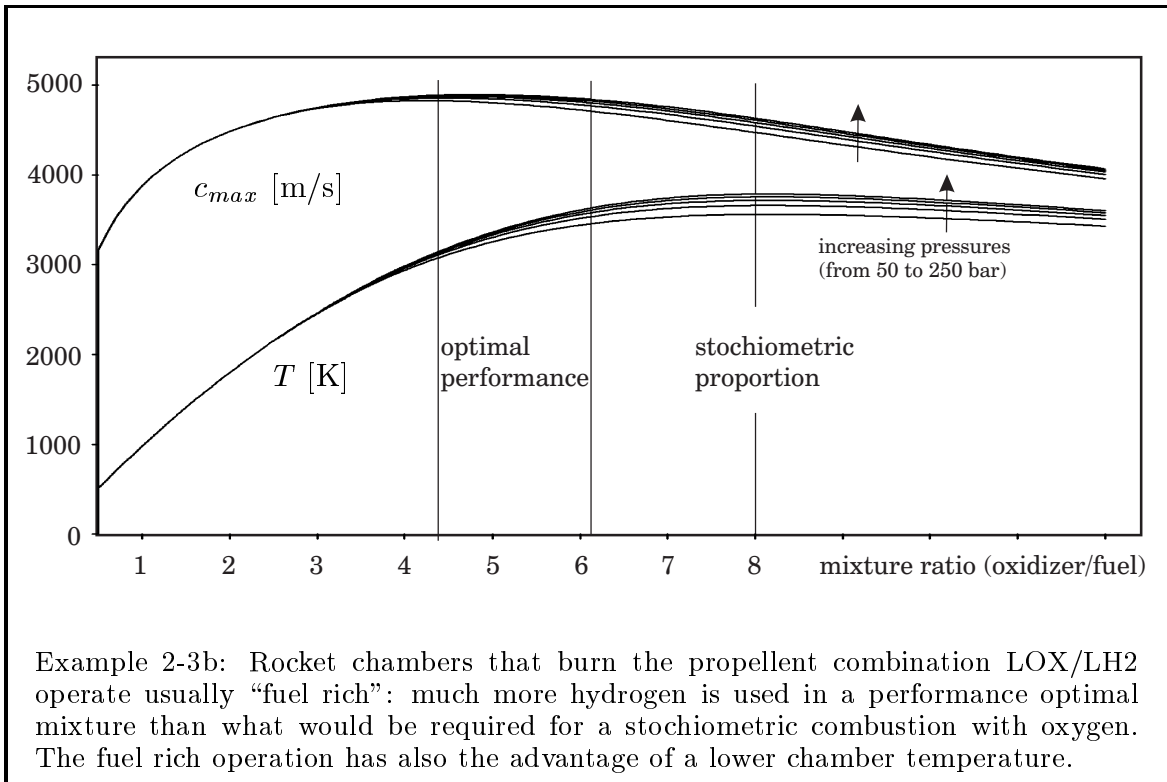
We have to divide the mixture ratio x by the factor 16, since 1 mol of oxygen weighs 16 times more than 1 mol of hydrogen. The combustion gas is regarded as a mixture of perfect gases; we can readily use equation (2-29) to replace the amount of a substance by the partial pressure (replace n_k by p_k). Now we have two equations for six unknown partial pressures (p_{H_2O} , p_{H_2} , p_{O_2} , p_H , p_O and p_{OH}). The four missing relationships are provided by the nonlinear equations (2-45), (2-46), (2-47) and (2-48). These equations relate the partial pressures depending on the temperature $T_{chamber}$. We can estimate $T_{chamber}$ and solve the equation system for the gas composition, and then use the equations (2-36) and (2-37) to improve the value for $T_{chamber}$ iteratively.



Performance optimization. Now we have seen how to compute the temperature and the composition of the chamber gas, however, ultimately we are interested in a performance optimization of the combustion process. Equation (2-35) allows us conclude that the specific chamber enthalpy $h = H/M$ should be as high as possible, we introduce as performance criterion the theoretical “maximum exhaust velocity” c_{max} , ignoring the velocity $c_{chamber}$ and the enthalpy H_{end} in the final area of the nozzle (such an exhaust velocity can actually never be achieved because this would require an infinite nozzle size and an expansion of the gas below the freezing point). Anyway:

$$c_{max} = \sqrt{\frac{2 H_{chamber}}{M}} = \sqrt{\frac{2 c_p T_{chamber}}{\mathcal{M}}} = \sqrt{\frac{2\kappa \Re T_{chamber}}{(\kappa - 1)}} \quad (2 - 53)$$

The mixture ratio influences the chamber temperature $T_{chamber}$ but also the specific heat ratio κ and the gas constant \Re (the universal gas constant $\Re_o = 8314.5 \text{ J/kmol K}$ divided by the mean molecular mass \mathcal{M} in kg/kmol of the gas). However, the optimal mixture ratio is also a function of the chamber pressure $p_{chamber}$. Numerical evaluation of the equation (2-53) for the propellant combination LOX/LH2 shows that optimal mixture ratios require “fuel-rich” combustion ($x \approx 4.5$ to 6.0 ; much more hydrogen is used than what would be required for the stoichiometric proportion $x = 8$). The combustion is incomplete because high temperatures inside the chamber cause the water molecules to dissociate, and it is a way to increase in the specific enthalpy by using more hydrogen (reducing the combustion temperature simultaneously) or by burning the gas at a higher pressure (a higher pressure shifts the equilibrium of the reaction $\text{H}_2 + 1/2 \text{O}_2 \rightleftharpoons \text{H}_2\text{O}$ into the direction of more water steam H_2O).

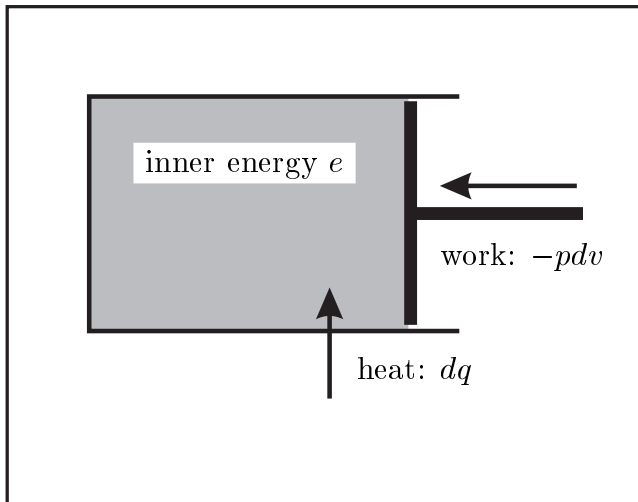


2.4. Hypersonic Nozzles

The hot fire gases stay just for a few milliseconds inside the combustion chamber before they enter the hypersonic rocket nozzle. A thermodynamic expansion process inside the rocket nozzle converts the heat energy of the fire gases into kinetic energy. Also the expansion of the gas is a quite complex process that involves many different chemical and physical phenomena (consider that the gas is still reacting inside the nozzle). The exhaust fumes must leave the final cross sectional area of the nozzle with a high relative velocity; the final gas flow should be uniform and parallel with a minimum content of heat energy. The thermodynamics of a one-dimensional flow process leads us to simple relationships; these equations are useful in practice when we want to analyze the performance of a thrust chamber. However, when we want to design the shape of the nozzle, we have to solve the Euler equations for a two-dimensional flow field (for example using the “method of characteristics”). A computation of the flow field inside the nozzle that considers also friction in the boundary layers near the walls requires the numerical integration of the Navier Stokes equation system. However, particularly because of still insufficient knowledge about reaction kinetics at high temperatures and the interaction with turbulent boundary layer effects, the problem cannot be considered as completely solved by computational fluid dynamics. These numerical computer solutions must be interpreted by experienced scientists.

2.4.1 Thermodynamic Phenomena inside the Rocket Nozzle

The specific enthalpy. We have already introduced the chamber enthalpy $H_{chamber}$ as a thermodynamic state variable that is useful to characterize the capability of the combustion gas to generate mechanical work. Let us now consider a certain amount of gas inside a special gas container: this container is capable to alter its volume, but it is closed in such a way that the gas cannot escape from it. A cylinder with a piston would be a perfect realization of such a gas container. Heat can cross the borders of the gas volume; however, the working gas itself stays always inside the container.



The gas should be a perfect gas, $p \cdot v = \mathfrak{R} \cdot T$; the state variables p (pressure), v (specific volume $v = 1/\rho$) and T (temperature) characterize the actual thermal situation for the gas (\mathfrak{R} is the individual gas constant). Now there are two ways how we can increase the energy content of the gas (its “inner energy”): we can press the piston into the cylinder and reduce the volume, or we can transmit heat into the gas across the borders of the gas container.

The state variables p , T and ρ are independent of the gas quantity in the container; therefore let us now use lower case letters to indicate mass specific quantities for the inner energy e , the enthalpy h , the volume v and the heat q . "Mass specific" means here now "per kilogram of the gas". The first law of thermodynamics (the energy conservation law) states the following condition for the inner energy of the gas:

$$de = -pdv + dq \quad (2 - 54)$$

Both, the work $-pdv$ and the heat dq enter the gas and increase its inner energy de . The infinitesimal mechanical work exerted on the gas by pressing the piston against the pressure of the gas is the product $-pdv$ (the negative sign comes from the fact that the volume is reduced, dv is negative). We can denote that this work is reversible, it means that the entire work comes back when we expand the gas and let the gas push the piston back to its original position. Term dq is infinitesimal heat that crosses the borders of the gas volume. Now we introduce the specific enthalpy $h = e + p \cdot v$, or:

$$dh = de + pdv + vdp \quad (2 - 55)$$

Enthalpy h and inner energy e are both state variables (as well as p , T , and v), which means that these values are not functions of the way on which the gas came to its actual state. We can measure these values (or compute them from measured values) disregarding the history of the gas. The heat q , however, is not a state variable: its actual value depends on the way how the gas has changed its state. Our intention is now to analyze the heat flow dq related to the change in temperature dT . The quantity dq/dT depends on the conditions how the change in state is performed:

$$dq = de + pdv \quad \rightarrow \quad c_v = \left[\frac{dq}{dT} \right]_{v=constant} = \frac{de}{dT} \quad (2 - 56)$$

$$dq = dh - vdp \quad \rightarrow \quad c_p = \left[\frac{dq}{dT} \right]_{p=constant} = \frac{dh}{dT} \quad (2 - 57)$$

The specific heat values c_v (at constant volume) and c_p (at constant pressure), according to the equations above, have now the dimension J/kg K (and not anymore J/mol K, in contrast to the definition 2-30). These specific heats are approximately constant (but actually they are functions of the temperature, compare 2-34). We can use the enthalpy definition $dh = de + d(pv)$ together with the perfect gas law $pv = \Re T$ to show that the specific heats stand in a certain relation to each other:

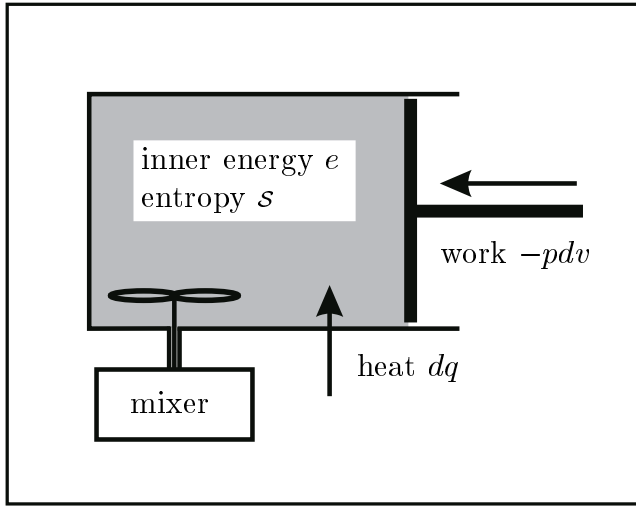
$$\frac{dh}{dT} = \frac{de}{dT} + \frac{d(pv)}{dT} \quad \rightarrow \quad c_p = c_v + \Re \quad (2 - 58)$$

The ratio κ of the specific heats is even more important than c_v and c_p when we examine the change in state of the hot gas that flows through a rocket nozzle. Thus:

$$\kappa = \frac{c_p}{c_v}, \quad \left(c_v = \frac{1}{\kappa - 1} \cdot \Re, \quad c_p = \frac{\kappa}{\kappa - 1} \cdot \Re \right) \quad (2 - 59)$$

The specific heat ratio κ is an approximately constant, dimensionless quantity (≈ 1.1 to 1.3 for the fire gas mixture in a rocket nozzle); κ is a typical propellant characteristic, its value depends mainly on the number of atoms in the gas molecules.

The specific entropy. Let us return to the example of the gas captured inside a cylinder by a piston mechanism. Compressing or expanding the gas is obviously a reversible process: the mechanical work done by the piston on the gas comes back when the original volume is re-established (provided that we execute an adiabatic change in state where all the heat remains inside the gas container, $dq = 0$). However, this is obviously not the case when we put the mechanical energy with a mixer into the gas: the energy conservation law is still valid and the mechanical



energy that drives the mixer increases the inner energy of the gas (the gas temperature grows), but the process is irreversible: the increased temperature will not automatically drop and drive the twirling-stick of the mixer in order to bring back the mechanical energy of the process. The mechanical energy has dissipated during this irreversible process, it heats up the gas and can be treated in the same way as if it were heat energy that had entered the gas container.

The irreversible mechanical work transferred from the mixer to the gas (and in the same way the heat energy dq that crosses the borders of the container) is stored inside the gas in a new state variable called entropy. The specific entropy dS is by definition:

$$dS = \frac{dq}{T} \quad (2 - 60)$$

The second law of thermodynamics declares an important property of this new state variable: the entropy of a closed system (where no heat or work crosses the borders of the system) can never shrink but just grow, in contrast to the energy which is constant inside a closed system. Thus, when we consider for example the whole universe as a closed system, we can conclude that its entropy will grow until finally all processes reach their equilibrium state (which means that the universe will come to an end). The fundamental importance of the entropy as a new state variable is the fact that it gives the coordinate “time” a unique direction: an irreversible adiabatic process can just run into the direction of growing entropy ($dS > 0$); and a cyclic process with heat produced by friction must discharge this heat to outside. Mathematics explains entropy using the statistical probability that a certain state will occur in the future: strictly speaking, it is not impossible that the entropy of a closed system shrinks, but it is extremely unlikely. Our objective is to calculate the flow process inside a rocket nozzle and therefore we can forget all the philosophical considerations about entropy. Instead of this we use the definition (2-60) to rewrite the equation (2-54) in the form:

$$Tds = de + pdv = dh - vdp \quad (2 - 61)$$

The expression Tds on the left side of this equation vanishes for an isentropic process.

The isentropic change in state. A change in state of a gas is called “isentropic” when its entropy remains constant, and this requires automatically that the process is adiabatic (without heat transfer through the walls) and reversible (without friction). We consider that $ds = 0$ and divide the equation (2-61) by the temperature T to find:

$$ds = \frac{de}{T} + \frac{p}{T}dv = \frac{dh}{T} - \frac{v}{T}dp = 0 \quad (2-62)$$

Let us remember that $de = c_v \cdot dT$, $dh = c_p \cdot dT$ and $\kappa = c_p/c_v$ (2-56, 2-57 and 2-59). The perfect gas law $pv = \Re T$ in its differential form becomes:

$$\frac{dp}{p} + \frac{dv}{v} = \frac{dT}{T} \quad (2-63)$$

Now we can transform the isentropic change in state (2-62) into a notation that expresses the pressure p as a function of the volume v and integrate this relationship:

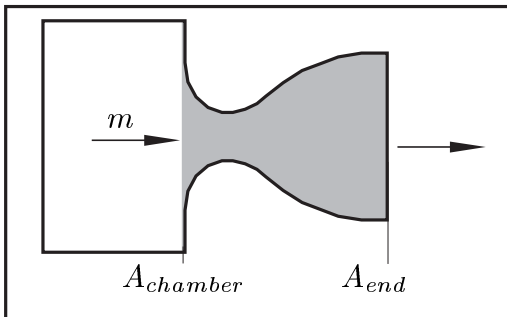
$$\frac{dp}{p} + \kappa \frac{dv}{v} = 0, \quad \rightarrow \quad p \cdot v^\kappa = \text{constant} \quad (2-64)$$

Term κ characterizes an isentropic change in state. The exponent in equation (2-64) accepts another value when the gas changes its state and the entropy is not constant. For example, the exponent equals 1 for an isothermal change in state ($T = \text{constant}$), it equals 0 during an isobaric change in state ($p = \text{constant}$) or ∞ during an isochoric change in state ($v = \text{constant}$). The general case is the so-called polytropic change in state, where the exponent can take an arbitrary value. For the expansion inside a rocket nozzle we will assume that the change in state takes place approximately at constant entropy (however, heat input by friction or by chemical reactions can shift the exponent during the expansion slightly into the direction of a smaller value):

$$\frac{p}{p_{chamber}} = \left(\frac{v}{v_{chamber}} \right)^{-\kappa} = \left(\frac{\rho}{\rho_{chamber}} \right)^\kappa = \left(\frac{T}{T_{chamber}} \right)^{\frac{\kappa}{\kappa-1}} \quad (2-65)$$

When we know just one state variable at a certain location inside the rocket nozzle (for example the temperature T , pressure p , specific volume v or the density ρ), then we can use the relationships (2-65) for the isentropic change in state to calculate the state of the gas at this location from the conditions at another location, for example from the conditions inside the combustion chamber.

The steady flow process. Up to now we have considered the change in state of gas inside a closed system (without mass flow across the borders of the system); now we want to extend the theory and incorporate that gas can enter and leave this container.



We consider the nozzle as a gas container: gas enters the nozzle through the combustion chamber area $A_{chamber}$ and leaves the nozzle through the exit area A_{end} . The expansion becomes a steady flow process when we disregard the special run-on (or shut-off) operation: the flow pattern is then constant in time, and the matter that enters (or leaves) the nozzle is the same for every instant.

Now we have to introduce the velocity c of the gas molecules as a new state variable. The gas velocity is actually a vector that has two components in a two dimensional flow field; however, let us now assume that the axial velocity component (parallel with the centerline of the nozzle) is everywhere much higher than the radial velocity component (perpendicular to the centerline of the nozzle). The gas velocity c becomes a scalar quantity when we neglect the radial component of the gas velocity vector. Strictly speaking, this assumption is just valid for the centerline of the nozzle, but the “one-dimensional streamline theory” has the advantage that the situation for the gas is the same for every location on a cross sectional area of the nozzle.

The mass flow rate m is constant for every sectional area A of the nozzle, therefore:

$$m = \frac{c \cdot A}{v} = \rho \cdot c \cdot A = \text{constant}, \quad \rightarrow \quad \frac{d\rho}{\rho} + \frac{dc}{c} + \frac{dA}{A} = 0 \quad (2-66)$$

Equation (2-66) constitutes the “matter conservation law” for a steady flow process.

Let us now equate the forces that act on the nozzle: gas that enters or leaves imparts a linear momentum to the open gas container, and the pressure in the entry area is different from the pressure in the exit area. The pressure forces on the walls cancel out when they act in radial direction; and therefore we just have to consider the pressure forces on the wall of the nozzle in axial direction. We can write:

$$m \cdot (c_{end} - c_{chamber}) + [A \cdot p]_{end} - [A \cdot p]_{chamber} = \int_{chamber}^{end} p \, dA \quad (2-67)$$

The same law written in a differential notation is also called “Euler equation”, thus:

$$m \, dc + d(pA) = p \, dA, \quad \rightarrow \quad c \, dc = -v \, dp = -\frac{dp}{\rho} \quad (2-68)$$

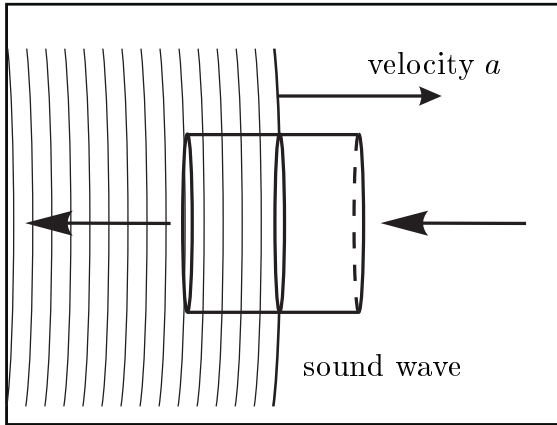
Finally, we have to reconsider the energy conservation law for a steady flow process. Energy forms which could be involved are the kinetic energy of the gas, its enthalpy, a change in the chemical composition of the gas, potential energy in case the gas moves inside a gravitational field, heat input or output and mechanical work done on or by the gas. When we neglect the small quantity of heat exchanged through the walls of the nozzle and when we ignore that the gas is still reacting inside the nozzle, then we just have to consider that the change in the enthalpy of the gas is transformed into kinetic energy. Then the energy conservation law for the nozzle flow process becomes:

$$h_{chamber} - h_{end} = \frac{c_{end}^2}{2} - \frac{c_{chamber}^2}{2}, \quad \rightarrow \quad dh + c \, dc = 0 \quad (2-69)$$

The conservation laws for mass (2-66), linear momentum (2-68) and energy (2-69) enable us to calculate the conditions everywhere inside the rocket nozzle from the conditions inside the combustion chamber (for example the pressure, velocity and temperature at a certain location). Note that we have used for the derivation of these conservation laws considerations concerning the integral forms of these equations. Differentiation of these equations is not always possible, for example, the differential forms are not valid at a discontinuity of the flow pattern (at a so-called shock wave).

The speed of sound and the Mach number. In gas dynamics there are two more quantities frequently used for the description of a flow field: the local velocity of sound (termed a) and the local Mach number (the quotient of actual velocity and speed of sound, $Mach = c/a$). The velocity of sound (or sonic speed) refers to the propagation of small pressure disturbances in a compressible medium, and in gas dynamics the sonic speed is directly a square root function of the local temperature, $a = \sqrt{\kappa \mathfrak{R} T}$.

To analyze the conditional equation for the velocity of sound let us consider a pressure disturbance propagating inside a gas medium. We imagine that a control volume (for example the shape of a cylinder) travels along with the sound wave; its constant cross sectional area is always parallel with the wave front. The conservation laws for matter (2-66) and linear momentum (2-68), applied to the steady flow process inside the control volume with constant cross sectional area ($dA = 0$), take now the form of:



$$\frac{d\rho}{\rho} + \frac{dc}{c} = 0$$

$$c \, dc = -\frac{dp}{\rho}$$

or together:

$$c^2 = a^2 = \frac{dp}{d\rho} \quad (2-70)$$

Provided that the pressure disturbance is small, we can assume an isentropic change in state. Then equation (2-64) determines the density ρ as a function of the pressure p :

$$\frac{dp}{p} + \kappa \frac{dv}{v} = \frac{dp}{p} - \kappa \frac{d\rho}{\rho} = 0 \quad (2-71)$$

The conditional equations for the velocity of sound and the Mach number follow as:

$$a = \sqrt{\left[\frac{dp}{d\rho}\right]_{isentropic}} = \sqrt{\kappa \frac{p}{\rho}} = \sqrt{\kappa \mathfrak{R} T} \quad (2-72)$$

$$Mach = c/\sqrt{\kappa \mathfrak{R} T} \quad (2-73)$$

The sonic speed is well-known to be approximately 334 m/s for the standard conditions on the surface of the earth ($\kappa = 1.4$, $\mathfrak{R} = 286.7$ J/kg K, $T = 288.15$ K); however, the gas inside the thrust chamber of a rocket is much hotter and the local velocity of sound is everywhere in the nozzle considerably higher than 334 m/s. Note that we used the assumption of a small perturbation with isentropic change in state for the derivation of the conditional equation for the speed of sound: this assumption is not valid anymore for strong perturbations caused by an explosion; and the shock wave caused by an explosion travels much faster than determined by the equation (2-72).

2.4.2 Velocity and Temperature in the Isentropic Nozzle

St-Vernant-Wantzel equation. Integration of the Euler equation (2-68) yields:

$$c_{end} = \sqrt{-2 \int_{chamber}^{end} v dp + c_{chamber}^2} \quad (2-74)$$

The assumption of an isentropic change in state of the gas, where $p \cdot v^\kappa = constant$, leads us to the St-Vernant-Wantzel equation (also called Zeuner-Wantzel equation):

$$\begin{aligned} c_{end} &= \sqrt{\frac{2\kappa}{\kappa-1} (pv)_{chamber} \left[1 - \left(\frac{p_{end}}{p_{chamber}} \right)^{\frac{\kappa-1}{\kappa}} \right] + c_{chamber}^2} \\ &= \sqrt{\frac{2\kappa}{\kappa-1} \Re T_{chamber} \left[1 - \left(\frac{p_{end}}{p_{chamber}} \right)^{\frac{\kappa-1}{\kappa}} \right] + c_{chamber}^2} \end{aligned} \quad (2-75)$$

The St-Vernant-Wantzel equation expresses the exhaust velocity c_{end} (the relative velocity of the exhausted gases in the final area of the nozzle) as a function of the pressure ratio $p_{end}/p_{chamber}$; the velocity in the chamber is usually neglected ($c_{chamber} \approx 0$). We can also use this equation to determine the velocity c as a function of the pressure p at any other location inside the nozzle when we simply omit the index “end”. We have already seen that, in order to achieve a high exhaust velocity, a high specific chamber enthalpy is desirable: the expression in front of the opening square bracket in equation (2-75) is two times the specific chamber enthalpy. For a calorically perfect gas ($c_p \approx constant$) we can integrate equation (2-57) to find $h = c_p \cdot T$. Thus:

$$h_{chamber} = c_p \cdot T_{chamber} = \frac{\kappa}{\kappa-1} \Re \cdot T_{chamber} \quad (2-76)$$

The individual gas constant of the chamber gas \Re is the universal gas constant $\Re_o = 8314.5 \text{ J/kmol K}$ divided by the mean molecular mass of the gas; and, considering this, it is often stated that the St-Vernant-Wantzel equation explains why a low mean molecular mass \mathcal{M} of the chamber gas is desirable for a high exhaust velocity. However, we can readily verify that the exhaust velocity c_{end} is a function of three factors only: the specific chamber enthalpy $h_{chamber}$, the isentropic exponent κ and the pressure ratio $p_{end}/p_{chamber}$. The mean molecular mass of the chamber gas is actually just important in case it has an influence on these main performance factors.

Temperatures in the rocket nozzle. The fire gases experience during the very small time interval of the expansion a considerable decrease of pressure and an increase of the specific volume. The state of the gas is determined by the perfect gas law, and when we assume an isentropic change in state of the gas (equation 2-65), we can readily calculate the temperature drop that accompanies the pressure drop:

$$T_{end} = T_{chamber} \cdot \left(\frac{p_{end}}{p_{chamber}} \right)^{\frac{\kappa-1}{\kappa}} \quad (2-77)$$

When we want to determine the temperature T inside the nozzle at a location different from the exit area, we may simply omit the index “end” and use the same relationship.

Duct area. The equation (2-65) explains the decrease of the gas density ρ during the isentropic expansion process inside the rocket nozzle; and we can rewrite it as:

$$\rho_{end} = \rho_{chamber} \cdot \left(\frac{p_{end}}{p_{chamber}} \right)^{\frac{1}{\kappa}} \quad (2-78)$$

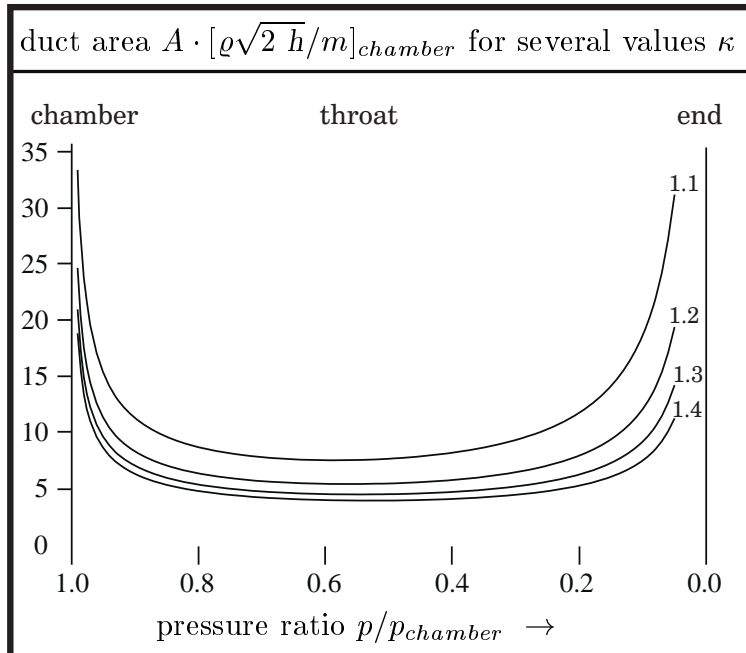
However, our intention is now to calculate the area A of the nozzle as a function of the pressure drop $p/p_{chamber}$. The matter conservation law ($m = \rho \cdot c \cdot A$, equation 2-66) indicates that we have to consider the density of the gas (2-78) and its velocity. We ignore in equation (2-75) the chamber velocity ($c_{chamber} = 0$) and conclude:

$$\begin{aligned} \frac{m}{A} &= \rho_{chamber} \cdot \left(\frac{p}{p_{chamber}} \right)^{\frac{1}{\kappa}} \cdot \sqrt{\frac{2\kappa}{\kappa-1} \Re T_{chamber} \left[1 - \left(\frac{p}{p_{chamber}} \right)^{\frac{\kappa-1}{\kappa}} \right]} \\ &= \rho_{chamber} \cdot \sqrt{\frac{2\kappa}{\kappa-1} \Re T_{chamber} \left[\left(\frac{p}{p_{chamber}} \right)^{\frac{2}{\kappa}} - \left(\frac{p}{p_{chamber}} \right)^{\frac{\kappa+1}{\kappa}} \right]} \end{aligned} \quad (2-79)$$

or:

$$A = \frac{m}{\rho_{chamber} \sqrt{2h_{chamber}}} \cdot \left[\left(\frac{p}{p_{chamber}} \right)^{\frac{2}{\kappa}} - \left(\frac{p}{p_{chamber}} \right)^{\frac{\kappa+1}{\kappa}} \right]^{-1/2} \quad (2-80)$$

Term m is the mass flow rate, a value that is usually predetermined for the combustion chamber of a certain rocket engine (as well as the specific enthalpy h , the density ρ and the pressure p inside the combustion chamber). The exponent κ is essentially a characteristic of the propellant. We can learn from equation (2-80) that the duct area A of the nozzle is a nonlinear function of the pressure drop $p/p_{chamber}$. According to the equation above the duct area is infinitely large ($A \rightarrow \infty$) for two occasions: first $p = p_{chamber}$ (this comes from the assumption that the velocity in the combustion chamber is zero) and then $p_{end} = 0$ (expanding the gas to an infinitely small pressure).



The equation is obviously not valid for the two boundary values, however, the numerical evaluation shows us that there exist a minimum duct area that is called the “throat of the nozzle”. Between the chamber and the throat there is the “convergent part of the nozzle”; between the throat and the exit area there is the “divergent part of the nozzle”. Hypersonic nozzles are also called Laval nozzles, named after the Swedish engineer Carl G.P.Laval (1845 - 1913).

Velocity and pressure in the throat area. Let us now take a closer look at the conditions for the gas in the throat area. It is easy to show that the local velocity of sound (the acoustic speed a) prevails in the throat of an isentropic rocket nozzle: term dA/A in the throat is zero by definition, and we can take the differential forms of the conservation laws for matter and linear momentum (2-66 and 2-68) to calculate:

$$\frac{d\rho}{\rho} + \frac{dc}{c} = 0 ; \quad c \, dc = -\frac{dp}{\rho} \quad \rightarrow \quad c_{throat}^2 = \left[\frac{dp}{d\rho} \right]_{A=constant} \quad (2-81)$$

For an isentropic process this is exactly the condition (2-70) for the sonic speed, thus:

$$c_{throat} = a = \sqrt{\kappa \mathfrak{R} T_{throat}} \quad (2-82)$$

$$Mach_{throat} = 1 \quad (2-83)$$

There is, however, a requirement for the presence of the velocity of sound in the throat area of an isentropic rocket nozzle: the local acoustic velocity (or critical velocity) is just established when the pressure in the combustion chamber is sufficiently high; or, in other words, the critical pressure ratio (or Laval pressure) $p_{throat}/p_{chamber}$ must be present in the hypersonic rocket nozzle. We find this pressure ratio when we equate the condition (2-82) with the throat velocity determined by equation (2-75), neglecting again the chamber velocity and considering the relationship (2-77). Thus:

$$c_{throat} = \sqrt{\kappa \mathfrak{R} T_{chamber} \left(\frac{p_{throat}}{p_{chamber}} \right)^{\frac{\kappa-1}{\kappa}}} = \sqrt{\frac{2\kappa}{\kappa-1} \mathfrak{R} T_{chamber} \left[1 - \left(\frac{p_{throat}}{p_{chamber}} \right)^{\frac{\kappa-1}{\kappa}} \right]}$$

The result is:

$$\left(\frac{p_{throat}}{p_{chamber}} \right)_{critical} = \left(\frac{2}{\kappa+1} \right)^{\frac{\kappa}{\kappa-1}} \quad (2-84)$$

The critical pressure ratio is exclusively a function of the isentropic exponent κ ; for example its value is 0.585 for $\kappa = 1.1$, 0.564 for $\kappa = 1.2$, 0.546 for $\kappa = 1.3$ and so on.

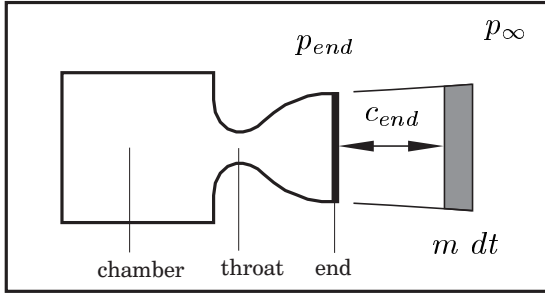
In case the pressure ratio $p_{throat}/p_{chamber}$ is greater than the critical pressure ratio, the velocity of sound can not be established in the throat of the rocket nozzle. Such a situation can occur when the pressure in the final nozzle area is too high; and, consequently, the chamber pressure is not high enough. Then the throat velocity is smaller than the sonic speed and the divergent part of the nozzle decelerates the gas flow (the divergent part of the nozzle operates as a diffuser). Everywhere in the nozzle the gas flows with subsonic speed. However, when the chamber pressure is high enough, it establishes the critical pressure and the critical velocity in the throat:

$$c_{throat} = \sqrt{\kappa \mathfrak{R} T_{throat}} = \sqrt{2\kappa \mathfrak{R} T_{chamber}/(\kappa-1)} \quad (2-85)$$

The pressure in the throat p_{throat} cannot become smaller than determined by the critical pressure ratio (2-84), and the velocity in the throat c_{throat} cannot become greater than determined by equation (2-85). The divergent part of the nozzle operates as a hypersonic nozzle: the gas flows at supersonic speed and accelerates while the duct area grows and the pressure drops. An expansion of the gas beyond the critical throat pressure in the hypersonic part of the nozzle requires a divergent duct area.

Optimal final pressure. We can use the St.Vernant-Wantzel equation (2-75) to calculate the velocity c_{end} of the gas as function of the pressure p_{end} in the final duct area of the rocket nozzle, and obviously the exhaust velocity is high when the final pressure is low. However, ultimately we are interested in a thrust optimization; and the pressure in the final nozzle area contributes to the thrust of the engine too.

The entire thrust of the engine consists of two parts: the momentum thrust $c_{end} \cdot m$ and the pressure thrust $A_{end} \cdot (p_{end} - p_{\infty})$, compare equation (2-3). Term p_{∞} is the ambient pressure (it ranges from $p_{\infty} \approx 1$ bar at the surface of the earth down to $p_{\infty} = 0$ bar in the vacuum of space). The effective exhaust velocity $c_{effective}$ is then:



$$c_{effective} = c_{end} + \frac{A_{end}}{m} (p_{end} - p_{\infty}) \quad (2-86)$$

(compare equation 2-3)

In order to find the optimal expansion as a function of the final pressure we have differentiated the equation (2-86) with respect to p_{end} . The conventional procedure is:

$$\frac{dc_{effective}}{dp_{end}} = 0 ; \quad \rightarrow \quad \frac{dc_{end}}{dp} + \frac{A_{end}}{m} + \frac{dA_{end}}{dp} \cdot \frac{p - p_{\infty}}{m} = 0 \quad (2-87)$$

Equation (2-87) is the conditional equation for the pressure p_{end} in the final duct area of a rocket nozzle with optimal thrust. The first two terms of the left hand side of this equation cancel out, because the conservation laws for matter (equation 2-66) and linear momentum (the Euler equation 2-68) state $A/m = v/c$ and $dc/dp = -v/c$. We can conclude that the “optimum expanded” nozzle operates with the ambient pressure in the final duct area, $p_{end} = p_{\infty}$. A rocket nozzle that operates with the ambient pressure in the final duct area is also called an “adapted nozzle”.

The expansion ratio. We have seen that it would be optimal to operate the nozzle always with the ambient pressure in the final duct area. However, then we would have to construct a expandable nozzle for a space launcher that climbs through the atmosphere from sea-level to space. Initially, the nozzle would be comparatively short, later in space it would be extremely big. Since it is obviously for practical reasons impossible (or at least too difficult) to construct a nozzle with variable size, the optimal size is a compromise solution. The expansion is characterized by the ratio ϵ :

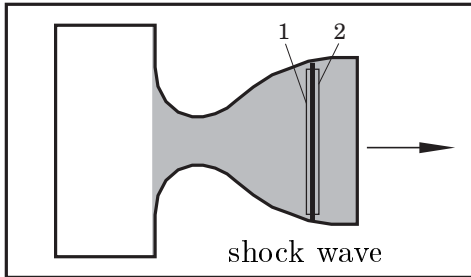
$$\epsilon = \frac{A_{end}}{A_{throat}} \quad (2-88)$$

Typical values for the expansion ratio ϵ range from 6 (comparatively short, ground level adapted nozzles) to 100 (comparatively long, vacuum adapted rocket nozzles). For ground level operation (boosters and first stages), nozzles operate often with overexpansion (overexpansion is possible up to about 40% of the ambient pressure).

2.4.3 Overexpanding Nozzles

Pressure shocks. Let us now consider a rocket nozzle with a large exit area A_{end} and a comparatively low throat pressure p_{throat} . The expansion of the exhaust gases in the divergent part of the nozzle cannot lead to a final pressure that is considerably lower than the ambient pressure (for example $p_{end} < 0.4 \cdot p_{\infty}$). In such a case, however, a separation of the gas flow from the nozzle wall takes place, accompanied by pressure shocks inside the nozzle which can transform the supersonic flow instantaneously into a subsonic flow. Pressure shocks inside hypersonic rocket nozzles may not occur during normal operation, but they play a role during startup and shutdown. These shock waves can cause destructive loads for the wall structure of the nozzle.

Conditions for the occurrence of a pressure shock. The sudden rise in pressure takes place inside a very thin front. In order to analyze this phenomenon with the one dimensional flow theory we apply the conservation laws for matter, linear momentum and energy (the integral forms of the equations 2-66, 2-67 and 2-69) to a constant area control volume placed over the thin shock front. The index “1” characterizes the conditions immediately before the shock; the index “2” the conditions after the shock:



$$\varrho_2 c_2 = \varrho_1 c_1 \quad (2-89)$$

$$\varrho_2 c_2^2 + p_2 = \varrho_1 c_1^2 + p_1 \quad (2-90)$$

$$\kappa \frac{p_2}{\varrho_2} + \frac{\kappa - 1}{2} c_2^2 = \kappa \frac{p_1}{\varrho_1} + \frac{\kappa - 1}{2} c_1^2 \quad (2-91)$$

To find the second of these equations we have to divide the equation (2-67) by the constant control volume area $A = m/(\varrho c)$; to find the third equation we have to use the perfect gas law $p = \varrho \mathfrak{R} T$ and replace the specific enthalpy h by the expression $c_p \cdot T$ (the specific heat at constant pressure is $c_p = \kappa \mathfrak{R}/(\kappa - 1)$, compare equations 2-59). The equations (2-89), (2-90) and (2-91) provide a nonlinear system for the calculation of the velocity c , the pressure p and the density ϱ immediately after the shock, assuming that these values are well-known immediately before the shock. The equation system has obviously the trivial solution that nothing happens in the control volume ($c_2 = c_1$, $\varrho_2 = \varrho_1$, $p_2 = p_1$), but it exists also another solution. This other solution is:

$$c_2 = c_1 \cdot \left(1 - \frac{2}{\kappa + 1} \left(1 - \frac{1}{Mach^2}\right)\right) \quad (2-92)$$

$$\varrho_2 = \varrho_1 / \left(1 - \frac{2}{\kappa + 1} \left(1 - \frac{1}{Mach^2}\right)\right) \quad (2-93)$$

$$p_2 = p_1 \cdot \left(1 + \frac{2\kappa}{\kappa + 1} (Mach^2 - 1)\right) \quad (2-94)$$

The expression $Mach^2 = c_1^2/(\kappa \mathfrak{R} T_1) = c_1^2 \varrho_1/(\kappa p_1)$ is the square of the Mach number immediately before the shock. The equations (2-92) to (2-94) are the well-known relationships of gas dynamics valid for a rectangular pressure shock inside a gas pipe.

Entropy change. As the next point we want to examine under which conditions pressure shocks can occur inside a nozzle. We can readily verify that pressure shocks cannot occur when the gas is flowing with sonic speed, because the equations (2-92) to (2-94) accept the trivial solution when the Mach number equals 1. Pressure shocks can also not occur when the gas flows with subsonic speed, because a shock is always accompanied by a change in the entropy of the gas; and we can see that the conditions for an isentropic change in state (equations 2-65) are not satisfied for a pressure shock. To integrate the conditional equation (2-61) for the entropy change we can calculate:

$$dS = \frac{dh}{T} - \frac{vdp}{T} = c_p \frac{dT}{T} - \Re \frac{dp}{p} = \frac{\Re}{\kappa - 1} \left(\kappa \frac{dT}{T} - (\kappa - 1) \frac{dp}{p} \right) \quad (2-95)$$

$$\rightarrow s_2 - s_1 = \frac{\Re}{\kappa - 1} \left(\ln \left(\frac{T_2}{T_1} \right)^\kappa - \ln \left(\frac{p_2}{p_1} \right)^{\kappa-1} \right) = \frac{\Re}{\kappa - 1} \ln \left(\left(\frac{p_2}{p_1} \right) \cdot \left(\frac{\rho_1}{\rho_2} \right)^\kappa \right)$$

Using (2-93) and (2-94), the entropy change for the rectangular shock wave becomes:

$$s_2 - s_1 = \frac{\Re}{\kappa - 1} \ln \left(\left(1 + \frac{2\kappa}{\kappa + 1} (Mach^2 - 1) \right) \cdot \left(1 - \frac{2}{\kappa + 1} \left(1 - \frac{1}{Mach^2} \right) \right)^\kappa \right) \quad (2-96)$$

The entropy change is positive for Mach numbers greater than 1 and negative for Mach numbers smaller than 1, indicating that subsonic pressure shocks cannot exist.

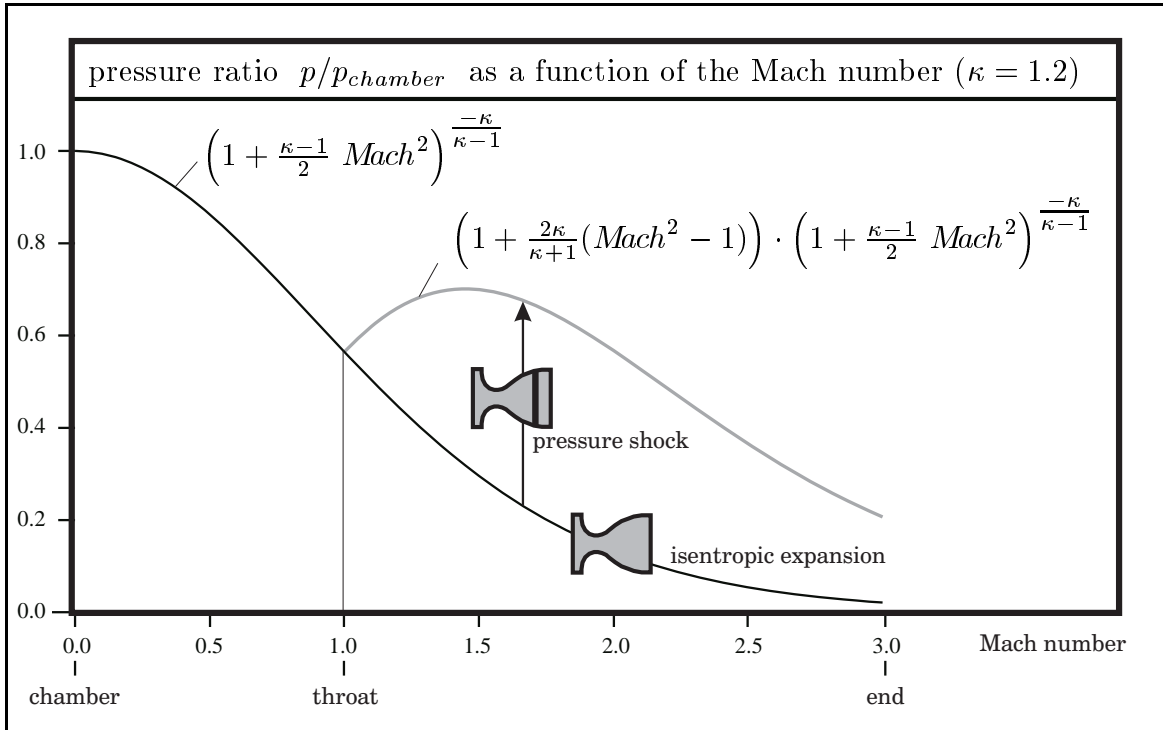
Pressure ratios in the overexpanding nozzle. The pressure ratio in a rocket nozzle with isentropic expansion can be expressed as a function of the Mach number. The relationship follows from the energy conservation law $c_p \cdot T_{chamber} = c_p \cdot T + c^2/2$, divided by the velocity of sound $a^2 = \kappa \Re T$ and transformed using the equation (2-77):

$$p = p_{chamber} \cdot \left(1 + \frac{\kappa - 1}{2} Mach^2 \right)^{\frac{-\kappa}{\kappa - 1}} \quad (2-97)$$

For example, the pressure in the final area of the nozzle is $p_{end} = 0.0212 \cdot p_{chamber}$ for an exponent $\kappa = 1.2$ and a final Mach number $Mach = 3$; and in this example an engine is adapted to the sea-level pressure $p_{end} = 1$ bar when operates with a chamber pressure of $p_{chamber} = 47$ bar. Equation (2-97) shows that the final pressure p_{end} diminishes proportionally when we start to reduce the chamber pressure. When the chamber pressure gets smaller than a certain value, a pressure shock in the final area of the nozzle takes place that increases the pressure ratio instantaneously:

$$p = p_{chamber} \cdot \left(1 + \frac{\kappa - 1}{2} Mach^2 \right)^{\frac{-\kappa}{\kappa - 1}} \cdot \left(1 + \frac{2\kappa}{\kappa + 1} (Mach^2 - 1) \right) \quad (2-98)$$

The equation above indicates that in any case a pressure shock will occur when we diminish the chamber pressure below 4.84 bar (in practice, however, the flow will separate earlier from the walls and oblique shock waves will take place inside the nozzle). The shock front travels upstream towards the throat when we continue to reduce the chamber pressure; however, it disappears when it reaches the throat. Finally, when the chamber pressure is smaller than the critical pressure (1.773 bar, equation 2-84), the gas flows through the nozzle with subsonic velocity at every location.



It is obvious that these pressure shocks may not occur during the normal operation. The example, however, illustrates what happens during start-up and shut-down: when a rocket engine designed for sea-level operation is ignited, the pressure in the thrust chamber grows rapidly and blows the pressure shock wave out of the nozzle; but the shock front enters the nozzle again from behind when the engine is stopped. The engine must be constructed to stand the additional loads of the shock wave. In case a ground-level adapted engine has been designed for variable thrust, its throttleability is limited because the chamber pressure may not be controlled below a certain value. The expansion ratio ϵ (equation 2-88) of nozzles designed for ground-level ignition is usually much smaller than the ratio ϵ of nozzles for vacuum operation.

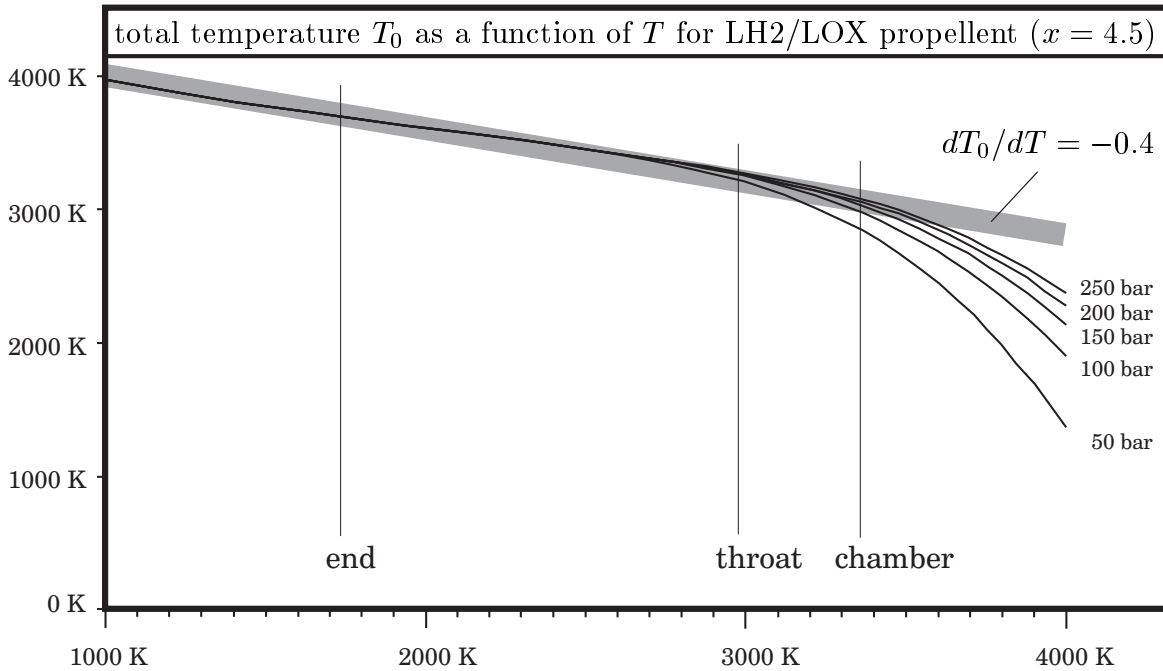
2.4.4 The Shifting Chemical Equilibrium

The reacting expansion process. The chemical reaction that burns the propellant takes place at high temperatures and high pressures (for example, liquid rocket engines operate with combustion chamber temperatures of about 3500 K at pressures of up to 200 bar); and, as a consequence of the high combustion temperature, a considerable portion of the propellant remains unreacted. The chamber gases are extremely hot when they enter the nozzle. We have seen that the pressure drop in the nozzle is accompanied by a temperature drop: the gases are considerably colder when they leave the nozzle (for example 1447 K, equation 2-77). The equilibrium for reversible chemical reactions is a function of the temperature (sometimes also a function of the pressure, compare equation 2-43); during the expansion it changes rapidly into the direction of a more complete combustion: the gases are still reacting inside the nozzle.

Heat input. The expansion becomes a “diabatic flow process” when we consider that heat is produced inside the flowing gas by the chemical reactions. The first step in our analysis is to calculate the heat input: the conservation law for the number of atoms (2-38) together with the chemical equilibrium for every individual reaction (2-41) allows us to calculate the chemical composition of the gas as a function of the temperature and pressure; then equation (2-36) allows us to calculate the heat of reaction $\Delta H_{reaction}$ as a function of the composition of the gas (or $\Delta h_{reaction}$, divided by the mass of the gas). The heat of reaction is a relative value: when we want to calculate the chamber temperature from this value, we have to consider the temperature of the gas before the combustion and the specific heat c_p of the gas after the combustion. However, our intention is now to find the heat released when the temperature drops. The change in the heat of reaction is the energy input for the diabatic flow process:

$$c_p \cdot T_0 = h_{chamber} + \Delta h_{reaction}(T, p) \quad (2 - 99)$$

According to the equation above, the value for the specific heat of reaction $\Delta h_{reaction}$ equals zero when the gas is at chamber conditions. Term T_0 , the stagnation point temperature (or total temperature) of the flowing gas, is well-known to be constant for an isentropic nozzle expansion process. For an expansion with heat input, however, the total temperature $T_0 = h/c_p$ is a function that grows with dropping temperature T :



$$T_0 \approx T_{chamber} + \frac{dT_0}{dT} \cdot (T - T_{chamber}) \quad (2 - 100)$$

As an example, the figure above shows T_0 as a function of T for the propellant combination LOX/LH2 (mixture ratio $x = 4.5$), the pressures range from 50 to 250 bar. We can see that T_0 is, independent of the pressure p , approximately a linear function of the temperature T (then we have $dT_0/dT \approx -0.4$ for temperatures below 3300 K).

Conservation laws for a diabatic flow process. Let us now reconsider the conservation laws for matter, linear momentum and energy (the equations 2-66 to 2-69):

$$\text{matter:} \quad m = \varrho \cdot c \cdot A = \frac{c \cdot A \cdot p}{\Re \cdot T} = \text{constant} \quad (2-101)$$

$$\text{linear momentum:} \quad c \, dc = -\frac{dp}{\varrho} = -\Re \, T \, \frac{dp}{p} \quad (2-102)$$

$$\text{energy:} \quad c_p \, T_0 + \frac{c_{chamber}^2}{2} = c_p \, T + \frac{c^2}{2} \quad (2-103)$$

These three conservation laws serve as an equation system for the calculation of the pressure p , the temperature T and the gas velocity c everywhere inside the nozzle from the conditions in the combustion chamber (we can use the perfect gas law $p = \Re T \varrho$ to eliminate the density ϱ and the definition $h = c_p \cdot T$ to eliminate the enthalpy h). Term m in equation (2-101) is the constant mass flow rate and A is the duct area. The expansion in the nozzle is an isentropic flow process when the flow field is free of pressure shock waves and the stagnation point temperature (or total temperature) T_0 equals the chamber temperature $T_{chamber}$ at every location inside the rocket nozzle (the idealized case of a “frozen equilibrium”). However, heat input from chemical reactions causes the total temperature T_0 to grow during the expansion in the rocket nozzle, and the expansion process is actually a diabatic flow process with heat input and growing entropy. The idealized case of a “shifting equilibrium” considers the heat input by chemical reactions during the expansion of the gas in the rocket nozzle.

Exhaust temperature. We can differentiate the energy conservation law (2-103) and insert the linear momentum conservation law (2-102) to find the relationship:

$$\frac{\kappa \, \Re}{\kappa - 1} (dT_0 - dT) = c \, dc = -\Re \, T \, \frac{dp}{p} \quad (2-104)$$

Term c_p is the specific heat of the gas at constant pressure, $c_p = \kappa \Re / (\kappa - 1)$. When the gas composition changes with the shifting equilibrium the value κ is nearly constant. Assuming the validity of equation (2-100) with a constant factor dT_0/dT we can write:

$$\frac{\kappa}{\kappa - 1} \left(1 - \frac{dT_0}{dT}\right) \frac{dT}{T} = \frac{dp}{p} \quad (2-105)$$

This relationship for a polytropic change in state can be integrated at once. Therefore:

$$T_{end} = T_{chamber} \cdot \left(\frac{p_{end}}{p_{chamber}}\right)^{\frac{\kappa-1}{\kappa(1-dT_0/dT)}} \quad (2-106)$$

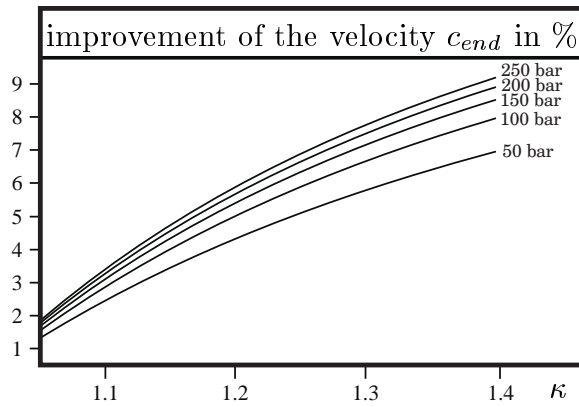
For example, for an isentropic expansion process ($T_0 = \text{constant}$) we find the nozzle exit temperature $T_{end} = 1447$ K (assuming $T_{chamber} = 3500$ K, $p_{chamber}/p_{end} = 200$ and $\kappa = 1.2$). A diabatic expansion process with heat input by a shifting chemical equilibrium, $dT_0/dT \approx -0.4$, however, gives for the same chamber conditions a considerably hotter nozzle exit temperature: $T_{end} = 1862$ K. We can estimate the actual nozzle exit temperature between these two boundary values, because neither the frozen chemical equilibrium (the adiabatic flow process) nor the shifting equilibrium (the diabatic process) represents the actual situation in the rocket nozzle accurately.

Exhaust velocity. We have seen that heat input by chemical reactions in the rocket nozzle raises the temperature of the exhausted gas considerably; ultimately, however, we are interested in finding the influence of the heat input on the thrust of the engine. Therefore we have to calculate the exhaust velocity c_{end} in the final area of the nozzle. The exhaust velocity follows immediately from the energy conservation law (2-103) when we insert the actual temperature T_{end} (according to the equation 2-106) and the total temperature T_0 (according to the equation 2-100). Then we can write down:

$$c_{end} = \sqrt{2 c_p (T_0 - T_{end}) + c_{chamber}^2} \quad (2-107)$$

$$= \sqrt{\frac{2\kappa}{\kappa-1} \Re T_{chamber} \left(1 - \frac{dT_0}{dT}\right) \left[1 - \left(\frac{p_{end}}{p_{chamber}}\right)^{\frac{\kappa-1}{\kappa(1-dT_0/dT)}}\right] + c_{chamber}^2}$$

Interesting is now the comparison of the exhaust velocity c_{end} for the shifting chemical equilibrium (2-107) with the exhaust velocity for the frozen equilibrium (the St.-Vernant-Wantzel equation 2-75). We may expect that the calculated performance of the nozzle increases when we consider that heat is transferred to the gas flow.



The figure on the left shows the improvement of the exhaust velocity as a function of κ for several pressure ratios ($p_{chamber}/p_{end}$ from 50 to 250 bar), assuming $c_{chamber} = 0$ and $dT_0/dT = -0.4$. The calculated performance improves by a few percent only, a disappointing result for a considerable amount of heat input to the gas flow.

Estimated performance. The actual performance of the nozzle is somewhere in the middle between the performance calculated using the frozen equilibrium and the performance calculated with the shifting equilibrium. The calculation with frozen equilibrium under-estimates the performance by a few percent; the shifting equilibrium over-estimates the performance. Comparison with experiments shows that for small engines the frozen equilibrium calculation is the better representative, for big engines, however, the shifting equilibrium calculation is more appropriate. The composition of the hot pressurized gas changes very fast during the expansion, but chemical reactions require finite time intervals. The reaction time depends essentially on the molar composition of the nozzle gas, its temperature and pressure. The calculation of the reaction velocity is usually very difficult (and therefore often not done in the design of rocket nozzles), because some of the rate constants are not accurately known for gas under extraordinary conditions. The velocity of chemical reactions is not the only problem that still resists an accurate computer calculation: the actual flow in the nozzle involves friction, heat transfer to the walls, turbulent boundary layer effects, multi-phase working fluid with small droplets, energetic noise and oscillations and so on. The calculated performance is anyway not accurate.

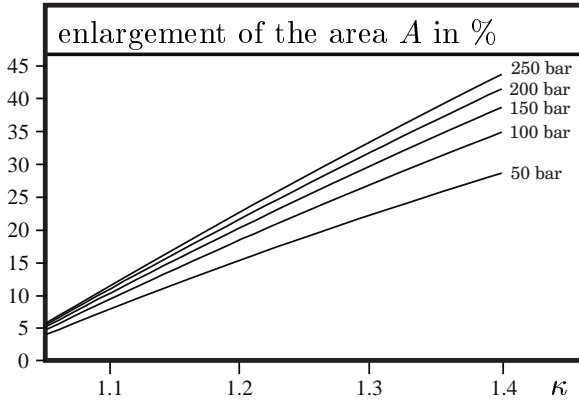
Duct area. Let us now use the matter conservation law (2-101) to calculate the duct area of the nozzle for the diabatic expansion as a function of the pressure ratio. We can readily transform this equation and insert the temperature T according to the equation (2-106) and the velocity c according to the equation (2-107). The intention is to compare the diabatic duct area with the adiabatic duct area (equation 2-80), therefore we neglect again the velocity in the combustion chamber ($c_{chamber} = 0$):

$$A = \frac{m \Re T}{p c} = \frac{m \Re}{p_{chamber}} \cdot \frac{\left(\frac{p}{p_{chamber}}\right)^{-1} \cdot T_{chamber} \cdot \left(\frac{p}{p_{chamber}}\right)^{\frac{\kappa-1}{\kappa(1-dT_0/dT)}}}{\sqrt{\frac{2\kappa}{\kappa-1} \Re T_{chamber} \left(1 - \frac{dT_0}{dT}\right) \left[1 - \left(\frac{p}{p_{chamber}}\right)^{\frac{\kappa-1}{\kappa(1-dT_0/dT)}}\right]}}$$

or:

$$A = \frac{m}{\rho_{chamber} \sqrt{2h_{chamber}}} \cdot \left[\left(1 - \frac{dT_0}{dT}\right) \left(\left(\frac{p}{p_{chamber}}\right)^{\frac{2-2\kappa dT_0/dT}{\kappa(1-dT_0/dT)}} - \left(\frac{p}{p_{chamber}}\right)^{\frac{\kappa-2\kappa dT_0/dT+1}{\kappa(1-dT_0/dT)}} \right) \right]^{-1/2} \quad (2-108)$$

When we compare the duct area for the diabatic nozzle expansion process ($dT_0/dT = -0.4$) with the duct area for the adiabatic process ($dT_0/dT = 0$) numerically, we find that the real nozzle must be bigger than the isentropic nozzle. The figure below shows the enlargement of the duct area as a function of κ for



several pressure ratios ($p_{chamber}/p_{end}$ from 50 to 250 bar). Even though the heat input has just a small influence on the gas velocity in the nozzle, it causes a comparatively high increase of the gas volume (a reduction of the density). In case the nozzle is designed using the isentropic expansion theory ignoring the influence of heat input, the mass flow rate m through the nozzle is actually smaller than expected: the heat input causes a “thermal blocking” of the flow through the isentropic nozzle.

Situation in the throat. In order to analyze the effect of heat input on the throat conditions in the rocket nozzle we make use of the Mach number, $Mach = c/\sqrt{\kappa \Re T}$. Remember that the Mach number equals 1 in the throat area of an isentropic nozzle. We differentiate the equation for the Mach number $c^2 = Mach^2 \cdot \kappa \Re T$ to find:

$$2 \frac{dc}{c} = \frac{d(Mach^2)}{Mach^2} + \frac{dT}{T} \quad (2-109)$$

The objective is now to express the nozzle area as a function of the Mach number; therefore we use again the conservation laws for matter, linear momentum and energy. The differential form of the energy conservation law (equation 2-105) considers already that heat energy is transferred to the gas flow by using the factor $dT_0/dT < 0$.

The differential forms of the conservation laws for the diabatic flow process are then:

$$\text{matter (2-101):} \quad \frac{dc}{c} + \frac{dA}{A} + \frac{dp}{p} = \frac{dT}{T} \quad (2-110)$$

$$\text{linear momentum (2-102):} \quad \frac{dc}{c} = \frac{-1}{\kappa Mach^2} \cdot \frac{dp}{p} \quad (2-111)$$

$$\text{energy (2-105):} \quad \frac{dp}{p} = \frac{\kappa}{\kappa - 1} \left(1 - \frac{dT_0}{dT}\right) \cdot \frac{dT}{T} \quad (2-112)$$

The expressions dc/c , dp/p and dT/T can readily be eliminated from the equations (system 2-109 to 2-112). The result is a relationship between dA/A and $d(Mach^2)$:

$$\frac{dA}{A} = \left[\frac{\kappa Mach^2 - 1}{Mach^2} - \frac{(1 + \kappa Mach^2) \cdot (\kappa - 1)}{(\kappa - 1) Mach^2 + 2(1 - dT_0/dT)} \right] \cdot \frac{d(Mach^2)}{2} \quad (2-113)$$

This relationship can be integrated (numerically and analytically) to find the duct area as a function of the Mach number, however, now we are mainly interested in what happens in the throat of the nozzle. For $Mach = 1$ the equation (2-113) simplifies to:

$$\frac{dA}{A} = \left[(\kappa - 1) \left(1 - \frac{\kappa + 1}{\kappa + 1 - 2dT_0/dT}\right) \right] \cdot \frac{d(Mach^2)}{2} \quad (2-114)$$

The expression in square brackets equals zero for an isentropic expansion process ($dT_0/dT = 0$), but the expression is greater than zero for a diabatic expansion process with heat input ($dT_0/dT < 0$, the specific heat ratio κ is always greater than 1). This indicates that dA/A is positive and that the sonic point $Mach = 1$ lies in the divergent part of the nozzle now. The condition for the throat area is $dA/A = 0$, and this shows that the Mach number is smaller than 1 in the throat of the nozzle:

$$Mach^2 = \frac{1 - dT_0/dT}{1 - \kappa dT_0/dT} \quad (2-115)$$

The Mach number $Mach = c/a$ as a substitute for the velocity c and the velocity of sound $a^2 = \kappa \Re T$ as a substitute for the temperature T play a key role in gas dynamics. The square of the velocity of sound a^2 (2-70), however, equals $dp/d\rho$ just in case of an isentropic expansion process. For an expansion with heat input (2-112) we have:

$$\frac{dp}{d\rho} = Mach_{throat}^2 \cdot a^2 = \frac{1 - dT_0/dT}{1 - \kappa dT_0/dT} \frac{c^2}{Mach^2} \quad (2-116)$$

To verify the relationship above we have to use the perfect gas law $p = \Re T \rho$ in its differential form $dp/p = dT/T + d\rho/\rho$. Finally we want to transform the differential equations (2-109), (2-111) and (2-112) in a way that they express the velocity c as a function of the Mach number (or the square of the Mach number). The result is:

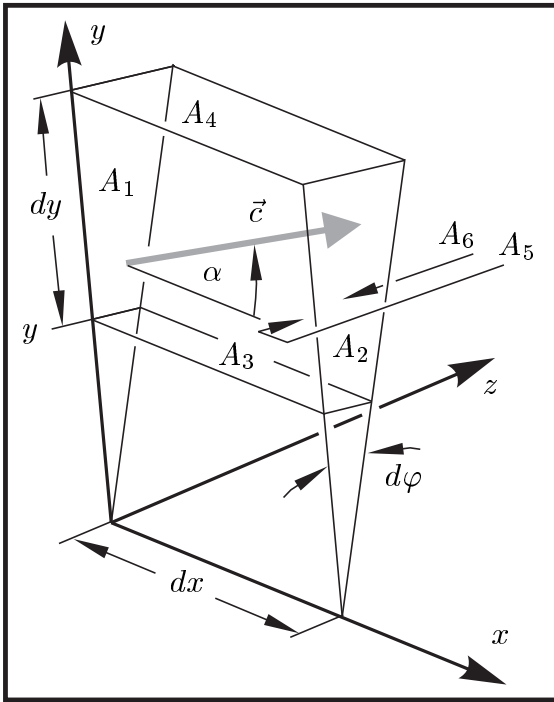
$$\frac{dc}{c} = \left[\frac{(1 - dT_0/dT)}{(\kappa - 1) Mach^2 + 2(1 - dT_0/dT)} \right] \frac{d(Mach^2)}{Mach^2} \quad (2-117)$$

We can readily verify that it is possible to integrate this relationship analytically, provided that the heat input factor dT_0/dT is constant (for example $dT_0/dT = -0.4$). We will use the equations (2-116) and (2-117) later to examine the shape of the nozzle.

2.4.5 Nozzle Geometry

The two-dimensional axisymmetric flow field. When we want to analyze the flow field of the hot gas inside a hypersonic rocket nozzle in order to design the shape of the nozzle we have to study the difficult subject of “computational fluid dynamics”. The computation of two (or three) dimensional flow fields is today usually performed by means of standard programs on powerful computers (such as finite element methods, for example). The “method of characteristics”, however, is another very elegant way to solve some problems of gas dynamics mainly with analytical transformations. Let us now examine the application of this important method to the computation of the axisymmetric flow through a hypersonic rocket nozzle.

Steady flow through a volume element. We remember that the one-dimensional streamline theory uses of the conservation laws for mass, linear momentum and energy in order to analyze the flow through the rocket nozzle. Therefore we start with the formulation of the conservation laws for mass and linear momentum, however, this time we consider just an infinitesimally small control volume inside the nozzle. The volume element is located in the flow field at the position x and y ; the x -coordinate is the centerline of the nozzle, the y -coordinate aims in radial direction. The gas velocity \vec{c} is now a vector with the axial velocity component u and the radial component v (the velocity in z -direction w vanishes in an axisymmetric flow field). The angle α is the angle between the velocity vector \vec{c} and the axis of the nozzle:



$$\vec{c} = \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} c \cdot \cos \alpha \\ c \cdot \sin \alpha \\ 0 \end{pmatrix} \quad (2 - 118)$$

$$c = \sqrt{u^2 + v^2}$$

Six surfaces confine the volume:

$$\begin{aligned} A_1 &= (y + 0.5 \cdot dy) d\varphi dy \\ A_2 &= A_1 \\ A_3 &= y d\varphi dx \\ A_4 &= (y + dy) d\varphi dx \\ A_5 &= dx dy \\ A_6 &= A_5 \end{aligned} \quad (2 - 119)$$

The control volume has not the shape of a cube but the shape of a prism: front and back face have the same area $A_1 = A_2$, the side face area A_5 equals A_6 , but the bottom face A_3 is smaller than the top face A_4 . Term $d\varphi$ is an arbitrarily small angle.

Conservation of matter. The law of conservation of matter (or mass) for a steady flow process states that the mass that enters the constant control volume at a certain instant is equal to the mass that leaves the volume at the same instant. Therefore:

$$\int_A \rho \vec{c} d\vec{A} = -[\rho u A]_{A_1} + [\rho u A]_{A_2} - [\rho v A]_{A_3} + [\rho v A]_{A_4} = 0 \quad (2-120)$$

with: $[\rho u]_{A_2} = [\rho u]_{A_1} + \frac{\partial \rho u}{\partial x} \cdot dx$; $[\rho v]_{A_4} = [\rho v]_{A_3} + \frac{\partial \rho v}{\partial y} \cdot dy$

The scalar vector product of velocity vector \vec{c} and surface vector $d\vec{A}$ (rectangular to the surface), multiplied by the density ρ , is exactly the mass flow rate through the control area. Note that no mass enters or leaves through the side faces A_5 and A_6 . When we evaluate the expressions in square brackets for the four control surfaces A_1 to A_4 (determined by the equations 2-119) we find a partial differential equation:

$$y \left(\frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} \right) + \rho v = 0 \quad (2-120)$$

or: $u \frac{\partial \rho}{\partial x} + v \frac{\partial \rho}{\partial y} + \rho \cdot \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = -\frac{\rho v}{y}$ (2-121)

Conservation of linear momentum. The conservation law for the linear momentum applied to the infinitesimally small control volume is a little more complicated than the matter conservation law: the equilibrium of forces is a vector equation that considers also pressure forces which act on the surfaces. We can write it down as:

$$\int_A (\rho \cdot \vec{c}) \cdot (\vec{c} \cdot d\vec{A}) = - \int_A p d\vec{A} = \quad (2-122)$$

$$\begin{aligned} & - \left[\begin{pmatrix} \rho u \\ \rho v \\ 0 \end{pmatrix} u A \right]_{A_1} + \left[\begin{pmatrix} \rho u \\ \rho v \\ 0 \end{pmatrix} u A \right]_{A_2} - \left[\begin{pmatrix} \rho u \\ \rho v \\ 0 \end{pmatrix} v A \right]_{A_3} + \left[\begin{pmatrix} \rho u \\ \rho v \\ 0 \end{pmatrix} v A \right]_{A_4} = \\ & + \left[\begin{pmatrix} p A \\ 0 \\ 0 \end{pmatrix} \right]_{A_1} - \left[\begin{pmatrix} p A \\ 0 \\ 0 \end{pmatrix} \right]_{A_2} + \left[\begin{pmatrix} 0 \\ p A \\ 0 \end{pmatrix} \right]_{A_3} - \left[\begin{pmatrix} 0 \\ p A \\ 0 \end{pmatrix} \right]_{A_4} + \left[\begin{pmatrix} 0 \\ 0 \\ p A \end{pmatrix} \right]_{A_5} + \left[\begin{pmatrix} 0 \\ p d\varphi A \\ -p A \end{pmatrix} \right]_{A_6} \end{aligned}$$

with: $[\rho u^2]_{A_2} = [\rho u^2]_{A_1} + \frac{\partial \rho u^2}{\partial x} \cdot dx$; $[p]_{A_2} = [p]_{A_1} + \frac{\partial p}{\partial x} \cdot dx$ and so on.

No matter crosses the side faces A_5 and A_6 , and therefore these surfaces do not contribute to the equilibrium of forces with a change in linear momentum $(\rho \cdot \vec{c}) \cdot (\vec{c} \cdot d\vec{A})$. The surface A_6 , however, is inclined against the y-axis (the angle is $d\varphi$) and thus it contributes to the equilibrium of forces with a pressure force $p d\varphi A$ in y-direction. Evaluation of the equation (2-122) brings us finally two partial differential equations:

$$\begin{aligned} y \left(\frac{\partial \rho u^2}{\partial x} + \frac{\partial \rho u v}{\partial y} \right) + \rho u v &= -y \frac{\partial p}{\partial x} \\ y \left(\frac{\partial \rho u v}{\partial x} + \frac{\partial \rho v^2}{\partial y} \right) + \rho v^2 &= -y \frac{\partial p}{\partial y} \end{aligned} \quad (2-123)$$

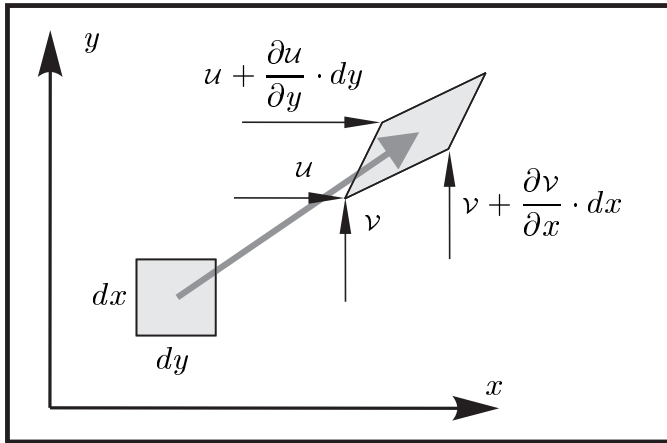
The equation system (2-123) is also called ‘‘Euler equation’’ for a steady flow field.

When we evaluate the equation (2-122) we find out that the contribution of the pressure force $pd\varphi A$ on the side face A_6 cancels out because the top surface A_4 is larger than the bottom surface A_3 . The two terms $\varrho u v$ and ϱv^2 come from the fact that the control volume grows with the y -coordinate: the Euler equations in the form of equation (2-123) for an axisymmetric flow field are not identical with corresponding equations for a plane flow field (a field with no change in the parameters along the z -axis). The matter conservation law (2-120) would also be different when we considered a plane flow field. We can eliminate the terms $\varrho u v$ and ϱv^2 from the equation system (2-123) when we insert the matter conservation law (2-120). Thus:

$$\begin{aligned} u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} &= -\frac{1}{\varrho} \cdot \frac{\partial p}{\partial x} \\ u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} &= -\frac{1}{\varrho} \cdot \frac{\partial p}{\partial y} \end{aligned} \quad (2-124)$$

Now these equations accept the same form as for the description of a plane flow field.

Conservation of the angular momentum. It is one of the essential statements of the “boundary layer theory”, developed by Ludwig Prandtl (1875-1953), that friction in a flow field plays a role only in thin layers next to the walls, and that the flow field distant from the walls is nearly free of friction. Therefore the expansion in a rocket nozzle is usually first calculated as a frictionless flow process, and then corrected considering the the boundary layer friction. We have used an infinitesimally small control volume to establish the conservation laws for matter and linear momentum, however, we have not considered that shearing forces might act on the surfaces: the result is the “Euler equation” for a frictionless flow field (incorporation of shearing forces would lead us to the “Navier-Stokes” equations). The important effect of the absence of shearing forces is that the flow field is “free of rotation”; it means that the forces cause the fluid elements to translate but not to rotate. Rotation free motion of a fluid element in a plane or axisymmetric flow field demands the following condition:



$$\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x} \quad (2-125)$$

The condition (2-125) for all elements in a rotation-free flow field as a consequence of the absence of shearing torques is actually the conservation law for the angular momentum (in a three-dimensional rotation-free flow field: $\vec{\omega} = 0.5 \nabla \times \vec{c} = 0$).

Transformation of the fundamental equations of gas dynamics. The objective is now to transform the conservation laws for matter (2-121), linear momentum (2-124) and angular momentum (2-125) in a way that we can eliminate the partial derivatives (∂x and ∂y) and replace them by total derivatives of other coordinates. The first step is to add the Euler equations (2-124) and incorporate the matter conservation law (2-121). The result is called “fundamental equation of gas dynamics”:

$$(u^2 - \frac{\partial p}{\partial \rho}) \frac{\partial u}{\partial x} + (v^2 - \frac{\partial p}{\partial \rho}) \frac{\partial v}{\partial y} + u v \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) = \frac{\partial p}{\partial \rho} \cdot \frac{v}{y} \quad (2-126)$$

The expression $\partial p / \partial \rho$ in the equation above equals the square of the velocity of sound a^2 just in case of an isentropic expansion; in case of an expansion with heat input, however, the term $\partial p / \partial \rho$ is more accurately determined by the equation (2-116). Note that the fundamental equation of gas dynamics for an axisymmetric flow in the form (2-126) is not identical with the corresponding equation for a plane flow field: the expression on the right hand side v/y would disappear in the description of a plane flow field. The next step is to replace the velocity components u and v by the flow velocity c and the flow angle α (compare equation 2-118). We can differentiate:

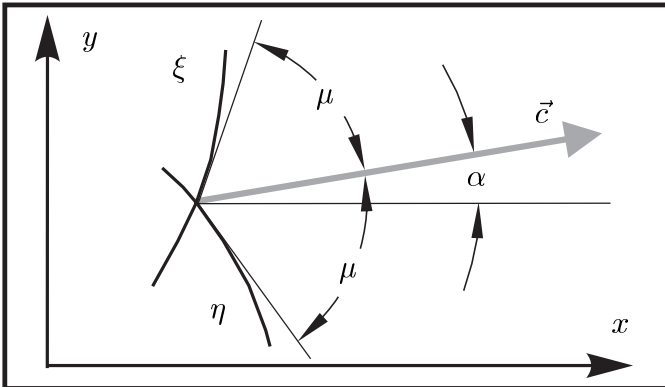
$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial c}{\partial x} \cos \alpha - c \sin \alpha \frac{\partial \alpha}{\partial x} ; & \frac{\partial u}{\partial y} &= \frac{\partial c}{\partial y} \cos \alpha - c \sin \alpha \frac{\partial \alpha}{\partial y} \\ \frac{\partial v}{\partial x} &= \frac{\partial c}{\partial x} \sin \alpha + c \cos \alpha \frac{\partial \alpha}{\partial x} ; & \frac{\partial v}{\partial y} &= \frac{\partial c}{\partial y} \sin \alpha + c \cos \alpha \frac{\partial \alpha}{\partial y} \end{aligned} \quad (2-127)$$

Then we insert the derivatives (2-127) into the equations (2-126) and (2-125) to find:

$$c \cdot \left[\sin \alpha \frac{\partial \alpha}{\partial x} - \cos \alpha \frac{\partial \alpha}{\partial y} \right] + \frac{c^2 - \partial p / \partial \rho}{\partial p / \partial \rho} \cdot \left[\cos \alpha \frac{\partial c}{\partial x} + \sin \alpha \frac{\partial c}{\partial y} \right] = \frac{c \sin \alpha}{y} \quad (2-128)$$

$$c \cdot \left[\cos \alpha \frac{\partial \alpha}{\partial x} + \sin \alpha \frac{\partial \alpha}{\partial y} \right] + \left[\sin \alpha \frac{\partial c}{\partial x} - \cos \alpha \frac{\partial c}{\partial y} \right] = 0 \quad (2-129)$$

The problem is to find directions in the x - y -plane where the partial derivatives $\partial / \partial x$ and $\partial / \partial y$ can be composed to total derivatives. The new coordinates are curves in the x - y -plane called η (right running characteristic) and ξ (left running characteristic).



The directions of these curves at a location x, y are determined by the angles $-\mu, +\mu$ with respect to the local flow direction α . Thus, at the location (x, y) the right running characteristic η has the inclination angle $\alpha - \mu$ with respect to the centerline of the hypersonic rocket nozzle; the left running characteristic ξ has the inclination angle $\alpha + \mu$.

The total derivatives of α and c with respect to the new coordinates η and ξ are then:

$$\begin{aligned}\frac{d\alpha}{d\eta} &= \frac{\partial\alpha}{\partial x} \cdot \frac{dx}{d\eta} + \frac{\partial\alpha}{\partial y} \cdot \frac{dy}{d\eta} &= \frac{\partial\alpha}{\partial x} \cos(\alpha + \mu) + \frac{\partial\alpha}{\partial y} \sin(\alpha + \mu) \\ \frac{dc}{d\eta} &= \frac{\partial c}{\partial x} \cdot \frac{dx}{d\eta} + \frac{\partial c}{\partial y} \cdot \frac{dy}{d\eta} &= \frac{\partial c}{\partial x} \cos(\alpha + \mu) + \frac{\partial c}{\partial y} \sin(\alpha + \mu) \\ \frac{d\alpha}{d\xi} &= \frac{\partial\alpha}{\partial x} \cdot \frac{dx}{d\xi} + \frac{\partial\alpha}{\partial y} \cdot \frac{dy}{d\xi} &= \frac{\partial\alpha}{\partial x} \cos(\alpha - \mu) + \frac{\partial\alpha}{\partial y} \sin(\alpha - \mu) \\ \frac{dc}{d\xi} &= \frac{\partial c}{\partial x} \cdot \frac{dx}{d\xi} + \frac{\partial c}{\partial y} \cdot \frac{dy}{d\xi} &= \frac{\partial c}{\partial x} \cos(\alpha - \mu) + \frac{\partial c}{\partial y} \sin(\alpha - \mu)\end{aligned}\quad (2-130)$$

Now we have to select the angle μ appropriately with the intention to combine the equation (2-128) with the equation (2-129). The composition of the partial derivatives to total derivatives is just possible when the angle μ satisfies the following condition:

$$\cot^2 \mu = \frac{c^2 - \partial p / \partial \varrho}{\partial p / \partial \varrho} \quad \left(= \frac{1 - \kappa dT_0/dT}{1 - dT_0/dT} Mach^2 - 1 \right) \quad (2-131)$$

In case of an isentropic expansion we have $\partial p / \partial \varrho = a^2$ and $dT_0/dT = 0$, and this means that in an isentropic flow field the angle μ is exactly the Mach angle. However, equation (2-116) represents the term $\partial p / \partial \varrho$ more accurately for a diabatic expansion.

We multiply equation (2-128) once by the factor $+\sin \mu$, the second time by $-\sin \mu$ and add the result every time to the equation (2-129), multiplied by the factor $\cos \mu$:

$$\begin{aligned}c \cdot \left[\cos(\alpha - \mu) \frac{\partial\alpha}{\partial x} + \sin(\alpha - \mu) \frac{\partial\alpha}{\partial y} \right] + \cot\mu \left[\cos(\alpha - \mu) \frac{\partial c}{\partial x} + \sin(\alpha - \mu) \frac{\partial c}{\partial y} \right] &= + \frac{c \sin\alpha \sin\mu}{y} \\ c \cdot \left[\cos(\alpha + \mu) \frac{\partial\alpha}{\partial x} + \sin(\alpha + \mu) \frac{\partial\alpha}{\partial y} \right] - \cot\mu \left[\cos(\alpha + \mu) \frac{\partial c}{\partial x} + \sin(\alpha + \mu) \frac{\partial c}{\partial y} \right] &= - \frac{c \sin\alpha \sin\mu}{y}\end{aligned}$$

$$\text{or with (2-130):} \quad \frac{d\alpha}{d\xi} + \frac{\cot \mu}{c} \frac{dc}{d\xi} = + \frac{\sin \alpha \sin \mu}{y} = + \frac{\sin \alpha \sin \mu}{\sin(\alpha - \mu)} \frac{dy}{y d\xi} \quad (2-132)$$

$$\frac{d\alpha}{d\eta} - \frac{\cot \mu}{c} \frac{dc}{d\eta} = - \frac{\sin \alpha \sin \mu}{y} = - \frac{\sin \alpha \sin \mu}{\sin(\alpha + \mu)} \frac{dy}{y d\eta} \quad (2-133)$$

The modification is now nearly complete, the only transformation that is still missing is the analytical integration of the expression $\cot \mu \cdot dc/c$ in the two equations above. We can integrate this expression when we consider that in the relationship (2-131) the angle μ is exclusively a function of the Mach number (assuming that the heat input factor dT_0/dT is constant, for example $dT_0/dT = -0.4$). The equation (2-117) relates also the expression dc/c to the Mach number; and the integration yields:

$$\nu(\mu) = \int \cot\mu \frac{dc}{c} = \mu - \frac{\pi}{2} + \sqrt{\frac{\kappa+1-2\kappa dT_0/dT}{\kappa-1}} \arctan \left(\sqrt{\frac{\kappa-1}{\kappa+1-2\kappa dT_0/dT}} \cot\mu \right) \quad (2-134)$$

Integration brings us the angle $\nu(\mu)$, an angle that is called "Prandtl-Meyer" angle.

Intersection of the characteristics. The analytical transformation of the conservation laws for matter, linear momentum and angular momentum onto the characteristics leads us to the differential equation system (2-132) and (2-133); the integration of the energy conservation law leads us to the Prandtl-Meyer angle (equation 2-134). The analysis is a little complicated; let us therefore summarize the final result:

$$\frac{d(\alpha + \nu(\mu))}{d\xi} = + \frac{\sin \alpha \sin \mu}{\sin(\alpha - \mu)} \frac{d \ln(y)}{d\xi} \quad \text{along} \quad \frac{dy}{dx} = \tan(\alpha - \mu) \quad (2-135)$$

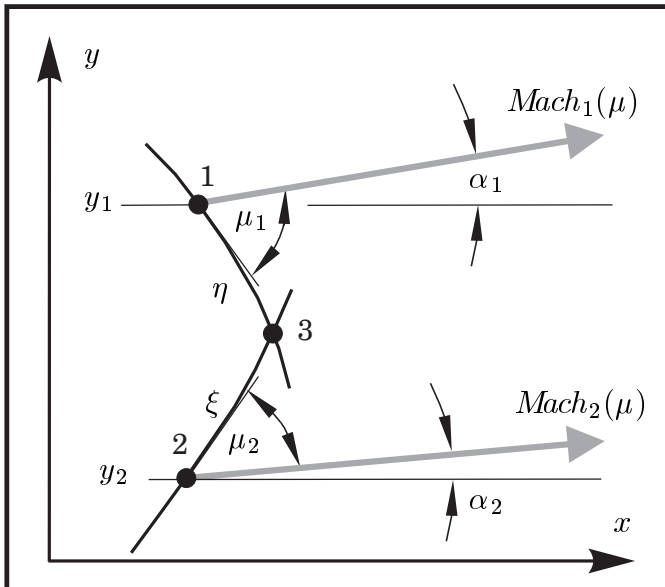
$$\frac{d(\alpha - \nu(\mu))}{d\eta} = - \frac{\sin \alpha \sin \mu}{\sin(\alpha + \mu)} \frac{d \ln(y)}{d\eta} \quad \text{along} \quad \frac{dy}{dx} = \tan(\alpha + \mu) \quad (2-136)$$

These equations describe the relationships between the flow direction (angle α) and the flow velocity (angle μ) along the right running characteristic η (2-135) and the left running characteristic ξ (2-136). When the values for α and μ are well-known at two neighboring locations inside the flow field (for example at the upper location 1 and the lower location 2), the equations (2-135) and (2-136) can be used to compute the corresponding values at another location (for example at a location 3 downstream).

The procedure works as follows: the angles α_1 and μ_1 determine at the location 1 the inclination $\tan(\alpha_1 - \mu_1)$ of the right running characteristic η ; and α_2 and μ_2 determine at the location 2 the direction $\tan(\alpha_2 + \mu_2)$ of the left running characteristic ξ . We assume that the characteristics are nearly straight lines and bring them to intersection at point 3. Then we replace in an approximation differentials by differences:

$$\Delta(\alpha + \nu(\mu)) = [\alpha + \nu(\mu)]_3 - [\alpha + \nu(\mu)]_1 = + \left[\frac{\sin \alpha \sin \mu}{\sin(\alpha - \mu)} \right]_1 (\ln(y_3) - \ln(y_1)) \quad (2-137)$$

$$\Delta(\alpha - \nu(\mu)) = [\alpha - \nu(\mu)]_3 - [\alpha - \nu(\mu)]_2 = - \left[\frac{\sin \alpha \sin \mu}{\sin(\alpha + \mu)} \right]_2 (\ln(y_3) - \ln(y_2)) \quad (2-138)$$

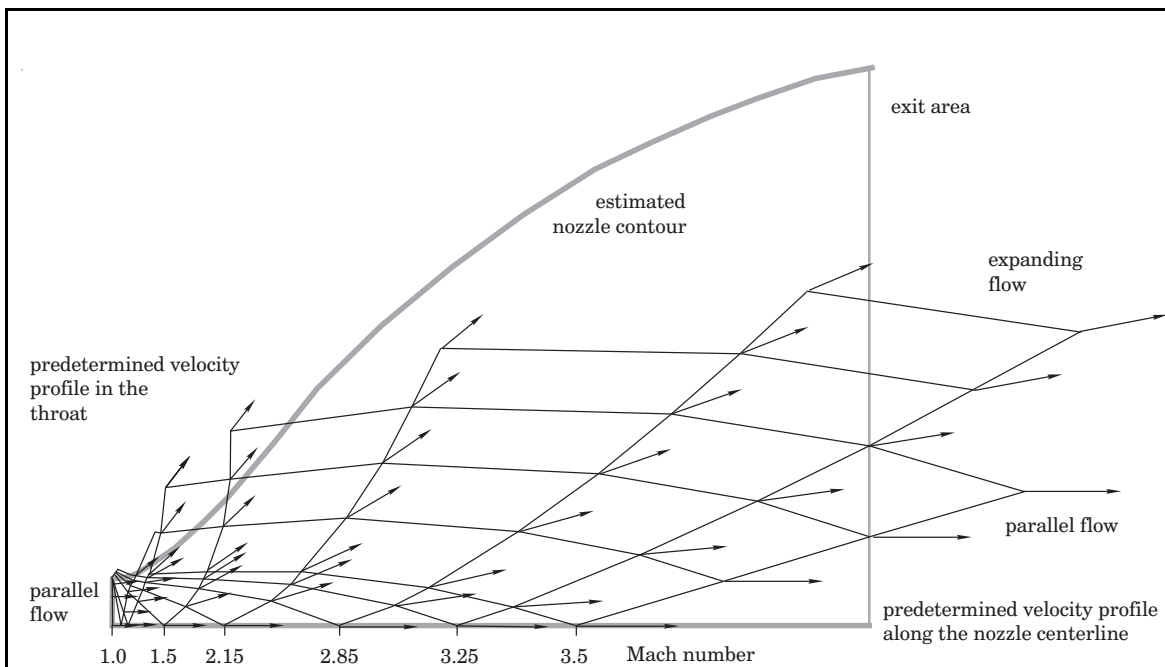


The equations for the differences above serve as a linear system for the calculation of the flow angle α and the Prandtl-Meyer angle $\nu(\mu)$ at the location 3 in the flow field; and then we find the angle μ at point 3 by solving the nonlinear equation (2-134). With the knowledge of the angles α and μ the situation is entirely determined at point 3; because the Mach number is a direct function of the angle μ (it follows from equation 2-131).

Construction of the flow field. Now we can use the method of characteristics to calculate the complete flow pattern, provided that we know the situation (that means flow angle and Mach number) along one of the borderlines of a two-dimensional field.

For example, we can compute the nozzle contour which belongs to a certain Mach number profile predetermined along the centerline of the nozzle; or we can compute the steady hypersonic flow field in a certain nozzle with divergent contour for a predetermined Mach number in the throat area (or immediately behind the throat area).

In every step we bring characteristics of two nearby points to intersection, and calculate flow angle and Mach number on the intersection point. Then we use the situation on this new point to proceed in the calculation with bringing the characteristics of the new point to intersection with the characteristics of another nearby point in the flow field. The method of characteristics transforms the problem of finding the solution to a system of partial differential equations into the problem of solving a system of many (comparatively simple) equations. However, the characteristics are not exactly straight lines, and for accuracy their grid should be closely meshed. The final calculation of the mesh of characteristics must be performed on a computer.



Example 2-4: The method of characteristics can be used to calculate the flow pattern inside a hypersonic rocket nozzle. The example shows the isentropic expansion in the divergent part of an axisymmetric rocket nozzle: the expansion begins with a parallel flow in the throat of the nozzle (Mach 1) and ends with a parallel flow at the exit of the nozzle (Mach 3.5). The flow field is determined for every point of the mesh of characteristics by the flow direction and the flow velocity (indicated by the direction and length of the arrows, the Mach angle between flow direction and characteristics).

2.5. Liquid Rocket Stages

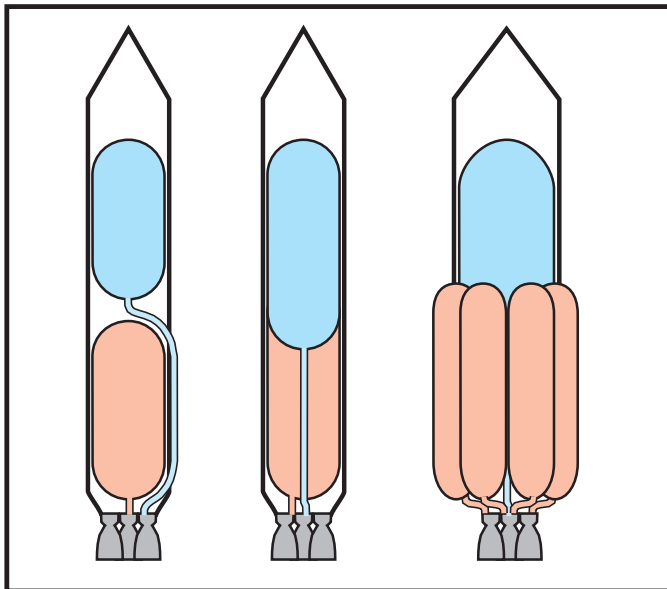
Propulsive maneuvers in spaceflight are always performed by means of rocket motors. We talk about liquid rocket engines when these motors operate with liquid fuel and liquid oxidizer. Different concepts are in use to feed liquids into thrust chambers. There the propellant is burned, expanded and exhausted with high relative velocity.

2.5.1 Tanks and Structures

Liquid propellents. A large variety of different liquids is theoretically usable as fuel or oxidizer in bipropellant (or tripropellant) systems; however, when we take a look at existing launch vehicles or spacecraft we can see that there are actually only three bipropellant combinations which are really in common use today. These three liquid propellents are: the high energetic combination LOX/LH₂; the lower energetic (but less voluminous) combination LOX/kerosene, and finally the storable, hypergolic combination nitrogen tetroxide/hydrazine (N₂O₄/N₂H₄, or derivatives).

Propellant accommodation. Regarding weight and structure firmness, spherical tanks are most efficient to store pressurized liquids. However, often ball shaped tanks do not meet the optimal launcher dimensions (slender-bodies), and therefore large propellant quantities are usually stored in cylindrical tanks with semi-spherical ends. Fuel and oxidizer can either be stored in two separate tanks, (usually arranged in a tandem configuration with external piping) or in a single tank with a partition that separates two compartments (usually a semi-spherical bulkhead with internal piping). Sometimes we find also stages which use several tanks in a parallel configuration.

During the flight the stresses and loads are not the same for all locations on the surface of a tank, and therefore the wall thickness of the tank structure material



(stainless steel or aluminum alloy) is usually not constant. The mass of a spherical tank is essentially a function of the tank material, the inner pressure and the storage volume. When we compare tanks of different size for a certain type of liquid propellant we might expect that tank mass is proportional to the tankage (propellant mass); however, in fact larger tanks have a lower specific weight: in practice it is more appropriate to assume that the tank weight is approximately proportional to the surface of the filling amount.

Space launchers store propellant in pressurized tanks, for structural firmness reasons and in order to avoid pump cavitation (typical pressures range from 0.7 to 3.4 bar). The pressure gas may not chemically react with the liquid, or condense or dissolve in the liquid. For example, the helium pressure gas can be taken from a ball-shaped high-pressure container near the engines, or combustion gas can be taken from the engines or from another gas generator. Tanks that store cryogenic liquids (liquid hydrogen and liquid oxygen) require heat insulation in order to minimize the propellant vaporization in the tanks and the formation of unwanted ice layers on the outer surface of the tank (ice layers do not only increase the lift-off mass of the launcher, pieces that break off during the initial flight phase can hit and damage the vehicle). Propellant lost by vaporization prior to launch has to be refilled shortly before lift-off. Filling cryogenic propellents into tanks is not easy: liquid or even gaseous pollutant can solidify in liquid hydrogen, and the vapor of liquid hydrogen is quite explosive.

Motion of the vehicle can cause sloshing (and vortexing) of the liquids in the tanks, particularly when the tank is partly empty. The sloshing is dangerous because it can interact in resonance with the attitude stabilization system of the vehicle and because it can uncover the tank outlet and allow gas bubbles to enter the discharge line. The motion of the liquid in the tanks is usually diminished by baffles (metal sheets).

It is unfortunately not possible to use the entire propellant charge of the tank for propulsion, a quite large amount (typically 2%) remains after burnout in the tanks and in the lines (wetting the tank walls or trapped in grooves, valves, instruments and so on). This residual propellant deteriorates the performance of the vehicle, and thus an effort is made in the construction of tanks to maximize their expulsion efficiency. It can be a problem to ensure that no gas enters the discharge lines of the tanks. Especially for the application in zero-gravity environment of space, several different devices have been developed in order to guarantee bubble free feeding of liquid propellents to the engine, including bladders, piston mechanisms, bellows and filters (metal screens with many small openings which attract the liquid by surface tension).

Lines and valves. The outlets at the bottom of the tanks are connected with the engines via pipes which convey the propellant components to the turbo-pumps. These pipes are usually made of stainless steel, heat insulation is required in case the pipes conduct cryogenic liquids. The flow through the lines is initiated and controlled by high-precision valves, such as drop ball valves (for switch on and off) or burst diaphragms (for one single start). Many liquid rocket engines use filters in their lines in order to prevent that dirt or other particles (debris from burst diaphragms) cause malfunction in the valves or regulators. Lines and valves in liquid rocket engines are intensively tested for utmost reliability, because a single malfunction of a valve or a leak in one line will usually have the failure of the whole mission as the consequence.

Swivel mounted thrust chambers (“nozzle gimbaling” for thrust vector control) require that the pipe lines are flexible or that pipes are joined by flexible fittings. For example, flexibility can be provided by two right-angle turns in the lines where special (pivotable) connections allow the pipes to deflect when the engine is swiveled.

2.5.2 Thrust Chambers

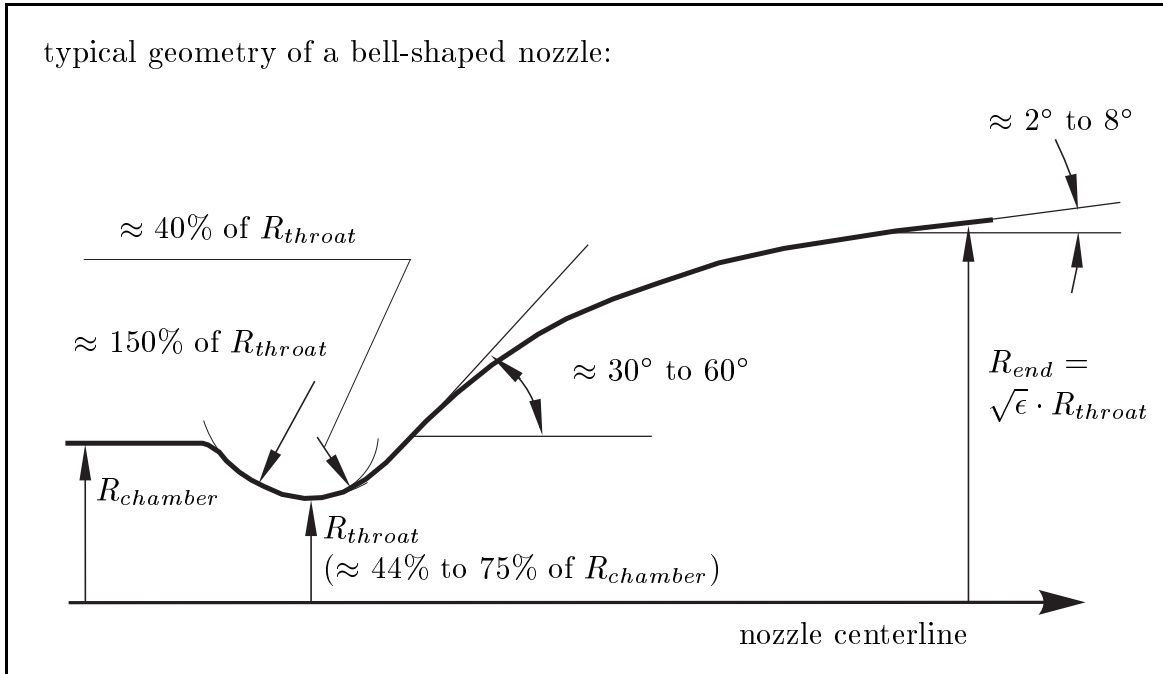
Injectors. The injector is usually a disk with many small orifices placed at the entry wall of the cylindrical combustion chamber. Its function is to introduce the propellant into the combustion chamber in the form of thin sprays which contain an optimal mixture of fuel and oxidizer. Propellant provided from the feed system in liquid form must vaporize during the injection; however, the propellant can also be provided in the form of gas when it comes from a precombustion chamber or when it has been used for chamber wall cooling (hydrogen is usually injected as a gas, and oxygen is vaporized during the injection by heat from the hydrogen). Several injector properties influence the efficiency and stability of the combustion: the number of holes in the injector face, their distribution and size, the injection velocity and the inclination of the jets of fuel and oxidizer and so on. The engineering of injectors for rocket motors is rather based on empirical approaches than on analytical understanding. The injector pressure drop for liquid propellents ranges from 5 to 10 bar, typically.

In case of LOX/LH2 engines or LOX/kerosene engines the electrical spark ignition system is usually integrated in the injector system. Engines that burn the hypergolic propellant combination N_2O_4/N_2H_4 ignite spontaneously without special igniter.

The chamber size. A cylindrical container serves as the combustion chamber. For weight reasons the combustion chamber should be short and as small as possible, however, its volume must be large enough to ensure a good combustion efficiency. The length of the chamber is compared with the volume $v_{chamber}$ for a particular propellant combination using the characteristic chamber length $L^* = v_{chamber}/A_{throat}$; typical values for L^* range between 0.8 and 1.8 meters for the propellant combinations LOX/LH2 or N_2O_4/N_2H_4 , or between 1.7 and 2.5 meters for LOX/kerosene.

The nozzle contour. The nozzle begins with a converging section at the exit wall of the combustion chamber (opposite to the injector face). The diameter of the well-rounded throat is smaller than the diameter of the combustion chamber (typical values for the throat diameter range between 55% and 80%). A “bell-shaped” diverging section of the nozzle is the best compromise: it is possible to permit large divergence angles right behind the nozzle throat (30° up to 60°), because the pressure is still high and the gas is able to expand rapidly; but the divergence angle at the end of the nozzle is small in order to minimize divergence losses (usually 2° to 8°).

The contour of a bell-shaped nozzle can be approximated by a circular throat section patched together with a parabolic section. Even though the flow field in the exhaust area of a rocket nozzle with bell-shaped contour is not exactly uniform and parallel, the kinetic energy losses are actually small (nozzles with the optimum contour would be too long). The expansion area ratio $\epsilon = A_{end}/A_{throat}$ determines the size of the nozzle, ground level adapted nozzles are shorter than vacuum adapted nozzles (for example the retired engines of Saturn-5: $\epsilon = 16$ for F1 and $\epsilon = 27.5$ for J2; or $\epsilon = 78$ for America’s SSME; and $\epsilon = 100$ for Japan’s upper stage engine LE-5A).



Cooling of the thrust chamber wall. The chemical reactions inside the thrust chamber of a large liquid rocket engine transform an enormous amount of energy per unit of time. Just a small part of this energy (between 0.5% and 5%) is transferred to the chamber walls; but without cooling the wall material would melt immediately. Thrust chambers of large liquid engines are regeneratively cooled: before the propellant enters the combustion chamber it is used to absorb the wall heating (the fuel component is usually used for cooling). The combustion temperature is consequently hotter and the heat energy that had entered the chamber walls is not lost for propulsion. The liquid that comes from the turbopump feeding system flows through a tubular cooling jacket (an assembly of many small pipes brazed together) before it proceeds to the injector. The diameter of the cooling tubes is not uniform along the nozzle surface: the highest heating rates occur at or near the throat and thus it is advisable to let the coolant flow faster at this location. The inlet manifold of the cooling jacket is often placed at the throat or at the combustion chamber to provide that fresh coolant absorbs the intensive heating. Then the coolant flows through small individual circumferential pipes, aggregates in the alternate tubes and proceeds directly to the injector. The efficiency of regenerative cooling of thrust chambers can be supported by other cooling methods, such as transpiration (film) cooling of highly stressed surfaces or dump cooling of the nozzle exit area (occasionally also radiation cooling). The pressure loss in the cooling jacket depends on the diameter of the tubes and ranges between 5% and 25% of the chamber pressure.

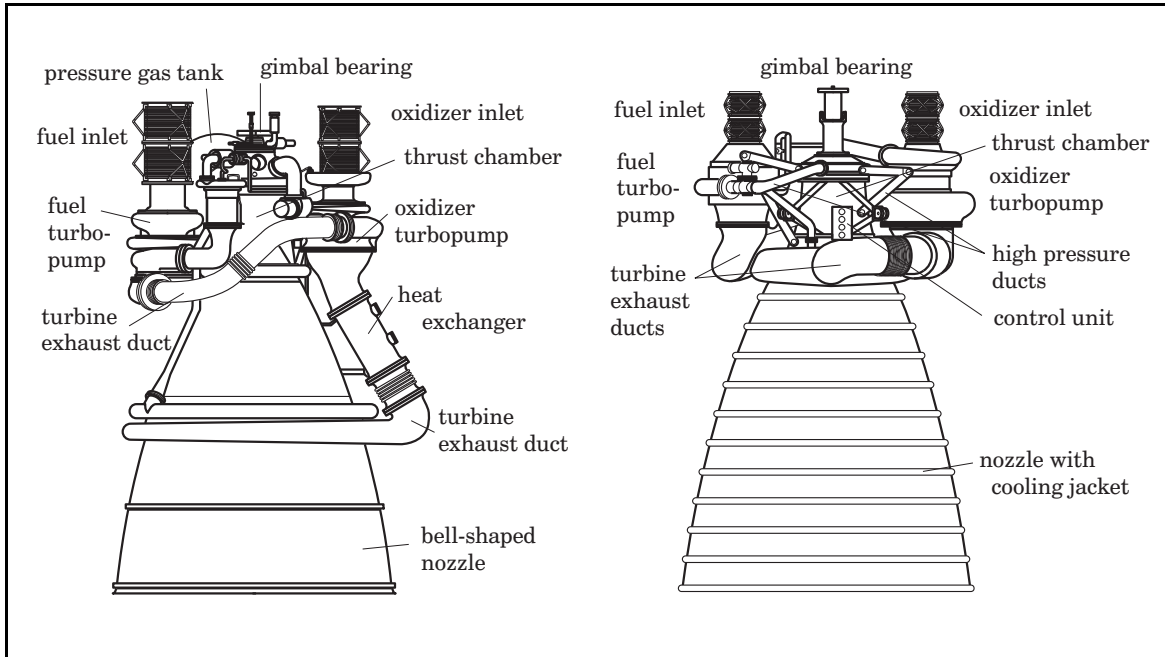
Hydrogen is an excellent regenerative coolant for rocket engines: the liquid increases its enthalpy considerably when it vaporizes under high pressure in the cooling jacket. The energy picked up by thrust chamber wall cooling can be used to drive turbopumps of the feed system (expander cycle engines), avoiding the necessity of a gas generator.

2.5.3 Turbopump Feed Systems

Pumps and turbines. The turbopump feed system is another main component of a powerful liquid rocket engine. The problem is to feed liquids with large flow rates and high pressures into the thrust chamber: these liquids are sometimes chemically aggressive or extremely cold; the feed system must be a lightweight construction with a high reliability; there is nearly no time for warming-up or checking-out the hardware; flow instabilities are a permanent threat due to the nearby presence of the thrust chamber as a strong source for vibrations, dynamic loads during the flight can also cause instabilities, bubbles in the liquids can cause pump cavitation, propellant leakage can form explosive gas and so on. The only advantage in comparison with other engines is that the operation time (or lifespan) is very short, usually below 10 minutes.

A key parameter in the selection of a turbopump feed system is the propellant density: pumps are basically constant volume flow machines, and the shaft power that is required for a certain pump head ($\Delta p/\rho$) is inversely proportional to the fluid density. Larger pumps are necessary for feeding rocket engines with low density propellant. When the engine operates with a propellant combination where oxidizer and fuel have approximately the same density (such as LOX/kerosene or nitrogen tetroxide/hydrazine for example), then the pumps for the two different liquids have approximately the same optimal shaft speed and it is possible to drive these pumps with the same turbine (without reduction gear). However, since the optimal turbine speed is usually higher than the pump speed, it will anyway save weight and turbine power to interpose a reduction gear between the turbine shaft and the pump shaft. When the fuel has a much lower density than the oxidizer (liquid hydrogen / liquid oxygen), the fuel pump needs more speed (and more than one stage) than the oxidizer pump, and it can be an advantage to use separate pump assemblies for the different liquids.

Pumps in liquid rocket engines are usually centrifugal pumps (occasionally also axial flow pumps), with carefully designed shaft seals to prevent leakage (explosive or corrosive propellant) and with carefully designed lubrication systems (in case of cryogenic propellant the liquid itself is the lubricant, oil would freeze at low temperatures). The pump power for a certain propellant is determined by the flow rate multiplied by the pressure raise; and in order to get a pump with a small size it is advisable to select the pump speed as high as possible (the highest pump speed is limited by the beginning of impeller blade cavitation). The flow rate is proportional to the speed (number of revolutions per minute), the pressure raise or pump head is proportional to the square of the pump speed and the consumed power is proportional to the cube of the pump speed. For example, the pumps in the large liquid rocket engine F1 (LOX/kerosene) of Saturn-5's first stage operated with a common shaft drive at a speed of 5600 rpm and consumed a power of 56000 hp. The more modern engine of the Space Shuttle SSME, however, uses for fuel and oxidizer different pump assemblies (each with precombustion and pre-pump): the 3-stage high-pressure hydrogen pump consumes 80000 hp shaft power, operates with a speed of 37500 rpm, and feeds hydrogen with a pressure of 476 bar into the cooling jacket of the thrust chamber.



The power to drive the pumps of large liquid rocket engines comes from axial gas turbines, they consist usually of a set of up to three (rotating and fixed) blade rings. A turbine transforms the enthalpy of a gas into shaft power by expanding the gas. The shaft power of a turbine is therefore proportional to its mass flow rate multiplied by the decrease of the enthalpy of the working gas, where the decrease of the enthalpy of the working gas can be calculated from the pressure drop and the temperature drop assuming approximately an isentropic expansion (where $p \cdot v^\kappa = \text{constant}$).

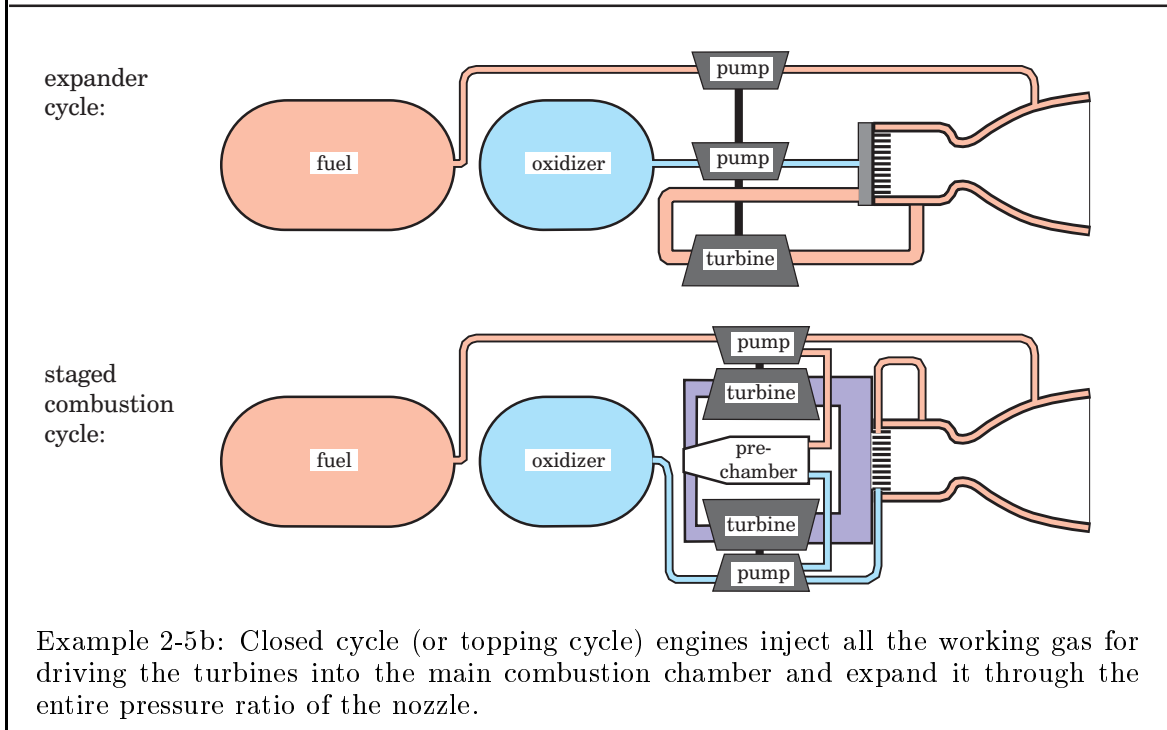
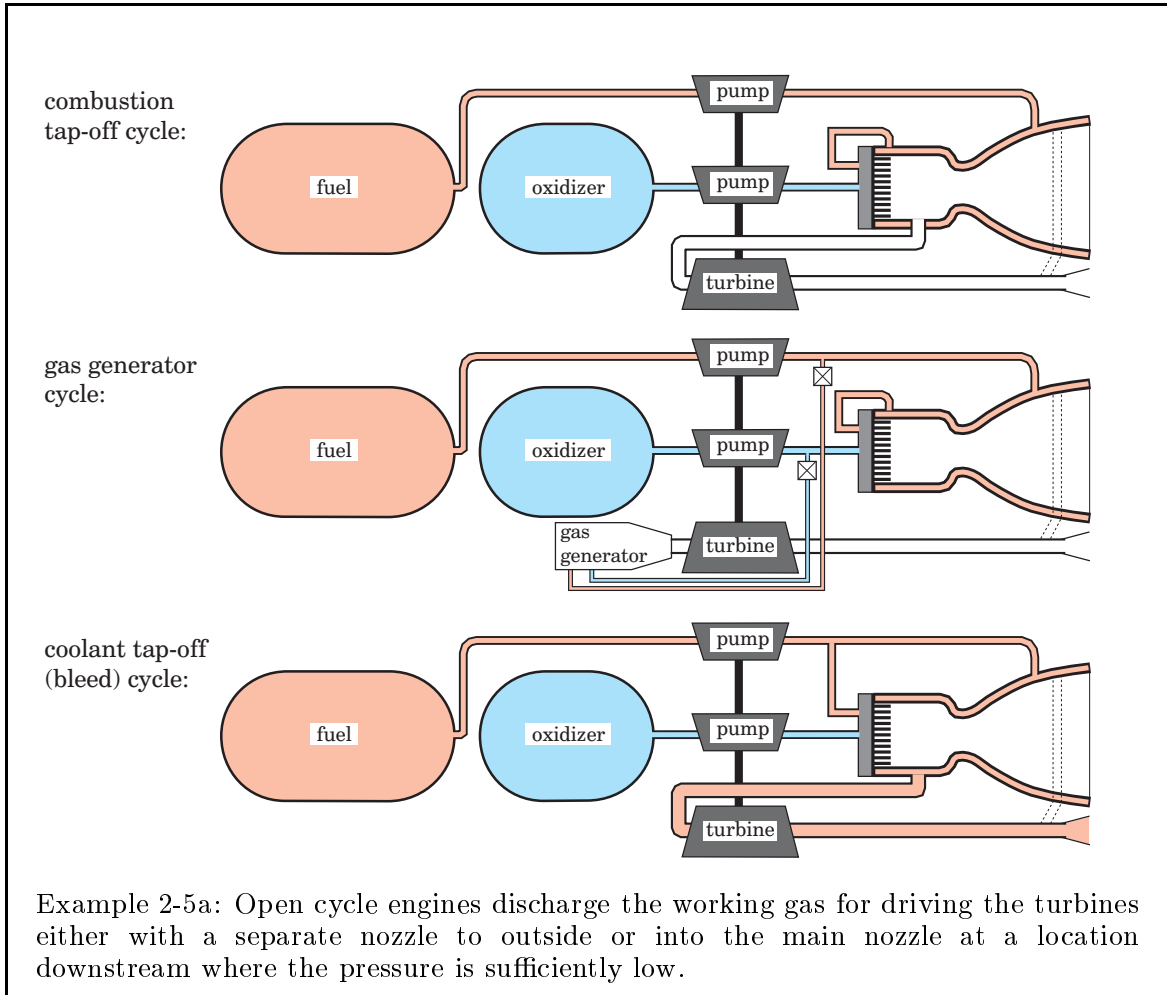
In liquid rocket engines there are various ways where the working gas comes from and where it goes to after the expansion: the working gas can be bleed gas from the combustion chamber, or it can be supplied by a separate gas generator, or it can be gas from thrust chamber wall cooling, or it can come from a pre-combustion chamber. After the expansion the gas can be discharged into the thrust chamber at a location where the pressure is considerably lower, or it can be discharged to outside by a separate exhaust nozzle, or, provided that its pressure is still considerably higher than the chamber pressure, the gas can be injected into the main combustion chamber. Different propellant combinations and different working principles of the feeding system require quite different turbine constructions: a high turbine inlet temperature is desirable for a low flow rate (single crystal material and special alloys allow turbine inlet temperatures between 1400 K and 1600 K). However, the necessity of blade cooling can be avoided when the working gas comes from a gas generator with a temperature below the melting point of the blade material. A high turbine speed is desirable for a good efficiency, but sometimes the construction of the engine demands that the pumps are connected directly to the turbine, and efficiency and the shaft speed are a compromise (turbines operate with an efficiency between 50% and 80%). The turbine power of large liquid rocket engines with high chamber pressure can reach 130000 hp and more, and this is usually more than the power of aircraft turbo-engines.

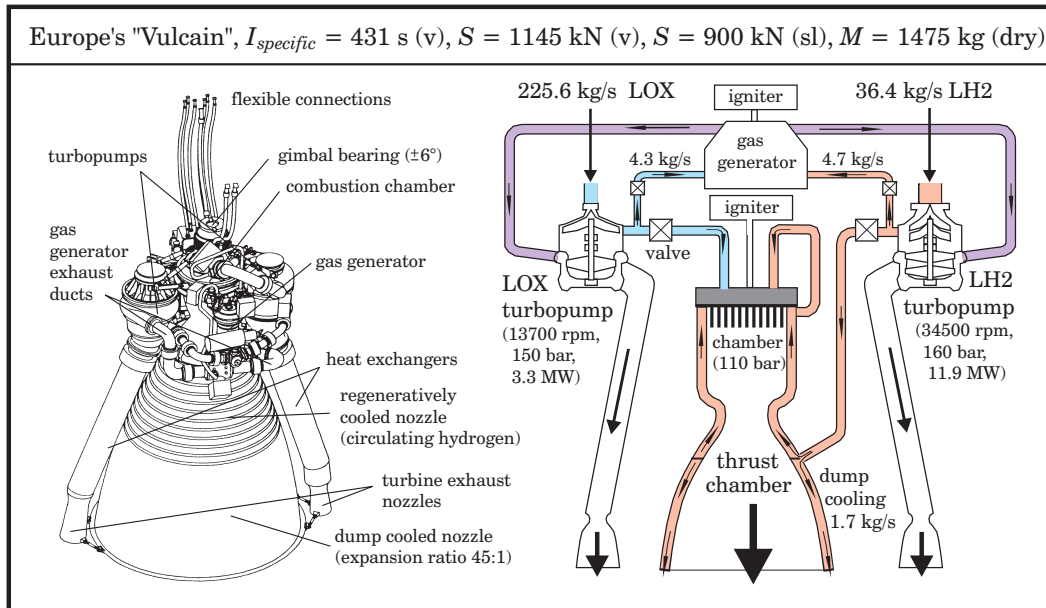
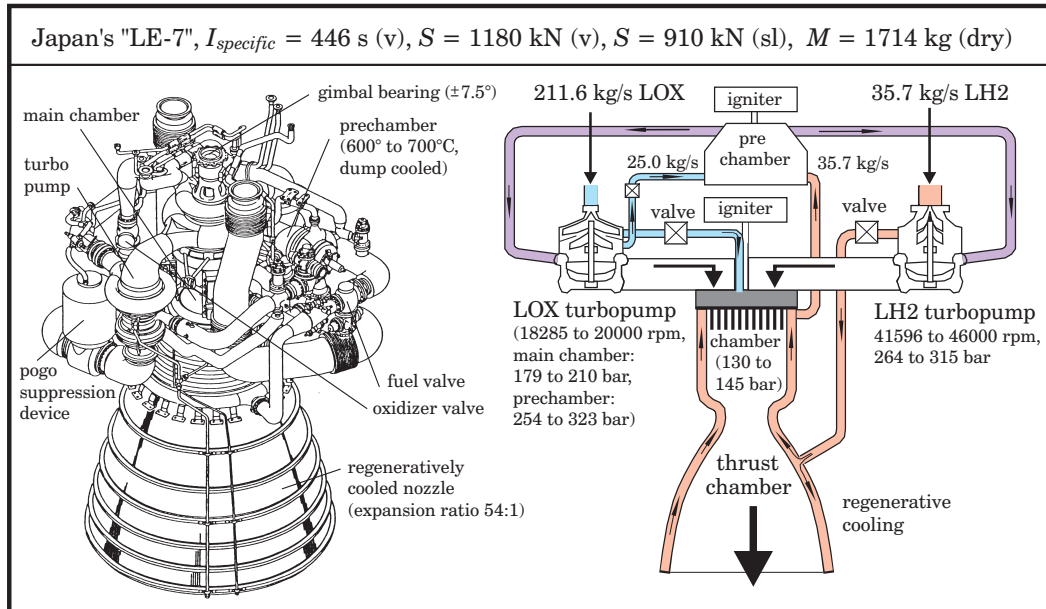
Gas generators. Often the pressure gas that is used for driving the turbines of a large liquid rocket engine is generated by a gas generator. Gas generators are similar to combustion chambers regarding their function, but the combustion is performed with the intention to generate power (and not thrust), and thus the propellant is usually burned with a different mixture at a lower temperature (for example 1500 K only). The lower temperatures avoid the necessity of cooling for turbine and gas generator.

Propellant feeding cycles. There are basically two different classes of turbopump feed system for large liquid rocket engines: so-called open cycles and closed cycles. Open cycle engines discharge the working gas of the turbopumps without expanding it in the main thrust chamber, usually in a separate nozzle and sometimes in the main nozzle at a location near the end where the pressure is considerably lower than the chamber pressure. Closed cycle engines (also called topping cycle engines) expand the working gas of the turbines in the main thrust chamber; they are more expensive but offer a slightly better performance because the energy content of the propellant is more efficiently used. Several sub-classes can be identified within these two main classes; also mixed forms are thinkable but have not been realized yet (for example rocket engines with a closed fuel cycle and an open oxidizer cycle). In common use are open gas generator cycle engines for all propellants and closed cycle engines with staged combustion for LOX/LH₂ propellant (occasionally LOX/kerosene).

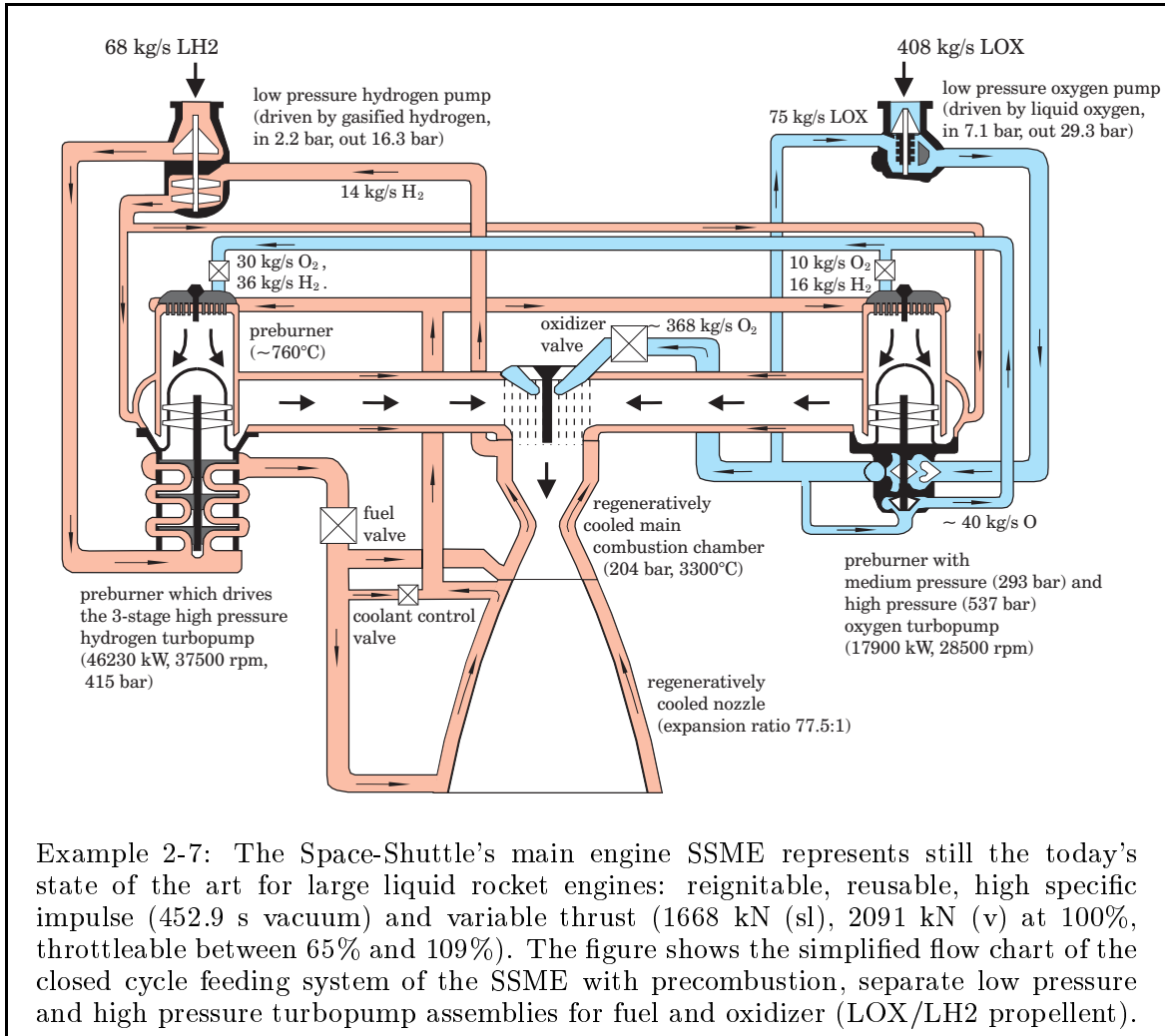
Open cycle engines. In so-called “combustion tap-off cycle engines” the pressure gas for driving the turbopumps comes directly from the main combustion chamber. However, the gas is usually not the optimal mixture for the turbine (it is too hot), and thus it must be cooled down with fuel or oxidizer before it enters the turbine. Better is the use of a separate gas generator, the so-called “gas generator cycle engine”, where it is easy to adjust the optimal mixture of fuel and oxidizer for the generation of turbine shaft power (the gas generator consumes between 1% to 5% of the propellant). The turbine exhaust can either be discharged into the main nozzle at a lower pressurized location downstream, or better into a special small hypersonic exhaust nozzle. Open cycle engines that use hot vaporized coolant from the thrust chamber cooling jacket for driving the turbopumps are called “coolant bleed cycle engines”. These cycles have the advantage of the absence of a gas generator, but the turbine power is limited to relatively low values by the maximum temperature of the cooling jacket.

Closed cycle engines. It is possible to feed the turbine exhaust gases directly into the combustion chamber, provided that the pressure raise of the pumps is comparatively high and the chamber pressure is comparatively low (in practice below 76 bar). The disadvantage of the working principle of “expander cycle engines” is that hydrogen, the best regenerative coolant, does not work efficiently at low chamber pressures (engines that burn LOX/LH₂ propellant operate with a chamber pressure between 100 bar and 250 bar). The solution to the problem is the use of a precombustion chamber, actually nothing else than a high-pressure gas generator. However, in contrast to gas generator cycle engines, these “staged combustion cycle engines” inject the total turbine exhaust gas into the main combustion chamber. These engines are more expensive but also the most efficient rocket engines with the highest specific impulse.





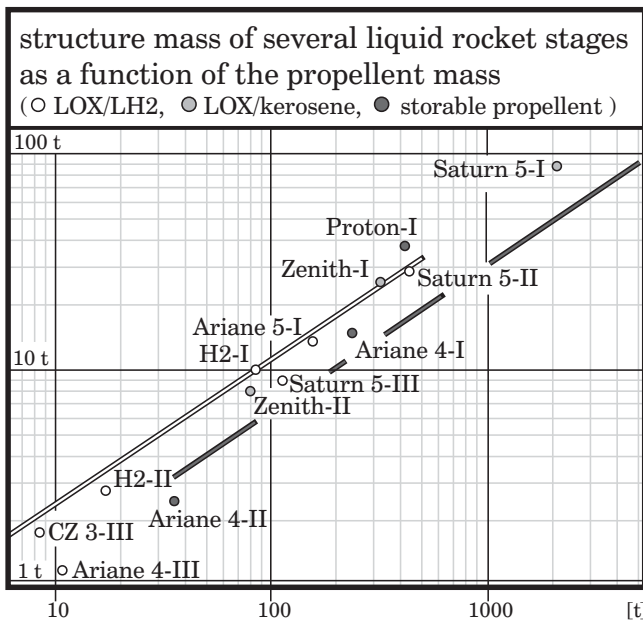
Example 2-6: Japan's LE-7 engine for the first core stage of the H2 launcher and Europe's Vulcain engine for the core stage of the Ariane-5 launcher are similar in size but quite different in their construction: both engines burn the high energetic propellant combination LOX/LH2, but the LE-7 engine uses a closed combustion cycle feeding system with a preburner while the Vulcain engine uses an open cycle feeding system with a gas generator and two turbopumps. The consequence of the different feeding cycle is that the more advanced LE-7 is slightly better in performance than the Vulcain engine, but the LE-7 is also the heavier and more expensive rocket motor.



Ignition and control. The liquid propellant flow is usually initiated by a solid propellant start grain (occasionally also by a small auxiliary pressure gas feeding system) that builds up the operational pressure in the gas generator or in the precombustion chamber (the reignitable engine of the Space-Shuttle SSME uses exclusively the tank pressure to initiate the propellant flow). The burning start cartridge ignites the propellant in the combustion chamber; engines which are made for several burning phases with non-spontaneously ignitable propellents (LOX/LH2 or LOX/kerosene) use usually electrical spark igniters near their injectors to activate the combustion. The control of the valves which regulate the propellant flow to the combustion chamber (or in the precombustion chamber) needs a carefully designed computer system: the building-up of pressure during the starting sequence must be quick and smooth, prohibiting an accumulation of unburned propellant in the thrust chamber which could cause an explosion of the engine. Some engines allow in some limits a variation of the thrust force, usually by a control of the chamber pressure and sometimes by a control of the mixture ratio (the optimal ignition mixture ratio deviates usually from the thrust optimal operational value, later the mixture ratio is controlled with the objective that at burnout both propellant components are completely consumed).

2.5.4 Liquid Engine Sub-Masses

A simple stage mass model. Let us now return to the beginning of this chapter: in section 2.1 we analyzed the theory of rocket propulsion, using a simple mass model (equations 2-9 to 2-13). There we assumed that for a certain rocket stage the structure mass is approximately proportional to the propellant mass and the engine mass is approximately proportional to the thrust force. However, the actual relationships are obviously more complicated: when we want to refine the liquid stage mass model for a more detailed consideration, we can also try the following nonlinear mass model. The model assumes that the structure mass is approximately proportional to the surface of the tanks, and that the specific engine mass gets smaller with growing thrust.

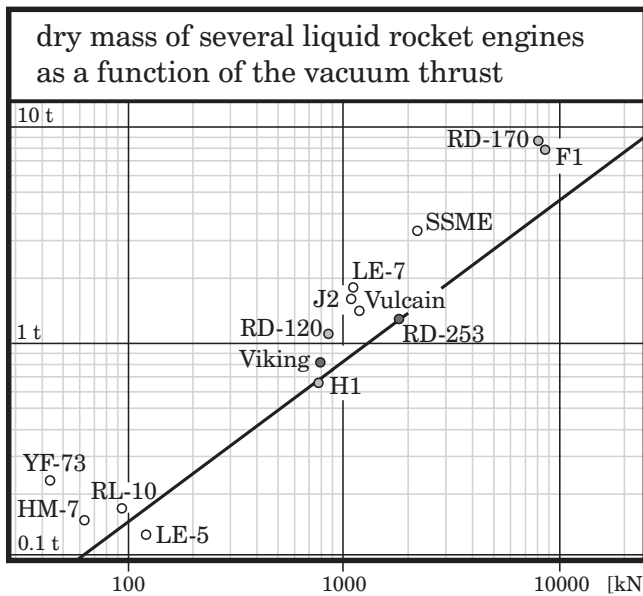


$$M_{structure} = C \cdot (M_{propellant}/\rho)^{0.666}$$

with $C=225$ for $\rho \approx 280 \text{ kg/m}^3$
and $C=350$ for $\rho \approx 1220 \text{ kg/m}^3$

(2 - 139)

We have to look at existing hardware to find the constants in the model above: the constant C in the equation (2-139) assumes approximately the value 225 for high energetic propellant (the white line in the diagram) and 350 for low energetic propellant (the dark line). Masses have the dimension of kgs.



$$M_{engine} = C \cdot S^{exponent}$$

with $C = 4.75$
and $exponent \approx 0.75$

(2 - 140)

The constants C and $exponent$ in equation (2-140) are actually not functions of the propellant type; they take for example the values $C = 4.75$ and $exponent \approx 0.75$ when the thrust S is specified in kN and the engine mass M_{engine} in kg (the line in the figure on the left).

The nonlinear model is still simple but usually more accurate than the linear model.

A more detailed engine mass model. For weight analysis the nonlinear model is better than the linear model of the section 2.1; however, for some other studies it is insufficiently detailed. For example, when we want to analyze the influence of mixture ratio or feeding cycle on the payload capacity of a new space launcher, we have to use a more detailed model. The following model considers not only thrust but also propellant type, feeding cycle, chamber pressure, and expansion ratio:

$$M_{engine} = 1.34 \cdot (M_{turbopumps} + M_{valves} + M_{chamber} + M_{injector} + M_{nozzle}) \quad (2-141)$$

with:

$$M_{chamber} = 0.75 \cdot S^{0.85}$$

$$M_{injector} = 0.25 \cdot S^{0.85}$$

$$M_{turbopumps} = C_{propellant} \cdot C_{turbopumps} \cdot (S \cdot p_{chamber})^{0.71}$$

$$M_{valves} = 0.02 \cdot (S \cdot p_{chamber})^{0.71}$$

$$M_{nozzle} = \epsilon \cdot S \cdot (0.00225 \cdot C_{nozzle} + (0.225 - 0.075 \cdot C_{nozzle})/p_{chamber})$$

$C_{propellant} = 0.19$ for high energetic propellant and 0.11 for low energetic propellant; $C_{turbopumps} = 0.5$ for engines with prepumps and 1.0 for engines without prepumps; and $C_{nozzle} = 1.0$ for regeneratively cooled nozzles and 0.0 for dump cooled nozzles. The masses M have again the dimension of kilograms, the vacuum thrust S of the engine is specified in kN and the chamber pressure $p_{chamber}$ in bar. Term ϵ is the expansion ratio. Equation (2-141) represents a so-called analytical/statistical model, which means it considers not only statistical data but also physical relationships.

engine	S [kN]	$p_{chamber}$ [bar]	ϵ	actual mass [kg]	model mass
LOX/LH2 propellant:					
SSME (prepumps):	2091	204	78	3177	3074
RL-10A-4:	92	40	84	168	220
J2:	1040	54	28	1578	1351
LE7:	1078	130	54	1741	2041
LE5:	122	39	130	244	324
Vulcain (dump cooling):	1145	110	45	1475	1851
HM-7B:	63	36	83	155	157
YF-73:	44	26	40	236	94
LOX/kerosene propellant:					
F1:	8200	69	16	8390	5737
H1:	735	38	28	680	787
RS-27A:	1032	48	8	1027	923
RD-170 (4 thrust chambers):	7904	242	37	8755	6779
RD-120:	834	161	106	1125	1224
storable propellant:					
Viking 5C:	752	58	11	826	759
RD-253:	1745	150	26	1280	2184

The model (2-141) gives also estimates for the masses of the sub-systems of a liquid rocket engine; however, like all mathematical models is often not really accurate.

2.6. Solid Rocket Motors

Solid rocket propulsion is ideal for military purposes; and, since the space industry is often involved in the fabrication of missiles, it does not surprise us that solid rocket motors are also manufactured for space applications. Solid rocket motors are often acclaimed for being comparatively safe and cheap, because they need no feeding system (and no other movable components except for a movable nozzle perhaps). However, the assumption that solid propulsion for space applications is cheap and safe is not backed up by experience; and in fact solid propulsion involves many disadvantages when it is applied to spaceflight. The performance of a solid rocket motors is usually lower than the performance of a liquid stage of the same size; the thrusting behavior of a solid motor is not uniform and the control is difficult.

2.6.1 Solid Propulsion for Space Applications

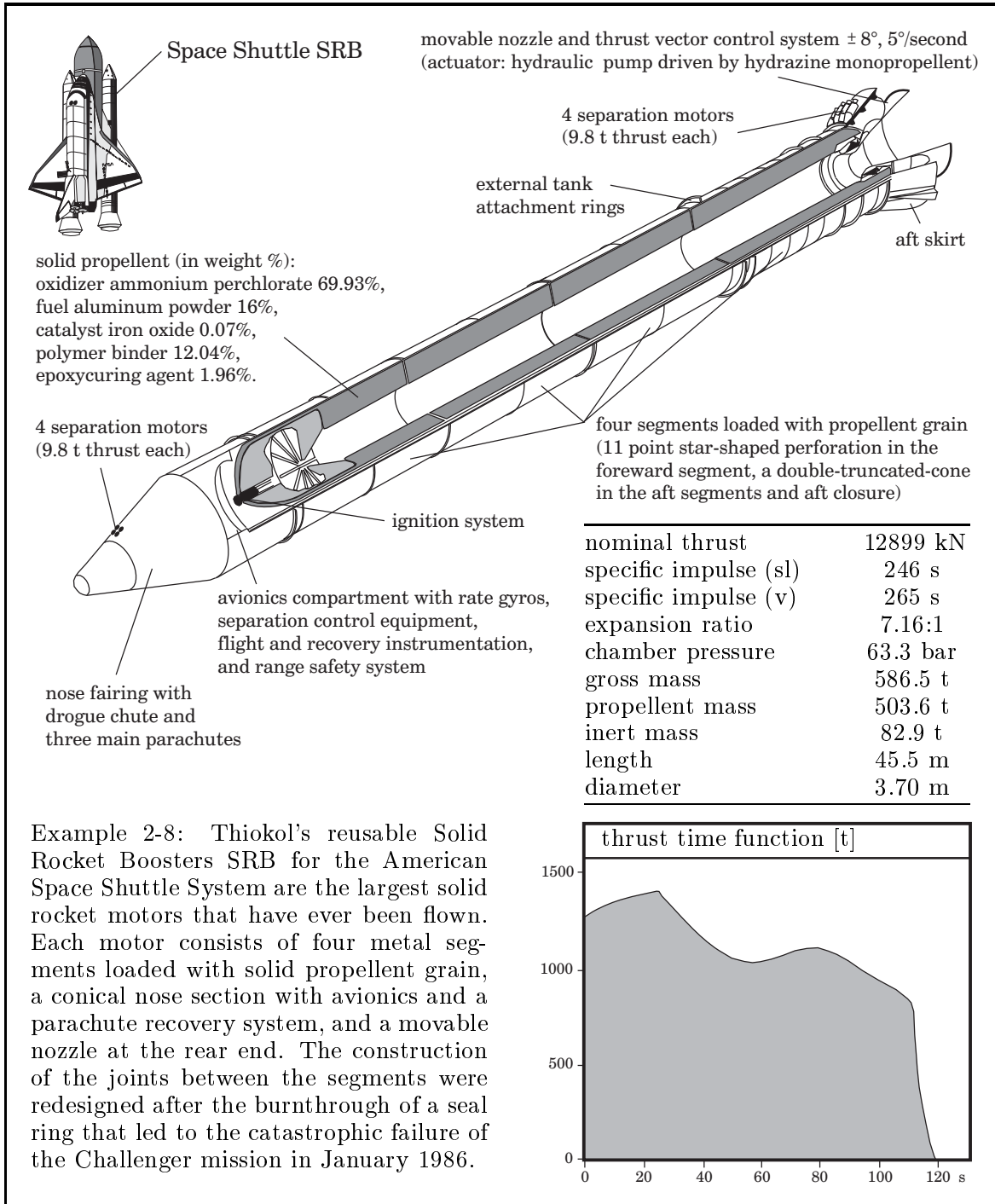
Composite solid propellant. Solid rocket motors use so-called composite (cast) propellant, named HTPB (an abbreviation for hydroxy-terminated polybutadiene) or CTPB (carboxy-terminated polybutadiene). It is a mixture of aluminum powder (fuel) and ammonium perchlorate (oxidizer), held together in one heterogeneous substance by a binder (for example polybutadiene). Solid propellant is “tailored” for a certain application: it means the mixture of fuel, oxidizer and binder varies substantially from one application to another and contains also several other ingredients. The material form resembles rubber or plastic, it burns irregular with a bright smoky flame and releases toxic exhaust gases which contain liquid and solid droplets.

Small solid strap-on booster motors. The disadvantages (low performance and difficult controllability) are less crucial when solid propulsion is applied to small, fixed-canted launch assist boosters. These motors weigh about 5 to 10 tons, their thrust is high in comparison with their weight (about five times as much), and the sea-level specific impulse ranges from 230 to 260 seconds. Typical examples are the PAP boosters of the Ariane-4 vehicle and the Castor IV boosters of the Atlas IIAS vehicle.

Big solid strap-on booster motors. The disadvantages of low performance and bad controllability are more serious when solid propulsion is applied to larger motors. However, ignoring this point the American Space Shuttle followed the concept of the Titan VI vehicle, and also the modern launchers of Europe Ariane-5 and Japan H2 use big solid strap-on boosters with steerable nozzles to assist the initial flight phase. The weight of these motors ranges from 70 tons (H2) to 590 tons (Space Shuttle).

Solid upper stages. Solid propulsion is also used in upper stages, and, since upper stages require a high exhaust velocity and a small burnout mass, a class of kick motors has been developed that uses cases of fiber material (wound-filament-reinforced plastic). Solid upper stage motors weigh between 1 and 10 tons, they often use spin stabilization and operate with a specific impulse of 290 to 300 seconds.

Small all solid launchers. There exist also some small space launchers which use solid propulsion exclusively, for example the US-Pegasus vehicle or the Japanese M5 vehicle. The poor performance of solid propulsion makes at least three stages necessary for orbital flight capacity (sometimes four or even five stages). However, some advantages are also involved when a launch vehicle uses solid motors exclusively: the propellant is storable for a long time interval, and the launch procedure needs comparatively little time for checking-out the hardware (tanking is not required).



Example 2-8: Thiokol's reusable Solid Rocket Boosters SRB for the American Space Shuttle System are the largest solid rocket motors that have ever been flown. Each motor consists of four metal segments loaded with solid propellant grain, a conical nose section with avionics and a parachute recovery system, and a movable nozzle at the rear end. The construction of the joints between the segments were redesigned after the burnthrough of a seal ring that led to the catastrophic failure of the Challenger mission in January 1986.

2.6.2 The Burning Rate of Solid Propellant

Burning rate and chamber pressure. The combustion velocity of a solid propellant grain is responsible for the chamber gas generation rate, and thus it influences the thrust of the motor directly. The reaction front propagates with a certain velocity into the burning solid propellant grain. This velocity, called “burning rate”, is an important propellant characteristic (a typical value for the burning rate of a solid propellant in a space application is for example 1 cm per second). The burning rate depends on the composition of the solid propellant, the combustion pressure and the temperature of the grain before the combustion (and also on some other factors). The burning rate can be increased by the addition of a catalyst to the propellant composition, or by imbedding silver or aluminum wires in the propellant grain.

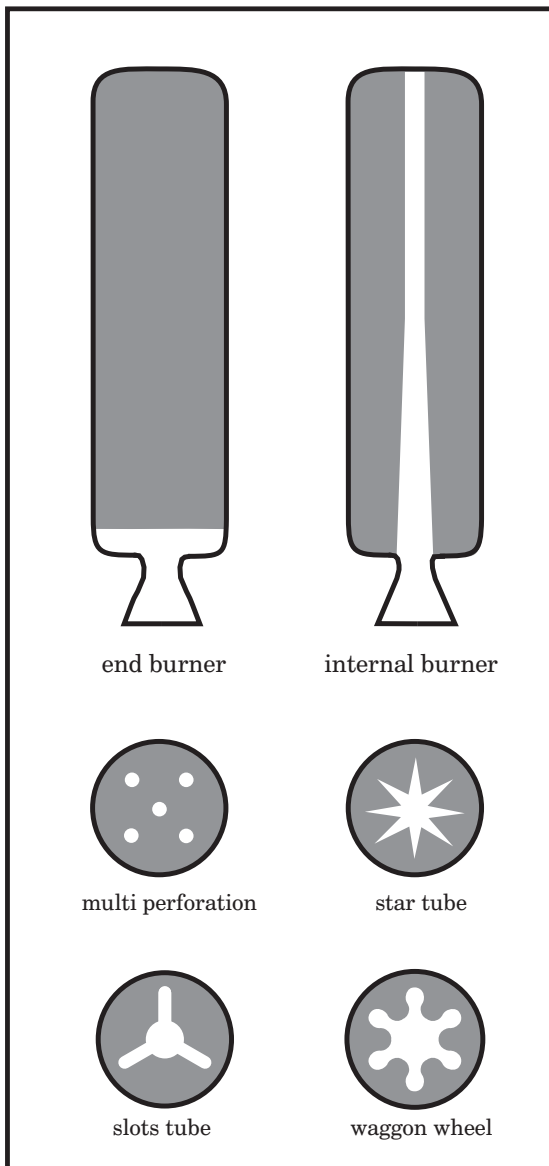
The empirical function $r = a \cdot p_{chamber}^n$ describes the influence of the chamber pressure $p_{chamber}$ on the burning rate r of a solid propellant approximately, where the term a is the “temperature coefficient” (indicating that this value is influenced by the grain temperature) and the term n is the “burning rate pressure exponent” or “combustion index”. The value for the pressure exponent must be smaller than 1, otherwise the motor would explode (detonate); typical values for n range between 0.2 and 0.8 (for example $n = 0.4$ for a HTPB propellant). The pressure sensitivity of the burning rate of a certain propellant cannot be calculated, it must be measured using special testing equipment (for example so-called “strand burners”). Flight loads and other accelerations can also cause a compression of the propellant; for example, an increase in the burning rate has been experienced with some spin stabilized upper stage motors.

Burning rate and grain temperature. The burning rate of a solid propellant is also sensitive to the temperature of the grain prior to combustion: the propellant burns faster when it is hot and slower when it is cold. The thrust of a solid launcher can vary by a few percent depending on the temperature. It can be dangerous, for example, that the sun heats the strap-on boosters of a space launcher uneven from one side, and that consequently one strap-on booster burns faster than the other one. Strand burners can be used to find the temperature sensitivity of the burning rate.

Erosive burning and surface cracks. Pressure and grain temperature have the most significant influence on the burn behaviour of a solid propellant, however, also the temperature of the burning surface is important. The burning rate is higher at locations near the exhaust, since there is hot combustion gas that flows quickly over the burning surface and, consequently, the heat transfer rate to the surface is high. This effect is called “erosive burning”, it amounts a few percent. Also defects in the surface (cracks and holes) cause an increase of the burning rate. The reason for surface cracks can be stresses during manufacturing, transportation, thermal cycling, or assembly. Solid propellant grain can accept a cumulative damage during the time prior to launch, and the number of cracks in the propellant grain can limit the storability of the propellant grain. The motor is not usable anymore after its “aging limit”, when a deterioration of the propellant grain prevents a reliable and safe operation.

2.6.3 Grain Configuration

Propellant geometry. The science that analyzes the behaviour of burning solid propellant as a function of the the grain configuration and the chemical composition is called “internal ballistics”. Solid propellant is a rubber-like (or plastic-like) substance; and the geometrical form of the grain has a main influence on the thrust-time function of a solid rocket motor (the “grain” is the shaped propellant mass inside the motor): the mass flow rate (or gas generation rate) of the motor is simply the burning rate multiplied by the area of the burning surfaces. The pressure in the motor, however, can react sensitively on small changes of the burning surfaces, because the burning rate itself is an increasing function of pressure. A constant pressure (or thrust) time function is called “neutral burning”; the burning behaviour is called “progressive” when the thrust increases in time and “regressive” when the thrust decreases.



The “end-burner” with cylindrical grain on the left, for example, provides an approximately neutral burning behaviour. This grain configuration has also the advantage that the volume inside the motor is efficiently used (the motor has a high “volumic loading fraction”), however, it is difficult to achieve a high thrust force when the motor case is a slender body, and the rear motor case (close to the nozzle) is exposed for an extended time interval to hot combustion gases. The grain configuration with an internal burning tube is an example for progressive burning behaviour. Progressive burning is usually a disadvantage in spaceflight, because it is wanted that the lift-off thrust is high (later the thrust may decrease). Regressive burning behaviour requires holes in the form of a star or a wagon wheel, sometimes several holes or slots in the holes and so on. The size of these holes varies usually along the length of the motor. Grooves and holes in radial direction can add a progressive or regressive behaviour to the burning, depending on the inclination of the holes with respect to the centerline of the grain. Wires or catalysts can be incorporated at certain points into the grain in order to predefine a thrust time function for the solid rocket motor.

Grain stresses and failures. Solid propellant behaves under high stress like a viscoelastic material: the deformation is not only a nonlinear function of the load but also a function of the time interval of the exposure to the load. The material does not return to its original form when the load is removed. Solid propellant is a nearly incompressible substance that can stand pure compression, but it can easily be damaged by tension or shearing forces. Excessive stress can cause cracking of the burning surface, which in turn can be the reason for an unwanted increase of the burning rate and consequence in a bursting of the motor. Both, the coefficient of thermal expansion and the modulus of elasticity of solid propellant, is quite different from the corresponding parameters of the case material: new, unwanted burning surfaces can result from a separation of the propellant grain from the walls. Pieces that break-off from the grain during the combustion do not only reduce the specific impulse of the motor but can also damage the throat when they are exhausted.

The manufacturing process of a solid propellant grain (mould casting or extruding) is not easy: attention has to be paid that the grain is essentially free of flaws (voids and surface cracks), that the grain has everywhere the predetermined chemical composition and that its burning behaviour is predictable. When the motor case itself is used as a casting mould, it must be observed during the curing process (heating in an oven) that the thermal expansion coefficient of the grain is much higher than the coefficient of the metal case, and that cooling down the motor can lead to tensile stresses in the grain or in the insulation between the grain and the motor case. When the grain is manufactured outside the motor case and installed later, it has to be assured that the bonding between grain and case is free of voids and that the seals of the motor case are tight under all conditions (a leak in one of the joints between the segments of a solid booster motor was responsible for the accident of the American Space Shuttle Challenger). When the motor case is made of wound-filament-reinforced-plastic material, the propellant grain itself can be used as a mandrel for coiling up the fibers. Attention has to be paid during all phases of the production that the grain is never mechanically over-stressed or overheated. The grain is a inflammable substance, and a inadvertent ignition will almost always result in a catastrophe.

The deformation of the grain under stress and the damaging effect of excessive stress are usually analyzed using finite element methods. The influence of cumulative damage, however, can still not satisfactorily be predicted by computer programs; and empirical methods have to be used to consider the nonlinear behaviour of the material.

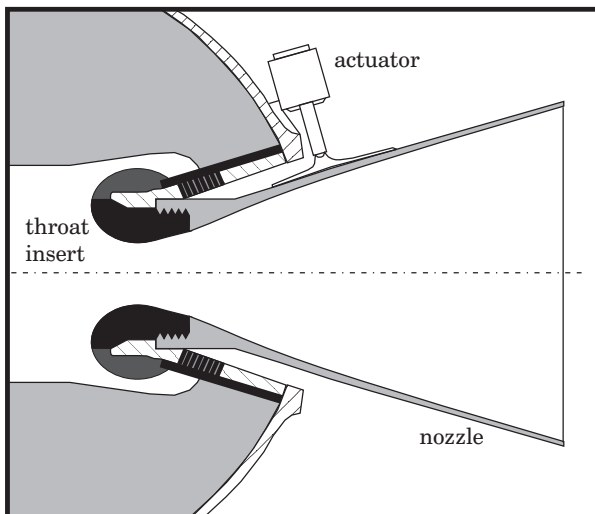
2.6.4 Solid Motor Cases

Metal cases. The motor case of a solid rocket motor is not only the combustion chamber, but usually also an essential part of the structure of the vehicle, and thus it has to stand not only internal but also external loads. Metal cases made of steel or titanium alloy have the advantages of a well-known stress resistance and a good temperature resistance; furthermore the material is rugged (appropriate for pasting the propellant on the case walls) and unproblematic during the whole production process.

Fiber material cases. The disadvantage is that metal cases for high chamber pressures are comparatively heavy, and thus modern constructions use more often now so-called “wound-filament-reinforced-plastic cases”. These plastic cases get their strength from many extremely thin fibers wrapped around the combustion chamber in precise pattern and glued together with a plastic binder (usually epoxy resin). Fibers can stand strong tension but only tension, therefore the filaments have to be aligned with the directions of the principal tension (that means at least two rectangular oriented filament layers are required). Shearing forces which occur during the normal operation have to be taken over by the weaker binder material, and thus the actual strength of the composite material is much lower than the strength of the fibers themselves (typically by a factor 3 to 5). Plastic cases require more heat insulation because the plastic softens at temperatures above 180°C , and the volume of the motor gets larger than with a metal case. Typical fiber materials are glass, Kevlar, graphite and carbon, where graphite or carbon fibers are preferred because of their best strength to density ratio. Plastic is sensitive to sunlight and air moisture, and mechanical properties can deteriorate when the case is stored for a longer time period.

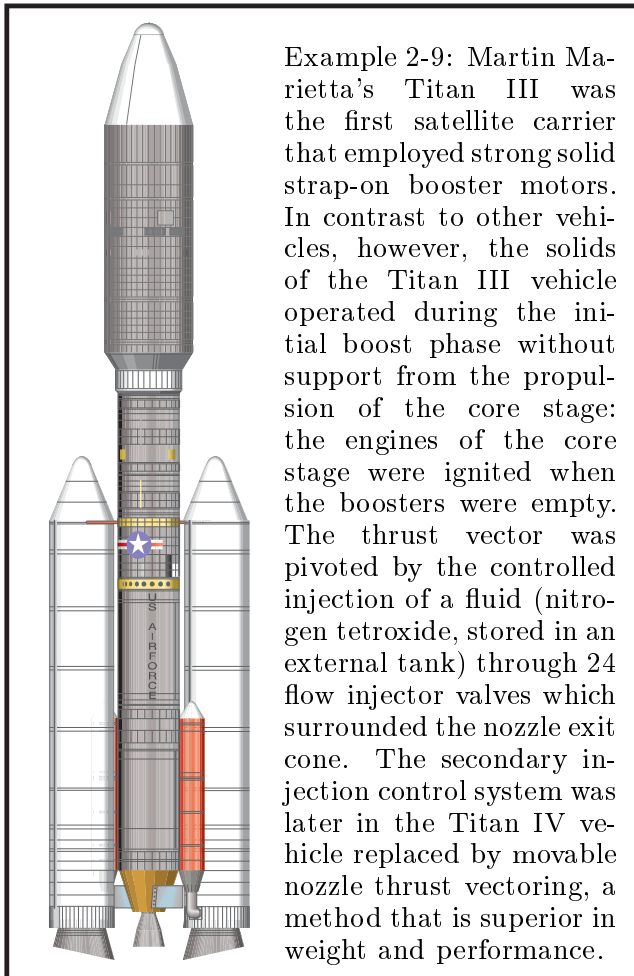
2.6.5 Solid Motor Nozzles

Nozzle shape. The combustion gas is expanded in a hypersonic nozzle which has a convergent part, a throat and a divergent part. Cooling method is usually heat radiation (there is no propellant that could be used for regenerative cooling) and, particularly in the throat region, material ablation (melting, vaporization and decomposition of material imbedded in the matrix of temperature resistant fibers serves as a heat sink). The throat erodes during the burn time interval; and abrasion can cause the throat diameter to grow by a few percent (where an increase below 5% is usually acceptable). Appropriate materials (such as carbon fibers in a carbon matrix) allow surface temperatures of 2300°C and more. Advances in material science can make considerable weight reductions and performance improvements possible, and the selection of the appropriate material is a key to success in the design of new solid rocket motor nozzles.



The solid rocket motor nozzle has rather the shape of a cone than the shape of a bell, because a bell-shaped nozzle would not stand the excessive erosion (the constant bombardment with liquid and solid droplets in the combustion gas). Nozzles with a small divergence cone angle are comparatively long and have a weight penalty, but nozzles with a large divergence cone angle involve divergence losses and a deterioration of the specific impulse (typical cone half angles $\approx 12^{\circ}$ to 20°).

Thrust vectoring. It is not always necessary to equip a solid rocket motor with a thrust vectoring capability: for example spin stabilized upper stages, rocket motors in winged vehicles, or small strap-on boosters do usually not need a control of the thrust direction. However, thrust vectoring capability is required when the solid motor is operative and has to take over the stabilization of the launch vehicle. Older solid rocket motors used often “secondary injection control”, but today the more modern constructions use nearly always “movable nozzles”, a thrust vectoring technology which is superior in weight and performance.



Example 2-9: Martin Marietta's Titan III was the first satellite carrier that employed strong solid strap-on booster motors. In contrast to other vehicles, however, the solids of the Titan III vehicle operated during the initial boost phase without support from the propulsion of the core stage: the engines of the core stage were ignited when the boosters were empty. The thrust vector was pivoted by the controlled injection of a fluid (nitrogen tetroxide, stored in an external tank) through 24 flow injector valves which surrounded the nozzle exit cone. The secondary injection control system was later in the Titan IV vehicle replaced by movable nozzle thrust vectoring, a method that is superior in weight and performance.

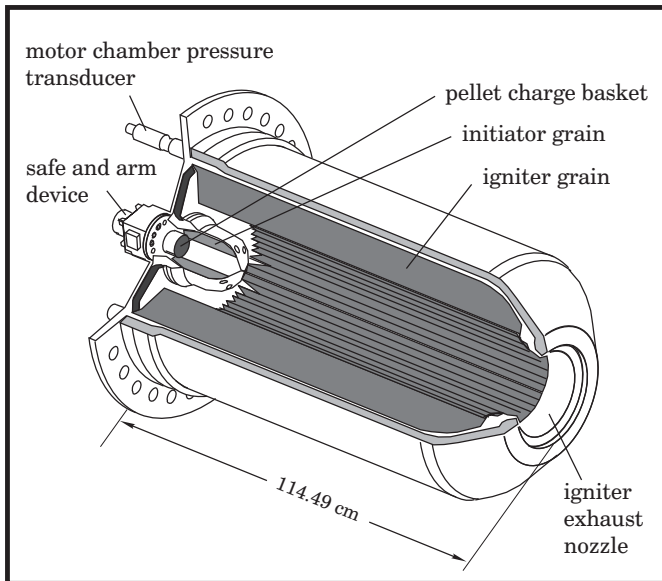
The “secondary injection control” influences the thrust direction by the appropriate injection of a fluid (gas from a gas generator, direct bleed from the chamber, or a stored liquid from an external container) through orifices in the divergent part of the nozzle. The additionally entering gas has the effect of forming oblique shock waves in the hypersonic nozzle which cause a slight deviation of the main gas flow. The injection of secondary liquids has the disadvantage of an excessive consumption of fluid when the control moments required for vehicle stabilization are large, and it has also the disadvantage of a slight reduction of the specific impulse. Hot gas injection thrust vector control is more attractive because this gas is sufficiently available in a solid rocket motor, but the conduction of hot gas in pipes and the control with valves is still not unproblematic from the material science point of view.

The disadvantages involved with secondary injection control can be avoided with the use of movable nozzles in solid rocket motors. However, the throat of a solid rocket motor nozzle is exposed to extreme stresses (temperature, pressure and mechanical abrasion), and the construction of a movable nozzle becomes even more problematic. The convergent part of a movable nozzle is connected with the combustion chamber via flexible joints, which can be sealed rotary joints or sealed flexible bearings. Strong mechanical actuators are necessary to pivot the nozzle. Common types of movable nozzle use the deformation of multi-layer bearing packs (with stacked elastomeric layers between metal sheets) that act as seals and as load transfer bearings.

2.6.6 Ignition and Control of Solid Rocket Motors

Ignition systems for solid rocket motors. The designer of the ignition system for a solid rocket motor is faced with the following three problems: the ignition may not at all occur at a moment when this is not wanted; then, immediately at the right moment, the system must inflame the entire grain surface that is intended to burn, and finally the system should weigh as little as possible. The igniter is usually a small solid rocket motor inside the motor case, placed at the front end of the free volume of the grain. The ignition occurs in stages: a signal triggers the electrical current that burns a bridgewire to ignite the booster charge; and then the booster charge inflames the main charge of the igniter (solid rocket propellant or explosive pellets). The hot combustion exhaust gases leave the igniter through one or several holes (actually sonic or supersonic nozzles) and inflame the propellant when they flow over the grain surface to the nozzle; or the basket that contains the pellets explodes and the pellets inflame the propellant when they impinge on the surface of the grain.

Protection against inadvertent ignition provides a so-called “safe and arm” device. This device is essentially a electrical switch that keeps the igniter circuit grounded in its safe position, supported by mechanical “safeguards” that prevent a motor ignition even in case of an unwanted release. Launch assist strap-on booster motors can be



ignited from outside through their nozzles with externally mounted igniters, which are then a part of the ground equipment and add no weight penalty to the structure of the empty rocket motor; and the motor is “safe” when the igniter is not present. However, also static electricity, induced current from electromagnetic radiation and mechanical shocks can be the energy source for an unintentional ignition. The figure on the left shows the igniter construction of the solid booster motors of the American Space Shuttle system.

Burnout control and reignition. The amount of solid propellant that remains unburned in the chamber is called slivers. With advanced grain design technology there are nearly no slivers today in modern solid rocket motors. However, the total impulse which a solid rocket motor develops in a certain mission is not accurately predictable, and thus efforts have been made to control the burnout instant by a sudden chamber depressurization (opening additional gas escape areas). It has also been tried to construct solid rocket motors for controlled burnout and re-ignition, but in summary we can say that all these tries have not proved to be worthwhile, because then the only great advantage of solid rocket motors for space applications gets lost: the simplicity.

2.7. Attitude and Trajectory Control Thrusters

Nearly all spacecraft need maneuvering systems in order to control their attitude and their trajectory. The attitude and trajectory control system is usually equipped with a certain number of small thrusters, their orientation is fixed with respect to the vehicle. For example, to be able to control the rotational motion separately from the translational motion, a satellite requires at least two times 6 thrusters; and a spacecraft that is intended to maneuver in the vicinity of a space station (rendezvous and docking) is often equipped with even more thrusters. Auxiliary propulsion systems provide small thrust forces between 0.02 and 400 N, typically. Like large rocket engines in main propulsion systems, auxiliary propulsion systems of spacecraft should be light weight constructions that consume propellant sparingly. However, the principal design aspects for auxiliary propulsion systems are: reliable working during a long operation time, and accurately predictable thrust for many small maneuvers. These features demand the application of pressure gas feed systems.

2.7.1 Gas Pressure Feed Systems

Pressure gas. The pressure gas is usually an inert gas, such as helium (sometimes nitrogen) stored in a spherical tank. It is used to feed the propellant into the thrust chamber by displacing it from the propellant tank (but the gas can also be used for other purposes, for example for the operation of valves). The amount of gas which is required for propellant feeding depends on the gas pressure, on the chamber pressure and on the propellant volume. A high chamber pressure leads to a comparatively high specific impulse of the propulsion system, but a high chamber pressure is also responsible for heavy propellant tanks. Therefore it is necessary to make a tradeoff between performance and weight. Spherical high pressure gas tanks operate with pressures between 70 and 700 bar, the tank pressure is always lower (for example between 10 and 90 bar). The chamber pressure is accurately controlled by high-precision valves. "Positive expulsion devices" in the tanks provide a mechanical separation of the pressure gas from the propellant (for example piston mechanisms, bellows, bladders and so on); they prevent that the pressure gas reacts with the propellant, dissolves in the propellant or enters the discharge line of the propellant tank.

Propellant. Essentially, there are three types of propellant for auxiliary thrusters: cold gas (any pressurized gas, for example hydrogen, helium or nitrogen), hydrazine N_2H_4 as monopropellant, and finally hydrazine (or MMH, a derivative of hydrazine) burned with nitrogen tetroxide N_2O_4 (or MON, a mixture of NO with N_2O_4).

Thrust chambers. The gas pressure feed system transfers the propellant via small pipes from the tank to the thrusters, which are usually mounted in the periphery of the spacecraft (fixed at rods or directly at the outer wall). The heat loads for the thrust chambers are comparatively small, particularly in case of pulsed operation: the radiation cooled nozzles of auxiliary thrusters require no special cooling system.

2.7.2 Auxiliary Propulsion Systems

Classification. A large variety of different thruster constructions for attitude and trajectory control is offered to satellite manufactures and launch service providers. The following table lists performance data of models from the German space industry:

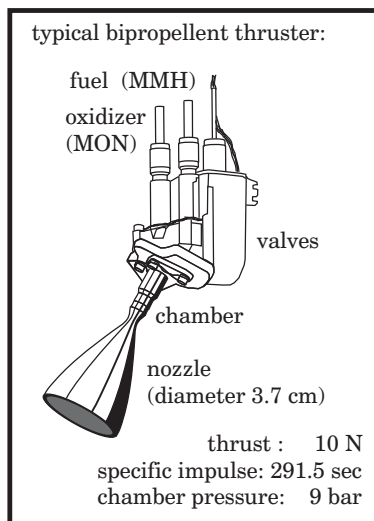
thruster name	thrust	chamber pressure	specific impulse	mass	length	nozzle diameter
cold gas (nitrogen,hydrogen or argon):						
CGT 1	0.02 N	4 bar	66 s	120 g	64 mm	
arcjet (hydrazine):						
HAJ 1	0.1 N	18 bar	530 s			
hydrazine monopropellant:						
CHT 0.5	0.5 N	22 bar	228 s	195 g	113 mm	4.8 mm
CHT 1	1.0 N	22 bar	228 s	377 g	172 mm	7.6 mm
CHT 2	2.0 N	22 bar	228 s	200 g	123 mm	8.6 mm
CHT 5	0.5 N	22 bar	229 s	220 g	129 mm	14 mm
CHT 10	10 N	22 bar	231 s	240 g	142 mm	19 mm
CHT 20	20 N	24 bar	235 s	360 g	195 mm	33 mm
CHT 400	400 N	26 bar	229 s	2700 g	325 mm	87 mm
MON/MMH bipropellant:						
S4	4.0 N	4 bar	285 s	290 g	115 mm	30 mm
S10	10 N	7 bar	287 s	350 g	138 mm	37 mm
S10/2	10 N	9 bar	292 s	310 g	138 mm	37 mm
S400/1	400 N	7 bar	303 s	2800 g	531 mm	248 mm
S400/2	400 N	10 bar	318 s	3400 g	531 mm	248 mm

Cold gas thrusters. The simplest system with the lowest specific impulse is the cold gas thruster; its advantages are high reliability (because of the simplicity), excellent controllability (because of extremely low thrust) and cold exhaust plumes. Typical applications for cold gas thrusters are manned maneuvering units for extra vehicular space activities and the precise control of a docking maneuver between a capsule and a space station. Since almost any pressure gas can be used as propellant, cold gas thrusters are attractive to utilize some waste gas as working fluid (for example carbon dioxide) in order to control a manned space vehicle during a long mission. The energy for propulsion comes exclusively from the pressure of the gas. The specific impulse of the system improves when gas with light weight molecules is used, but storing pressure gas in tanks is easier when the gas molecules are heavier (theoretical values range from 284 seconds for hydrogen to 76 seconds for nitrogen).

Resistor jets and arc jets. The specific impulse of an auxiliary thruster is improved when the working gas is heated up before the expansion; and heating up the gas can be performed electrically with a resistor or a light arc. In space however, electrical energy is usually not disposable in large amounts, and therefore resistor jets or arcjets use often a monopropellant as working fluid. For example, the propellant hydrazine decomposes in the thrust chamber and releases additional energy for propulsion.

Hydrazine monopropellant thrusters. The usual way to control the attitude and the trajectory of a satellite with auxiliary propulsion is the application of hydrazine thrusters. The propellant hydrazine is frequently used in bipropellant systems using nitrogen tetroxide as oxidizer, however, the liquid releases energy when it decomposes over a heated catalyst bed (iridium). The gaseous reaction products are nitrogen, hydrogen and ammonia. In comparison with a cold gas system the hydrazine thruster has the advantage of a better specific impulse (typical values range from 220 to 250 seconds, for cold gas systems from 50 to 75 seconds only). The position keeping subsystem of an earth orbiting satellite requires a Δv capacity of 100 to 200 m/s for every year in space, typically, and thus a quite large portion of the initial satellite mass amounts to propellant (30% to 50%). Hydrazine thrusters are controlled by precision valves often in pulse mode (multiple small impulses for accurately reproducible Δv). A thrust regulation can readily be achieved by a reduction of the chamber pressure.

Bipropellant thrusters. The specific impulse of hydrazine control thrusters can be increased to values between 270 and 320 seconds when an oxidizer is used (nitrogen tetroxide or MON). Hydrazine burned with nitrogen tetroxide is a so-called hypergolic combination, it means that the fuel ignites spontaneously when it is mixed with the oxidizer. These auxiliary propulsion systems are then actually nothing else than usual rocket engines for storable propellant, however, they use a pressure gas feeding system instead of a turbopump feeding system. The figure below shows the sketch of a typical 10 N bipropellant (MON/MMH) thruster for the control of a satellite.



The application of a pressure gas feeding system is sensible for the accurate control of many comparatively short thrusting maneuvers during a long operation time period (ten years or more), but tanks for pressure gas feeding systems are heavier than tanks for turbopump feeding systems. Thus, for a single maneuver that requires a large Δv capacity, the application of an engine with turbopump feeding system is preferable. In comparison with monopropellant systems, bipropellant systems are more economical regarding propellant consumption, and thus bipropellant thrusters are now frequently used in modern satellites for attitude and trajectory control maneuvers which require comparatively much propulsion.

Some complex attitude control systems operate bipropellant together with monopropellant thrusters, where the bipropellant thrusters are used for the more propulsive maneuvers and the monopropellant thrusters are used for more precise maneuvers. The reaction control system of the American Space Shuttle, for example, operates with many thrusters (38 primary and 6 vernier thrusters), but the vehicle is also equipped with a complex orbital maneuvering unit that uses large gimballed engines for the initialization of the descent maneuver and for orbital transfer maneuvers.

3. Launch Services

It is well-known that space launchers are very expensive, but space is not accessible without launchers; and thus the availability of a launch facility is a key position for the success of a space programme.

The third chapter of this book is concerned with the technique of launcher design. When a space launcher system is optimized during the preliminary design phase, the optimization objective (“application”) is quite a diffuse quantity. It includes all missions the new launch system will have to perform when it is finally operational. We will have a look at the tasks for launchers in the first section of this chapter. Many practical engineering aspects have to be considered during the preliminary design phase (for example: how many stages, what kind of propellant and so on). We will put emphasis on these practical engineering aspects in the second section. The cost/performance ratio is probably the most important design aspect, at least for commercial launch services. Even when the technical parameters “costs” and “performance” are not accurately defined, we can make mathematical models of costs and performances. In the third section we will evaluate the quality of commercial launch services using such a cost/performance model. The technique of space launcher optimization is demonstrated in the fourth section, taking as an example the European Ariane launcher family in the scenario of strong market competition at the end of the last century. Finally, in the fifth section of this chapter, we will see that even today the use of modern technology does not bring improvements automatically.

The models we use in this chapter are based on material which is provided by the space launcher dictionary of S.J.Isakovits [“International Reference Guide to Space Launch Systems”, AIAA, Washington D.C., 1991 and 1999], the annual volumes of Interavia’s “space directory”, A.Wilson [“Jane’s Space Directory”, Jane’s Information Group, London, 1997-1998] and the ESA launch vehicle catalogue [developed by Aerospatiale under ESA contract number 8152/88/F/BL, Paris, rev.11, Dec. 2000].

3.1. Applications for Space Launchers

Most launchers which exist today are “by-products” of ballistic missile programmes. National prestige and military interest are usually the reasons for their development. When the missile system is operative, it seems to be comparatively easy to modify the weapon for the transportation of military satellites. When the launch system is operative, it seems to be lucrative to offer it also for commercial satellite launches. A scientific research programme often serves just as an alibi to get public funding for a national launcher programme. However, not all space launchers are originated in military programmes: the American Saturn-5 is an example for a launcher development as a “tool” for a scientific space programme; the European Ariane-5 launcher is an example for a launcher development with commercial interest as the main reason.

3.1.1 Target Orbits

The launcher system should be optimized for the transportation of space payloads. Here the word “space” means the selection of some quite different earth orbits. Most payloads go to one of the following space destinations:

Low earth orbit LEO. A launcher has the best payload capacity when it departs from a launch site near the equator into eastern direction to reach a low earth orbit. Thus, for many payloads the low earth orbit is simply the “cheapest” orbit (micro-gravity research, bio-medical experiments, astronomical observations and so on).

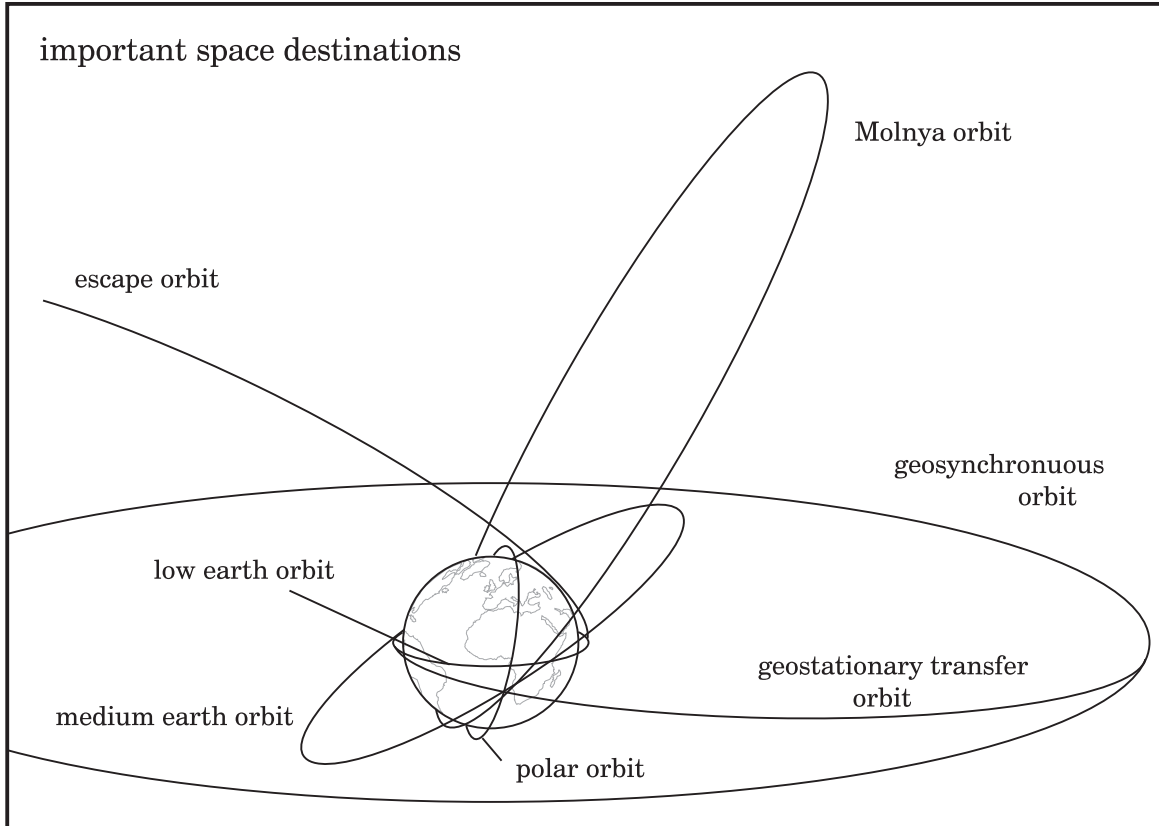
Polar earth orbit PEO. Many big military spying satellites with heavy optics go to polar orbit because they have to observe the complete surface of the earth. Low earth orbits with high inclinations are also used by non-military earth observation satellites (environmental surveillance, discovery of earth resources, meteorological monitoring and so on). The oblateness of the earth causes a gyroscopic effect on polar orbits; this perturbation depends on the inclination and can be used to synchronize a polar orbit with the direction of the sunlight (to establish a sun-synchronous orbit SSO).

Medium earth orbit MEO. Mobile telecommunication systems and satellite navigation systems use direct links to satellites with non-directional antennae (without directing a parabolic antenna towards the satellite). For global coverage these orbits have to be inclined with respect to the equator. When a lower orbital altitude is used, the data transmission requires less electrical power, but then more satellites are necessary for permanent availability of the system (the altitude should not be too low, otherwise satellites will have to be replaced frequently due to short orbital lifetimes).

Geostationary transfer orbit GTO. The geostationary earth orbit offers the advantage that fixed antennae can be used for the communication between ground stations and data relay satellites. Since the orbit has the altitude of 35775 km above the equator, data transmission requires more electrical power; and the orbit cannot be used for communication when the ground stations are near the north pole or near the south pole. Usually, launchers deliver GEO-satellites into the transfer orbit (circularization is done by satellite-integrated-propulsion or by a separate kick stage).

Molnya orbits. Russia uses communication satellites on high eccentric, inclined orbits to cover the north polar regions of the country. With an orbital period of 12 hours these Molnya orbits are not geostationary orbits, but at least the satellites stay some hours during well-known time intervals over the desired geographical region.

Earth escape trajectories. Some scientific probes leave the gravitational field of the earth on escape trajectories. When a planet is the target, the launcher has to bring the space probe to a hyperbolic orbit. When the moon is the target, the launcher has to bring the payload to a (more or less) parabolic escape trajectory. Escape orbits are also the destination for the “payload” of another exotic project: we should think of the recent collision of pieces of a broken comet with planet Jupiter and install a global defense system against those asteroids on an earth impact orbit.



3.1.2 Commercial Market for Satellite Launch Services

Typical annual launch frequency. When we regard the situation at the end of the 20th century (year 2000), we count 130 payloads as a typical number for the world-wide annual launch frequency. About 40% of these satellites are military payloads. Most of the spying satellites have to be replaced frequently, because they fly at low altitudes for a better optical resolution. The remaining 60% have civilian character, but only half of them are commercial satellites (in case of a public financed satellite project the launch system is usually predefined by the government). Thus, a typical number is 40 satellites per year open for launch service market competition. From these 40 commercial satellites about 15 (better say 10 to 20) go to geostationary position, the principal share of the rest (maybe 25 satellites) goes to low earth orbit.

The GTO market. The geostationary position GEO is typical for commercial satellites, where most of them are telecommunication services. Satellites are ideal for broadcasting over extended geographical regions. For fixed-link telecommunication services, however, terrestrial alternatives seem to be superior in price and quality (fiber optics). GEO satellites weight between 1000 kg and 3000 kg, typically. The typical weight of GEO communication satellites increases slowly, but probably there remains a market for smaller GEO satellites (with masses below 1000 kg).

The LEO market. The principal market share of commercial LEO satellites consists of private mobile telecommunication facilities. Satellite based global telephone services compete with ground based (or aircraft based) alternatives. The technology changes rapidly, and improved electronics can quickly generate an over-capacity: the market is not at all stable, the annual launch frequency may increase or decrease considerably. Satellite based telephone services make use of many small satellites. These systems are preferably installed by multiple launching (several satellites share one big launcher); replacing out-of-order satellites should be done by single launching.

Market for earth observation and navigation. Extraordinary military implications make an open commercial market for both applications improbable. It can be expected that also in the future global satellite positioning systems are installed and operated by military organizations, and that surface imaging will always be censored.

Replacement of satellites. The service life of satellites is limited. The satellite needs propulsion for stabilization of the orbit, otherwise it would slowly drift away from its position (essentially because of drag exerted by the rest atmosphere, and gravitational forces from the sun and the moon). The position keeping sub-system consumes propellant equivalent to a Δv capacity of 100 m/s to 200 m/s per year, corresponding to a propellant mass fraction of 30% to 50% and a life-span of 5 to 10 years, typically. It would be easy to extend the service life of satellites simply by providing more propellant to the position keeping subsystem. However, the technology changes rapidly and electronic and computer equipment is out-of-date just after a few years; and an extended satellite lifetime means a reduced instrumental payload.

Probably, a longer service life will become desirable with more established technology. Solar electric propulsion could greatly extend the operation time of satellites; but then also other parts of the satellites will have to be improved (the radiation in space deteriorates the solar cells, secondary batteries have a limited number of discharge/recharge cycles, micro-meteorites destroy optical surfaces and so on).

3.1.3 Influences on the Commercial Launch Market

Ballistic missiles. A ballistic missile is not an optimal construction for the launching of satellites. Its “application” is quite different from the application of a launcher. The missile is extremely expensive and its construction is a military secret. However, inter-continental ballistic missiles can be used for launching small satellites. Sometimes satellite launches are offered at a dumping price when a missile system is “phased out” (due to disarmament or when a new system is going to be installed).

The space shuttle. Experience shows that the American space shuttle is too big and too expensive for commercial satellite launches (the only sensible application for the space shuttle is the erection and operation of the international space station ISS). A permanent lunar base or a manned Mars mission require even a bigger launcher. When an unmanned launcher derived from shuttle technology is available, it can offer launches at a very low specific price, and create a new sector of space applications.

3.2. General Space Launcher Design Aspects

Theoretically, launch system optimization is a nonlinear optimization problem. Some design parameters are predetermined, some others are open for optimization. The ones open for optimization have to be adjusted appropriately. There are “continuous” parameters (engine thrust, propellant mixture ratio, stage size and so on) and “integer” parameters (the number of stages, the number of engines and so on). Essentially, the parameters open for optimization have to be adjusted in a way that the launcher gets the maximum payload capacity for a certain reference orbit. For maximum payload capacity the launcher has to fly on an optimum trajectory. Since every modification requires another optimum trajectory, launcher optimization is the problem of nonlinear parameter optimization based on trajectory optimization.

Practically, many other design aspects have to be considered too. The launcher is nearly never a completely new development, existing components have to be used. The launcher must have good payload capacity, but it should also have a high success rate, it should deliver accurately into target orbit, it should not pollute the environment and so on. The launching should stress the satellite as little as possible. Very important is the availability of the launch service (this is the time interval between contract and launch) and the flexibility of the launcher concerning payload mass and target orbit. In this section we will generalize some launcher design aspects.

Engineering is a difficult subject when commercial, political and emotional aspects are involved. Technical problems can “simply” be solved by using mathematics: when our calculations are obviously correct, everyone will trust in our conclusions. We leave this strong mathematical base when we discuss “non-technical” questions. Consider, for example, the design of a car. We want to find out what is a good car, and what guarantees a market success. Therefore we optimize it and consider just technical aspects. Then the car should have a good price/performance ratio, this means low fuel consumption with medium motor power for average highway speed, the car should have plenty of space for luggage and place for many persons. The importance of every design aspect depends on the “opinion” of the customers (a different opinion is not necessarily wrong). Obviously, the optimization result is a family car, it does not at all fit for sports cars; but psychological aspects are also important, and the “image” of a sports car sells quite good. Neither can we ignore psychology in astronautical engineering: when a company wants to sell a space plane project to tax-paying public, the design aspect “good looking” is also very important.

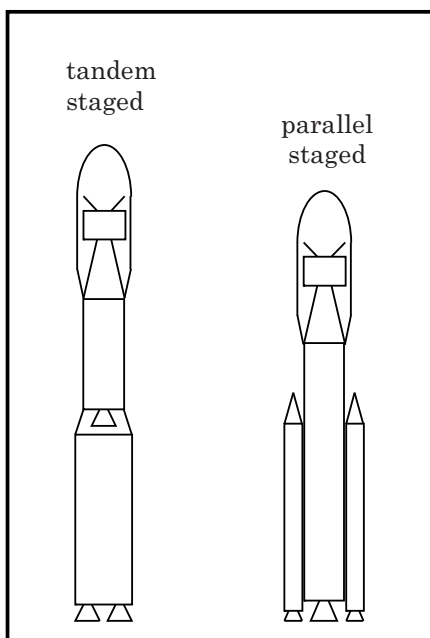
3.2.1 Performance Optimization.

The growth factor. The initial mass of a launcher divided by its payload mass for a given target orbit is called the “growth factor” of the launcher. Usually, the optimization objective in launcher design is “minimization of the growth factor”. It is assumed that a space launcher with a small growth factor has also low payload-specific launch costs. Such a point of view is at least approximately correct.

Optimal number of stages. Usually, space launchers consist of several stages. The staging principle is applied to discharge empty tanks (propellant is the main share of the initial mass, it is consumed during the mission). It is necessary that launchers consist of “enough” stages, otherwise they would have insufficient payload fractions; but on the other hand, costs and complexity increase with the number of stages. Ruppe’s formula gives an answer to the question “how many stages are optimal”: consider the difficulty of the mission (represented by the Δv capacity required to reach the orbit) and “state of the art” (represented by the average exhaust velocity c of the propellant), then form the integer value of the quotient Δv divided by c . It determines approximately the optimal number of stages. When “high technology” is used, the exhaust velocity c is 4.3 km/s, typically. We find “two stages” as optimum for LEO transportation ($\Delta v \approx 9$ km/s) and “three stages” as optimum for GTO transportation ($\Delta v \approx 12$ km/s). When “antiquated technology” is used, $c = 3.0$ km/s, and one more stage is necessary to have the optimal stage number.

Sometimes it is stated that “single stage to orbit” launchers (so-called SSTO systems) are optimal for LEO transportation; but a single stage launcher has just a sufficient payload capacity when it is supported by strap-on boosters (and then it is a 1+1/2 stage system, actually). A single stage launcher without boosters has an extremely bad growth factor (or cannot reach the orbit), even when lightweight tanks are used.

Optimum size of the stages. Every stage should share approximately the same portion of the Δv requirement of the mission as all the other stages. For example, in case of a two-stage LEO launcher ($\Delta v \approx 9$ km/s), each stage should raise the velocity by about 4.5 km/s. Therefore the first stage has to be bigger than the second stage. The mass of the launcher at the ignition of a stage divided by the total mass of the launcher at burnout of this stage should be approximately the same for every stage.



Parallel staging. Two stages can operate either one after the other in a tandem arrangement or burn together in a parallel arrangement, where a core stage is supported by strap-on booster motors. While a tandem arrangement is better from the energetic point of view, parallel arrangement is better for safety reasons (all engines are ignited before lift-off). Sometimes parallel staging offers also the option of a better system flexibility, when more (or fewer) booster motors can be used for a launch. In a parallel arrangement the thrust acceleration of the booster stage has to be higher than the thrust acceleration of the core stage, otherwise the core stage wastes energy to accelerate the booster motors. The booster stage must be separated before burnout of the core stage, otherwise the two stages operate as only one stage.

Circularization in GEO position. The transfer from GTO to GEO is called “circularization” of the orbit. The upper stage of a GTO launcher can perform this maneuver when its engine is constructed for shut-down and re-ignition. More in common practice is the other possibility to establish a circular orbit: a satellite can use propulsion from its position keeping subsystem. This method avoids that the upper stage enters the orbit (energetically it is the same as using one more stage).

Possibilities to support the lift-off phase. Today, conventional strap-on booster motors represent the only established launch assist facility; alternatives are not in use; but naturally, some other launch assist methods are imaginable:

- refilling consumed propellant during the lift-off phase;
- launching from a high mountain (for example Kilimanjaro, Kenia);
- launching from a vessel, to be flexible in the launch location (“sea-launch”);
- launching from a catapult on a mountain uphill slope;
- providing additional propulsion using rope drives;
- launching from pipes, where exhaust gases generate additional propulsion;
- providing additional propulsion energy via radiation (laser energy).

All alternatives seem to be practicable, but probably these options are not cheaper than the use of booster motors (in every case the cost saving potential is very small).

3.2.2 The Appropriate Propellant

High energetic liquid propellant LH₂/LOX. The combination of liquefied hydrogen as fuel and liquefied oxygen as oxidizer is also called “cryogenic propellant” (referring to the fact that these liquids are gases under normal ambient conditions). The combination LH₂/LOX has a high content of chemical energy per unit of mass; consequently LH₂/LOX rocket engines operate with high specific impulse (exhaust velocities up to 4.5 km/s). The high performance is the greatest advantage of this propellant: space launchers which use LH₂/LOX as main propellant do not need as many stages as launchers which use low energetic propellant (when one “high energetic” stage replaces two “low energetic” stages). Another advantage is that propellant and exhaust gases are non-toxic. The disadvantage is that this propellant has a low density (it requires large and heavy tanks) and cannot be stored for an extended time period (hydrogen gas disappears through seals and tank walls). Launchers have to be tanked shortly before the mission. In case the start is some hours behind schedule, the propellant lost by evaporation has to be refilled. LH₂/LOX cannot be used in upper stages when these stages ignite much later. However, the advantages override the disadvantages: LH₂/LOX is probably the best propellant for space launchers.

Kerosene/LOX. The fuel is the more “problematic” partner in the combination: when we replace hydrogen by kerosene, we get the less explosive and better storable propellant combination kerosene/LOX (which is also often used in space launchers). The lower energy content reduces the exhaust velocity to just 3.0 km/s, typically; but the much higher density allows the construction of more compact rocket stages.

Storable liquid propellant. When we talk about storable liquid propellant, we mean the combination of the fuel hydrazine N_2H_4 (or the hydrazine derivatives MMH or UDMH) and the oxidizer nitrogen tetroxide N_2O_4 . The propellant is hypergol, this means it will ignite spontaneously when the fuel is mixed with the oxidizer. Hydrazine decomposes when it flows over a catalyst bed; rocket stages need no complicated ignition system when they use hydrazine as fuel. Chemical aggressiveness makes the propellant to an unlovely toxic substance. Besides storability the propellant combination hydrazine/nitrogen tetroxide has the advantage of a high density. The density is about three times higher than the density of LH2/LOX propellant: lighter and more compact tanks can be used for the construction of space launchers. The advantages make N_2H_4/N_2O_4 propellant to a real alternative to LH2/LOX propellant, even though its exhaust velocity is much lower (typically just 2.9 km/s).

Solid propellant. When solid propellant is used in military missiles, it has many advantages in comparison with liquid propellant: solid rocket stages are storable for long time periods, they are compact, shock resistant, it is comparatively easy to transport them, and they need a short time of preparation before use. Solid propellant missiles can accelerate extremely fast. When solid propellant is used in space launchers, its application involves mainly disadvantages: the performance is comparatively bad, the structures are heavy, and thrust vector control in direction and magnitude is impossible or at least difficult. The irregular burn behaviour stresses the payload and leads to a low flight precision. Sometimes the thrust is too high and a reduction would be desirable, but such a thrust reduction would lead to a lower specific impulse. When big solid rocket stages are applied, toxic exhaust gases are bad for the environment near the launch site. For space applications the disadvantages override the advantages: we can conclude that solid propulsion should not be used for astronomical activities.

In space transportation there is just one exception where solid motors are sensible: they can be used as small strap-on boosters to assist the lift-off phase. High thrust is required to boost the heavy vehicle; low performance is not a serious problem since the motors are discharged in the early flight phase. When the attitude stabilization is performed by the core stage, the boosters need no thrust vector control system.

The situation is different when big solid boosters are used (or when the booster motors operate without the core stage). Now the solid boosters have to be equipped with thrust vector control. This can either be done by “secondary injection” (influencing the thrust direction by the appropriate injection of a liquid through orifices into the divergent part of the nozzle) or by “nozzle gimbaling” (making a swivel-mounted nozzle with a flexible throat). The modern technique of nozzle gimbaling is better in performance and therefore more often used now. Experience shows that then these big solid booster motors are not cheaper or more reliable than their liquid alternatives.

Sometimes spin-stabilized solid “kick motors” are used as launcher upper stages. Light-weight composite carbon-fiber material for structures gives these stages a comparatively good performance; but inflexibility in the application and bad injection accuracy make the use of solid kick motors in space launchers to a wrong strategy.

3.2.3 Optimal Launcher Size

Launcher families. Sometimes, the weight of the satellite is smaller than the payload capacity of the launcher. Then the launcher flies on a non-optimal trajectory to annihilate its over-performance (or it uses ballast load or it uses less propellant). Compared with a launcher of exactly the right size, the launch costs are higher now. When the satellite is too big, the launcher cannot be used any more, and the launch service competitor loses the business. A “launcher family” offers the option to adapt the performance of the launcher in a way that it matches the size of the satellite.

Booster motors for selection. The performance of a launcher can be upgraded by using strap-on booster motors to assist the lift-off phase. It has to be made sure that the launcher is not over-stressed when many strong boosters are applied. In a similar manner the performance of a launcher can be downgraded, simply by omitting some boosters when they are not necessary for the launch of a small satellite. Now it has to be seen that the launcher is still strong enough to lift-off from the launch pad. If the launcher is too heavy to lift-off without boosters, there is sometimes the option to charge it with a smaller amount of propellant (so-called propellant “off-loading”).

Several compatible upper stages. Another way to adapt the payload capacity of a launcher is using a selection of upper stages with different performances. However, it is important that upper stages with a lower performance are also considerably cheaper. Otherwise the launch service customer has to pay for the launch of a small satellite almost the same price as for the launch of a big satellite.

Multiple launching. In comparison with smaller launchers, bigger launchers have higher absolute costs but lower specific costs. The costs per kilogram payload reduce with the size of a launcher. In a rough approximation we can assume a square root function as a basis: a launcher with double payload capacity is not twice as expensive, it costs just about 1.414 times more. When we use a big space launcher for launching two or more small satellites, we can expect a reduction of the specific launch costs. The cost saving potential of this so-called “multiple launching” is considerable:

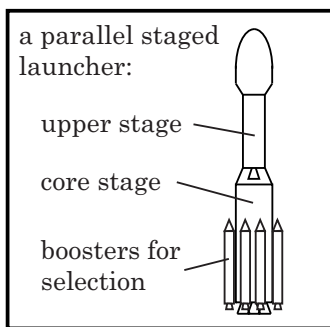
satellites	absolute transportation costs	specific transportation costs
1	100%	100%
2	141%	71%
3	173%	58%
4	200%	50%

Multiple launching is practicable but it includes also some severe disadvantages. First, all satellites have to go to more or less the same orbit; and second, all satellites have to fit into the payload bay. However, all satellites must be ready at the same time. This is the main problem when different launch service customers are involved. If one satellite is missing, the space launcher is in a fatal way too big for the business. Experience shows that double launching is a commercial option; but it does not follow automatically that there is also a market for the launching of three or four satellites.

3.2.4 Flight Safety Aspects

Qualification and quality. Satellites are extremely expensive (a satellite costs two or three times more than the launcher, typically). The customers for satellite launch services have no patience or understanding when fatal launch failures happen often. The design aspect “flight safety” is of high importance to the launch service providers. Launchers undergo extreme certification procedures to guarantee a low failure rate. Usually, the qualification process is so rigorous that the quality of a launcher cannot be increased any more by the application of even stricter qualification procedures; but quality is not only a function of qualification procedures.

Safety by using a parallel staged launcher concept. A high flight safety can also be achieved by system built-in reliability. A parallel staged launcher is considered to be safer than a tandem stage system. The manned US space shuttle is a single stage system supported by two big solid booster motors. First, the liquid main engines of the orbiter are ignited and checked-out. Then, if these motors work properly, the boosters are ignited and the vehicle lifts-off. The start is interrupted when problems in the liquid engines are detected (then these engines are simply switched-off again). Today, many new launcher developments follow the safety strategy demonstrated by the American space shuttle (for example Europe’s Ariane-5 and Japan’s H2). Except for upper stage engines, all engines are ignited before the lift-off takes place. Particularly because of improved flight safety, the single stage launcher supported by strap-on boosters is the best launcher concept. The advantages are:



- the flight safety is greatly improved;
- it is easy to build a launcher family;
- the thrust of all engines is used for the lift-off;
- the body of the vehicle is comparatively short.

The principle involves the following disadvantages:

- the engine checkout consumes propellant before lift-off;
- the nozzles must be adapted to the pressure at sea-level;
- the core stage wastes energy to accelerate the boosters.

The boosters must have a higher thrust acceleration and a shorter burn time than the core stage, otherwise the staging principle is violated. Nozzles for operation at sea-level are shorter than nozzles for operation in the vacuum of space; the shorter nozzles mean a performance reduction for the engines of the core stage. To avoid this disadvantage it is sometimes considered to equip the engines with extensible nozzles, but such a risky and difficult construction has never been used in a space launcher (probably extensible nozzles are more problematic than helpful).

For GTO-launches an upper stage is required, which has usually a better performance because the nozzle of its engine is made for operation in the vacuum of space. The GTO upper stage should not be too small in comparison with the first core stage, and then the first stage should not reach a low earth orbit (it should be suborbital).

Redundant core stage engines. The parallel staging concept assures that all main engines are operative when the US space shuttle lifts-off. If in a later flight phase problems in one of the three main engines are detected, the problematic engine can simply be switch-off. Now the maximum engine thrust is not required any more, and the vehicle can fly with a lower thrust level. The space shuttle is able to continue its mission on another trajectory, where the entire propellant is consumed by the remaining two engines. The safety principle of using redundant engines is also used by other space launchers (for example it was used by the American Saturn-5 launcher, where it was not necessary to terminate the mission in case of an engine problem).

Redundant upper stage engines. We have seen that fatal flight failures by engine ignition problems can be avoided when a parallel stage launcher is used for LEO transportation: all engines are checked-out at lift-off. For the transportation of GEO satellites we need at least one upper stage. One or several engines have to be ignited during the flight (sometimes more than one time, in case of restartable engines). Usually, single engines are used for the propulsion of upper stages (since the gravity losses are small for orbital flight, these upper stage engines are comparatively small). Instead of using only one single engine, we could equip the upper stage alternatively with two or three even smaller engines. This would offer the following advantages:

- the flight safety is improved when one of these engines is redundant;
- now less strict qualification tests are required, this makes the engines cheaper;
- the interstage between core stage and upper stage becomes shorter now.

It is a disadvantage that several small engines are more expensive than a single engine; but without redundancy, an engine malfunction will lead to a total launch failure.

The statistics shows that upper stage ignition problems can lead to fatal flight failures. For the following consideration the European Ariane-4 launcher family serves as an example: its hypergolic second stage never failed to ignite, while year 1991, after 35 successful launches, ignition problems of the cryogenic third stage were responsible for 3 of 5 launch failures. Obviously, the ignition of the third stage is a safety problem. Now we consider to equip the third stage of the launcher with a redundant engine. Since the thrust of one engine alone would suffice for the injection of the satellite, the entire weight of the additional engine is now at the expense of payload capacity. The HM-7B upper stage engine weighs 155 kg and costs about 3 million euros. Thus:

launcher	additional upper stage engine	launch costs	payload
Ariane-40	155 kg (ca.2.9millUS\$2000)	+4.6%	-8.2%
Ariane-44P	155 kg (ca.2.9millUS\$2000)	+4.0%	-5.2%
Ariane-44L	155 kg (ca.2.9millUS\$2000)	+2.6%	-3.7%

In any case, the redundant upper stage engine increases the launch costs and reduces the payload capacity. When the payload is small, the use of a spare engine is expensive and deteriorates the capacity considerably; but the penalty for the additional engine gets smaller when the payload gets bigger (compared with the size of the upper stage); and engine problems will not automatically cause the loss of the expensive satellite.

3.3. Evaluation of Commercial Launch Services

Consider the situation at the beginning of the 21st century: worldwide there are just a few launch service providers which compete in the satellite transportation market; and new competitors from Russia, China and Japan press into the shrinking market. We might expect that the launch service providers use “optimized vehicles” to survive in this scenario of strong competition. Surprisingly, we see many differences when we take a closer look at the construction of some different launch vehicles. These differences have historical reasons and involve operational advantages and disadvantages.

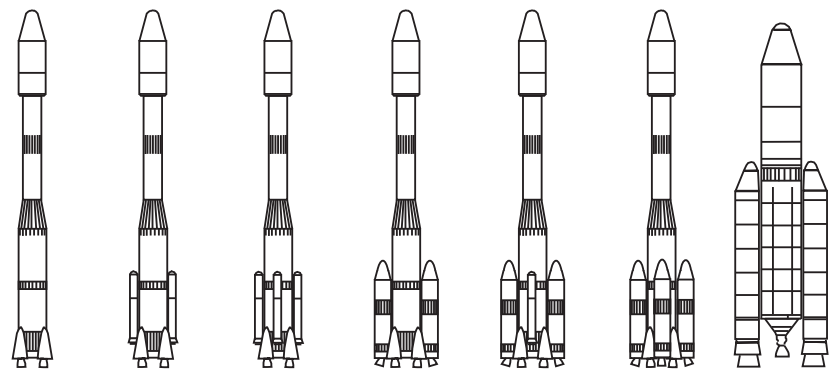
3.3.1 Launch Vehicles

ESA’s Ariane-4 launcher family. The original intention for the development of Europe’s commercially most successful space launcher in 1973 was simply to close the gap in space technology between Europe and the USA. Therefore a quite “conservative” design was selected: first and second stage use N₂O₄/UDMH propellant (essentially with the same type of engine), the high energetic LOX/LH₂ upper stage brings the satellite accurately into the geostationary transfer orbit. Later the Ariane vehicle was completed by solid and liquid strap-on booster motors. The launcher did not close the gap in space technology between Europe and US-America, however, it won more than 50% of the commercial market for GTO-satellite transportation. Three technical reasons are responsible for this success: first, the launch site in Kourou is much better than America’s launch site in Cape Kennedy (it brings 18% more payload); second, the availability of different booster motors allows the composition of a launcher family (flexible in payload capacity); and third, the accurate non-restartable liquid upper stage is optimal for commercial GTO transportation. The production of Ariane-4 was discontinued after the first test flights of its successor (the Ariane-5 launch vehicle); and Ariane-4 performed its last flight in February 2003.

ESA’s Ariane-5 launcher. Already in 1985 Europe planned to replace its launcher by a bigger vehicle, because that time it was expected that the manned space shuttle “Hermes” (another space project which was later dropped) would require a new transport facility. A single stage vehicle supported by two strong solid booster motors was considered as safer and cheaper: safer because it avoids that engines are ignited during the flight, and cheaper because it needs fewer components. The new launcher, named Ariane-5, should use LOX/LH₂ as main propulsion. It was also expected that the vehicle, equipped with an appropriate upper stage, would give better specific costs for GTO satellite transportation when the option of “multiple launching” was used frequently (the simultaneous launch of two or three satellites with one launcher). However, the transportation of a manned vehicle to a low earth orbit is a wrong “nominal mission” for a commercial GTO satellite launcher, and Ariane-5 suffers still from this wrong initial concept: the launcher is too big for commercial satellite launches, the core stage is too large in comparison with the upper stage, and the use of two big solid booster motors is not appropriate for the composition of a launcher family.

Ariane-4	boosters		1.stage	2.stage	3.stage
	PAL	PAP	L-220	L-33	H-10
propellant	UDMH	Solid	N2O4/UDMH		LOX/LH2
stage mass [t]	43.5	12.6	251	38.5	12.1
propellant mass [t]	39.0	9.5	233	35.2	10.8
thrust (sl/v) [kN]	667/737	720/	2708/3034	/785	/62.7
sp.impulse (sl/v)[s]	248/278	241/	248.5/278.4	/293.5	/444.2
diameter [m]	2.17	1.07	3.8	2.6	2.6
height [m]	18.6	11.5	25.4	11.6	9.9

Ariane-5	boosters	core-stage	upper stage
	P-230	H-155	L-7
propellant	Solid	LOX/LH2	MMH/N2O4
stage mass [t]	2 * 265	170	8.12
propellant mass [t]	2 * 230	155	7.2
thrust (sl/v) [kN]	2 * 6360/	800/1000	/27.3
sp.impulse (sl/v)[s]	245/273	/430	/316
diameter [m]	3.03	5.4	5.4
height [m]	30.0	29.0	4.5



Ariane version	40	42P	44P	42L	42P2L	44L	5
lift-off mass [t]	240	339	358	400	420	470	710
GTO payload [t]*	1.9	2.6	3.0	3.2	3.7	4.2	6.8
price [millUS\$1991]	60-70	62-72	65-75	85-95	90-100	110-120	100-110?

* from Kourou

Example 3-1: ESA's Ariane-4 launcher, designed as a starting model for the European space activities, won the commercial market against the competition from America. Three technical reasons were responsible for the success: the vehicle was optimized for satellite transportation to the commercially interesting geostationary transfer orbit; its launch location was better; and a launcher family flexible in price and capacity was available. The successor model Ariane-5, however, is a vehicle that is optimized for the transportation of a manned space glider to low earth orbit; but the vehicle is ill-conditioned for the commercial satellite transportation market: it is too large; the first core stage has not the correct size in comparison with the small upper stage; and the big solid booster motors prohibit the composition of a launcher family. It can be expected that the wrong nominal mission during the preliminary design phase of the new launcher will discontinue the European success in space transportation.

US-America's Delta-II launch vehicle. In 1960 already the Douglas Aircraft Company started with the development of a new launcher from components of military programmes (the Thor ballistic missile and the Vanguard vehicle). In the following years the Delta vehicle became a reliable system with a good flight history, the launcher of the USA that was most often used for the transportation of small GTO satellites. However, the Delta II had too many components (up to nine launch assist motors, just six ignite at lift-off); and the use of a spin-stabilized solid upper stage was certainly a disadvantage for the transport of commercial satellites. The company McDonnell Douglas (later a part of Boeing) continued the development of the Delta launcher family with the construction of two high energetic stages: the second stage was replaced in 1998 (Delta-III) and the first stage in 2002 (Delta-IV). The Delta IV launcher is now a completely new vehicle in the payload class of ESA's Ariane-5 or Martin Marietta's Titan-IV, optimized for the transport of commercial satellites.

US-America's Atlas-IIA launch vehicle. The history of this commercial GTO launcher began with the modification of the Atlas ICBM in 1958 and the incorporation of the Centaur upper stage. Atlas II, America's high technology space launcher, was a good space launcher when we compare the payload mass with the lift-off mass. The first stage and the second stage used the same tank (burnout of the first stage was defined by the instant when the two strong liquid booster engines were dropped); the high energetic Centaur upper stage could be ignited two or even three times to bring a geostationary satellite to its final destination orbit. Small solid strap-on boosters were available to increase the payload capacity (Atlas-IIAS). In 1994 the Atlas-Centaur programme of the company General Dynamics was sold to the competitor Martin Marietta (later Lockheed-Martin) due to a consolidation phase in the US aerospace industry; and it was decided to continue the development to match the competition from Europe. The new Atlas-V launcher is essentially a version with stretched tanks and improved engines, using the high performance RD-180 engine from Russia in the first stage and a single RL-10 engine in the Centaur upper stage.

US-America's launchers Titan IV and Titan II. Also Lockheed Martin's Titan launcher was originally a military missile programme (the development started in 1961 using components of the Titan ICBM). In 1964 a two stage version (Titan-II) brought the Gemini capsule to orbit; the use of two big solid solid booster motors was introduced in 1965 (Titan-IIIC). Later a selection of different upper stages was made available for the launcher (Centaur, Transtage, IUS, PAM); the control of the big solid boosters was changed from secondary injection to gimbaled nozzles; and the Titan-IV vehicle became the biggest commercial launcher of the USA. A four stage version of Titan IV was used for GTO transportation (the core stage ignites after booster burnout). In comparison with Atlas, the launcher represents more the other strategy of cost saving: simply make it bigger than invest into high technology. However, the main applications for the launcher were always the transport of heavy military spy satellites to low earth orbit. The launcher was not at all cheap and just used when its capacity was really needed (finally, in 2002, the company Lockheed Martin decided to stop the production of Titan in favor of the Atlas launch vehicle).

Delta 6925/7925	Castor IV	GEM	Thor-1	Thor-2	Star 48B	
propellant	solid	solid	LOX/RP1	N2O4/Aer.	Solid	
stage mass [t]	9 * 11.7	9 * 13.0	101.7	6.997	11.7	
propellant mass [t]	9 * 10.1	9 * 11.7	96.1	6.076	10.1	
thrust (sl/v) [kN]	9 * 427.1/	9 * 435/	911/	/42.43	/66.44	
sp.impulse (sl/v)[s]	237.3/265.7	245.7/273.8	263.2/295.0	/319.4	/292.6	
diameter-height [m]	1.0-11.2	1.0-13.0	2.44-26.1	2.44-5.97	1.25-2.04	
Atlas IIA /IIAS	Castor IV	Atlas (Sustainer)	3:Centaur			
propellant	solid	LOX/Ker.	LOX/Ker.	LOX/LH2		
stage mass [t]	4 * 11.6	167.1	-	19.0		
propellant mass [t]	4 * 10.1	140.9+ca.15	ca.15	16.7		
thrust (sl/v) [kN]	4 * 433/478	1840+269/	269/	/185		
sp.impulse (sl/v)[s]	229/	262.0/292	220.4/312	/448.9		
diameter-height [m]	1.0-9.12	3.05-24.9	-	3.1-8.7		
Titan IV	SRM	SRMU	Stage-1	Stage-2	Centaur	
propellant	solid	solid	N2O4/Aer.	N2O4/Aer.	LOX/LH2	
stage mass [t]	2 * 317	2 * 350	163	39.6	23.86	
propellant mass [t]	2 * 273	2 * 313	155	35.1	20.32	
thrust (sl/v) [kN]	2 * /7000	2 * /7500	/2410	/462	/146	
sp.impulse (sl/v)[s]	266/271.6	/285.6	/302	/316	/444	
diameter-height [m]	3.11-34.1	3.20-34.3	3.05-26.4	3.05-10.0	4.3-9.0	
launcher	D-6925	D-7925	Atlas IIA	Atlas IIAS	Titan III	Titan IV
boosters	CastorIV	GEM		CastorIV	SRM	SRMU
upper stage	Star 48B	Star 48B	Centaur	Centaur	Transtage	Centaur
lift-off mass [t]	218	230	187	234	680	860
GTO payload [t]	1.45	1.82	2.81	3.49	4.3	(8.6)
LEO payload [t]*	3.99	5.05	6.76	8.39	14.5	21.6
price [millUS\$1991]	45-50	45-50	80-90	110-120	130-150	227
* 200 km circular from Cape Canaveral, Delta and Titan without upper stage						
<p>Example 3-2a: The American launchers Delta-II, Atlas-II and Titan-IV could not compete with the European Ariane-4 launcher family on the commercial GTO market (because of heavy upper stage structures, too many components, and a launch site far away from the equator). However, due to a consolidation phase in the American aerospace industry, the launcher fleet has now been reduced in number and modernized to match the competition from Europe, Japan and China in the 21st century.</p>						

Delta IV	GEM	Stage 1 (common core)	Stage 2 (4 m fairing)	Stage 2 (5 m fairing)
propellant	solid	LOX/LH2	LOX/LH2	LOX/LH2
stage mass [t]	19.3	218.0	23.130	30.840
propellant mass [t]	17.0	200.0	20.407	27.200
thrust (sl/v) [kN]	606/627	2886/	/110.1	/110.1
sp.impulse (sl/v)[s]	/273.8	365/	/462.4	/462.4
diameter-height [m]	1.5-15.2	5.13-36.6	4.0-12.2	5.1-13.7

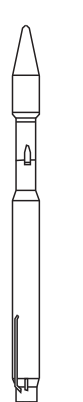
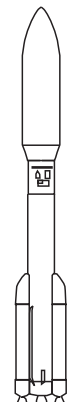
Delta IV version	M	M (5,2)	M (4,2)	M (5,4)	H
boosters	-	2 GEM	2 GEM	4 GEM	2 Core
fairing size [m]	4	5	4	5	5
lift-off mass [t]	246	293	287	335	698
GTO payload [t]	4.2	4.6	5.8	6.5	13.1
LEO payload [t]*	8.12	7.98	10.42	11.47	23.0
price [millUS\$2005]	70	80	90	100	140

* 200 km circular from Cape Canaveral

Example 3-2b: In order to match the competition from Europe, Japan and China in the 21st century, America started new launcher developments. The Delta IV launch vehicle consists of a high energetic core stage (propelled by the strong, newly developed LOX/LH2 engine named RS-68) and a modified Centaur upper stage (propelled by a single RL-10 engine). A selection of 2 or 4 GEM solid booster motors is available to support the lift-off phase. The strongest version of the Delta launcher family uses as its first stage three common core stages in a parallel arrangement.

Global players. At the end of the last century, the American space industry finally responded to the competition from Europe and started the development of new launch vehicles. With their precursors the Delta IV and Atlas V vehicles have actually not much more in common than the names. Both vehicles are now two-stage liquid launchers supported by solid strap-on booster motors. Two upper stages with different fairing size are available for both vehicle types; and all these upper stages are modified Centaur stages which use high energetic propellant. Following the recipe for success of the retired Ariane-4 launcher, a variety of vehicles with different capacity can be composed. It can be expected that the Ariane-5 launcher will not be able to compete with these new, optimized space launcher families on the commercial market.

Atlas V	solid strap-ons	Stage 1 (common core)	Stage 2 (Centaur)	Stage 2 (Centaur)
propellant	HTPB	LOX/RP1	LOX/LH2	LOX/LH2
stage mass [t]	40.8	310.0	18.960	23.220
propellant mass [t]	38.8	284.4	16.780	20.410
thrust (sl/v) [kN]	1134/	3820/	/99.0	2 * 99.0
sp.impulse (sl/v)[s]	275/	311/	/451	/451
diameter-height [m]	1.55-17.7	3.8-32.4	3.0-10.2	3.0-11.7

Atlas V version	401	501	511	521	531	541	551
boosters	0	0	1	2	3	4	5
lift-off mass [t]	334	337	378	419	460	500	541
GTO payload [t]*	5.0	4.1	4.9	6.0	6.9	7.6	8.2
LEO payload [t]*	12.5	10.3	12.05	13.95	17.25	18.75	20.05
price [millUS\$2005]	90	85	90	95	100	105	110

* GTO with only one RL-10 engine; LEO with two RL-10 engines

Example 3-2c: The newest versions of the Atlas launcher family are propelled by a single RD-180 engine (a two chamber derivative of the four chamber RD-170 engine from Russia). Two different payload fairings are available, and there is the option to choose from of 0 to 5 solid strap-on boosters in order to adapt the capacity of the launcher to size and weight of the satellite. LEO-launches are performed by using a stretched version of the Centaur upper stage with two RL-10 engines. The strongest version of the family is not being marketed commercially (the “Atlas V Heavy” vehicle is composed of three first core stages coupled together, like the Delta IV H launcher).

In order to meet a possible competition from states of the former Soviet Union, the American industry founded joint ventures to market the Russian space launchers. In parallel with offering the Atlas vehicle to launch service customers, the company Lockheed-Martin has taken over the marketing of Russia’s Proton launcher (actually a competition to the Atlas vehicle). The company Boeing is the main contractor for the Delta launcher family, but Boeing participates also in the company “Sea Launch”, responsible for the marketing of Ukraine’s Zenit launcher (actually a competition to the Delta launcher family). Even though the adaptation of Soviet technology in the American industry was certainly a reasonable idea, it can be expected that not all of the launch vehicles will survive in a scenario of strong market competition.

Russia's Soyuz launcher. The A-system (Vostok A1, Molniya A2e, Soyuz A2) is the workhorse of the Russian space programme. An early version brought in 1957 the first satellite "Sputnik" into space, later the vehicle evolved into several variants and became with a flight rate of more than 60 launches per year the world's most frequently used space launcher. The yearly flight rate dropped to about 10 to 20 flights after the disintegration of the Soviet Union; but the planning is to continue the use of this reliable launcher in a cooperation between France and Russia (a joined venture Starsem was founded in 1996). The Soyuz vehicle is a two-stage launcher that utilizes LOX/kerosene as propellant; the core stage is supported by four strap-on booster motors with the same propulsion; and a set of different upper stages is available. When we compare the vehicle with western models we see differences in the construction: the liquid engines use four thrust chambers fed by the same turbo pump, it makes the engines more compact without reducing the reliability. The interstage between first stage and upper stage consists of a framework without coating sheets: the engine ignites before the upper stage has been separated from the first core stage.

Russia's Proton launcher. Operational since 1965, the D-system Proton is presently the largest Russian space launcher. The first stage is propelled by six single chamber gimbaled liquid engines; and six fuel tanks are assembled in a way around the central oxidizer tank that the vehicle looks like a core stage supported by strap-on booster motors. However, Proton is a conventional three stage vehicle that uses storable N₂O₄/UDMH propellant; its nominal mission is to bring heavy payloads to low earth orbit. A fourth stage with LOX/kerosene propellant is available for high energetic missions; but the launch location in Tjuratam (Baikonur) is extremely bad for GEO launches (just 14% of the Protons LEO performance reaches GEO). Russia had to face severe financial problems after its turn to democracy, and thus the Russian space agency Glavkosmos offered the experienced Proton launcher at a dumping price to the commercial market (35-70 millUS\$1990). Now an agreement between the USA and Russia regulates the commercialization of Proton in a joined venture of the Russian space agency and the aerospace company Lockheed-Martin.

Ukraine's Zenit launcher. The J-system Zenit was developed in parallel with the Energia heavy launcher (its first stage is more or less identical with Energia's strap-on boosters); and today it is the newest large space launcher of the former Soviet Union, operational since 1985. Zenit has a comparatively small launch history of just about two flights per year. The vehicle is a two stage LEO launcher that uses LOX/kerosene propellant; however, an upper stage is available for GTO launches (a derivative of Proton's fourth stage). After the end of the Soviet Union it was expected that Russia would not use this launcher anymore for launching heavy spy satellites, and the Ukraine started offering the vehicle to the commercial satellite transportation market. Initially, a joined venture offered unsuccessfully launch services with Zenit from Australia's Cape York site (nearer to the equator than Baikonur), but later the company Sea Launch offered successfully launch services from a modified oil platform. The company Sea Launch is a joined venture between the Ukraine (launcher), Russia (upper stage), Norway (launch platform) and Boeing (integration and launch).

Soyuz A2	boosters	stage 2	stage 3	(Molniya)		
propellant	LOX/ker.	LOX/ker.	LOX/ker.	LOX/ker.		
stage mass [t]	4*44.5	105.4	25.2	7.4		
propellant mass [t]	4*39.2	96.4	21.4-22.9	5.5		
thrust (sl/v) [kN]	4*813/992	779/997	/298	/67		
sp.impulse (sl/v)[s]	245/310	265/311	/330	/340		
diameter [m]	2.68	2.15-2.95	2.66	2.4		
height [m]	19.6	28.0	6.7	3.2		
Proton D-1e	stage 1	stage 2	stage 3	stage 4		
propellant	N2O4/UDMH	N2O4/UDMH	N2O4/UDMH	LOX/RP1		
stage mass [t]	455.6	165.6	55.6	19.95		
propellant mass [t]	410.2	150	50	17.3		
thrust (sl/v) [kN]	8844/9810	/2400	/630	/85		
sp.impulse (sl/v)[s]	285/316	/316	/316	/351.8		
diameter [m]	7.4 (4.1+2*1.8)	4.00	4.00	3.5-4.0		
height [m]	20.2	13.7	9.1	5.5		
Zenit J-1	stage 1	stage 2	stage 3			
propellant	LOX/ker.	LOX/ker.	LOX/ker.			
stage mass [t]	352.7	89.9	17.65			
propellant mass [t]	318.8	80.6	15.0			
thrust (sl/v) [kN]	7259/7911	/912	/86.5			
sp.impulse (sl/v)[s]	/337	/350	/351.8			
diameter-height [m]	3.9- 32.9	3.9-11.0	3.7-5.5			
launcher	Kosmos	Tsyklon	Vostok	Proton	Zenit	Energia
stages	2 (C1)	2 (F1) 3 (F2)	4 boosters 2 core	3 (D1) 4 (D1e)	2 or 3	1 core 4 boosters
lift-off mass [t]	120	180-190	290	705	459-466	2400
LEO payload* [t]	1.8	4.0	4.7 (A1)	20 (D1)	13.7	88
GTO payload [t]	-	-	-	5.5 (D1e)	4.3	-
* 200 km circular from Tjuratam						
<p>Example 3-3: After the disintegration of the former Soviet Union, Russia and the Ukraine entered the market for commercial satellite transportation with an immense space experience and an excellent launcher fleet. However, according to their financial situation, these countries continued space activities just at a considerably lower level. Companies in America and Western Europe took over the management for the commercialization of the best Soviet space launchers: Vostok, Proton and Zenit.</p>						

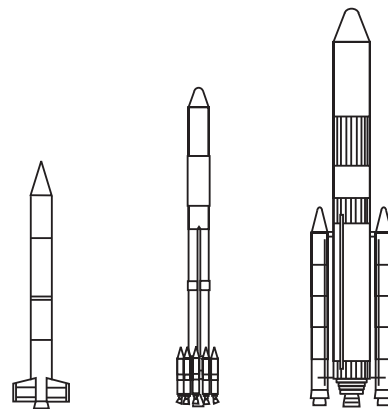
Japan's H1 space launch vehicle. Japan's capability to put payloads into orbit was developed from American technology, starting in 1975 with the N1 launcher which was initially nothing else than licensed versions of the American Delta launcher. An improved version of the N1 launch vehicle, the N2, was used successfully in the time period between 1981 and 1987 with a typical launch rate of two flights per year. However, a commercialization of the launcher was impossible due to the American license contract; and, like the American Delta-II launcher, the N2 was not a very good construction for the launch of satellites: it was a complicated system that consisted of many quite different components (9 small solid strap-on booster motors, an LOX/kerosene first stage, an N2O4/Aerozine second stage, and a solid kick stage). Japan replaced the second stage of the N2 launcher by a self-developed high energetic stage (with LOX/LH2 propellant) and named the successor H1. However, the H1 vehicle was not much better than the N2 vehicle, and it was still commercially restricted due to the license contract. The solution to this problem was the development of a completely new launcher based on the self-developed cryogenic propulsion technology. The H1 was not used anymore after the start of the testing programme for the H2.

Japan's H2 launch vehicle. The first and successful test flight of Japan's indigenous large space launcher took place in 1994. The H2 vehicle is apparently a launcher concept similar to the European Ariane-5 launcher: it is a two stage vehicle that uses LOX/LH2 propellant, supported by two solid strap-on booster motors; and, following the example Ariane-5, the transportation of manned space glider (the later dropped project "Hope") was initially a motive for its development. However, in contrast to the Ariane-5 launcher, the H2 vehicle is not "too large" for the commercial market of satellite transportation. The two solid strap-on booster motors are considerably smaller than the solid booster motors of the Ariane-5 launcher, and the liquid engine of the first core stage is sufficiently strong to push the smaller vehicle. The upper stage has the appropriate size in comparison with the payload mass. The H2 launcher is a much better vehicle than Europe's Ariane-5, and it is optimized for the commercial market. It is comparatively easy to upgrade the performance with more boosters, the use of up to six boosters can bring the launcher into the payload class of Ariane-5. It is also easy to downgrade the performance with the use of smaller boosters. However, the launch site inside Japan (Kagoshima 31.25 North) is bad for the launch of geostationary satellites (in comparison with an equatorial launch site it reduces the capacity by about 23% and makes it necessary that the second stage is re-ignitable).

Japan's M-V launch vehicle. In parallel with the development of liquid propulsion technology from the licensed American Delta launcher, Japan started the development of small solid launchers for scientific applications (the launchers of the so-called Mu-series are not derivatives of a missile programme). The first successful orbital flight took place in 1970. The all solid M-V launcher is the newest version of the Mu-series, operational since 1997; and, with about the double performance than its predecessor, the M3-SII vehicle, the M-V is the biggest all-solid launcher of the world. Three stages are necessary to bring a payload to low earth orbit, an optional fourth stage can be utilized for the injection of small probes into interplanetary space.

M-V	stage 1	stage 2	stage 3	(stage 4)
propellant	solid	solid	solid	solid
stage mass [t]	83.6	34.5	11.0	1.430
propellant mass [t]	71.5	31.1	10.0	1.312
thrust (sl/v) [kN]	/3780	/1240	/290	/52
sp.impulse (sl/v)[s]	/276	/288	/301	/298
diameter [m]	2.5	2.5	2.2	1.2
height [m]	13.8	6.8	3.6	1.5
H1	boosters Castor II	stage 1 Thor	stage 2	stage 3
propellant	solid	LOX/Ker.	LOX/LH2	solid
stage mass [t]	9 * 4.48	85.8	10.6	2.2
propellant mass [t]	9 * 3.73	81.4	8.8	1.84
thrust (sl/v) [kN]	9*220/	775/	/103	/77
sp.impulse (sl/v)[s]	235/282	253/284	/447.8	/291
diameter [m]	0.79	2.44	2.49	1.34
height [m]	7.25	22.9	10.32	2.34
H2	boosters	core-stage 1	core-stage 2	
propellant	solid	LOX/LH2	LOX/LH2	
stage mass [t]	2 * 70.15	98.1	19.7	
propellant mass [t]	2 * 59.2	86.3	16.7	
thrust (sl/v) [kN]	2 * 1560/	843/1080	/121.6	
sp.impulse (sl/v)[s]	/273	/445	/452	
diameter [m]	1.81	4.0	4.0	
height [m]	23.4	29.0	10.9	

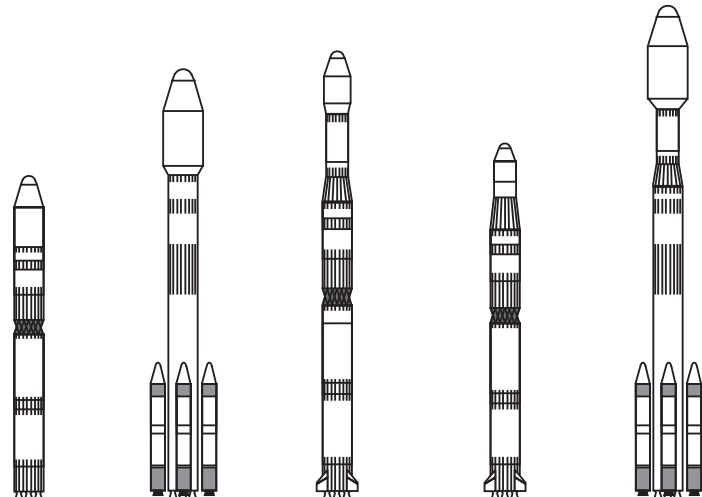
Example 3-4: Japan's high-tech space launchers, the large high energetic H2 vehicle and the smaller all-solid M-V vehicle, are examples for launcher developments mainly stimulated by commercial interest. The H2 vehicle has now replaced the H1 vehicle, which had been restricted on the commercial market because it was actually a derivative of the American Delta-II launcher. The M-V vehicle is the newest version of the Mu-series, it serves for launching low earth orbit satellites and smaller scientific deep space probes.



launcher	M-V	H1	H2
stages	3-4	2-3 core	2 core
	-	9 boosters	2 boosters
lift-off mass [t]	138	140	264
LEO payload [t]	1.8	3.2	10.5
GTO payload [t]	1.2	1.1	4.0
height [m]	30.8	40.3	50.0
price [millUS\$1998]	55-60	-	165-170

China's Long March launcher family. The people's republic of China started space transport activities in 1970 with a small two stage liquid launcher CZ-1, derived from a military missile programme. The bigger vehicle CZ-2 appeared in 1975; CZ-3, a three stage version of the launcher, was used in 1984 for the launch of China's first GTO satellite. Later the vehicles were equipped with liquid strap-on boosters (CZ-2E/HO) and enlarged in all stages to meet the competition from Europe. China has now increased its launch rate from initially once per year to about six per year, experimenting frequently with new variants of the "Long March" launcher. In 2003 a Long March 2F vehicle was used to bring China's first taikonaut to orbit.

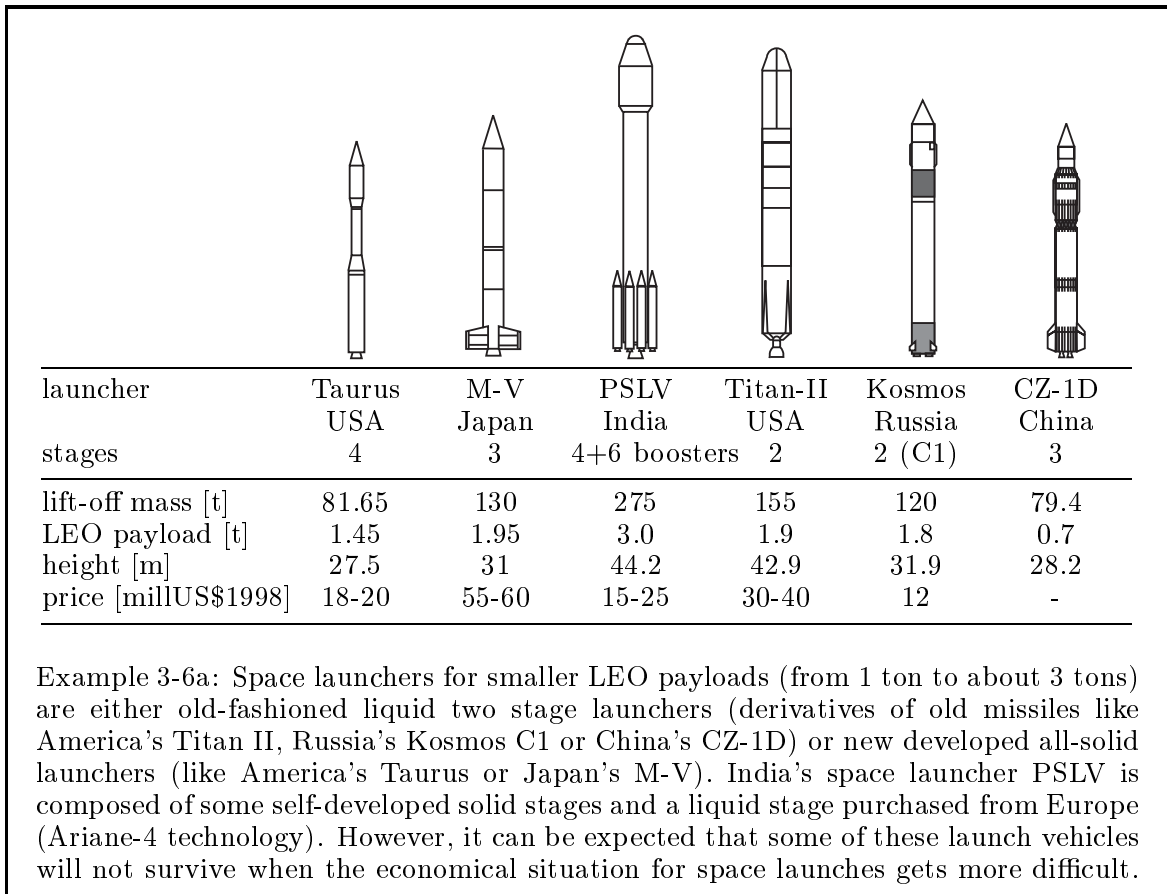
Long March	boosters	1:L140-L180	2:L35	upper stage	
propellant		N2O4/UDMH		LOX/LH2	N2O4/UDMH
stage mass [t]	41	151-197	39	10.5	15.15
propellant mass [t]	38	142-187	35	8.5	14.15
thrust (sl/v) [kN]	741/	2788-2964/	/904	/44.1	/100
sp.impulse (sl/v)[s]	/289	259/290	/296	/425	/295
diameter [m]	2.25	3.35	3.35	2.25	2.9
height [m]	16	20.2-23.7	7.5	7.48	1.92



	CZ-2C	CZ-2E	CZ-3A	CZ-4	CZ-2E/HO
launcher	CZ-2C	CZ-2E	CZ-3A	CZ-4	CZ-2E/HO
boosters	-	4	-	-	4
upper stage	-	(solid)	LOX/LH2	storable	LOX/LH2
lift-off mass [t]	191	464	240	249	474
LEO payload [t]	3.2	9.2	-	4.0	13.6
GTO payload [t]	-	(3.4)	2.5	1.1	4.5
height [m]	35.2	51.2	52.3	42.0	57.8
price [millUS\$1998]	20-25	44-55	45-55	20-30	55

Example 3-5: The Chinese launcher family "Long March" is an example for the try of an under-developed country to get a commercial success in a "high-technology" sector. The launcher fleet is similar to the Ariane-4 family, regarding the size of the stages, the propellant and the composition of the vehicles. However, the launch site in Xichang, 28° North, is worse than Europe's Kourou; and missing trust in the political and economical stability of the country can be an obstacle for the market success.

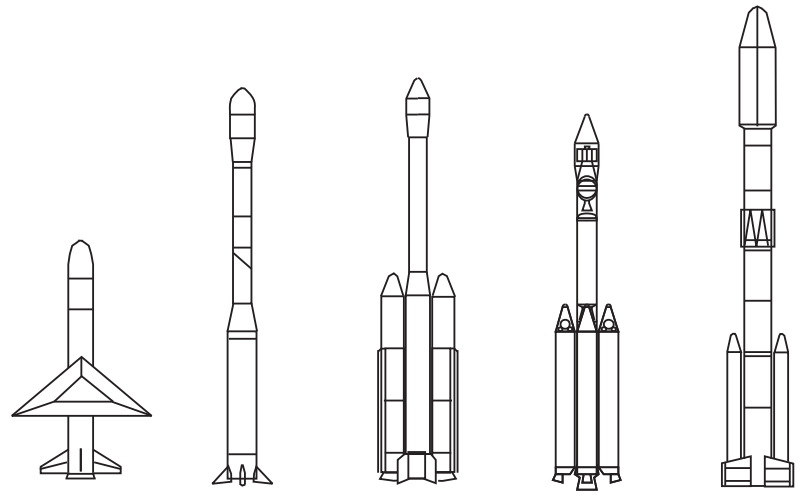
Commercial LEO launch vehicles. Many payloads for a low earth orbit are big military observation satellites in the class between 10 to 20 tons; their transportation requires comparatively large launch vehicles. However, for a space launcher it is much easier to access a low earth orbit than the geostationary transfer orbit, and therefore it cannot surprise us that lot of smaller launchers exist which can only fly to low earth orbit. Even though this space destination is actually of smaller commercial interest, the expectation of future satellite-based, private telecommunication networks has stimulated the hope for a commercial applicability of LEO launchers. After the turn to democracy Russia continued to use its approved Cosmos launcher, a two stage liquid propellant vehicle, operational since 1964. The US company Lockheed-Martin had been offering the Titan-II vehicle without solid strap-on boosters until the production of the Titan-IV stopped. In China a re-development of the CZ-1 launch vehicle was seriously considered at a time when the CZ-1 had already been phased-out. The American Taurus vehicle of the company Orbital Sciences Corporation is a surface launched version of the airborne Pegasus vehicle that uses a first stage of 50 tons instead of an aircraft. The M-V vehicle is Japan’s newest all-solid satellite launcher. Besides the “great space nations” (America, Russia, Europe, Japan and China), particularly India has a reasonable launcher development programme. India’s PSLV vehicle is a development for the satellite transportation market; however, the heavy structures of the solid stages, the bad launch history, and the liquid propulsion purchased from France do not make the marketability of the Indian launcher probable.



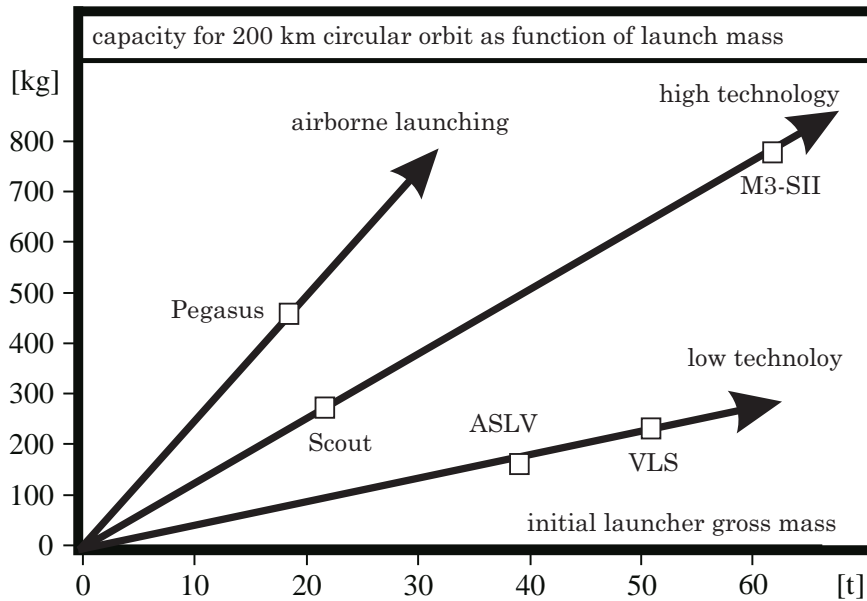
Small LEO launchers. There is also a commercial market for the launch of small payloads from 100 kg up to 1 ton to low earth orbit (so-called LightSats or MicroSats). Some typical all-solid small LEO launch vehicles are listed in the following table:

US Pegasus	(Taurus)	1.-stage	2.-stage	3.-stage	
stage mass [t]	(49.000)	14.020	3.400	0.984	
propellant mass [t]	(44.540)	12.150	3.025	0.782	
thrust (sl/v) [kN]	(/1547)	/486.7	/122.8	/34.56	
sp.impulse (sl/v)[s]	(271/)	/295.3	/295.5	/291.1	
diameter-length [m]	(2.34-8.76)	1.27-9.40	1.27-2.31	0.97-1.47	
US Scout G-1		1:Algol III	2:Castor II	3:Antares II	4:Altair III
stage mass [t]		14.735	4.819	1.637	0.322
propellant mass [t]		12.702	3.718	1.285	0.273
thrust (sl/v) [kN]		414.8/	/267.4	/80.8	/25.8
sp.impulse (sl/v)[s]		229/259	/280	/295	/288
diameter-height [m]		1.14-9.40	0.79-6.31	0.76-3.51	0.51-1.48
India's ASLV	1:AS0	2:AS1	3:AS2	4:AS3	5:AS4
stage mass [t]	2*10.5	ca.11.8	ca.3.7	ca.1.4	0.984
propellant mass [t]	2*8.637	8.9	3.2	1.060	0.317
thrust (sl/v) [kN]	2*422/	/702	/304	/90.7	/35
sp.impulse (sl/v) [s]	/253	/259	/275.9	/277	/281
diameter-height [m]	1.0-11.0	1.0-10.0	0.8-6.4	0.82-2.44	0.66-1.4
Brazil's VLS		1:4*S-43	2:S-43TM	3:Stage	4:Stage
stage mass [t]		35.10	8.82	5.70	0.99
propellant mass [t]		28.90	7.14	4.37	0.82
thrust (sl/v) [kN]		1110	/302	187	30.4
sp.impulse (sl/v) [s]		230/263	/279	270	282
diameter-height [m]		4*1.0-9.0	1.0-9.0	1.0-6.0	1.0-1.7
Japan's M3-SII	B:SB-735	1:M-13	2:M-23	3:M-3B	
stage mass [t]	2*5.0	34.7	13.1	3.590	
propellant mass [t]	2*4.0	27.1	10.4	3.280	
thrust (sl/v) [kN]	2*284/327	114.7/127.5	379/542	/132	
sp.impulse (sl/v) [s]	/266	/266	/282	/294	
diameter-height [m]	0.74-8.25	1.41-14.75	1.41-6.19	1.5-2.68	

However, the market for the transportation of small payloads to low earth orbit is not at all stable, and thus the available launcher models are changing continuously (for example, the table above lists the most representative models for the year 1995, but the Pegasus vehicle is the only vehicle that was still operational in the year 1998). Today several other vehicles are in the state of proposal, development or operation. Russia is offering some of its inter-continental ballistic missiles for satellite launches. Israel's space launcher Shavit is a non-commercial all solid launcher, developed as derivative of the Jericho-2 ballistic missile for small payloads of about 160 kg into circular LEO with retrograde orbit inclination. Europe developed its own launch capacity in this class of carriers already in the middle of the sixties, the French Diamant, but this launcher hasn't been used any more since 1975 (the development of a small Ariane-4 derived launch vehicle was rejected because of high specific costs).



solid launcher	Pegasus	Scout	ASLV	VLS	M3-SII
	USA	USA	India	Brazil	Japan
stages	3	4	5	4	3+2 boosters
lift-off mass [t]	19.0	21.75	39.0	50.8	61.7
payload mass [kg]	455	270	163	230	780
height [m]	15.5	22.9	23.5	19.0	27.8
price [millUS\$1992]	7-12	10-12	12?	10?	29



Example 3-6b: Probably, there remains a small market for the launch of Microsats in the 21 century. “Airborne launching” is better concerning operational flexibility and costs; and probably the future market for the transportation of small satellites will be served by airborne launchers like the US Pegasus vehicle (its payload capacity is about twice as good as payload capacity of surfaced launched rivals with the same size). Some new space countries offer launch services with small solid launchers derived from military missile programmes, but the performance of the vehicles is usually bad.

3.3.2 Launcher Costs

Cost models in space transportation. In the preceding section of this chapter we came to know a selection of space launchers. Some of the vehicles are still in use, the others have been offered at least for some time for commercial satellite transportation. Due to historical and political reasons the space launchers are quite different constructions regarding size, number of stages, type of propellant and so on; and, naturally, these technical differences have an influence on the marketability of the vehicles. A cost/performance analysis makes the quality of a space launcher more transparent.

The costs are certainly one of the most important design factors for a space launcher. Here the term “costs” is considered as a technical parameter of a launch vehicle, the “price” of a launch service is not the same thing. The reason for making a cost model is to make the quality of launch vehicle constructions comparable (for example for the optimization of a new development or for the evaluation of an existing vehicle). The launch prices which are quoted in the documents of the launch service providers describe the amount of money that the provider wants to get for the service (this can include, for example, dumping prices below the real costs or intentional high prices when the service is actually not available yet). We have to pay attention to the fact that the quoted launch prices do not always describe the real costs of the vehicles.

Good and simple cost models for astronautical engineering take just the important technical aspects under consideration which have a noticeable influence on the costs. It follows from experience that complicated models are not more accurate or more reliable. The idea to add up the costs of every small part of a vehicle (every screw, metal sheet, wire and so on) will fail to work for obvious reasons. We also do not use this type of cost model in every day life: when a salesman, for example, plans a business trip, he thinks in advance how much money he will need for the travel. However, he does not consider in advance the price of every telephone call, taxi drive, business diner and so on. The salesman simply counts the days, considers to which country he travels and what he has to do there (maybe he assumes a luxury factor depending on the standards). Then he considers how much money he usually spends for a day in this country (or what other salesmen spend there) and calculates the costs for the trip. A model that considers only important influence factors brings quite realistic results.

Thus, our space launcher cost model will just take important influence factors under consideration: the number and the size of the rocket stages, the thrust of the liquid engines, and the type of propellant that is used. On the easiest way we evaluate the costs for every stage individually and then add up the costs for all the stages. The result in terms of million US dollars should match the quoted launch prices of the often used European and American vehicles (we assume that the quoted prices of the newcomers Russia, China and Japan do not always represent the real production costs for the vehicles). These quoted prices change also with the year of the production: the dollar is subjected to a small inflation (a few percent per annum), but it is also a characteristic aspect for the business that the real launch prices drop every year.

Costs of liquid rocket stages. A square root function of the mass describes approximately the costs of liquid rocket engines (and other equipment in astronautics): an engine with the double mass is about $\sqrt{2}$ times more expensive. The mass of the engine is an exponential function of the thrust; the exponent is approximately 0.75:

$$\begin{aligned} \text{liquid engine production costs [millUS\$2000]} &\approx 0.276 \cdot \sqrt{\text{mass [kg]}} \\ \text{liquid engine mass [kg]} &\approx 26.7 \cdot (\text{thrust [t]})^{0.75} \end{aligned} \quad (3-1)$$

The constant in the equation above depends on the year of production; for example the constant 0.276 describes the production costs in million US\$ for the year 2000. The production costs for the structures can be estimated in a similar manner. Thus:

$$\begin{aligned} \text{structure production costs [millUS\$2000]} &\approx 0.186 \cdot \sqrt{\text{mass [kg]}} \\ \text{structure mass [kg]} &\approx \text{constant} \cdot (\text{propellant volume [m}^3\text{)})^{0.666} \end{aligned} \quad (3-2)$$

for cryogenic propellant: constant = 225, density = 280 [kg/m³];
 for storable propellant: constant = 350, density = 1220 [kg/m³].

Thus, it is assumed that the structure mass is approximately proportional to the surface of the tanks (the constants in the equation above consider statistical data). Finally, the control equipment of the stages has to be taken into account: strap-on boosters without thrust vector control are certainly cheaper than stages with gimbaled engines; it is assumed that they cost approximately 70% of a steerable stage. We get the following simple equation for the cost estimation of a liquid rocket stage:

$$\begin{aligned} \text{costs [millUS\$2000]} &\approx \text{factor} \cdot (1.43 \cdot (\text{stage thrust [t]})^{0.375} \\ &\quad + \text{constant} \cdot (\text{propellant mass [t]})^{0.333}) \end{aligned} \quad (3-3)$$

with: constant = 4.2 for cryogenic propellant
 constant = 3.2 for storable propellant
 factor = 1.0 for stages with thrust vector control
 factor = 0.7 for stages without thrust vector control

Equation (3-3) refers to the entire vacuum thrust, also in the case when several engines propel the stage (then it is not appropriate to give the stage a cost penalty, several smaller engines instead of a single big engine are certainly used to save costs). A cost calculation for liquid stages of the retired European Ariane-4 launcher shows the application of this simple cost model. We get the estimated production costs as:

L220:	Costs= 1.43 · 303. ^{0.375} + 3.2 · 233. ^{0.333}	= 31.8 [millUS\$2000]
L33:	Costs= 1.43 · 80. ^{0.375} + 3.2 · 35.2 ^{0.333}	= 17.8 [millUS\$2000]
H10:	Costs= 1.43 · 6.4 ^{0.375} + 4.2 · 10.8 ^{0.333}	= 12.1 [millUS\$2000]
PAL:	Costs= 0.7 · (1.43 · 74 ^{0.375} + 3.2 · 39.0 ^{0.333})	= 12.6 [millUS\$2000]

The cost model considers only thrust, propellant mass and controllability of a rocket stage; and it makes a distinction between high energetic and storable propellant. The cost model estimates the production costs of liquid rocket stages for commercial space launcher applications (the development costs are about 10 to 50 times higher).

Costs of solid rocket stages. The equation (3-3) for the estimation of the costs of liquid rocket stages cannot be used to estimate the production costs of solid motors: with solid propulsion it is comparatively easy to generate high thrust, but on the other hand the solid propellant is quite expensive. The control of a solid stage is not easy: a solid rocket stage consists sometimes of many small strap-on motors with fix-canted nozzles; sometimes the stage is a big motor with a gimbaled nozzle and sometimes it is a small spin-stabilized upper stage. For the cost estimation it is appropriate to consider just the total amount of propellant and the controllability of the stage (considering strap-on boosters as a single stage). Let us try the following approach:

$$\text{solid stage production costs} \approx \text{constant} \cdot (\text{propellant mass [t]})^{\text{exponent}} \quad (3-4)$$

A price of 10 to 12 million dollars was quoted in 1991 for the experienced American Scout G1 launcher, a price of 7 to 12 million dollars was quoted at the same time for the smaller airborne launcher Pegasus (but in 1998, after the retirement of Scout, a price of 12 to 15 million dollars was quoted for the vehicle). The retired American Titan launcher is actually the only vehicle that has commercial experience with big solid boosters: a comparison of the prices of Titan-II (without boosters) and Titan-IV (with boosters) indicates that big gimbaled solid rocket motors are very expensive. Let us try the model for some solid stages with $\text{constant} = 1$ and $\text{exponent} = 0.750$:

$$\text{costs [millUS\$2000]} \approx \text{factor} \cdot (\text{propellant mass [t]})^{0.750} \quad (3-5)$$

with: factor = 1.0 for stages with thrust vector control

factor = 0.7 for stages without thrust vector control

Scout G1:	Costs=	$12.7^{0.75} + 3.7^{0.75} + 1.3^{0.75} + 0.3^{0.75}$	=	11.0 [millUS\$2000]
Pegasus:	Costs=	$12.1^{0.75} + 3.0^{0.75} + 0.8^{0.75}$	=	9.6 [millUS\$2000]
Taurus:	Costs=	$44.5^{0.75} + 9.6$	=	26.8 [millUS\$2000]
9 Delta-6925 boosters:	Costs=	$0.7 \cdot (9 \cdot 10.1)^{0.75}$	=	20.6 [millUS\$2000]
2 Titan-IV boosters:	Costs=	$(2 \cdot 313)^{0.75}$	=	125.1 [millUS\$2000]
2 US-Shuttle boosters:	Costs=	$(2 \cdot 230)^{0.75}$	=	178.4 [millUS\$2000]
2 Ariane-5 boosters:	Costs=	$(2 \cdot 502)^{0.75}$	=	99.3 [millUS\$2000]

We have to pay attention to the exponent in the equation above: the value 0.70 would also fit for small solid launchers, but the value 0.70 as exponent would determine much lower costs for big solid strap-on boosters. However, the production costs of big solid rocket motors are a critical point in the cost assessment of space launchers: shortly before the termination of the Titan programme the company Lockheed-Martin specified the costs for the Titan-IV vehicle as 350 to 450 million US-dollars, but only as 30 to 40 million dollars for the vehicle Titan-II without big solid boosters. The company Arianespace stated in the year 2002 a purchase price of 100 million euros for the Ariane-5 vehicle, where 45 million euros were on the account of the two big booster motors. There is obviously a considerable discrepancy between these quoted figures. However, we also know that the space shuttle is now in its operational phase much more expensive than it was promised at the time of its development. Our cost model states that small solid stages are cheaper than liquid stages of the same size, but it states also that big gimbaled solid boosters are quite expensive.

Costs of a space launcher. Now we can make use of our cost model and calculate the estimated costs of a launcher, therefore we have simply to add up the costs of all stages as they are determined by the equations (3-3) and (3-5). We can compare the results with the quoted launch prices which we can find in the literature. For example, launch prices are quoted in S.J.Isakowitz [“International Reference Guide to Space Launch Systems”, AIAA, Washington D.C., 1991 and 1999]; Interavia’s “space directory”, A.Wilson [“Jane’s Space Directory”, Jane’s Information Group, London, 1993]:

Price in million US\$	cost model	Isakowitz(1991)	(1999)	Jane’s (1993)
Europe’s launcher family:				
Ariane 40:	61.7	60-70	65-85	40-50
Ariane 42P:	68.1	62-72	70-90	55-65
Ariane 44P:	72.4	65-75	80-100	65-80
Ariane 42L:	86.9	85-95	80-100	90-...
Ariane 42P2L:	93.3	90-100	90-110	...
Ariane 44L:	112.1	110-120	100-125	...
Ariane 5:	138.1	110	150-180	...-110
US launchers:				
Scout:	11.0	10-12	-	10-12
Pegasus:	9.6	7-12	12-15	6?
Taurus:	26.8	-	18-20	-
Delta 6925:	61.8	45-50	-	
Delta 7925:	64.2	45-50	50-60	
Atlas IIA:	64.3	80-90	75-85	60 (Atlas-1)
Atlas IIAS:	75.5	110-120	90-105	
Titan II:	42.6	43	30-40	
Titan IV-Centaur:	185.2	227	350-450	110 (Titan III)
Space-Shuttle:	232.1	245	300	375 (STS-27)
New competitors:				
Proton D1-e:	103.4	35-70	90-98	28
Zenit:	85.5	65-70	75-95 (SL)	
M3-SII:	23.4	29	-	31
M-V:	44.6	38	55-60	-
H1:	49.2	90	-	90
H2:	77.1	100-120	165-170	
CZ-2:	47.2	20	20-25	30 (CZ-2E)
CZ-3:	58.2	33	35-40	25-35

We can see that the cost model represents approximately the quoted launch prices, even though it considers just the number of the stages, their size and performance. The cost model ignores other items like reusability of components, use of several engines in the same stage, type of control system, interstages and adapters, experience of the manufacturers, number of manufacturers, different nationalities and so on. However, also the quoted prices do not always represent the real costs of the vehicles: when a newcomer offers a launch service for “half the price” this does not mean that the vehicle is “twice as good” (for example China’s launcher CZ-3); and when a newcomer quotes high costs for a launch service that is actually not available on the market this does not mean that the vehicle is bad (for example Japan’s launcher H1). Then the model is actually better because it considers costs as a technical parameter.

3.3.3 Launcher Payload

The performance as a quality factor of a launcher. Our intention is to make the quality of different launch vehicles comparable; therefore we have introduced a cost model in the preceding pages of this chapter. Our cost model is not at all accurate, but cost models are usually not accurate: the costs of a space project depend sensitively on many parameters which are never well-defined (capacity of the industry, number of parallel projects, experience of the manufacturers, cost of manpower and so on). However, the performance of a space launcher (its payload capacity for a certain orbit) is a parameter which can be calculated accurately by the appropriate trajectory optimization software. Launch service providers present the capacities of their vehicles usually in the form of diagrams or tables, where the payload mass is a function of the altitude and the inclination of the destination orbit. When we want to make the performances of different space launchers comparable, we have to calibrate the quoted performances and transform the payload capacity to representative reference orbits.

LEO capacity. The low earth orbit LEO is the easiest space destination: a space launcher has the highest payload capacity when it departs from the launch pad in easterly direction to attain a circular low earth orbit. We will take the altitude of 200 km above sea-level as the reference altitude and the geographical latitude of the individual launch site as the reference inclination (eastbound launches). When the capacity of a launcher is specified for an equatorial low earth orbit with another altitude, we can use the following simple model for the transformation of the payload capacity:

$$\Delta v \approx 0.59 \cdot (\text{orbit altitude}[\text{km}] - 200)$$

$$\text{model payload mass/quoted payload mass} = e^{-\Delta v/1500} \quad (3-6)$$

The velocity requirement Δv that is necessary to transfer from one circular low earth orbit to another one is nearly a linear function of the vertical deviation (provided that the difference in the altitude is small, for example less than 500 km); and this velocity requirement can be used to calculate the modeled capacity for the 200 km orbit. Trajectory simulations show that a hypothetical exhaust velocity of 1500 m/s, used in the Ciolcovskij equation (3-6), is an appropriate value that fits for many launchers.

GEO capacity. The geostationary transfer orbit GTO is the terminal for many launch services, but then the destination for the satellite payload is always the geostationary earth orbit GEO. The quality of a launch service is rather a function of the GEO capacity than a function of the GTO capacity, because the inclination of the launch site has a strong influence on the velocity requirement Δv for the maneuver that brings the payload from GTO to GEO. The transfer from GTO to GEO is either performed by the satellite itself, or by a separate kick stage, or by the upper stage of the launcher (the last option requires that the upper stage is re-ignitable). When we want to compare the quality of different launch services we have to decide for one of the three options. For our model we will assume satellite integrated propulsion: then the launcher has the best capacity, and this option is more in common practice.

Then we can use a simple model to find the GEO capacity from the GTO capacity:

$$\Delta v \approx 1477 + 0.45 \cdot (\text{launch site latitude [deg]})^2$$

$$\text{GEO payload mass/GTO payload mass} = e^{-\Delta v/2500} \quad (3 - 7)$$

Within the limited accuracy of the model, the Δv requirement that is necessary to transfer from GTO to GEO is just a function of the inclination of the transfer orbit. The constant 1477 m/s in the equation (3-7) represents the velocity requirement for a coplanar transfer from an elliptical orbit (35775 km \times 200 km) to the geostationary position; the constant 0.45 takes under consideration that the final velocity in GEO amounts 3074 m/s. The inclination of the transfer orbit is the same as the latitude of the launch site (or slightly greater). We assume that the value 2500 m/s in general is appropriate to characterize the exhaust velocity of satellite integrated propulsion.

Launch location loss factor. Finally we have to consider the influence of the geographical latitude of the launch site: a good quality launcher can have a bad performance when it departs from a launch site far away from the equator. We have already seen that the influence is particularly interesting for GTO launch services, but let us first consider LEO services: the earth rotation has for space transportation the positive effect that the launch site has already a velocity with respect to inertial space when the launcher departs from the launch pad. The payload capacity of a space launcher increases when we assume that the vehicle is launched exactly on the equator into an equatorial low earth orbit. Then the following model is approximately valid:

$$\Delta v \approx 0.07 \cdot (\text{launch site latitude [deg]})^2$$

$$\text{equatorial LEO payload mass/LEO payload mass} = e^{\Delta v/1500} \quad (3 - 8)$$

For the launch of geostationary satellites we have to consider the smaller effect of the earth rotation and the larger effect of the inclination change in geostationary position. We have seen that the Δv improvement for a launch exactly on the equator is for both effects approximately a quadratic function of the latitude of the launch site; the factors 0.07 (equation 3-7) and 0.45 (equation 3-8) have to be added:

$$\Delta v \approx 0.52 \cdot (\text{launch site latitude [deg]})^2$$

$$\text{equatorial GEO payload mass/GEO payload mass} = e^{\Delta v/2500} \quad (3 - 9)$$

Again, the exhaust velocity is characterized by the value 1500 m/s for LEO launches and 2500 m/s for GEO launches. The following table shows the results of the model:

equatorial-launch payload improvement for			GEO-launches	LEO-launches
Europe's Kourou	5.23° N,	52.78° W	+0.6%	+0.1%
America's Cape Canaveral	28.5° N,	80.5° W	+18.4%	+3.8%
Russia's Tjuratam	45.6° N,	63.2° E	+54.1%	+10.2%
Japan's Kagoshima	31.25° N,	131.09° E	+22.5%	+4.7%
China's XiChang	27.88° N,	102.3° E	+17.5%	+3.7%
India's Sriharikota	13.78° N,	80.25° E	+4.0%	+0.9%
Brazil's Alcântara	2.28° S,	44.38° W	+0.1%	+0.0%

3.3.4 Launcher Quality Comparison

Specific costs. Let us start the launcher quality comparison with the analysis of the specific costs, the aspect that is most important for the customers of launch services. The table below lists the modeled specific costs for some commercial launch services (the absolute launch costs divided by the payload capacity for LEO and for GEO); but it must be emphasized that neither the cost model (equations 3-3 and 3-5) nor the performance model (3-6 and 3-7) is really accurate. However, the figures for the different vehicles deviate so much that we are at least able to recognize some trends:

Launch vehicle:	payload capacity			modeled costs [millUS\$2000/ton]		
	LEO	GTO	GEO	absolute	LEO	GEO
Europe:						
Ariane40	(4.9)	1.9	1.05	62	(12.6)	62
Ariane42P	(6.1)	2.6	1.43	68	(11.1)	47
Ariane44P	(6.9)	3.0	1.65	72	(10.4)	44
Ariane42L	(7.4)	3.2	1.76	87	(11.8)	49
Ariane42L2P	(8.3)	3.7	2.04	93	(11.2)	46
Ariane44L	(9.6)	4.2	2.31	112	(11.7)	50
Ariane-5 (without L7)	24.7			130	5.8	
Ariane-5 (with L7)		6.2	3.41	138		40
USA:						
Scout	0.270			11	40.7	
Pegasus	0.455			10	22.0	
Taurus	1.45			27	18.6	
Delta 7920	5.050			59	11.7	
Delta 7925		1.82	0.87	64		73.6
Atlas IIA	6.7	2.81	1.34	64	9.6	47.8
Atlas IIAS	8.4	3.49	1.67	75	8.9	44.9
Titan II	1.9			43	22.6	
Titan IV	21.6			170	7.9	
Titan IV Centaur		8.6	4.11	185		45.0
new competitors:						
Proton D1e	20.0	5.5	2.10	103	5.2	49.0
Zenit (Baikonur)	13.7	4.3	1.64	86	6.3	52.4
Zenit (sea-launch)	15.1	4.7	2.53	86	5.7	34.0
M-V	1.95			45	23.1	
H2	10.5	4.0	1.86	77	7.3	41.4
CZ-2	3.2			47	14.7	
CZ-3		2.5	1.20	58		48.3

We can see that the specific costs decrease substantially with the size of the vehicle: a bigger launcher has lower specific costs. Europe's Ariane-5 and Russia's Proton are optimized for the transportation of heavy LEO satellites. We can also see that the market accepts a much higher specific price when a large vehicle is not required for the launch of a small satellite. Europe's Ariane-4 family offered moderate specific launch costs for GEO satellites with a weight between one and two tons. The modeled launch costs indicate that the vehicles of the new competitors from Japan, China and Russia are serious rivals for the established launch services from US-America and Europe.

Flight safety. The reliability is the next very important aspect that decides the quality of a space launcher. Typical reasons for failures are malfunctions of the guidance system, engines that fail to ignite or to restart, and stage separation problems. Consequently, a space launcher should not be composed of too many stages; and in-flight engine ignition should be avoided. A new developed vehicle requires a test phase before the operational phase, and some of the test flights are usually failures. Later the success rate increases because the management learns from the mistakes. The “launch history” characterizes the experience of the launch service providers; for example, the following table lists the launch history for the year 1990:

launch record 1990	first flight	total number	failures	failure rate
Europe's Ariane(1-4)	1979	40	5	12.5%
America's Scout	1960	113	14	12.4%
America's Pegasus	1990	1	0	0.0%
America's Thor-Delta	1960	201	12	6.0%
America's Atlas	1958	245	32	13.1%
America's Titan	1964	172	12	7.0%
US Space Shuttle	1981	38	1	2.6%
Russia's Kosmos	1964	377	6?	1.6%
Russia's Tsyklon	1966	201	2?	1.0%
Russia's A-system	1957	1401	43?	3.1%
Russia's Proton	1965	187	23	12.3%
Russia's Zenit	1985	13	1	7.7%
Russia's Energia	1987	2	0	0.0%
Japan's M series	1966	20	2	10.0%
Japan's (Delta) N1,N2,H1	1975	22	0	0.0%
China's CZ	1970	25	2	8.0%
India's SLV Series	1979	6	4	66.7%

Flight precision. The satellite saves time and propellant when the launcher brings it accurately into the destination orbit; and therefore the flight precision is another important quality factor of a space launcher. The following table lists the data specified by some GTO launch service providers for their launch vehicles:

GTO accuracy (3σ)	upper stage	pericenter	apocenter	inclination
Ariane-4 (H10)	LOX/LH2	200±1.0 km	35786±52 km	7.0±0.018°
Delta 6925	solid	200±6.0 km	35775±1100 km	28.5±0.6°
Atlas-Centaur	LOX/LH2	167±1.9 km	35912±52 km	26.4±0.020°
Proton	LOX/ker.	5500±400 km	35786±400 km	0-25±0.5°
Zenit (Baikonur)	LOX/ker.	200 km	35775±300 km	51.0±0.7°
Zenit (Sea-Launch)	LOX/ker.	200 km±10 km	35775±80 km	0±0.2°
H2	LOX/LH2	250 km±10 km	36225±250 km	28.5±0.08°
CZ-3	LOX/LH2	200±1.1 km	35786±94 km	31.1±0.07°

The final stage that delivers the satellite into the target orbit should be a high precision liquid propulsion upper stage; a spin stabilized solid kick stage is not accurate.

Environment for the payload. Noise, shock loads, vibrations, axial and lateral accelerations during the launch phase mean stress to the satellite; and therefore a moderate payload environment is another important quality factor for a space launcher. The following table lists some data provided by the launch service providers:

payload load factors	axial/lateral	shock loads [g]	sound level [dB]
Europe's Ariane(1-4)	4.5 g /0.2 g	2000 g (1500-4000 Hz)	142
America's Pegasus	8.5 g /3.5 g	325 g (500-5000 Hz)	133.5
America's Delta 7925	6.0 g /2.0 g	4100 g (1500 Hz)	139-144
America's Atlas	6.0 g /2.0 g	2000 g (1500 Hz)	137-138
America's Titan IV	6.5 g /1.5 g	2000 g (5000 Hz)	139.3
US Space Shuttle	3.2 g /2.5 g	5500 g (4000 Hz)	140
Russia's Proton	6.0 g /3.0 g	2500 g (1500-5000 Hz)	144
Russia's Zenit	6.0 g /0.6 g	500 g (1000-2000 Hz)	140
Japan's H1 (3 stages)	9.2 g /2.0 g	2000 g (1500-4000 Hz)	141
Japan's H2	5.0 g /2.0 g	2000 g (800-2500 Hz)	141
China's Long March	5.8 g /0.8 g	2000 g (1500-2000 Hz)	137-142

The launch contract. When an operator of commercial satellites is going to install a new transponder he is faced with the problem to select a launch service provider. The technical aspects (costs, flight safety, accuracy and payload environment) are certainly very important; however, also "non-technical" points have to be considered. Some parts of the contract are negotiable between the provider of a launch service and the customer, depending on the actual political and economical situation.

A very important aspect is the time interval between the signing of the contract and the actual launch date. Satellites are extraordinarily expensive; and the capital costs (interest rates) are so high that even hours count. It is an important question how many days before the actual launch date the satellite must be delivered to the launch service provider; and whether it is possible to access the satellite after the delivery. Established providers offer usually short time intervals between contract and launch; but sometimes the launcher is "booked-out" or the annual launch rate is limited. However, over-capabilities can also lead to "last minute" flight opportunities.

It is also important how flexible a launch contract is, regarding changes according to the wishes of the customer. It can be the case that the satellite is not ready at the right time, or that it is at the launch date bigger than it was originally planned (therefore a launcher family with a variety of payload shrouds should be available).

Finally, it should also be negotiated what happens in case of unexpected problems. Another flight free of charge can be offered in case of a fatal flight failure. A "money return guarantee" can be negotiated for case that, due to political reasons, the launch vehicle is suddenly not available anymore (a revolution can happen in an under-developed country; an environmental protection group can create problems and prohibit launches; a strong national lobby can consider the foreign satellite as an unwanted competition; military can provide obstacles for the launch and so on).

3.4. Optimization of Space Launchers

The development of a new space launcher is extremely expensive, and therefore a new space launcher is very often not a completely new development, but it has to utilize existing components. The optimization of a launch vehicle in a preliminary design phase has to take available engines and rocket stages under consideration, even when these components do not fit perfectly for the new application. This section demonstrates the technique of space launcher optimization; the situation that was present for the European Ariane launcher family in the year 1992 is used as an example.

3.4.1 The Concept Ariane-5 Commercial

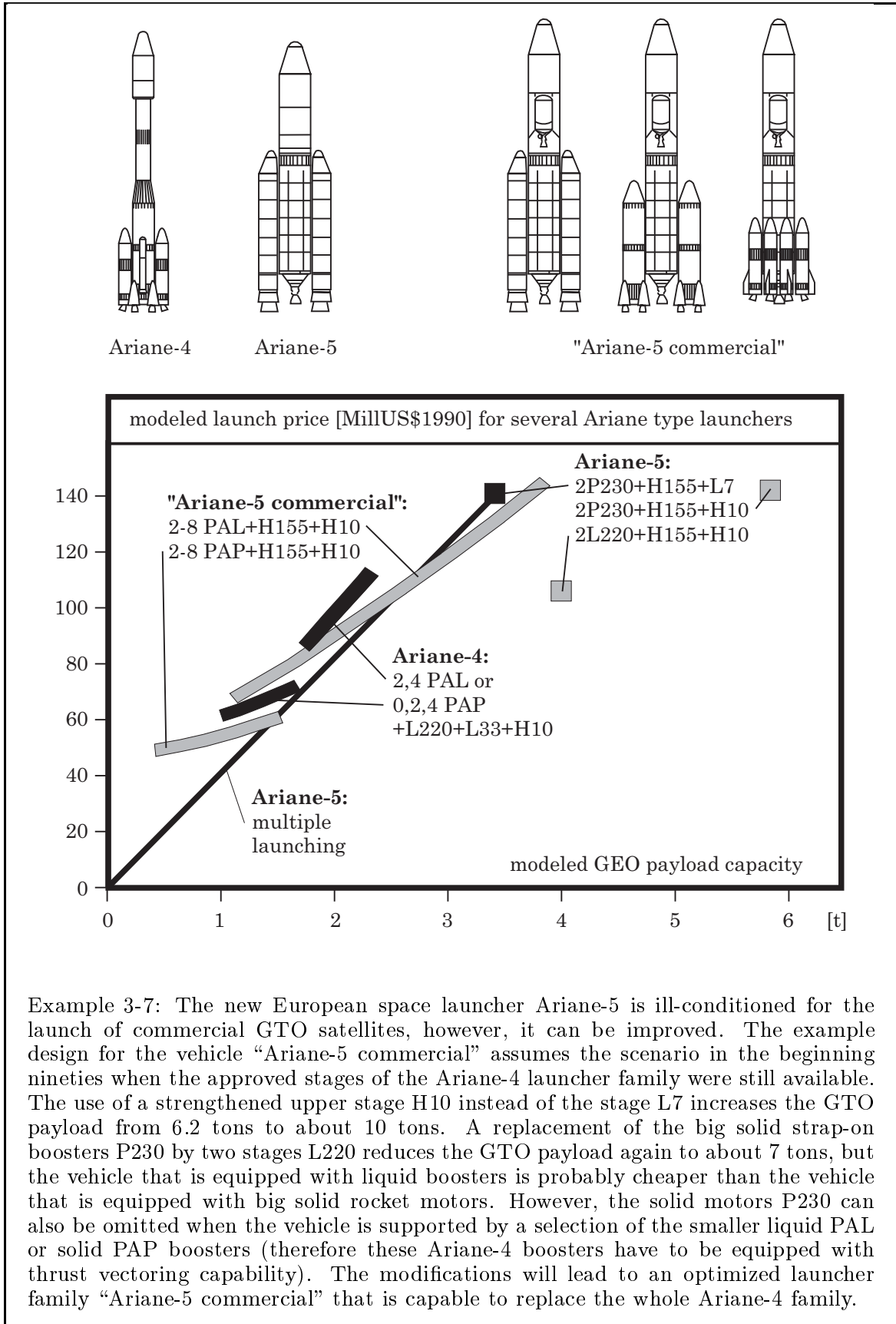
The Ariane-5 design concept. Let us go back to Europe in the beginning nineties when the Ariane-4 was the most successful launcher on the commercial market. That time it was expected that in the future a scenario of stronger market competition would prevail in the business of commercial satellite launch services. However, the decision to replace the Ariane-4 vehicle by the new development Ariane-5 had been based on the expectation that the new launcher would be cheaper and better; and Ariane-4 was phased out before the operational Ariane-5 had proved superiority.

It was planned to terminate the production of the Ariane-4 stages early in order to liberate the capacity for the production of Ariane-5. However, Ariane-5 is a typical LEO launcher, it is ill-conditioned for the commercial market of launching GTO satellites: the nominal mission in the preliminary design phase was the transportation of a big manned space glider to low earth orbit (20 tons Hermes), therefore a high energetic single stage system was selected, equipped with only one liquid engine but supported by two big solid booster motors. All engines ignite before lift-off, first the engine of the core stage and then, when the core stage engine works normally, the solid boosters. Since the solid boosters are strong and liquid engines are expensive, the core stage was equipped with a comparatively weak engine (named “Vulcain”). Quite good as far as it goes, but then the project Hermes was dropped and Ariane-5 got a new nominal mission: the transportation of commercial GTO satellites. Therefore the vehicle required an upper stage. Since it was anyway “too big” for the new nominal mission, the capacity was downgraded with an upper stage (“L7”) that was too small in comparison with the rest of the vehicle (storable propellant was selected for the upper stage in order to avoid that the stage has to be tanked shortly before lift-off). To avoid that the large core stage reaches an orbit during the mission, an unusual flight maneuver became necessary to bring the stage to a controlled re-entry in front of the coast of Peru. However, the most serious disadvantage in the launcher concept was the use of two big solid strap on boosters. Even when these motors are cheaper and more reliable than liquid rocket stages of the same size, they prevent the composition of an Ariane-5 launcher family (flexible in payload capacity); but probably it is not true that big gimbaled solid motors are cheap and reliable: the experience is not at all promising (America’s Titan-IV and the Space Shuttle).

Improvement of the Ariane-5. Let us now think about how we can improve the ill-conditioned Ariane-5 launcher concept using components of the Ariane-4 vehicle. A replacement of the L7 upper stage by the H10 upper stage is comparatively easy: it requires small structural modifications (essentially a thickening of the hydrogen tank, let us assume 10% more structure weight). Trajectory simulations show that the core stage is now sub-orbital and performs a re-entry over the Indian Ocean without special trajectory restrictions; and the GTO capacity increases from 6.2 tons to 10.6 tons.

However, with a capacity of 6 tons the Ariane-5 launcher was already too big; and a capacity of more than 10 tons would make simultaneous launches of two or even three satellites necessary all the time. In order to reduce the capacity for the launch of a single satellite we have to remove the big solid strap-on booster motors (“P230”). For example, we can try to replace the P230 motors by two L220 stages. The “L220” stage is the first stage of Ariane-4, it has about the same size as the solid boosters P230. Our cost model states that a L220 stage is cheaper than a P230 motor (32 million US\$ instead of 49 million US\$). Since the L220 stage is weaker than a P230 solid motor, we have to remove propellant from the core stage (so-called propellant “off-loading”). When we remove 30 tons propellant from the H155 core stage and simulate the trajectory of the vehicle, we find a GTO payload capacity of 7.4 tons. The vehicle with liquid strap-ons has approximately the same specific costs as the vehicle with solid strap-ons (≈ 13 million US\$/ton GTO); and, assuming that the cost model is correct, the vehicle is better because it has a smaller payload capacity.

However, the capacity of the vehicle is with 7.4 tons still too high for the launch of a single GTO satellite; and the use of two big strap-on booster motors prevents the composition of a launcher family. The solution to the problem is to replace the two big boosters by a selection of many smaller solid or liquid strap-on booster motors (the solid boosters PAP or the liquid boosters PAL of Ariane-4). The new vehicle is then actually an Ariane-4 vehicle, where the single high energetic stage H155 of Ariane-5 replaces the first stage L220 and the second stage L33. The larger diameter of the H155 stage allows the use of upto eight solid or liquid strap-on motors; however, the use of the small boosters confronts us with three technical problems: first, we have to introduce structural modifications to mount the smaller boosters to the stage H155. Second, the small boosters are not equipped with thrust vector control: either the core stage has to take over attitude stabilization and flight control alone, or the booster have to be equipped with thrust vectoring capability (a comparison with America’s Delta-II launcher, where the core stage with 95 tons thrust is able to control six fixed canted boosters with all together 300 tons thrust, indicates that upto four boosters can be used without thrust vectoring capability). The third problem is that the fully tanked vehicle is “too heavy” to depart from the launch pad: we have to remove a certain amount of propellant from the core stage, depending on the number of booster motors. The low lift-off thrust leads to lucrative conditions for the flight control subsystem, which are also necessary when the thrust vector control system of the core stage alone takes over guidance and stabilization. We can calculate the capacity of the vehicle with a trajectory optimization program.



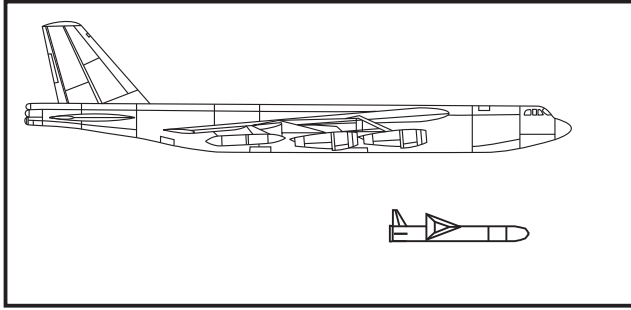
The following table compares the performance of our hypothetical Ariane-5 version (“Ariane-5 commercial”) that uses a variable number of solid PAP or liquid PAL strap-on boosters with the original Ariane-5 vehicle that uses two big boosters P230 (the cost model was used to estimate the specific launch costs):

Launch vehicle:	payload capacity			modeled costs [millUS\$2000/ton]		
	LEO	GTO	GEO	absolute	LEO	GEO
Ariane-5:						
2P230+H155	24.7			130	5.8	
2P230+H155+L7		6.2	3.41	138		40
2P230+H155+H10		10.6	5.84	142		25
2(L220-30)+H155+H10	15.2	7.4	4.04	106	6.9	26
“Ariane 5 commercial”:						
2PAP+(H155-82)+H10		0.81	0.45	49.1		110.3
3PAP+(H155-68)+H10		1.35	0.74	51.3		69.3
4PAP+(H155-59)+H10		1.71	0.94	53.4		56.8
5PAP+(H155-51)+H10		2.01	1.10	55.4		50.4
6PAP+(H155-44)+H10		2.29	1.26	57.2		45.4
7PAP+(H155-42)+H10		2.52	1.38	59.0		42.6
8PAP+(H155-37)+H10		2.76	1.52	60.7		39.9
2PAL+(H155-49)+H10	4.7	2.05	1.13	68.1	14.5	60.3
3PAL+(H155-36)+H10	6.5	3.00	1.65	80.8	12.5	48.9
4PAL+(H155-27)+H10	8.1	3.84	2.11	93.5	11.5	44.3
5PAL+(H155-20)+H10	9.8	4.68	2.57	106.2	10.9	41.3
6PAL+(H155-14)+H10	11.4	5.47	3.01	118.9	10.5	39.5
7PAL+(H155-8)+H10	12.9	6.24	3.43	131.6	10.2	38.4
8PAL+(H155-2)+H10	14.4	6.97	3.83	144.3	10.1	37.7
Ariane4:						
Ariane40	(4.9)	1.9	1.05	62	(12.6)	62
Ariane42P	(6.1)	2.6	1.43	68	(11.1)	47
Ariane44P	(6.9)	3.0	1.65	72	(10.4)	44
Ariane42L	(7.4)	3.2	1.76	87	(11.8)	49
Ariane42L2P	(8.3)	3.7	2.04	93	(11.2)	46
Ariane44L	(9.6)	4.2	2.31	112	(11.7)	50

We can see that the new vehicle “Ariane-5 commercial” supported by a selection of PAP or PAL boosters has enough capacity to replace the old Ariane-4 launcher family: it covers the entire range of payloads and offers comparable or even better costs. Assuming that the cost model is correct, the H155 stage equipped with a single Vulcain engine is cheaper than the L220 stage and the L33 stage with all-together five Viking engines. The strongest version of the “Ariane-5 commercial” is supported by 8 liquid PAL boosters; this vehicle has a GTO capacity of approximately 7 tons, offers moderate launch costs and is capable to replace Ariane-5. However, also with these modifications the vehicle is still not the best vehicle for launching GTO satellites, because the core stage H155 is “too heavy”, or the Vulcain engine is “too weak” (propellant off-loading is always necessary). To construct the best GTO launcher, Europe’s space industry will have to develop a stronger version of the Vulcain engine.

3.4.2 The Airborne Ariane-4 Launcher

The benefits of airborne launching. The first flight of the American Pegasus launcher in 1990 proved the feasibility of airborne launching (the small winged Pegasus vehicle used a B-52 bomber as launch platform). The great advantage is that the airborne launcher is independent of a launch infrastructure in an equatorial region of the earth; and another great advantage is that the airborne launched vehicle has about the double payload capacity than a surface launched vehicle of the same size. These two advantages override obviously the only disadvantage that a big (and expensive) jet aeroplane is necessary for airborne launching.



These two advantages override obviously the only disadvantage that a big (and expensive) jet aeroplane is necessary for airborne launching.

An airborne launcher composed of Ariane-4 stages. Let us go back again to Europe in the year 1992 when the stages of the commercially successful Ariane-4 launch vehicle were still in production. At that time a LEO launch capacity was more or less not available in Europe; and the smallest launcher of this family, Ariane-40, had already a GTO capacity of 1.9 tons. The launch of smaller GTO satellites was just possible with the method of “double launching”, a technique which is sometimes not appropriate for a launch service customer. Let us now consider how we can compose a small airborne Ariane launcher from components of the approved Ariane-4 stages.

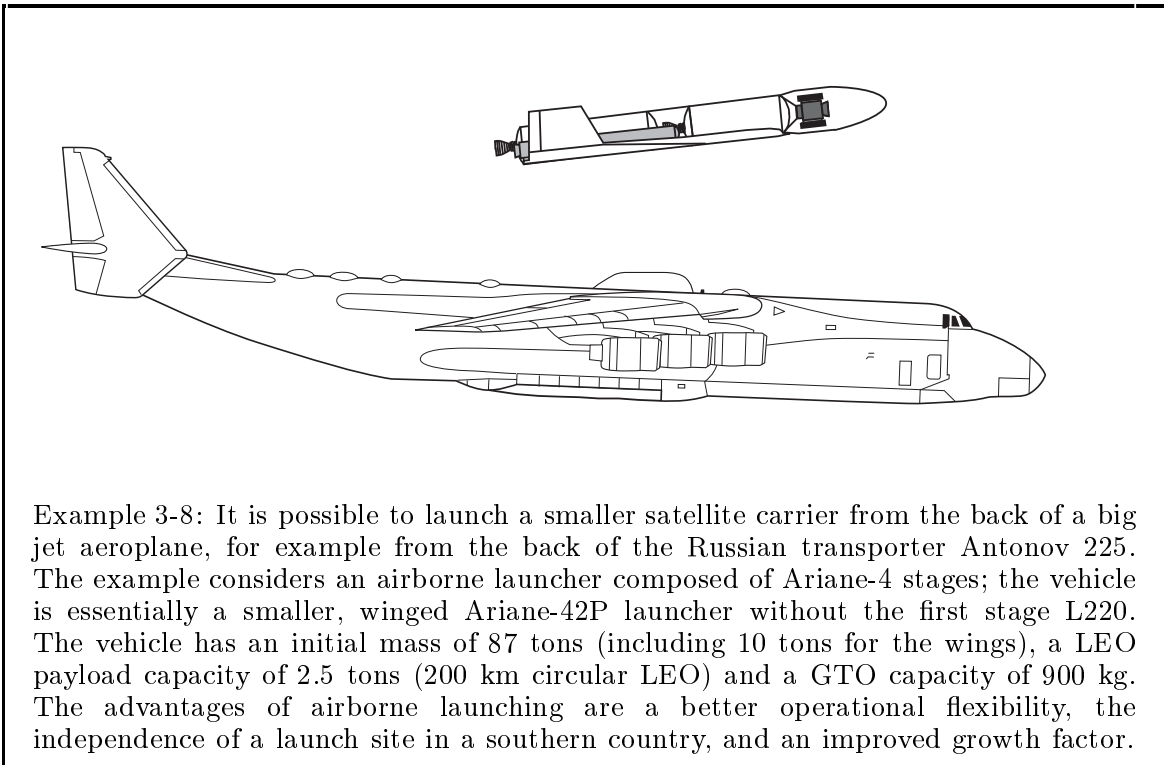
Four components of the Ariane-4 launcher are usable for airborne launching: PAP, PAL, L33 and H10. The first stage L220 can be ignored, since it is simply too heavy. The boosters PAP and PAL are constructed for ground level ignition; we can extend the nozzles for the ignition at a higher altitude to give them a better specific impulse. The stage L33 cannot be ignited at 11 km altitude (at the cruising altitude of jet airplanes); but the ignition should not present a problem at an altitude of about 18 km when the stage is used as a second stage of an airborne launcher. Both boosters PAP and PAL are not equipped with thrust vector control: wings are necessary for the aerodynamic stabilization and for the initial pull-up maneuver. It is the best that we integrate the wings in the first stage and jettison them together with the empty stage. The stages cannot stand high lateral accelerations, therefore we have to limit the the lift force during the initial “pull-up” maneuver. The carrier aircraft must be a big (civil or military) jet aeroplane; it must have sufficient capacity to carry the external load and it must have some capacity left over for cruising and for mission control. The airplane carries the space launcher on its back; the separation of the launcher from the airplane is performed when the airplane flies with normal cruising conditions (for example Mach 0.8 at 11 km altitude). When the airplane arrives at the appropriate launch location above the equator, it separates the launcher and dives away from it; then, after some seconds, the launcher ignites its rocket motor.

Four different airborne launcher concepts. Let us now consider how we can construct an airborne launcher from stages of the Ariane-4 launcher. Four different assemblies seem to be sensible: the first concept, denoted as “L33+H10”, is a two stage launcher composed of the stages L33 and H10. The other concepts are three stage launchers, because the ignition of the L33 stage takes place after burnout of the booster motors. The second concept “2PAP+L33+H10” uses two solid boosters as a first stage; the third concept “4PAP+L33+H10” uses four solid boosters; and the fourth concept “2PAL+L33+H10” replaces the solid boosters by two liquid boosters.

To calculate the ascent trajectories of the four launcher concepts we assume the following conditions: the separation from the aircraft is performed when the configuration is flying above the equator exactly into easterly direction at an altitude of 11 km and with a speed of 270 m/s (Mach 0.8); the climb angle is 3° . Immediately after the separation the launcher is flying with a constant angle of attack of 10° ; a short time interval later the lift limitation becomes effective to limit the lift force for the pull-up maneuver. We assume that the maximum lift force is proportional to the initial weight of the launcher concept: a launcher that is heavier needs also heavier wings. These wings weigh 13% of the initial mass of a launcher concept; they generate a lift force that may not exceed the value of the initial weight multiplied by the factor 1.75. The wings are dropped together with the first stage; their construction includes also some additional structure to protect the stages L33 and H10 against aerodynamic pressure and heating. We have to bear in mind that the stages of the Ariane-4 launcher are not constructed to stand high lateral forces. It is necessary that the wings introduce the lift force carefully into the structures of the stages, otherwise the lift force would break the stages. Thus, the wings are not only fins, they carry and protect the whole launcher. Nearly no lateral acceleration is allowed anymore after the separation of wings; therefore we restrict the product of the aerodynamic pressure and the angle of attack to the small value of 500 N/m^2 (the angle of attack is very small during the flight phase of the second stage). Finally, we assume a fairing of 1 ton to protect the payload against the aerodynamic loads, the fairing is dropped at L33 burnout. When we simulate the ascent trajectories of the four different assemblies with a trajectory optimization program we find the payload capacities for LEO services and for GTO services. The following table lists the computed results:

		H10	H10	H10
third stage				
second stage	H10	L33	L33	L33
first stage	L33	2PAP	4PAP	2PAL
total stage mass [t]	50.15	75.27	100.39	137.7
wing mass [t]	6.52	9.88	13.05	18.00
fairing mass [t]	1.00	1.00	1.00	1.00
payload GTO (130 km * 35775 km) [t]	-	0.905	1.226	1.515
payload LEO (200 km circular) [t]	0.897	2.461	-	-
total launcher mass [t]	58.56	87.05	115.66	158.2
GEO-payload [t]	-	0.49	0.67	0.84
modeled costs [millUS\$2000]	30.0	36.4	40.7	55.4
GEO specific costs [millUS\$2000/ton]	-	73.5	60.1	66.3

The GTO payload vanishes for the concept that we called L33+H10; this means that the two stage vehicle can only be used for the transportation to low earth orbit. This concept is anyway not very promising because it is not possible to ignite the engine of the stage L33 at a flight altitude of 11 km (or we would have to shorten the nozzle of the Viking engine). The concept 2PAP+L33+H10, with two solid booster motors integrated into the wings, is much better. The booster motors propel the vehicle during the phase of the initial pull-up maneuver; the first winged stage is separated at an altitude of 18 km where an ignition of the engine of the L33 stage is possible without a modification of the nozzle. Naturally, it is also possible to equip the first winged stage with remote control and design it as a reusable part of the space transport system. The initial weight of the configuration 2PAP+L33+H10 is with 87 tons certainly within the limits of the lift capacity of big aeroplanes; and its payload capacity fits in the interesting segment of the commercial market of launch services. A small airborne launched Ariane vehicle can serve for launching satellites which are too small for the other members of the Ariane-4 family. Launched exactly above the equator on a position with the appropriate degree of longitude, the small airborne launcher has the remarkable GTO capacity of 900 kg. The payload increases to 1.22 tons when we double the size of the solid boosters (or when we use four solid boosters, the concept 4PAP+L33+H10). However, we get an ill-conditioned launcher when we replace the two solid boosters by just one liquid booster: the assembly becomes longer and more fragile, and one Viking engine alone is too weak to push the heavy vehicle. When we use liquid boosters the right number is two: concept 2PAL+L33+H10 gets the considerable GTO payload of 1.5 tons, however, now the initial weight is with about 160 tons already at critical limits.



Example 3-8: It is possible to launch a smaller satellite carrier from the back of a big jet aeroplane, for example from the back of the Russian transporter Antonov 225. The example considers an airborne launcher composed of Ariane-4 stages; the vehicle is essentially a smaller, winged Ariane-42P launcher without the first stage L220. The vehicle has an initial mass of 87 tons (including 10 tons for the wings), a LEO payload capacity of 2.5 tons (200 km circular LEO) and a GTO capacity of 900 kg. The advantages of airborne launching are a better operational flexibility, the independence of a launch site in a southern country, and an improved growth factor.

3.5. The Promise of Modern Technology

US-America developed the reusable space-shuttle launch system in succession to the expensive Apollo project. That time people thought that the concept of reusability would make space transport cheaper and safer likewise; but today the experience with the American space shuttle spoils that view. Now people search for new ways to make space transport cheaper and safer: novel technology “aerospace planes” and ballistic “single stage to orbit” vehicles are proposed. It is expected that “air-breathing” engines make such a fantastic performance possible. Apparently, the concepts seem to have their origin rather in science-fiction novels than in engineering handbooks.

3.5.1 The Concept of Reusability

Reusable space launchers. When we think it over, we realize that there are serious arguments against reusability in space transportation. The reusable system has the following disadvantages in comparison with an expendable system of the same size:

- it has a considerably smaller payload capacity;
- the costs for its development are much higher;
- there are maintenance and repair costs;
- it is much more complex due to the necessity of a return flight;
- and a catastrophic flight failure means a much larger loss.

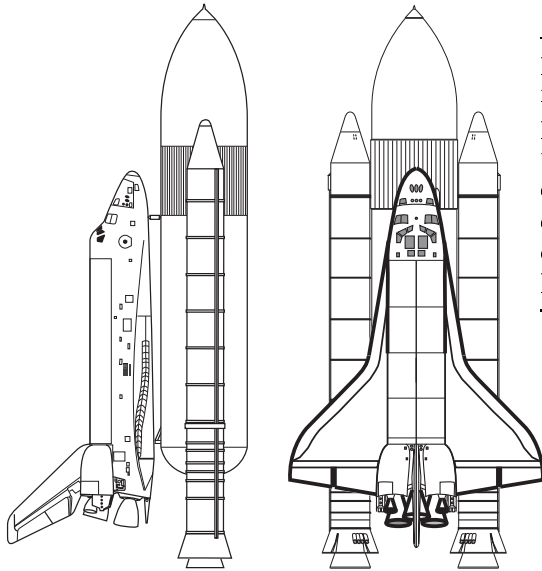
To override all these disadvantages, reusability offers actually just one advantage:

- the reusable system is produced only once for many flights.

At least when we consider the transportation of satellites, the concept of reusability is a false doctrine. It involves high risks and offers just a small cost saving potential. Every flight of the less efficient reusable system is charged with new financial burdens. We save the production costs, but now we have depreciation costs, costs for returning the high investment of the expensive development, costs for recovery, repair, and check-out, and sometimes also costs for the astronauts in case the vehicle needs a crew. A reusable space shuttle system is just sensible for a huge national space programme (for example, when the logistics of a space station needs a manned return vehicle). The reusable shuttle with its special performance and flexibility is an ideal tool for manned space activities; but it makes no sense to develop an expensive space shuttle and use it just a few times every year for the transport of some conventional satellites.

Reusable components of the American space transportation system STS are the winged cabin (the shuttle orbiter itself), the three liquid main engines (they are recovered with the orbiter), and the two big solid booster motors (they are captured by a special ship after parachute watering). The only expendable component of the American shuttle system is the huge external LH₂/LOX tank. The retired Russian space shuttle looks similar but is a quite different system. Strictly speaking, Russia’s shuttle is not a reusable space transportation system: it is composed of the expendable launcher Energia and its payload, the non-powered reusable orbiter Buran.

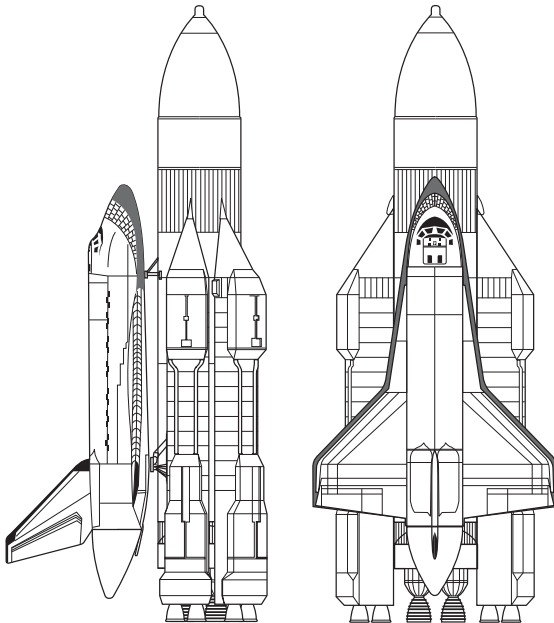
US Space Transportation System STS



	boosters	core stage
propellant	Solid	LH2/LOX
ignition mass [t]	2·590	750
propellant mass [t]	2·502	721
thrust [t]	2·1200(sl)	642(v)
<i>c</i> (vacuum) [m/s]	2600	4465
<i>c</i> (sea level) [m/s]	2414	3561
diameter [m]	3.7	8.4
height [m]	45.5	47.0

launch mass \approx 2040 tons,
LEO payload mass \approx 25 tons,
orbiter mass \approx 85 tons,
first flight in 1981.

Russia's Energia K-1/Buran System



	boosters	core stage
propellant	ker./LOX	LH2/LOX
ignition mass [t]	4·355	905
propellant mass [t]	4·320	820
thrust [t]	4·741(sl)	800(v)
<i>c</i> (vacuum) [m/s]	3306	4439
<i>c</i> (sea level) [m/s]	3032	3473
diameter [m]	3.9	8.0
height [m]	40.0	60.0

launch mass \approx 2400 tons,
LEO payload mass \approx 30 tons,
orbiter mass \approx 75 tons,
Energia LEO payload \approx 90 tons,
just two test flights: 1987, 1988.

Example 3-9: Apparently, the retired Russian space shuttle Energia/Buran is a copy of the American space shuttle STS; but actually it is a quite different and much better concept. First, the Russian shuttle uses liquid boosters instead of solids, and the same motor type is also used as the first stage of the Zenit satellite launcher. Second, the LH2/LOX core stage is equipped with its own engines: automatically, there is the option of using the launcher Energia without Buran as a heavy cargo lifter (with considerably increased payload capacity when simply more boosters are added). The orbiter Buran can be used for manned missions and unmanned missions as well.

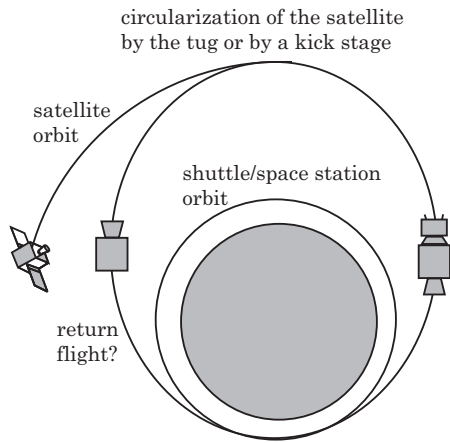
Reusability of space launcher components. Now let us consider that the space transportation system is not completely reusable, but just some of its components:

It is easy to recover strap-on boosters (or first stages), because they are separated when the launcher is still comparatively slow. For a return with soft splash-down in the ocean they need some heat protection, aerodynamic stabilization and parachutes. Such a procedure is carried out for the solid boosters of the American space shuttle; but the saving of costs is questionable: the reusable boosters need parachutes, a ship has to find them, fish them and transport them back. It also costs money to repair the boosters and refill them with new solid propellant. We may consider the option of booster recovery also for new commercial launchers, but we should not expect great cost reductions. The American Atlas-II launcher discharged its two strong liquid engines after about three flight minutes (just the engines were separated, not the tank). Recovery of the engines was not done although it would have been comparatively easy.

“Winged recovery” of the first stage with horizontal landing is another possibility. Certainly the winged recovery is a feasible option, but it is also a more complex procedure than the “ballistic watering”. The rocket departs rapidly from its launch location; and a non-powered coast flight back to the launch site is impossible, even with wings. The stage requires propulsion for the return flight, or the stage has to land on a distant airport. We should just consider the construction of a winged first stage when we are convinced that it saves costs (or we construct it for its own sake). Despite many proposals, up to now no serious attempt for realization has been made.

A recovery of the faster (orbital) second stage is also feasible. Particularly, this is sensible for a manned system which needs a return cabin anyway. The vehicle needs heat protection because it is exposed to high aerodynamic loads on its reentry trajectory. The American space shuttle serves as an example, but when we look at it closely, we see that actually just a small part of the orbital stage is recovered with the orbiter. The big external tank is expendable; just the expensive main engines are reusable.

We can also consider the concept of reusability for upper stages. An upper stage could work as a reusable “space tug”. It could be operated from a space station in low earth orbit, with propellant tanked from a huge cargo lifter (the “gas station” in space). The tug would transport the satellite to the destination orbit and then return back to the space station for maintenance and refuelling (or for the transport back to earth in the cargo bay of a space shuttle); but the tug needs propellant for the return flight, and this deteriorates its satellite payload capacity considerably. Detailed calculations show that the reusable operation of an upper stage is just cheaper for near-distance transportation (for far-distance transportation the payload is simply too small). The target orbit has to be lower than 10000 km altitude, or the tug has to deliver the satellite into a geostationary transfer orbit; but also then the cost saving potential is very small (below 10%). These results will only change in favor of reusable space tugs if, in the future, space shuttles are much cheaper and satellites much more numerous. Since also then the cost advantages are just small, reusable space tugs are disapproved of satellite transportation in the foreseeable future.

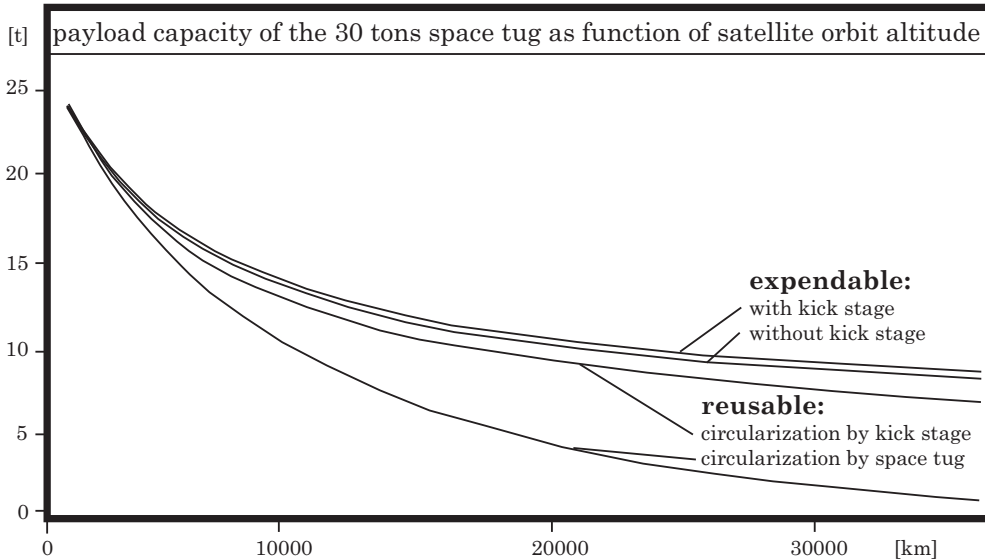


“space tug” missions:

Considered are transfer orbits between a circular low earth orbit at 200 km altitude and a coplanar, circular satellite orbit. The 30 tons initial mass includes the tug, propellant, the satellite and the kick stage (if there is one).

“space tug” spacecraft:

initial thrust acceleration = 2.5 m/s^2 ;
 engine mass[kg] = $0.15 \cdot (\text{thrust[kp]})^{0.750}$;
 tank mass[kg] = $225 \cdot (\text{volume[m}^3])^{0.666}$;
 propellant density: 280 kg/m^3 ;
 exhaust velocity = 4316 m/s .



Example 3-10: The reusable upper stage or “space tug” makes the most serious disadvantage of the concept of reusability in astronautics evident: compared with an expendable system of the same size, the payload is considerably smaller now. The tug, operated from a space station or shuttle orbit at 200 km altitude, needs propellant for its return flight. Detailed calculations verify that the reusable space tug is just sensible for “short distance traffic”; for “long distance traffic” it has an insufficient payload capacity. The plot compares a reusable space tug with an expendable upper stage, both systems have the total mass of 30 tons. The circularization of the satellite orbit is either performed by the transfer system itself or by a separate (expendable) kick stage. For the geostationary orbit the reusable tug has a capacity of 1.01 tons only, while the expendable stage has still a capacity of 8.40 tons. With a kick stage, the capacity increases to 6.73 tons for the reusable tug and to 8.56 tons for the expendable stage.

3.5.2 Single Stage to Orbit Vehicles

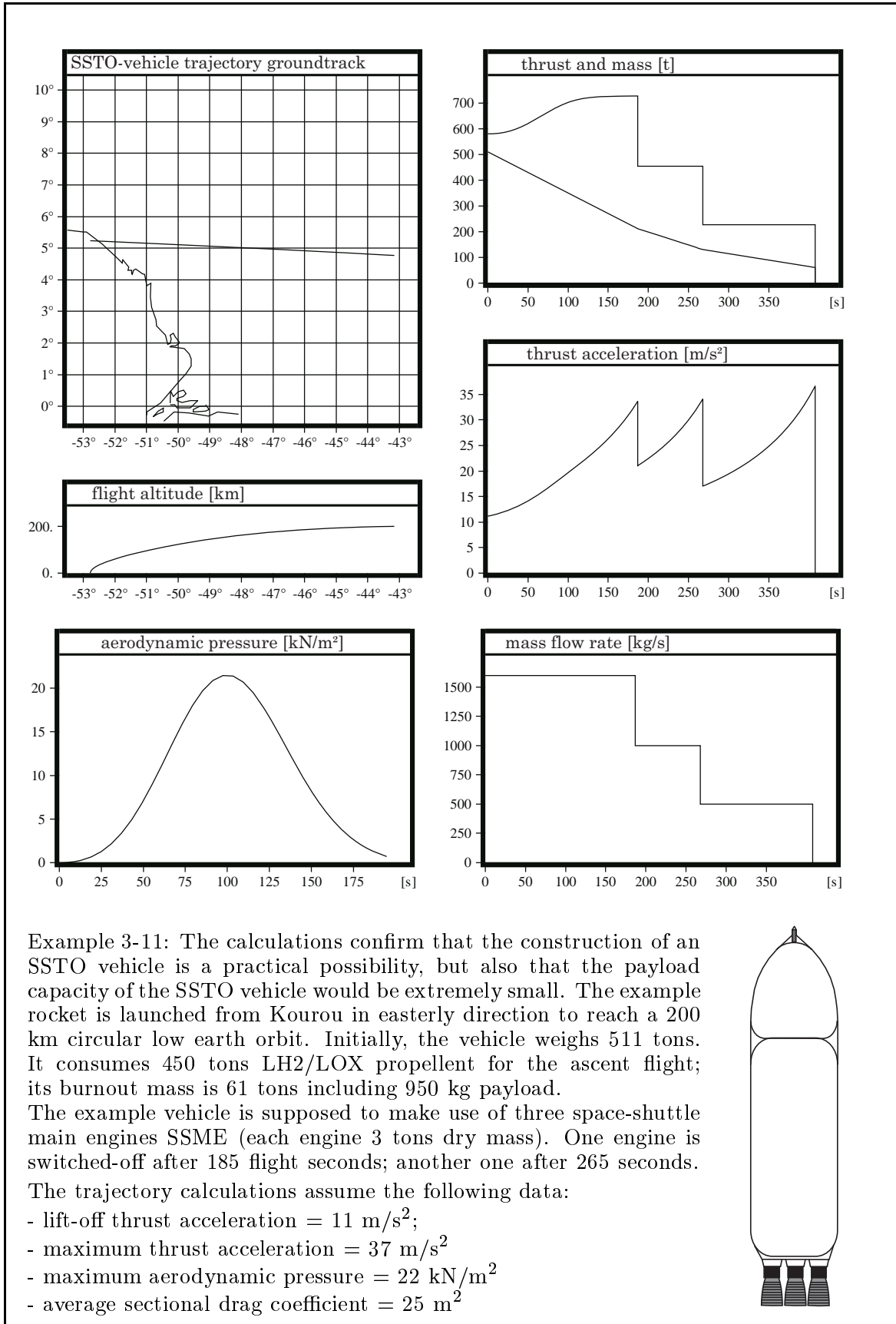
SSTO vehicles based on conventional technology. As we have seen, the staging principle for space launchers is applied to discharge empty tanks during the mission; but every stage is a complicated sub-system. The integration of the stages before the flight is expensive, and the separation during the flight involves safety risks. No wonder that it is often considered to construct so-called single stage to orbit (“SSTO”) vehicles. With modern technology it is possible to build extremely light weight structures and highly efficient engines. The SSTO principle for a launcher would greatly reduce the complexity, and consequently improve reliability and costs.

When we have a closer look at the SSTO principle, we realize that such a vehicle would transport mainly empty stage mass to orbit. A mass fraction of 12% (burnout mass / ignition mass) is a typical value which represents today the state of the art for big high performance rocket stages. The burnout mass includes the empty tanks (6%), residual propellant (2%), engines strong enough for the lift-off (2%), and structure to introduce the thrust force into the vehicle (2%). We can compare these values with the core stage of Russia’s Energia launcher (13.1%), with the second stage of America’s Saturn-5 (11.5%), and with the external tank of the space shuttle system (6.3%). For the construction of an SSTO vehicle the mass fraction must be better than 12%, otherwise the vehicle would have no payload capacity, even for a low earth orbit.

A reduction of the structure weight improves the payload capacity of a single stage launcher much more than the payload capacity of a two-stage launcher (but also the two-stage vehicle gets better when the structure weight is reduced). If we are able to reduce the burnout mass fraction from 12% to 11%, for example, this would give the SSTO vehicle already a reasonable payload capacity for the flight to a low earth orbit; but still serious arguments remain against the single stage launcher concept:

- it does not allow the composition of a launcher family with variable capacity;
- the performance is bad for polar orbits (or an upper stage is necessary);
- the performance is also bad for “high” low earth orbits (above 200 km);
- other target orbits require an upper stage anyway;
- weight reduction is probably not cost-neutral but cost-driving;
- the reliability is deteriorated when lower safety factors are assumed;
- the lift-off flight requires strong liquid engines, they are expensive.

We can avoid all these disadvantages when we simply support the lift-off flight of the single-stage-to-orbit vehicle by using a selection of small strap-on booster motors. The launcher supported by a variable number of small (fix-canted) boosters is always the better concept, also when great improvements in structure mass and engine performance are feasible. It is more flexible concerning payload capacity and target orbit; and a reduction of the structure weight is just helpful but not desperately required. The single stage concept is just sensible when the Δv requirement of the mission is significantly below 9 km/s (for example for short range ballistic missiles or sub-orbital sounding rockets). An SSTO launcher should not be constructed for its own sake: space activities far behind low earth orbit will always require multi stage rockets.



Example 3-11: The calculations confirm that the construction of an SSTO vehicle is a practical possibility, but also that the payload capacity of the SSTO vehicle would be extremely small. The example rocket is launched from Kourou in easterly direction to reach a 200 km circular low earth orbit. Initially, the vehicle weighs 511 tons. It consumes 450 tons LH₂/LOX propellant for the ascent flight; its burnout mass is 61 tons including 950 kg payload.

The example vehicle is supposed to make use of three space-shuttle main engines SSME (each engine 3 tons dry mass). One engine is switched-off after 185 flight seconds; another one after 265 seconds.

The trajectory calculations assume the following data:

- lift-off thrust acceleration = 11 m/s²;
- maximum thrust acceleration = 37 m/s²
- maximum aerodynamic pressure = 22 kN/m²
- average sectional drag coefficient = 25 m²

The reusable ballistic SSTO vehicle. In 1991, the American aerospace company McDonnell Douglas proposed the realization of a single-stage reusable vertical ascent and landing vehicle, named “Delta Clipper” or “DC-X”. A model of the concept (scaled 1/3) flew in summer 1993 as a technology demonstrator. The model demonstrated its lift-off and landing capability. It did not at all prove that it can also attain orbit; but the only problem in the concept is the ability to reach the orbit.

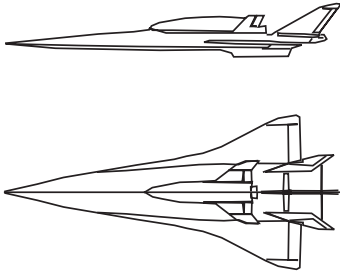
We know that throttleable rocket engines can give a space vehicle “hovering capability”. We have known this since 1969, when the Lunar Module of the American Apollo project demonstrated this capability by landing on the surface of the moon. Thus, the test flights of the “Delta Clipper” vehicle proved nothing actually. The managers of an aerospace company must know that it is impossible to build a completely reusable single-stage-to-orbit vehicle on the basis of conventional technology. We may believe that promoters of new aerospace projects are usually convinced that their favorite concept is feasible, but the managers of “Delta Clipper” took advantage of the show effect when they demonstrated these meaningless hover tests. Their intention was simply to betray the general public to get more funding for an infeasible project. Finally, after years of consuming public money, they will simply have to say that they did not get enough support to complete the nice “Delta Clipper” project successfully.

3.5.3 Aerospace Planes

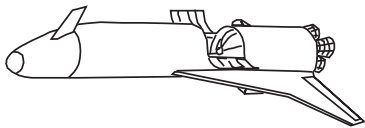
Using an aircraft as a first stage. It is possible to use a conventional jet aircraft as a “launch platform” for small satellite launchers. This was proved by the first successful flight of the Pegasus vehicle in April 1990. The three-stage Pegasus was dropped from a pylon under the right wing of a B-52 bomber at 13 km altitude. With a mass of just 18.5 tons the airborne launcher was able to transport a payload of 408 kg into 463 km circular low earth orbit. When we want to launch the same payload as usual from the surface, we need a launcher of approximately double size.

An airborne launcher has a much better growth factor than a surface launcher, furthermore the airborne launcher needs no launch site in a region near the equator. Both advantages promise a considerable cost saving potential. However, now the mission is a little more complicated because of the separation of the rocket from the aircraft. The airborne launcher requires some lift for the initial pull-up maneuver: a winged first stage or jettisonable wings are necessary. An aircraft is quite expensive, particularly when it cannot be used for any other purpose than for airborne launching.

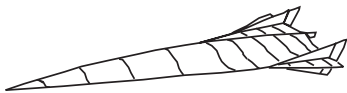
The capacity of Pegasus is still too small for the interesting commercial market: Pegasus is unable to reach the important geostationary transfer orbit GTO. However, the charge capacity of a big (civil or military) jet aircraft is by far not fully occupied when the aircraft carries a rocket of the size of Pegasus. Airplanes like Boeing’s 747 or Lockheed’s Galaxy can carry a payload of more than 100 tons. Russia’s transport aircraft Antonov 225 is specified for even 250 tons external payload capacity.



In 1985 Germany's aerospace industry proposed the "Saenger" launch vehicle, a two stage aerospace plane with a lift-off weight of about 360 tons and a LEO payload capacity of about 8 tons. The first winged stage should use air-breathing engines, the second (also winged) stage should use conventional rocket engines. The first stage was predicted to have a cruising capacity of about 6000 km, on its back the second stage as an external payload of approximately 112 tons. The separation of the second stage should take place at an altitude of 32 km, speed Mach 7. The second stage was planned as a cargo version or as a manned shuttle version, alternatively.



The HOTOL ("Horizontal Take-off and Landing") proposal of the British aerospace industry was originated in 1984, at about the same time as the German Saenger proposal. HOTOL was supposed to be a single stage, fully reusable unmanned vehicle with just 250 tons lift-off weight. It was expected to get the performance of 7 to 11 tons payload (and 4000 km cross range capability) from novel inter-cooled air-breathing engines. Later a so-called "Interim HOTOL" configuration was proposed, where then the vehicle was supposed to be launched from the back of the Russian transport aeroplane Antonov 225.



The objective of the American proposal X-30 (National Aero-Space Plane NASP) was to build a horizontal take-off and landing vehicle, as initial step for velocities of Mach 5 to Mach 15, as second step for orbital flight. In 1986 the companies McDonnell Douglas, General Dynamics and Rockwell were interested in the body of the vehicle, the companies Rocketdyne and Pratt-Whitney were interested in the propulsion of the vehicle (a ramjet-scamjet-rocket compound motor, with liquid hydrogen as propellant). As in 1992 the trust in the feasibility of the NASP project started to drop off, the company Boeing proposed the construction of a huge hypersonic carrier aircraft as a first stage for the NASP vehicle.

Example 3-12: At the end of the 20th century a large number of space plane concepts appeared at aerospace congresses all over the world. The most famous names are Saenger, HOTOL and NASP; but also students presented papers with proposals for new aerospace-planes as diploma works. The performance and cost data of the proposed vehicles improved inflationary in time. The Japanese laboratory NAL analyzed several concepts similar to Saenger and HOTOL. The Russian proposal TU-2000 RASP (Russian Aerospace Plane) was nearly an identical copy of the NASP concept. Many publications did not give the feeling that the projects were seriously considered.

Air-breathing engines for space transportation. Usually, the proposals for new-technology aerospace planes involve the use of airbreathing engines. The operation principle of these engines is called “duct engine principle”, it is different from the operation principle of rocket engines: now the vehicle has to carry just the fuel, oxidizer and working fluid are taken from the ambient atmosphere. Conventional jet engines work in accordance to this principle. Duct engines are used in jet aircraft because they have a much higher propellant consumption efficiency than rocket engines.

In comparison with rocket engines, air-breathing engines have also disadvantages for space transportation. The working fluid (the air) must leave the nozzle faster than it enters the inlet, otherwise there is no thrust. In contrast to a rocket propulsion vehicle, a duct propulsion vehicle cannot fly faster than the relative velocity of its exhaust gases. The available propulsion energy limits the operation velocity for duct engines; and because there is no air available in the vacuum of space, an aerospace plane needs a rocket engine for the orbital flight anyway. Preferably, the spaceplane engine is a rocket-airbreathing compound motor. The stagnation point temperature increases considerably with the velocity: when the vehicle flies with hypersonic speed, the gas is already extremely hot at the entrance of the combustion chamber. Then the engine needs a heat exchanger for inter-cooling the gas, or, even better, it operates as a ramjet, first with subsonic and later with supersonic combustion. We have to face difficult technical problems when we want to build an engine for a wide velocity range; but such an engine could translate the reusable winged aerospace-plane into reality.

Spaceplane technology. The technology for spaceplanes is not available today. It is not even sure if it is possible to realize spaceplanes at all. Many proposals exist; but when we look closer at a particular proposal, we see often that the specifications are high above the best data which have ever been reached by airplanes or rockets. Often very optimistic assumptions are made for the masses of fuselage, wings and tanks (but these spaceplanes have to stand the high stresses of the re-entry flight). Everything depends sensitively on the realization of the new air-breathing engines. As long as the feasibility of the engines is not proved, we cannot seriously evaluate the performance of the vehicles and make a detailed cost propagation. The task is very difficult, and solving a detail problem does not prove that a concept is feasible; and even if an aerospace-plane project is feasible, it is still not sure whether it makes space transportation really cheaper than it is today. Considering these facts we shall study future spaceplanes in form of scientific research programmes (it is too early for an industrial development). We will learn a lot about hypersonic flight, but maybe the projects will not lead to new (and cheaper) space transport systems. The commercial disaster of the American space shuttle shows that it is irresponsible to expect great changes or even a “cost-breakthrough” from non-existing technology.

Initially, the public interest in new spaceplane projects was high; but the success failed to appear. Since nothing was shown after years of spending public money (except for some illustrations, models and flight simulations), the public acceptance to spend money for spaceplanes declined in time; and there is still the controversial discussion on spaceplane projects; it is just clear that spaceplanes are very expensive.

3.5.4 The Future of Space Transportation

Future aerospace planes. The prediction of the future is quite a delicate undertaking because life is - from the mathematical point of view - a “chaotic process”: the future depends sensitively on some events which we simply cannot foresee (what will happen to astronautics, for example, when some day “little green men” show up). However, it is shortsighted when we do not to consider the effect of fatal shuttle accidents on the manned American space programme. It is possible that the international space station project ISS is terminated when the NASA loses another space shuttle after the Columbia accident in January 2003. However, also when the technological and economical environment develops positively in the next decades we cannot expect the realization of aerospace planes similar to Saenger, Hotol or NASP. Today it is not clear whether such projects are feasible at all, and from “feasible” to “cost efficient operation” there is still a long way. The new propulsion technology is too difficult to be solved within the next 25 years. Thus, the manned space programme will be served basically by the American shuttle and by conventional rockets like the Russian Soyuz launcher. Even when US-America develops a new shuttle, conventional satellite launchers are probably the best way to transport satellites in the foreseeable future.

The future of the public interest in spaceplanes. Space planners consider it as very important to influence the public opinion in favor of astronautical activities. Many articles are written for space magazines and many presentations are given at space congresses to propagate how important astronautics is to mankind. Audience of this “space advertisement” are usually astronautical engineers or space planners which are in favor of astronautics anyway. However, the propaganda gives a more detached audience the feeling that there is a huge space industry which simply wants to make money. Someone who uses in his house a satellite dish for watching video programmes on TV has probably little doubt that it is sensible to bring a satellite into orbit (nothing succeeds like success). However, it is not sure that he will also accept an inefficient space launcher as a national prestige project. After years of spending money for aerospace-planes, practically nothing else than illustrations came out (“space art”). The German Saenger concept was proposed in 1985. A decade later the project was still controversially discussed at astronautical congresses by space engineers; and all “hardware” which “Mr.Taxpayer” could see was a model of the vehicle (with a size of about two meters). Then the project was terminated because it was obvious for everybody that the public money was incorrectly investigated.

Considering this point of view, the most serious problem that astronautics suffers today is not the absence of realistic and sensible projects, it is the profusion of superfluous or unrealistic proposals. Sensible and realistic proposals have to fight with absurd proposals, the battles are carried out not in the public but in the space industry with the public as audience. It is easy to make proposals, but it seems to be extremely difficult to push these projects through against the resistance of space planners who are naturally in favor of their own projects, which are sometimes unrealistic and driven by hubris and science fiction. Since in such a scenario nothing will be shown over periods of years, “Mr.Taxpayer” is going to lose his interest.

3.5.5 Infeasible Ways to Access Space

In engineering it is allowed to analyze (nearly) everything, also impossible projects. In astronautics many ideas are in discussion how to access space easier than today. The intention is to translate today's science-fiction into reality in the future. The following list discusses some of these ideas just as "mental exercises". In every case it is evident that the way will not soon lead to a new space access opportunity:

Orbital tower satellite. This idea is to stretch a geostationary satellite, one side towards the earth and the other side away from the earth. Finally, the satellite touches the earth with its inner side. Now it represents a 144560 km high building. The tower satellite is pretty thick in the middle to stand the high centrifugal forces (even when the best material is used). An elevator can transport payloads into space.

Space railroad. The intention is to accelerate space payloads on an electrically propelled carriage, avoiding the use of chemical propellant. Wheels cannot be used any more for speeds above Mach 1, so the idea is to build a huge inclined catapult for a magnetic-field monorail. The space railroad needs an enormous amount of electrical power when it gets very fast (power is the product of velocity and thrust force). Space payloads are simply discharged when the carriage attains orbital velocity.

Space cannon. Space cannons have been tested unsuccessfully. Obviously, they offer no advantages in comparison with conventional rockets. The idea is to shoot very robust payloads into space, using a long gun. The projectile has to stand high accelerations inside the barrel and high aerodynamic stresses when it exits the muzzle. At the moment the projectile runs through the culmination point of its trajectory, a rocket motor is ignited for the circularization of the orbit of the payload.

Nuclear propulsion. Also nuclear rocket motors were tested seriously (but not successfully) at the beginning of the age of astronautics; but unsolvable safety concerns were always involved, and today, in default of political acceptance, nuclear propulsion is not an option any more. The nuclear rocket motor works with much higher specific impulse, but it is also much heavier and requires radiation protection. The idea is to use a nuclear reactor to heat up the exhaust gases (energy is needed to accelerate the exhaust gases in the thrust chamber). Conventionally, this propulsion energy is stored in a chemical form in the propellant, but a nuclear process can liberate much more energy per unit of mass than a chemical process.

Metallic hydrogen. It is expected that hydrogen changes to a metallic phase when it is exposed to pressures above 4 million bar (today, pressures of about 2 million bar can be realized in laboratories). It is not known whether this "metallic hydrogen" can be produced at all, and whether the costs of the production are acceptable. Either it is clear whether the metallic hydrogen is stable (probably it is not stable); but when metallic hydrogen is "storable" and its production is "cheap", it is maybe the ideal propellant for novel-technology single-stage-to-orbit aerospace planes.

4. Spacecraft Trajectories

The accurate observation of the orbital motion of planets in the seventeenth century was the reason that the medieval geocentric cosmic system was finally displaced. Therefore we can say that trajectory dynamics is the first modern natural science. Sir Isaac Newton (1643-1727) formulated his three laws for the behaviour of a mass particle. The subject of celestial mechanics was ideal for studying the motion of a restless mass; and there was also a practical importance for the navigation of ships. The movement of planets is not subjected to energy dissipation, and even with its huge size a planet can be considered as a particle when the trajectory is what matters.

Thus, trajectory analysis belongs to particle dynamics, the oldest field of mechanics (the subject of dynamics is the study of forces; kinematics is the science of motion). Particle dynamics treats exclusively the translational part of the motion and ignores completely the rotational part. Any moving mass is considered as a material body with infinitely small dimensions. The objective of trajectory dynamics in astronautics is the determination of the translational motion of a satellite or space launcher. In the field of astronautics it is possible to separate the translational motion from the rotational motion with a high degree of accuracy (an exception is the problem of attitude stabilization for large space stations). A flying object is subjected to gravity, thrust and aerodynamic forces; the “equations of motion” consider all these forces and determine the locomotion of the flying object as a function of time.

Comprehensive introductions to the mechanics of spaceflight you can find in the books of W.T.Thomsom [“Introduction to Space Dynamics”, Dover Publications Inc., New York, 1986], W.E.Wiesel [“Spaceflight Dynamics”, McGraw-Hill, New York, 1989] and M.H.Kaplan [“Modern Spacecraft Dynamics and Control”, John Wiley and Sons, New York, 1976] (lot of other literature exists which can be recommended as well).

4.1. Vectors and Coordinate Systems

Conveniently, vectors are used to describe the motion of a vehicle that flies in three-dimensional space. Vectors are ideal for the determination of physical quantities which have beside their scalar value (magnitude or amount) also a certain direction. The actual position of a flying vehicle can be defined using its location vector \vec{r} . It is the arrow that points from a well-known (fixed) location to the moving vehicle. Since the vehicle is moving its location changes in time. The time derivative of the location vector $d\vec{r}/dt$ is called velocity \vec{v} , again a vector. It must be emphasized that vectors are independent of coordinate systems. For the definition of vector operations it is not necessary to designate a certain coordinate system. Addition of vectors can be declared as composition, putting one vector behind the other. Multiplication of a vector with a scalar value is nothing else than scaling the length of the vector. Consequently, vector subtraction is nothing else than the addition of two vectors, where the second vector is turned into the opposite direction (multiplied by -1).

4.1.1 Moving Coordinate Systems

Vector resolution. Thus, vectors are arrows in space with a length and a direction; but they are independent of the coordinate system that we take as a basis. However, we need a coordinate system to resolve (decompose) a vector to get its component notation (for example $\vec{x} = (x_1, x_2, x_3)$). Just the component notation of a vector refers to a certain coordinate system: the same vector has a different component notation when we use another coordinate system with axes that point into other directions.

It is important to mention that we cannot treat all physical quantities as vectors just because they possess magnitude and direction: necessary is that vector composition (and decomposition) is possible; and that vector summation is commutative (independent of the order) and associative (independent of grouping in any order). An example for a quantity which cannot be treated as a vector is the angular rotation of a rigid body: the rotation has a value (the angle) and it has a direction (the rotation axis); but two rotations around different axes are usually not commutative. Rotations are not vectors, because the body assumes a different attitude when we exchange the order of the rotations.

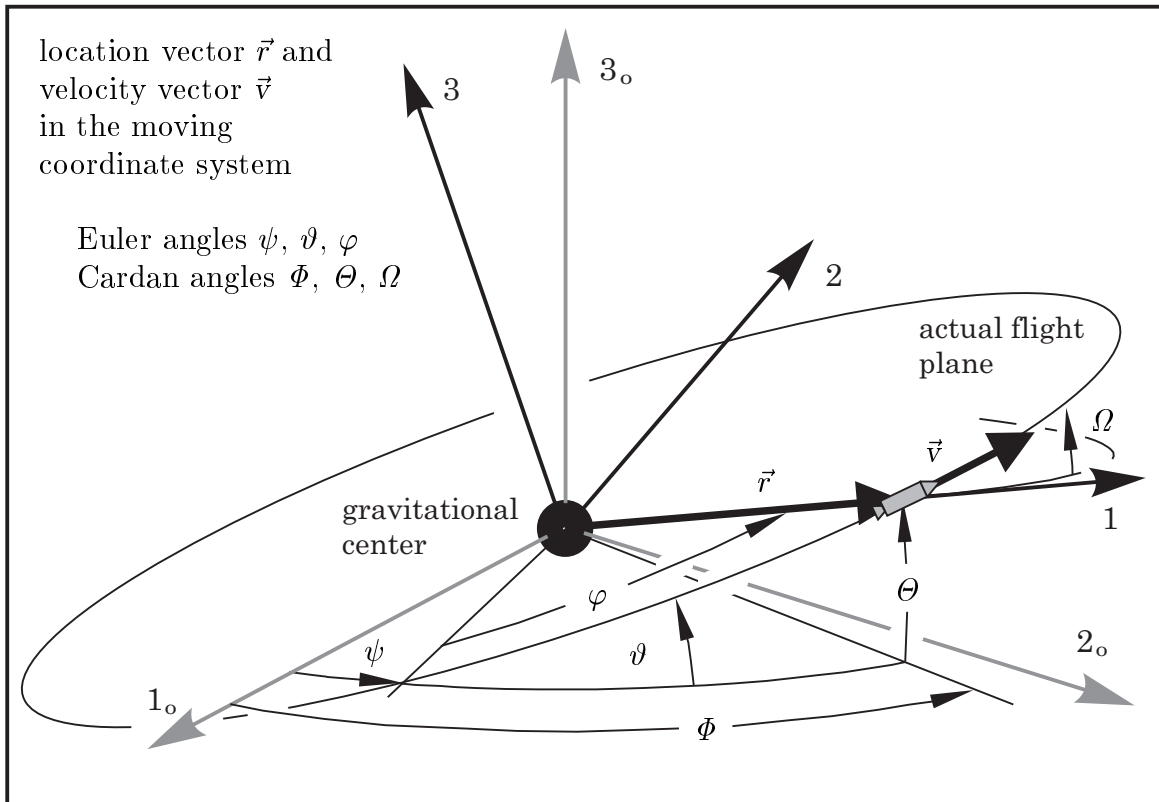
Point of application. A force, for example, is a physical quantity which has a magnitude and a direction; and we are allowed to use a vector to describe these properties. However, we cannot simply ignore that the force has also a “point of application”, the location where it is effective. Obviously, it makes a difference whether a body is pushed on the left or on the right side. Thus, actually two vectors are necessary to describe this physical quantity called “force”: a vector for its magnitude and its direction, and a location vector. The location vector is the vector which points from a known location (usually a location fixed in inertial space) to the point of application.

Cartesian coordinate systems. Nearly always so-called Cartesian coordinate systems are used in technical mechanics, named after René Descartes (1596–1650): all three axes are scaled in the same way, these axes are mutually perpendicular, and the “right hand” rule is valid (spread your right hand in a way that the thumb, the forefinger and the middle finger form a frame system: the orientation of the 1-axis corresponds to your thumb, the orientation of the 2-axis to your forefinger and the orientation of the 3-axis to your middle finger). Two vector operations are defined in a Cartesian coordinate system, applicable to the component notation of two vectors: the scalar “dot” product is useful for the calculation of the angle between the vectors; and the “cross” product is useful for the determination of the vector perpendicular to the other two vectors. Incorrectly both schemes are called “multiplication” of vectors.

Orientation angles. A Cartesian coordinate system is not necessarily an inertially fixed system: many problems are better solved when we use a moving Cartesian coordinate system. The reason is that vector equations themselves are independent of a coordinate system, but their component notation is a function of the attitude of the coordinate system. We can simplify the component notation of a vector equation considerably when we use a coordinate system with a suitable orientation in space.

Generally, the attitude of a moving system with respect to an inertial system is entirely determined by a set of three angles. For this are widely used the so-called Euler angles $(\psi, \vartheta, \varphi)$, named after the great mathematician Leonhard Euler (1707–1783); or alternatively so-called Cardan angles (Φ, Θ, Ω) , referring to the construction of Cardanic gyro bearings, named after Girolamo Cardano (1501–1576, even though he has actually not invented them). Notations with other angles are also possible, but other angles are not in common use in particle dynamics or in gyro dynamics.

Location and velocity of a flying spacecraft. Now we will make use of a moving Cartesian coordinate system. We place the origin of the system at the center of gravitation (the center of the celestial body) and determine the attitude of the system using the location vector \vec{r} and the velocity vector \vec{v} of the spacecraft: the location vector lies always on the 1-axis and the velocity vector lies always in the 1-2 plane. The coordinate system changes its orientation with the displacement of the vehicle.



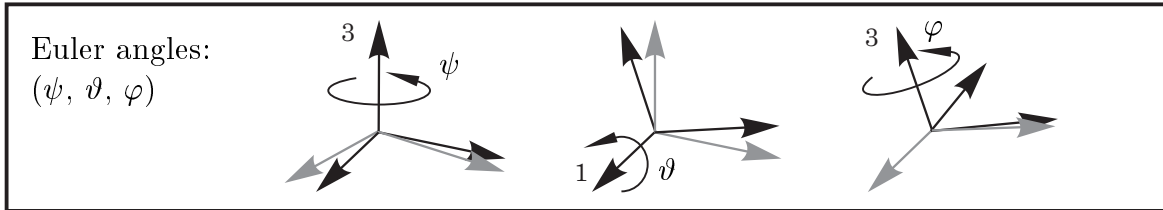
When we designate the length of the location vector \vec{r} by the letter r and term the resolution of the velocity vector \vec{v} vertically u and horizontally v , then we have:

$$\vec{r} = \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} \quad ; \quad \vec{v} = \begin{pmatrix} u \\ v \\ 0 \end{pmatrix} \quad (4 - 1)$$

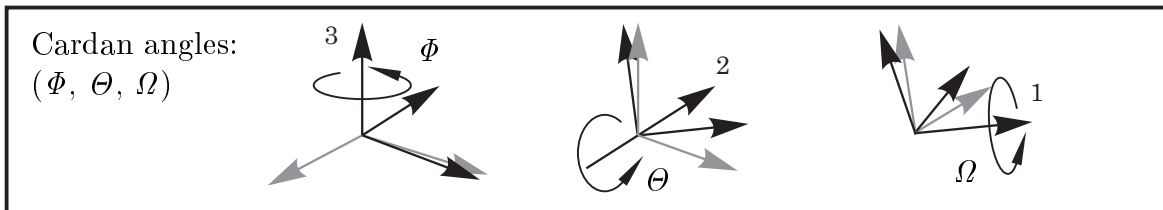
The notation of both vectors would be much more complex if we used components of an inertial system. Obviously, the moving system allows us a simple formulation.

4.1.2 Coordinate Transformation

Transformation of vector notations. Sometimes it is the case that we know a vector in the moving system (for example the location vector \vec{r} , equation 4-1); and we want to find the component notation of the same vector in the inertial system (termed \vec{r}_o). Then we can find the inertial system notation of the vector when we “multiply” the vector with an appropriate transformation matrix (in this case a “rotation matrix”). However, note that the vector is not rotated: we simply transform the “moving-system-notation” of a vector \vec{x} to its “inertial system notation” \vec{x}_o . Using Euler or Cardan angles alternatively, the transformation matrix becomes:



$$\begin{aligned} \vec{x}_o &= \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\vartheta & -\sin\vartheta \\ 0 & \sin\vartheta & \cos\vartheta \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \vec{x} \\ &= \begin{pmatrix} \cos\psi \cos\varphi - \sin\psi \cos\vartheta \sin\varphi & -\cos\psi \sin\varphi - \sin\psi \cos\vartheta \cos\varphi & \sin\psi \sin\vartheta \\ \sin\psi \cos\varphi + \cos\psi \cos\vartheta \sin\varphi & -\sin\psi \sin\varphi + \cos\psi \cos\vartheta \cos\varphi & -\cos\psi \sin\vartheta \\ \sin\vartheta \sin\varphi & \sin\vartheta \cos\varphi & \cos\vartheta \end{pmatrix} \cdot \vec{x} \end{aligned} \quad (4-2)$$



$$\begin{aligned} \vec{x}_o &= \begin{pmatrix} \cos\Phi & -\sin\Phi & 0 \\ \sin\Phi & \cos\Phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\Theta & 0 & -\sin\Theta \\ 0 & 1 & 0 \\ \sin\Theta & 0 & \cos\Theta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Omega & -\sin\Omega \\ 0 & \sin\Omega & \cos\Omega \end{pmatrix} \cdot \vec{x} \\ &= \begin{pmatrix} \cos\Phi \cos\Theta & -\sin\Phi \cos\Omega - \cos\Phi \sin\Theta \sin\Omega & \sin\Phi \sin\Omega - \cos\Phi \sin\Theta \cos\Omega \\ \sin\Phi \cos\Theta & \cos\Phi \cos\Omega - \sin\Phi \sin\Theta \sin\Omega & -\cos\Phi \sin\Omega - \sin\Phi \sin\Theta \cos\Omega \\ \sin\Theta & \cos\Theta \sin\Omega & \cos\Theta \cos\Omega \end{pmatrix} \cdot \vec{x} \end{aligned} \quad (4-3)$$

Each one of the two complex rotation matrices above is the result of three individual coordinate transformations: a coordinate system that has initially the same attitude as the inertial system must be rotated consecutively three times to coincide with the moving system. The system represents an “intermediate system” after every rotation; the last intermediate system is finally coincident with the moving system.

Thus, three times we rotate the coordinate system around one of its axes, which is fixed with respect to inertial space at the moment when we rotate (“intermediate systems” of the Euler angles and “intermediate systems” of the Cardan angles are not the same). The orientation of the moving system is reached after three rotations:

- rotation 1: angle ψ (3-axis), or angle Φ (3-axis).
- rotation 2: angle ϑ (1-axis), or angle Θ (2-axis);
- rotation 3: angle φ (3-axis), or angle Ω (1-axis);

The next step is that we establish the three individual transformation matrices. When a vector is resolved into its component notation and transformed from one coordinate system into the other one, then the transformation does not alter the component that aims into direction of the rotation axis. The other two components change, we can compose the new values using the sine function and cosine function of the rotation angle. Finally, we can find the complete transformation matrix by a multiplication of three individual matrices (according to the rules of matrix multiplication).

Vector transformation into a notation of the moving system. The inverse procedure is the transformation of the “inertial system notation” of a vector into its “moving system notation”, for example to calculate the vector \vec{x} when \vec{x}_o is known. We can find these inverse rotation matrices with the same geometrical consideration:

$$\begin{aligned}
 \vec{x} &= \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\vartheta & \sin\vartheta \\ 0 & -\sin\vartheta & \cos\vartheta \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \vec{x}_o \\
 &= \begin{pmatrix} \cos\psi\cos\varphi - \sin\psi\cos\vartheta\sin\varphi & \sin\psi\cos\varphi + \cos\psi\cos\vartheta\sin\varphi & \sin\vartheta\sin\varphi \\ -\cos\psi\sin\varphi - \sin\psi\cos\vartheta\cos\varphi & -\sin\psi\sin\varphi + \cos\psi\cos\vartheta\cos\varphi & \sin\vartheta\cos\varphi \\ \sin\psi\sin\vartheta & -\cos\psi\sin\vartheta & \cos\vartheta \end{pmatrix} \cdot \vec{x}_o \\
 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Omega & \sin\Omega \\ 0 & -\sin\Omega & \cos\Omega \end{pmatrix} \begin{pmatrix} \cos\Theta & 0 & \sin\Theta \\ 0 & 1 & 0 \\ -\sin\Theta & 0 & \cos\Theta \end{pmatrix} \begin{pmatrix} \cos\Phi & \sin\Phi & 0 \\ -\sin\Phi & \cos\Phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \vec{x}_o \\
 &= \begin{pmatrix} \cos\Phi\cos\Theta & \sin\Phi\cos\Theta & \sin\Theta \\ -\sin\Phi\cos\Omega - \cos\Phi\sin\Theta\sin\Omega & \cos\Phi\cos\Omega - \sin\Phi\sin\Theta\sin\Omega & \cos\Theta\sin\Omega \\ \sin\Phi\sin\Omega - \cos\Phi\sin\Theta\cos\Omega & -\cos\Phi\sin\Omega - \sin\Phi\sin\Theta\cos\Omega & \cos\Theta\cos\Omega \end{pmatrix} \cdot \vec{x}_o
 \end{aligned} \tag{4-4}$$

A rotation matrix consists of nine elements; we can see it as a composition of three “row vectors” or three “column vectors”. The matrix is called “orthogonal matrix” when these row vectors and column vectors are unit vectors (length 1) which form a perpendicular frame system. It is easy to verify that all the rotation matrices which we consider here are “orthogonal matrices”. Just in this special case we can find the inverse matrix simply by transposing the matrix (by exchanging column vectors for row vectors, or row vectors for column vectors). Generally, finding the inverse matrix is a much more complicated procedure than simply forming the transposed matrix.

Transformation of Euler angles to Cardan angles. Let us consider now that the attitude of the moving coordinate system is determined because the Euler angles $(\psi, \vartheta, \varphi)$ are well-known, and that it is the task to find the corresponding set of Cardan angles (Φ, Θ, Ω) . The equation (4-4) contains implicitly a set of nine equations to perform the transformation: every element of the “Cardan angle matrix” corresponds to an element of the “Euler angle matrix”. However, two problems are involved: both, Cardan angles and Euler angles, determine uniquely a certain attitude of the coordinate system, but neither Cardan angles nor Euler angles are uniquely determined when the attitude of the coordinate system is given (for example, change the angle Φ by $\pm 180^\circ$ and you will find different angles Θ and Ω which belong to the same orientation). Then, both Euler angles and Cardan angles are singular for some certain orientations of the moving system. Euler angles have their singularity at the “equator” ($\vartheta = 0^\circ$ or $\vartheta = 180^\circ$); Cardan angles have their singularity at the “poles” ($\Theta = \pm 90^\circ$). For a unique determination of the Cardan angles we have to impose a limitation on the angle Θ , for example $-90^\circ < \Theta < +90^\circ$. Then we get:

$$\begin{aligned}
 \sin \Theta &= \sin \vartheta \cdot \sin \varphi \\
 \cos \Theta &= +\sqrt{1 - \sin^2 \Theta} \\
 \sin \Omega &= \sin \vartheta \cdot \cos \varphi / \cos \Theta \\
 \cos \Omega &= \cos \vartheta / \cos \Theta \\
 \sin \Phi &= (\sin \psi \cos \varphi + \cos \psi \cos \vartheta \sin \varphi) / \cos \Theta \\
 \cos \Phi &= (\cos \psi \cos \varphi - \sin \psi \cos \vartheta \sin \varphi) / \cos \Theta
 \end{aligned} \tag{4-5}$$

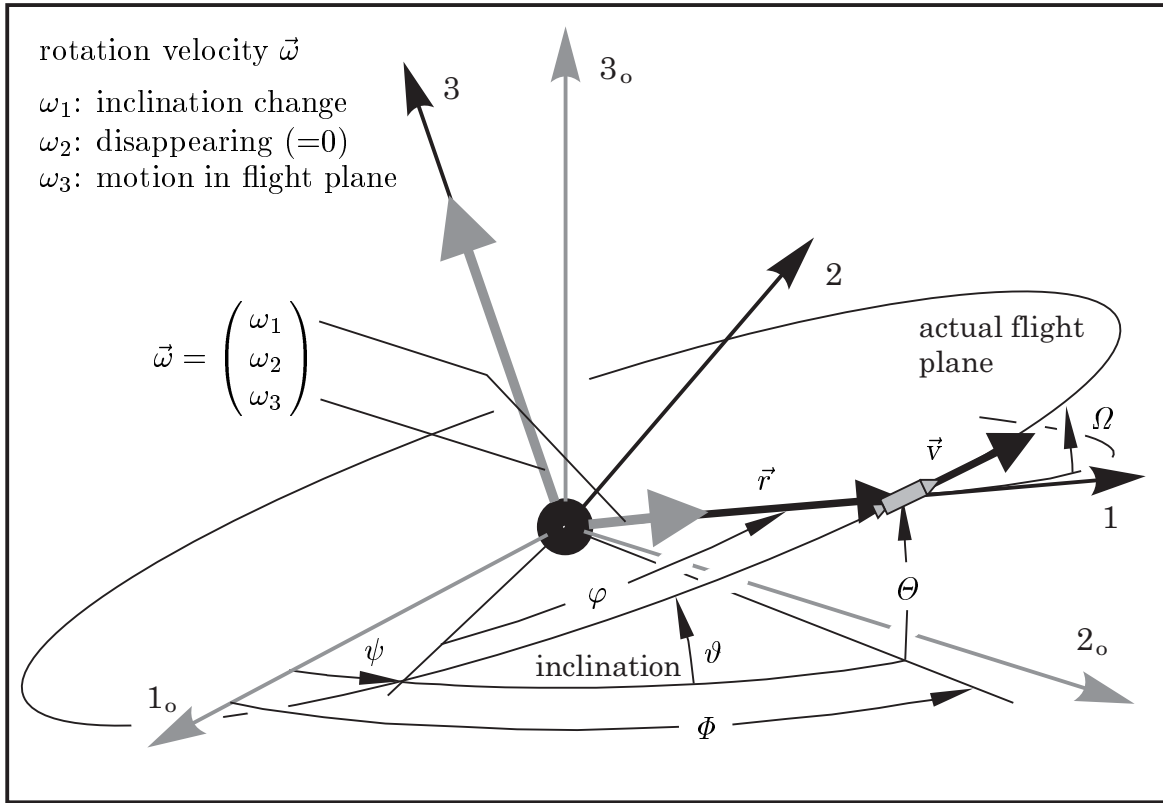
Transformation of Cardan angles to Euler angles. The inverse transformation involves similar problems. In orbital mechanics the inclination angle ϑ is restricted for a unique determination of the Euler angles, conventionally we define $0^\circ < \vartheta < 180^\circ$. The set of equations for transforming Cardan angles into Euler angles becomes finally:

$$\begin{aligned}
 \cos \vartheta &= \cos \Theta \cdot \cos \Omega \\
 \sin \vartheta &= +\sqrt{1 - \cos^2 \vartheta} \\
 \sin \varphi &= \sin \Theta / \sin \vartheta \\
 \cos \varphi &= \cos \Theta \cdot \sin \Omega / \sin \vartheta \\
 \sin \psi &= (\sin \Phi \sin \Omega - \cos \Phi \sin \Theta \cos \Omega) / \sin \vartheta \\
 \cos \psi &= (\cos \Phi \sin \Omega + \sin \Phi \sin \Theta \cos \Omega) / \sin \vartheta
 \end{aligned} \tag{4-6}$$

Euler and Cardan angles in the spherical triangle. When three planes intersect a sphere with the unit radius “1” right through the center, the intersection curves are circles (also with the radius “1”). Segments of these circles form a “spherical triangle” on the surface of the sphere. A spherical triangle is formed by the Euler angles and the Cardan angles $\Phi - \psi$, Θ and φ , with an angle of 90° between side Φ and side Θ (the side lines of a spherical triangle can be considered as angles, because the sphere has the unit radius). We can use the laws of spherical geometry to verify the relationships (4-5) and (4-6) when we observe the rectangular spherical triangle which is formed by the Euler angles and the Cardan angles on the surface of a sphere.

4.1.3 Rotation Velocity

Vector representation of the rotation velocity. We have seen that we cannot treat rotations as vectors, because their chronological order is not commutative. However, we can treat the time derivatives of rotation angles as vectors. The rotation velocity (or angular velocity) of a rigid body (or a moving coordinate system) is defined as a vector, conventionally termed $\vec{\omega}$. Its length (magnitude) is the time derivative of the rotation angle, its direction is the orientation of the actual spin axis. The spin axis is defined as the line of all locations which are at the moment in rest. Every location which is not on the spin axis is in motion, the velocity is proportional to rotation velocity and proportional to the distance from the spin axis. The vector triple (1 rotation velocity, 2 orthogonal distance, 3 velocity) forms a “right-hand” frame system. Let us now consider the special case of our moving coordinate system:

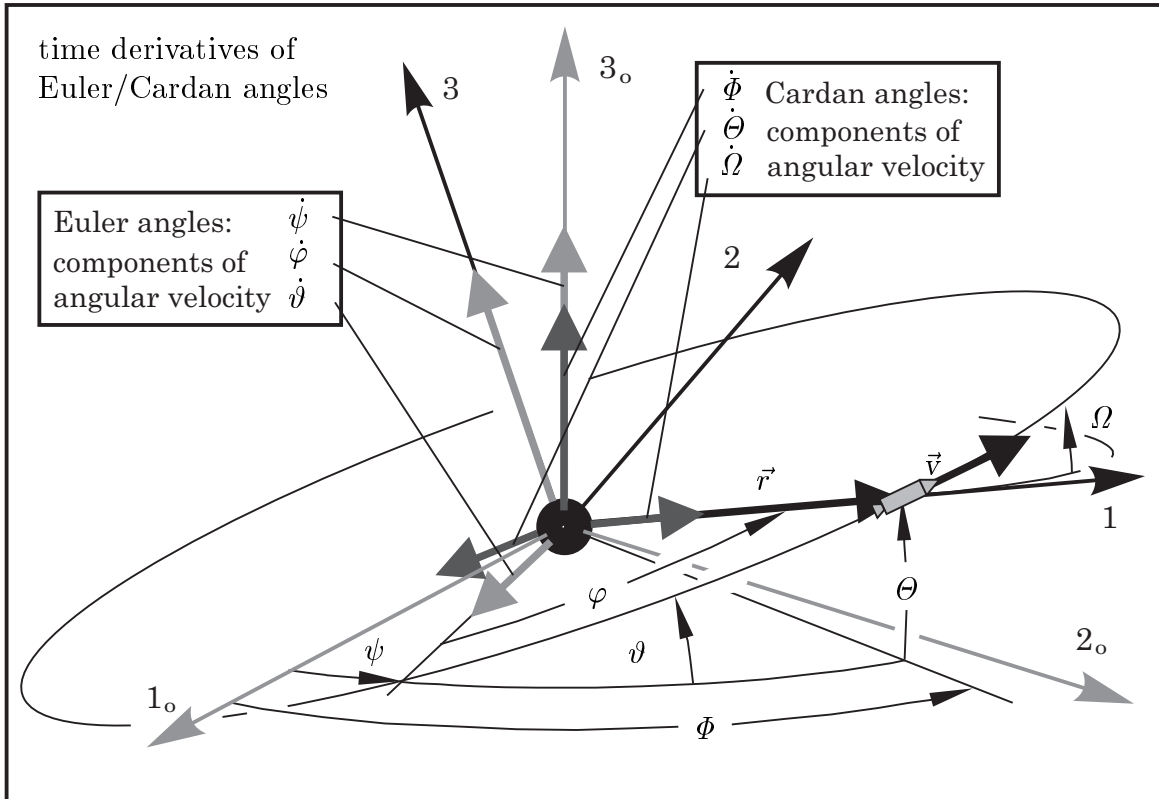


The attitude of the moving system is determined by the location vector $\vec{r} = (r, 0, 0)$ (on the 1-axis) and velocity vector $\vec{v} = (u, v, 0)$ (in the 1-2 plane). We can conclude that the rotation velocity vector $\vec{\omega}$ has no component in the direction of the 2-axis. The horizontal velocity component v generates a rotation of the coordinate system v/r on the 3-axis; an inclination change maneuvers rotates the system on the 1-axis. Thus:

$$\vec{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} \omega_1 \\ 0 \\ v/r \end{pmatrix} \quad (4-7)$$

Note that the vector $\vec{\omega}$ is written in a component notation of the moving system.

Time derivatives of Euler and Cardan angles. Our next problem is to find the orientation of the moving coordinate system as a function of time. The time derivatives of the Euler angles (or the Cardan angles) are components of the angular velocity vector $\vec{\omega}$ of the moving system. We can resolve $\vec{\omega}$ into three oblique-angled individual rotation velocity vectors, therefore we can take either the derivatives of the Euler angles $\dot{\psi}$, $\dot{\vartheta}$, $\dot{\varphi}$, or alternatively the derivatives of the Cardan angles $\dot{\Phi}$, $\dot{\Theta}$, $\dot{\Omega}$. These vectors have the same orientation as the rotation axes to which they belong.



Correctly composed, the angular velocity of the moving system can be expressed as:

$$\vec{\omega} = \begin{pmatrix} \omega_1 \\ 0 \\ v/r \end{pmatrix} = \begin{pmatrix} \dot{\psi} \sin \vartheta \sin \varphi + \dot{\vartheta} \cos \varphi \\ \dot{\psi} \sin \vartheta \cos \varphi - \dot{\vartheta} \sin \varphi \\ \dot{\varphi} + \dot{\psi} \cos \vartheta \end{pmatrix} = \begin{pmatrix} \dot{\Phi} \sin \Theta + \dot{\Omega} \\ \dot{\Phi} \cos \Theta \sin \Omega - \dot{\Theta} \cos \Omega \\ \dot{\Phi} \cos \Theta \cos \Omega + \dot{\Theta} \sin \Omega \end{pmatrix} \quad (4-8)$$

A transformation brings us two equivalent sets of first order differential equations:

$$\begin{aligned} \dot{\psi} &= \omega_1 \frac{\sin \varphi}{\sin \vartheta} & \dot{\Phi} &= \frac{v}{r} \frac{\cos \Omega}{\cos \Theta} \\ \dot{\vartheta} &= \omega_1 \cos \varphi & \dot{\Theta} &= \frac{v}{r} \sin \Omega \\ \dot{\varphi} &= \frac{v}{r} - \omega_1 \sin \varphi \cot \vartheta & \dot{\Omega} &= \omega_1 - \frac{v}{r} \cos \Omega \tan \Theta \end{aligned} \quad (4-9)$$

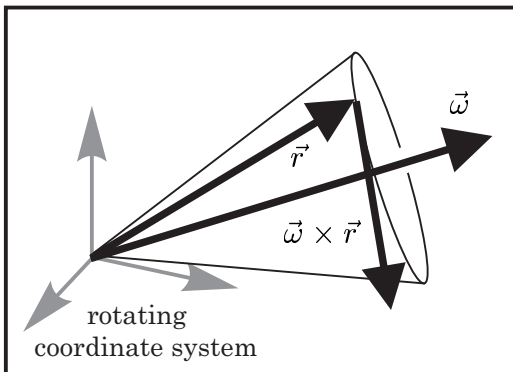
We can find the attitude of the coordinate system by the integration of nonlinear differential equations, provided that the vector $\vec{\omega}$ is a determined function of time.

4.1.4 Vector Differentiation in a Rotating Coordinate System

Relative changes. The geometrical interpretation of a vector equation is nothing else than a closed polygonal line, when we put always one vector behind another one. This vector polygon is independent of a coordinate system; we are free to select every coordinate system for resolving the vector equation into the component notation. The resolution of a vector equation into components brings us three scalar equations.

Often, motion is treated in dynamics; and the use of a rotating system as alternative to an inertial system can reduce the complexity of a vector component notation considerably. All (Cartesian) coordinate systems are legal to resolve the vectors, and this includes moving coordinate systems. However, resolving a vector equation into components of a rotating system involves one problem: the total time derivative of a vector cannot be formed any more by a simple differentiation of its three components. We have to take into account that, in a rotating system, vectors change in relation to the system, and vectors change because the system is moving. For example, consider the location vector of an object that is in rest with respect to the inertial space. The total time derivative of the location vector is the vector of the inertial velocity; in the example it must be a null-vector indicating that the object has no velocity. The inertial velocity must be a null-vector in every coordinate system. In a rotating coordinate system, the component notation of a constant vector (not moving with respect to inertial space) is a function of time, not constant considering a change in time. We never get a null-vector when we simply differentiate the components of this vector. Differentiation of the components of the location vector yields the vector of the velocity relative to the rotating system. More important is the vector of the inertial velocity. To get it, we have to consider not only changes of the vector relative to the system but also changes which come from the fact that the system itself is changing.

Absolute changes. Fortunately, the calculation of the total time derivative of a vector in a rotating system is not a complicated procedure when we know the rotation velocity vector $\vec{\omega}$ of the system. The total time derivative (with respect to inertial space) consists of the relative time derivative (with respect to the moving system, differentiation of the three components) and the cross product of the angular velocity vector and the vector itself. For example, applied to the vectors \vec{r} and \vec{v} the rule is:



$$\begin{aligned} \left(\frac{d\vec{r}}{dt} \right)_{inertial} &= \left[\frac{d\vec{r}}{dt} \right]_{relative} + \vec{\omega} \times \vec{r} \\ \left(\frac{d\vec{v}}{dt} \right)_{inertial} &= \left[\frac{d\vec{v}}{dt} \right]_{relative} + \vec{\omega} \times \vec{v} \end{aligned} \quad (4-10)$$

Relative velocity and inertial velocity. The scheme of equation (4-10) is of fundamental importance to dynamics, since it applies to all other vector quantities too (not only to the location vector \vec{r} and the velocity vector \vec{v}). The relation is a geometrical formula. The time derivative of a vector signifies the change of this vector with respect to time; but two different changes can be meant: the change of the vector in relation to the coordinate system, or the change of the vector in relation to inertial space. Differentiation of the three components of the vector yields the change in relation to the coordinate system. The change of the vector caused by the rotation of the coordinate system equals the cross product of rotation vector and vector itself. Addition of both changes yields the absolute change, the change of the vector with respect to inertial space. It must be emphasized that equation (4-10) is a vector equation (theoretically it can be written in components of any coordinate system). Since the relative change is formed by the differentiation of the three components of the vector, just a component notation of the moving system makes any sense. In scheme (4-10) all vectors must be resolved into components of the moving system.

The total differentiation of the location vector \vec{r} with respect to time must bring us components of the inertial velocity vector $\vec{v} = (u, v, 0)$. Therefore we can conclude:

$$\begin{pmatrix} u \\ v \\ 0 \end{pmatrix} = \begin{pmatrix} \dot{r} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \times \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \dot{r} \\ r\omega_3 \\ -r\omega_2 \end{pmatrix} \quad (4-11)$$

The velocity vector \vec{v} consists of the components u and v ; component u is the relative velocity of the vehicle; and component v is introduced by the rotation of the coordinate system. Equation (4-11) confirms that the vector \vec{v} has no component in 2-direction, and that term v/r is the third component of the rotation vector:

$$u = \dot{r} \quad (4-12)$$

$$v = r \omega_3 \quad (4-13)$$

$$0 = \omega_2 \quad (4-14)$$

Acceleration with respect to inertial space. We have to apply scheme (4-10) a second time to calculate the absolute acceleration $\ddot{\vec{r}}$ of the spacecraft. This time we differentiate the inertial velocity vector $\vec{v} = (u, v, 0)$; the acceleration vector becomes:

$$\ddot{\vec{r}} = \begin{pmatrix} \dot{u} \\ \dot{v} \\ 0 \end{pmatrix} + \begin{pmatrix} \omega_1 \\ 0 \\ v/r \end{pmatrix} \times \begin{pmatrix} u \\ v \\ 0 \end{pmatrix} = \begin{pmatrix} \dot{u} - v^2/r \\ \dot{v} + uv/r \\ v \omega_1 \end{pmatrix} \quad (4-15)$$

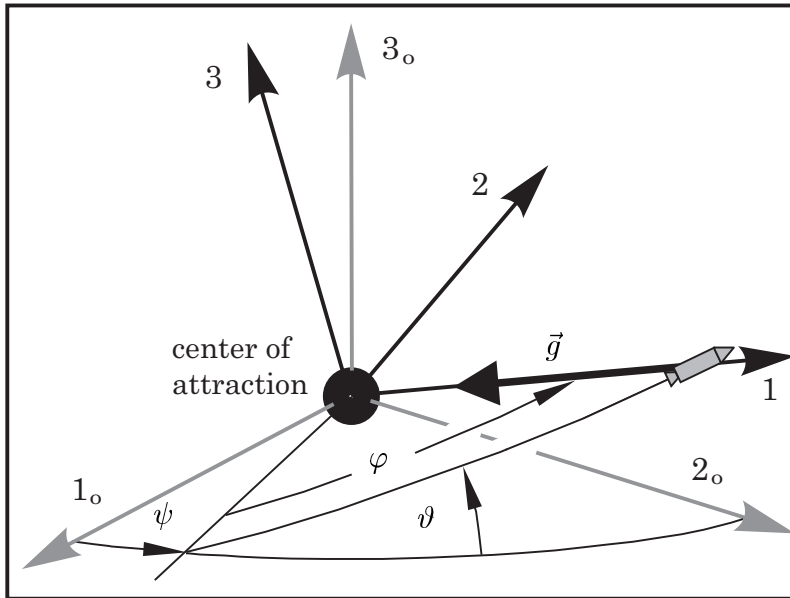
The acceleration with respect to a moving system is quite unimportant in dynamics; because the only acceleration which we actually notice inside a moving vehicle is the acceleration with respect to inertial space. Sir Isaac Newton formulated his laws for the acceleration of a particle that moves in inertial space. It may sound confusing: the equation (4-15) represents the acceleration with respect to inertial space, but the vector $\ddot{\vec{r}}$ is written using a component notation of the moving coordinate system.

4.2. Mechanics of Rocket-Powered Flight

The equations of motion of a mechanical system are a set of differential equations, derivable by the application of Newton's laws, for example. These equations describe the behaviour of the system under the influence of external forces and restrictions. There are four forces which act on a flying vehicle: gravity, drag and lift, and thrust.

4.2.1 The Gravitational Force

Attraction by the gravity of a celestial body. Every vehicle that moves in the central gravitational field of a celestial body experiences a gravitational attraction. The attractive force is inversely proportional to the square of the distance of the attracted vehicle from the center of the celestial body. At any location in the central gravitational field, the gravitational force points towards the center of gravitation. The vector of the gravitational acceleration \vec{g} is the gravitational force divided by the mass of the vehicle; the component notation of \vec{g} is quite simple when we make use of our moving coordinate system. The vehicle moves always on the 1-axis, located by the vector $\vec{r} = (r, 0, 0)$. The origin of the coordinate system is assigned to the center of gravitation. Conventionally, symbol γ is used to characterize the gravitational constant; γ is proportional to the mass of the celestial body. We denote vector \vec{g} as:



$$\vec{g} = \begin{pmatrix} -\gamma/r^2 \\ 0 \\ 0 \end{pmatrix}$$

$$g = \gamma/r^2 \quad (4-16)$$

$$\text{sun: } \gamma_{\odot} \approx 1.3271244 \cdot 10^{20} \text{ [m}^3/\text{s}^2\text{]}$$

$$\text{earth: } \gamma_{\oplus} \approx 3.9866625 \cdot 10^{14} \text{ [m}^3/\text{s}^2\text{]}$$

$$\text{moon: } \gamma_{\text{☾}} \approx 4.89804 \cdot 10^{12} \text{ [m}^3/\text{s}^2\text{]}$$

The gravitational field of an oblate celestial body with irregular mass distribution is not exactly symmetrical (for example the planet earth has continents and oceans). Thus, long-term orbit prediction and some other problems of orbital mechanics have to consider the perturbations which come from irregularities in the gravitational field. However, a representation of the vector of the gravitational acceleration in the form of equation (4-16) is sufficiently accurate for many trajectory computation problems.

Potential energy. Energy is the ability to work. Work is produced when a force acts upon a particle that moves through a distance in the direction of this force (the force is multiplied with the infinitesimal distance and integrated). Consider a spacecraft (mass M) that moves inside a central gravitational field, at the distance r_1 from the center of the celestial body. The celestial body attracts the vehicle (gravitational acceleration $g = \gamma/r^2$, gravitational force $M \cdot g$). Potential energy is stored in the mass M when the spacecraft moves against the gravitational force (on any way) from the distance r_1 to the higher distance r_2 . The equivalent work W is a negative value, it indicates that energy is consumed to transport the mass to a higher altitude. Thus:

$$W = \int_{r_1}^{r_2} \left(-\frac{M \cdot \gamma}{r^2}\right) dr = M\left(\frac{\gamma}{r_2} - \frac{\gamma}{r_1}\right) \quad (4 - 17)$$

The maximum potential energy which can be stored is the potential energy at infinite distance ($r_2 \rightarrow \infty$). We can define the field of (mass specific) potential energy as:

$$e_{potential} = -\gamma/r \quad (4 - 18)$$

The potential energy has a negative sign everywhere inside the field of gravitation. When the vehicle moves away from the center of gravity, its potential energy grows; but the absolute value diminishes since the potential energy is negative everywhere. It is unimportant which way the spacecraft takes to get from the lower altitude to the higher altitude when we consider just the change in the potential energy; and it is also not necessary that the lower and the higher altitude are located on the same radial in the central gravitational field. The potential energy assumes its highest value at an infinite distance from the gravitational center ($e_{potential} = 0$ for $r \rightarrow \infty$).

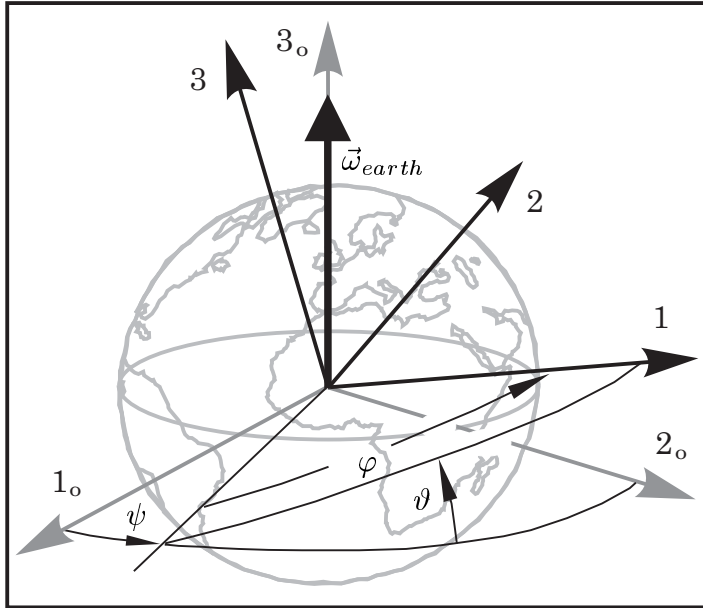
Assume that the vector $\vec{x} = (x_1, x_2, x_3)$ defines a location, and $r = \sqrt{x_1^2 + x_2^2 + x_3^2}$. We find the vector of the gravitational acceleration \vec{g} when we differentiate the potential field of gravity $e_{potential}$ with respect to the coordinates \vec{x} . We can write:

$$\vec{g} = - \begin{pmatrix} \partial(-\gamma/r)/\partial x_1 \\ \partial(-\gamma/r)/\partial x_2 \\ \partial(-\gamma/r)/\partial x_3 \end{pmatrix} \quad (4 - 19)$$

The gravitational acceleration vector is the gradient of the potential energy field; because of this, the gravitational force is called a “conservative force”. This means that the energy which is necessary for the transportation of the mass from a lower altitude to a higher altitude is not lost (at least in the sense of mechanics). When the mass falls back from the higher location to the lower location, all energy comes back (for example in form of kinetic energy when the mass is accelerated, or in form of heat when the energy is dissipated). The other forces which act on the flying vehicle are drag, lift and thrust. Compare that these forces are not conservative forces, because we cannot formulate them as gradient vectors of potential fields. For example, when the drag force converts mechanical energy into heat, mechanical energy gets lost (heat is an energy form which cannot be converted into kinetic energy without losses). The reason is that there is no potential drag field with a drag force as gradient vector.

4.2.2 Aerodynamic Forces

Velocity of the atmosphere. Ascending space launchers and returning space gliders experience quite strong aerodynamic forces. What we call “wind” is usually the motion of the atmosphere in relation to the surface of the earth. Space vehicles are much faster than the velocity of the wind, usually; and for the calculation of nominal “reference” trajectories we are allowed to consider a calm atmosphere with no wind. However, even when we consider no wind, the atmosphere has a velocity with respect to inertial space: the atmosphere is carried around with the rotation of the earth.



We consider the earth as a gyro with a stabilized rotation axis. The vector of the earth rotation velocity $\vec{\omega}_{earth}$ points from the middle of the earth to the geographical north pole ($\omega_{earth} = 360^\circ/\text{day}$). An inertial coordinate system, not carried around with the daily rotation of the earth, is centered in the middle of the earth. Its 3_o-axis points always towards the north pole; and the earth's equator lies in the 1_o-2_o plane. The system is an inertial reference for the moving system.

Our problem is to resolve the rotation vector of the earth $\vec{\omega}_{earth}$ in a component notation of the moving system, using Euler angles or Cardan angles alternatively. With help of the transformation matrix (4-4) we may write down:

$$\vec{\omega}_{earth} = \omega_{earth} \begin{pmatrix} \sin \varphi \sin \vartheta \\ \cos \varphi \sin \vartheta \\ \cos \vartheta \end{pmatrix} = \omega_{earth} \cdot \begin{pmatrix} \sin \Theta \\ \sin \Omega \cos \Theta \\ \cos \Omega \cos \Theta \end{pmatrix} \quad (4-20)$$

The velocity of the atmosphere $\vec{v}_{atmosphere}$ (carried around with the earth rotation) is a function of the geographical latitude. We can find it when we form the vector “cross” product of the earth rotation vector $\vec{\omega}_{earth}$ and the location vector \vec{r} . Thus:

$$\begin{aligned} \vec{v}_{atmosphere} &= \vec{\omega}_{earth} \times \vec{r} \\ &= r \cdot \omega_{earth} \begin{pmatrix} 0 \\ \cos \vartheta \\ -\cos \varphi \sin \vartheta \end{pmatrix} = r \cdot \omega_{earth} \begin{pmatrix} 0 \\ \cos \Omega \cos \Theta \\ -\sin \Omega \cos \Theta \end{pmatrix} \end{aligned} \quad (4-21)$$

Equation (4-21) defines the velocity of the atmosphere with respect to inertial space. The vector is written using a component notation of the moving coordinate system.

Relative velocity of the atmosphere. Now we have to find the velocity of the atmosphere with respect to the flying vehicle, in order to calculate the atmospheric forces. The difference between the velocity of the atmosphere and the velocity of the vehicle is called “incident wind”. We use the term \vec{w} to describe this vector quantity:

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} -u \\ r \omega_{earth} \cos \vartheta - v \\ -r \omega_{earth} \sin \vartheta \cos \varphi \end{pmatrix} = \begin{pmatrix} -u \\ r \omega_{earth} \cos \Omega \cos \Theta - v \\ -r \omega_{earth} \sin \Omega \cos \Theta \end{pmatrix}$$

$$w = \sqrt{w_1^2 + w_2^2 + w_3^2} \quad (4-22)$$

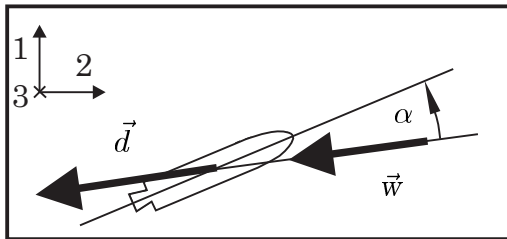
Atmospheric density. Aerodynamic forces are well-known to be approximately proportional to the aerodynamic pressure Q , defined by the equation $Q = \frac{1}{2} \cdot \rho w^2$. Consequently, the aerodynamic forces are also functions of the local air density $\rho(r)$. Sometimes we can approximate $\rho(r)$ using an exponentially declining function of r :

$$\rho(r) = \rho_o e^{-(r - R_{earth})/h_{scale}} \quad (4-23)$$

with: $\rho_o = 1.225 \text{ kg/m}^3$; $R_{earth} = 6378. \text{ km}$; $h_{scale} = 8.333 \text{ km}$

Assumed is a spherical earth with the radius R_{earth} . The term ρ_o defines the air density at “sea-level”; and h_{scale} is the atmospheric “scale-height”. In reality, the earth is oblate and the atmospheric density depends on the latitude of the geographical region and its climate. In fact it can deviate from the reference given by the exponential approximation (4-23) considerably, particularly when the flight altitude is high. Experience in ascent trajectory computation shows us that inaccuracies in the air density have just a small influence on the calculated payload capacity of a space launcher. Accurate aerodynamics is more important in reentry trajectory calculations; where inaccurately defined air density causes uncertainties in the time functions of the aerodynamic drag and aerodynamic heat flux. When a highly accurate course of these functions is needed, it is better to use a more accurate determination of the atmospheric density (for example, we can consider the scale-height h_{scale} as a function of the altitude and take the actual values from a “local atmospheric density” table).

Deceleration by the drag force. We introduce the angle of attack α to resolve the atmospheric force into two rectangular components. The component that is aligned with the incident air flow \vec{w} is called “drag” (the other component is called “lift”):



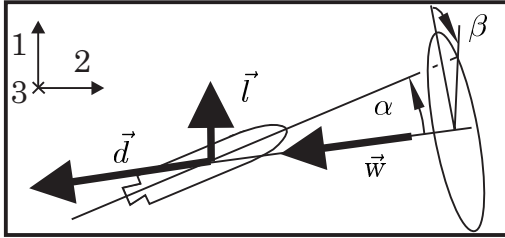
$$\vec{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \frac{1}{2} \rho \frac{C_D A}{M} w \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

$$d = \frac{1}{2} \rho \frac{C_D A}{M} w^2 \quad (4-24)$$

The drag deceleration d is approximately proportional to the aerodynamic pressure.

Acceleration by the lift force. Two angles define the attitude of the flying vehicle with respect to the incident air flow: the angle of attack α and the banking angle β . The “nose” of the vehicle is “up” for $\beta = 0$; full banking to one side means $\beta = \pm 90^\circ$. Also the lift acceleration is approximately proportional to the aerodynamic pressure:

$$\vec{l} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \frac{1}{2} \rho \frac{C_L A}{M} \frac{w}{\sqrt{w_1^2 + w_2^2}} \begin{pmatrix} w w_2 \cos \beta - w_1 w_3 \sin \beta \\ -w w_1 \cos \beta - w_2 w_3 \sin \beta \\ (w_1^2 + w_2^2) \sin \beta \end{pmatrix}$$



$$l = \frac{1}{2} \rho \frac{C_L A}{M} w^2 \quad (4-25)$$

$$\text{for } w_3 \ll w : \vec{l} \approx \frac{l}{w} \begin{pmatrix} w_2 \cos \beta \\ -w_1 \cos \beta \\ w \sin \beta \end{pmatrix}$$

You can readily verify the equations above: form the scalar vector product of drag and lift, it vanishes for any angle β , because the lift is always rectangular to the drag.

Aerodynamic coefficients. Term C_D in equation (4-24) is the drag coefficient; term C_L in equation (4-25) is the lift coefficient. In the equations the aerodynamic coefficients are multiplied by the reference area A and divided by the mass of the vehicle M . This reference area A is just used to get the coefficients without dimension, a widely used convention in aerodynamics. Usually, drag coefficients take a cross sectional area as reference; and lift coefficients take the wing area. However, often aerodynamic coefficients are specified where it is not clear to which area they refer (space launchers may have many sectional areas, usually launchers have no wings). When the aerodynamic coefficients are specified without dimensions it is also necessary to specify the reference area A to which they refer (but it is not necessary to specify a reference area when the coefficients have the dimension of square meters).

The aerodynamic coefficients C_D and C_L are functions of the angle of attack α . The lift is nearly a linear function of α ; and the drag is a quadratic function of α (provided that α is small). The lift vanishes for $\alpha = 0$, but the drag does not. Hence:

$$C_D = C_{D_0} + C_{D\alpha} \cdot \alpha^2 \quad (4-26)$$

$$C_L = C_{L\alpha} \cdot \alpha \quad (4-27)$$

The values of C_{D_0} , $C_{D\alpha}$ and $C_{L\alpha}$ are influenced by the shape and size of the vehicle. The coefficient C_{D_0} describes the drag for a flight with a zero angle of attack; the coefficient $C_{D\alpha}$ describes the drag introduced by the angle of attack; and the coefficient $C_{L\alpha}$ describes the lifting capability of the flying vehicle.

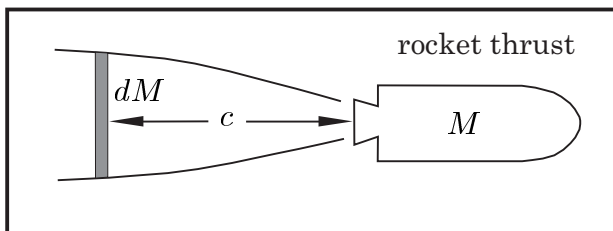
Also the so-called Mach-number (the ratio of air speed and velocity of sound in the ambient atmosphere) has a major influence on the drag and the lift of the vehicle. Usually, this influence on the aerodynamic coefficients C_{D_0} , $C_{D\alpha}$ and $C_{L\alpha}$ is provided in the form of drawings and tables (sometimes also in form of inaccurate equations).

Inaccuracy of aerodynamic coefficients. The equations for the determination of the aerodynamic forces are quite inaccurate. Sometimes, the knowledge about the aerodynamic behaviour of a space vehicle is hardly more than a rough estimate. The density ρ of the ambient air as function of the flight altitude is inaccurately described by equation (4-23). Also equations (4-26) and (4-27) are approximations, they require that the angle of attack is small (for example $\alpha < 20^\circ$). In case of convenience, it is allowed to replace terms α and α^2 by expressions using trigonometric functions (for small angles α we have approximately $\alpha \approx \sin \alpha$ and $\alpha^2 \approx 2(1 - \cos \alpha)$).

Fortunately, for the computation of ascent trajectories it is not always required that we know the aerodynamic behaviour of a space launcher accurately. The drag force has usually a small influence on the nominal payload of a space launcher, because the essential acceleration phase begins when the launcher has already left the atmosphere. The lift force which appears during the aerodynamic flight phase of space launchers, however, is quite harmful to the stabilized attitude and the structural integrity. Therefore, space launchers follow ascent trajectories which require nearly no lift (for example, with an angle of attack smaller than 1°); and the lift force can be ignored when it may not exceed a certain very small value. The lift force cannot be neglected for airborne launchers or winged space gliders which use lift to neutralize weight.

4.2.3 Thrust Acceleration

Thrust magnitude. A rocket motor expels exhaust fumes to generate the thrust force. This thrust force is proportional to the rate of expenditure of propellant m , it is also proportional to the effective exhaust velocity c (in relation to the vehicle). The thrust $c \cdot m$ divided by the rocket mass M constitutes the thrust acceleration, termed s . Even when the thrust force remains constant, the thrust acceleration s increases in time t , because the rocket loses propellant mass. Therefore we can denote:



$$m = -\frac{dM}{dt} \quad (4-28)$$

$$s = \frac{c m}{M} \quad (4-29)$$

Rocket engines are usually not throttleable; and the mass flow rate m is constant. For space launcher engines the effective exhaust velocity c is a function of the ambient aerostatic pressure. The “vacuum” value c_{vacuum} is better than the “sea-level” value $c_{sealevel}$ (up to 25%). Both values depend on the construction of the rocket motor. The exhaust velocity c is approximately a linear function of the air density $\rho(r)$:

$$c = c_{vacuum} - (c_{vacuum} - c_{sealevel}) \frac{\rho(r)}{\rho_0} \quad (4-30)$$

Usually, chemical rocket engines which are designed for the application in the vacuum of space operate with constant exhaust velocity c and constant mass flow rate m .

Thrust of parallel operating rocket engines. We may consider several engines in parallel operation as a single engine when we compute the trajectory of a spacecraft. For example, when two motors (I and II) are operative, we can devise a representative thrust and replace the equations (4-28) and (4-29) by the following relationships:

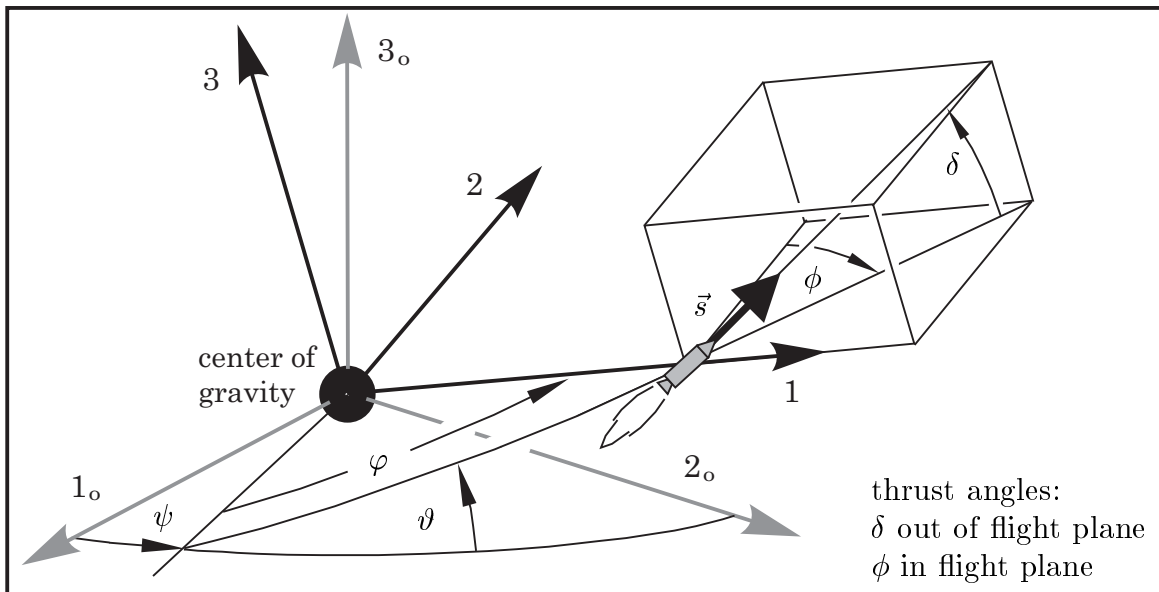
$$\begin{aligned} m_I + m_{II} &= -\frac{dM}{dt} \\ s &= \frac{c_I m_I + c_{II} m_{II}}{M} \end{aligned} \quad (4-31)$$

The representative mass flow rate m and representative exhaust velocity c follow as:

$$\begin{aligned} m &= m_I + m_{II} \\ c &= (c_I m_I + c_{II} m_{II}) / (m_I + m_{II}) \end{aligned} \quad (4-32)$$

However, when we throttle one engine by a reduction of the mass flow rate, this can change the representative exhaust velocity c (assuming c is different for the engines). Switching-off one engine will change instantaneously the representatives m and c .

Thrust direction. Two thrust angles are necessary to define the thrust direction:

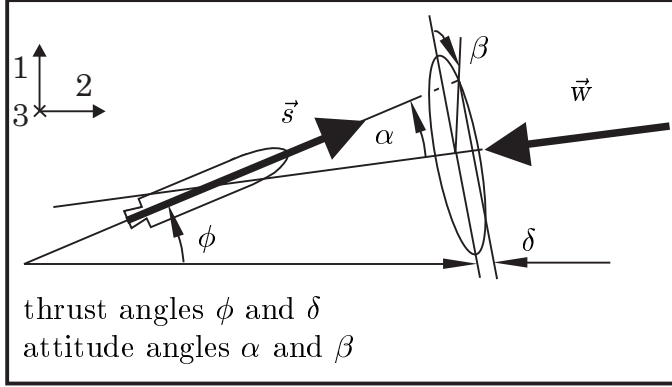


We use the thrust angles ϕ (in flight plane) and δ (out of flight plane) to resolve the thrust vector \vec{s} into a component notation of the moving coordinate system. To get a uniquely determined thrust direction we limit ϕ to $\pm 180^\circ$ and δ to $\pm 90^\circ$:

$$\vec{s} = s \begin{pmatrix} \sin \phi \cos \delta \\ \cos \phi \cos \delta \\ \sin \delta \end{pmatrix} \quad (4-33)$$

The thrust direction is usually restricted during the atmospheric phase of the ascent trajectory of a space launcher (the thrust direction is usually not restricted on orbit).

Thrust direction aligned with the attitude of the vehicle. The steering system of a space launcher has to stabilize every time the appropriate attitude of the vehicle while it ascends through the atmosphere. For example, the vehicle could use aerodynamic fins for attitude stabilization. However, in practice another control method is always preferred for space launchers: so-called “thrust vector control”. The launcher balances its attitude by control torques which come from an appropriately deflected thrust direction. The thrust vector deviates at the right moment just for a few seconds from its nominal line of action (the centerline of the vehicle).



Space launchers cannot control the thrust direction independent from their attitude. For the atmospheric phase of the flight it is better to replace the thrust angles ϕ and δ by expressions which use the attitude angles α and β (with respect to the incident air flow). The launcher has to follow a trajectory where the lift force is strictly limited.

When we use α (the angle of attack) and β (the banking angle), we can limit the lateral acceleration by constraining α . Therefore we resolve the thrust direction into two rectangular components (in the opposite drag direction and in the lift direction):

$$\begin{pmatrix} \sin\phi\cos\delta \\ \cos\phi\cos\delta \\ \sin\delta \end{pmatrix} = -\frac{\cos\alpha}{w} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} + \frac{\sin\alpha}{w\sqrt{w_1^2+w_2^2}} \begin{pmatrix} w w_2 \cos\beta - w_1 w_3 \sin\beta \\ -w w_1 \cos\beta - w_2 w_3 \sin\beta \\ (w_1^2 + w_2^2) \sin\beta \end{pmatrix}$$

$$\approx -\cos\alpha \begin{pmatrix} w_1/w \\ w_2/w \\ 0 \end{pmatrix} + \sin\alpha \begin{pmatrix} \cos\beta w_2/w \\ -\cos\beta w_1/w \\ \sin\beta \end{pmatrix} \quad (4-34)$$

When we want to find the attitude angles α and β from given values of ϕ and δ , we can transform the equation above: $\cos\alpha$ follows when we form the dot vector product of thrust direction and wind direction. Finally we get the following equations:

$$\begin{aligned} \cos\alpha &= (w_1 \sin\phi\cos\delta + w_2 \cos\phi\cos\delta + w_3 \sin\delta)/w \\ \sin\alpha &= \sqrt{1 - \cos^2\alpha} \end{aligned} \quad (4-35)$$

$$\begin{aligned} \cos\beta &= (w_2 \sin\phi - w_1 \cos\phi)\cos\delta / (\sqrt{w_1^2 + w_2^2} \sin\alpha) \\ \sin\beta &= (w_3 \cos\alpha + w \sin\delta) / (\sqrt{w_1^2 + w_2^2} \sin\alpha) \end{aligned} \quad (4-36)$$

However, the use of the thrust angles ϕ and δ is preferable for the final flight phase (when the launcher has left the atmosphere and moves in vacuum of space). The angles ϕ and δ refer to the rotating coordinate system; the attitude angles α and β refer to the direction of the air flow. This “incident wind” does not exist on orbit.

4.2.4 Equations of Motion

Application of Newton's law. When a particle is in motion, the product of mass and velocity vector is called “linear momentum” vector; and the time derivative of the linear momentum vector is equivalent to all the external forces (Newton's law). The mass of the vehicle is not constant because the motor expels propellant mass; and the time derivative of the mass of the vehicle, multiplied by the velocity of the exhaust fumes relative to the vehicle, constitutes the thrust force. We can follow that the change of the velocity in time (the acceleration of the location \vec{r} , equation 4-15) is equivalent to the sum of four accelerations: gravitational acceleration \vec{g} (4-16), thrust acceleration \vec{s} (4-33), drag deceleration \vec{d} (4-24) and lift acceleration \vec{l} (4-25). The equations of motion of a flying vehicle are defined by the differential equation:

$$\ddot{\vec{r}} = \vec{g} + \vec{s} + \vec{d} + \vec{l} \quad (4-37)$$

Resolved into components of the moving system, the equation (4-37) takes the form of:

$$\begin{pmatrix} \dot{u} - v^2/r \\ \dot{v} + uv/r \\ v \omega_1 \end{pmatrix} = \begin{pmatrix} -\gamma/r^2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} s \sin \phi \cos \delta \\ s \cos \phi \cos \delta \\ s \sin \delta \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} + \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} \quad (4-38)$$

The motion of the vehicle “in the actual flight plane” is determined by the integration of the first two components of the vector equation above. We have to consider the relationship $\dot{r} = u$ (equation 4-12). The “motion of the actual flight plane” is determined by the integration of the kinematical Euler equations (system 4-9), where we can alternatively use Euler angles $(\psi, \vartheta, \varphi)$ or Cardan angles (Φ, Θ, Ω) . The third component of the vector equation (4-38) determines the first component of the rotation vector $\vec{\omega}$ of the moving coordinate system (as defined by equation 4-8):

$$\omega_1 = \frac{s \sin \delta + d_3 + l_3}{v} \quad (4-39)$$

Finally we get:

$$\begin{aligned} \dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} + s \sin \phi \cos \delta + d_1 + l_1 \\ \dot{v} &= -\frac{u v}{r} + s \cos \phi \cos \delta + d_2 + l_2 \\ \dot{r} &= u \end{aligned} \quad (4-40)$$

$$\dot{\psi} = \omega_1 \frac{\sin \varphi}{\sin \vartheta}$$

$$\dot{\Phi} = \frac{v}{r} \frac{\cos \Omega}{\cos \Theta}$$

$$\dot{\vartheta} = \omega_1 \cos \varphi$$

$$\dot{\Theta} = \frac{v}{r} \sin \Omega$$

$$\dot{\varphi} = \frac{v}{r} - \omega_1 \sin \varphi \cot \vartheta$$

$$\dot{\Omega} = \omega_1 - \frac{v}{r} \cos \Omega \tan \Theta$$

Equation system (4-40) provides two sets of six first-order differential equations to find the acceleration of a vehicle that flies in three-dimensional space, substituting the three second order differential equations of the vector equation (4-37).

Trajectory integration. The system of the equations of motion (4-40) is valid for many different types of space-flight trajectories. With a few exceptions, analytical solutions are not available. Trajectory integration has to be performed numerically on a computer, employing an integration scheme for a system of differential equations. We start from a well-known initial state and simulate the motion of the vehicle by finding step-wise always the next state, considering the influence of forces which act on the vehicle. The thrust is not zero in case of powered flight ($s \neq 0$); and we can just integrate the trajectory when we know the behaviour of the thrust vector (when s , δ and ϕ are determined functions of the time or the location). Usually, this is not the case, but the thrust vector is (at least in some limits) free for optimization. Trajectory optimization is an important but also quite complicated mathematical problem: the solution depends much on the particular conditions for an individual problem. Trajectory optimization problems are treated in other chapters of this book.

4.2.5 Mechanical Energy of a Trajectory

Kinetic energy. The kinetic energy of a particle (mass M) is the energy which is stored in its motion. To calculate it, we have to substitute the force by the time derivative of the linear momentum $M \cdot dv/dt$, consider $v \cdot dt$ as coordinate and integrate. The kinetic energy becomes $E = 1/2 \cdot Mv^2$, where v is the length of the velocity vector $\vec{v} = (u, v, 0)$. We relate the kinetic energy to the mass of the vehicle M and use the smaller symbol e to denote the specific kinetic energy (pay attention not to confuse e with the symbol e reserved for the Euler number $e = 2.71828\dots$). Thus:

$$e_{kinetic} = \frac{1}{2}(u^2 + v^2) \quad (4-41)$$

Specific energy of the trajectory. The total mechanical energy of the trajectory consists of kinetic energy (4-41) and potential energy (4-18); therefore we can write:

$$e = e_{kinetic} + e_{potential} = \frac{1}{2}(u^2 + v^2) - \frac{\gamma}{r} \quad (4-42)$$

When we form the time derivative \dot{e} of the total specific energy and consider the equations of motion (4-40), we find the following relationship:

$$\begin{aligned} \dot{e} &= u\dot{u} + v\dot{v} + u\gamma/r^2 \\ &= u \cdot (s \sin \phi \cos \delta + d_1 + l_1) + v \cdot (s \cos \phi \cos \delta + d_2 + l_2) \\ &= \vec{v} \cdot \vec{s} + \vec{v} \cdot \vec{d} + \vec{v} \cdot \vec{l} \end{aligned} \quad (4-43)$$

Thus, the change of the energy in time equals the sum of three scalar vector products: velocity times thrust acceleration, velocity times drag deceleration and velocity times lift acceleration. Since the gravity is a conservative force it does not change the energy of the flying vehicle (application of gravity transforms just kinetic energy into potential energy, or vice versa). However, thrust, drag and lift are not conservative forces. When they are effective, they change the mechanical energy e of the flying vehicle.

4.2.6 The Angular Momentum Vector

Angular momentum with respect to the gravitational center. The vector cross product of the location vector \vec{r} and the linear momentum vector $M\vec{v}$ of a flying vehicle (mass M) is termed “angular momentum” \vec{H} of the trajectory, $\vec{H} = \vec{r} \times M\vec{v}$.

The angular momentum \vec{H} is a vector, like the linear momentum $M\vec{v}$. However, the angular momentum refers to a reference point, in contrast to the linear momentum: it is always the point where the location vector \vec{r} starts. When we take the center of gravity as reference point, the specific (mass related) angular momentum \vec{h} becomes:

$$\vec{h} = \vec{r} \times \vec{v} = \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} u \\ v \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ r \cdot v \end{pmatrix}$$

$$h = r \cdot v \quad (4-44)$$

Like all vectors, the angular momentum \vec{h} is independent of a coordinate system (the length h of the angular momentum vector depends just on the reference point). Equation (4-44) denotes the component form of \vec{h} in the moving coordinate system.

Time derivative of the angular momentum vector. According to the “angular momentum law”, another fundamental law of mechanics, the torques of all the external forces must be equivalent to the total time derivative of the angular momentum vector (provided that the reference point is not accelerated). The forces thrust, drag and lift generate torque vectors in the center of the coordinate system, just the gravitational force does not. Therefore, the following relationship must be valid:

$$\left(\frac{d\vec{h}}{dt} \right)_{inertial} = \left[\frac{d\vec{h}}{dt} \right]_{relative} + \vec{\omega} \times \vec{h} = \vec{r} \times (\vec{s} + \vec{d} + \vec{l}) \quad (4-45)$$

When we resolve the vector equation (4-45) into a component notation of the moving coordinate system, we have to observe the rule of vector differentiation (4-10). Thus:

$$\left(\frac{d\vec{h}}{dt} \right)_{inertial} = \begin{pmatrix} 0 \\ 0 \\ \dot{r}v + r\dot{v} \end{pmatrix} + \begin{pmatrix} \omega_1 \\ 0 \\ v/r \end{pmatrix} \times \begin{pmatrix} 0 \\ 0 \\ r v \end{pmatrix} = \begin{pmatrix} 0 \\ -\omega_1 r v \\ \dot{r}v + r\dot{v} \end{pmatrix} \quad (4-46)$$

$$= \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} s \sin \phi \cos \delta + d_1 + l_1 \\ s \cos \phi \cos \delta + d_2 + l_2 \\ s \sin \delta + d_3 + l_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -r (s \sin \delta + d_3 + l_3) \\ r (s \cos \phi \cos \delta + d_2 + l_2) \end{pmatrix}$$

We learn nothing from the first component of this vector equation. The second component confirms equation (4-39). The third component brings us the relationship:

$$\dot{h} = \dot{r}v + r\dot{v} = r(s \cos \phi \cos \delta + d_2 + l_2) \quad (4-47)$$

The result (4-47) can be verified using the equations of the motion (system 4-40). The angular momentum law is another way to establish the equations of the motion.

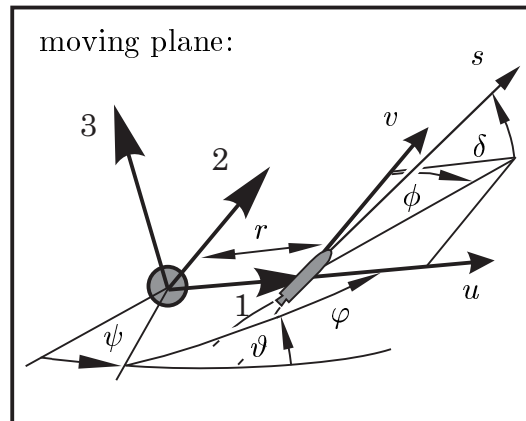
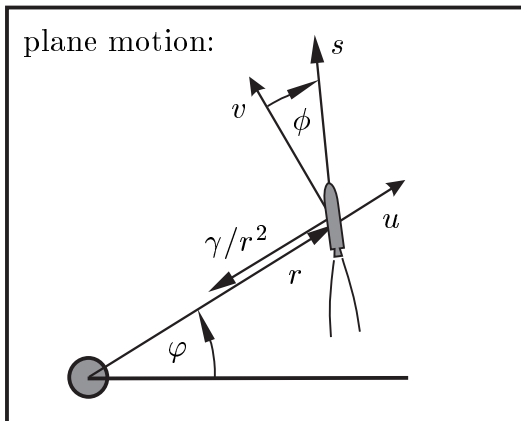
4.3. Orbital Flight

The density of the ambient air flow diminishes rapidly while the rocket climbs through the atmosphere. Finally the vehicle moves in the vacuum of space. We talk about “coast arcs” (sections of conic orbits) when we consider flight intervals with switched-off engine. When the thrust is switched-off, the three-dimensional trajectory of a spacecraft that moves in space near a celestial body becomes a plane orbit (the motion takes place in the “actual flight plane”). A spacecraft that coasts in the gravitational field of a celestial body moves on a circular, elliptic, parabolic or hyperbolic orbit.

4.3.1 Conic Orbits

Equations of motion for orbital flight. Consider a spacecraft which moves in the vacuum of space near a celestial body. The motion is exclusively influenced by the gravitational acceleration \vec{g} of the celestial body and the thrust acceleration \vec{s} of the rocket motor. Without drag and lift, the equation system (4-40) simplifies to:

$$\begin{aligned} \dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} + s \sin \phi \cos \delta & \dot{\psi} &= \frac{s \sin \delta \sin \varphi}{v \sin \vartheta} \\ \dot{v} &= -\frac{u v}{r} + s \cos \phi \cos \delta & \dot{\vartheta} &= \frac{s \sin \delta}{v} \cos \varphi \\ \dot{r} &= u & \dot{\varphi} &= \frac{v}{r} - \frac{s \sin \delta}{v} \sin \varphi \cot \vartheta \end{aligned} \quad (4-48)$$



System (4-48) describes the “plane motion” in a “moving plane”. The time derivatives of the Euler angles ψ and ϑ vanish when the motor is switched-off ($s = 0$), this indicates that node angle and inclination angle remain constant on coast arcs. The orientation of the actual flight plane is stabilized with respect to inertial space. We can use polar coordinates to describe the plane motion by a system of first-order differential equations (r, φ , vertical velocity $u = \dot{r}$ and horizontal velocity $v = r\dot{\varphi}$).

Coast arcs. The solution to the system of differential equations for coast arcs ($s = 0$) is well-known: the spacecraft moves on a conic orbit. We can readily express the components of the velocity u and v and the distance r as functions of the path angle φ :

$$\begin{aligned} u &= \frac{\gamma \varepsilon}{h} \sin(\varphi - \varphi_p) \\ v &= \frac{\gamma}{h} (1 + \varepsilon \cos(\varphi - \varphi_p)) \\ r &= \frac{h^2}{\gamma(1 + \varepsilon \cos(\varphi - \varphi_p))} \end{aligned} \quad (4-49)$$

Terms h , ε and φ_p are the integration constants; term h (4-44) is the value of the angular momentum of the orbit with respect to the center of gravity. Term ε is called “numerical eccentricity” of the orbit. The angle $\varphi - \varphi_p$ is named “true anomaly” of the orbital position, where φ_p is the path angle of the pericenter (radius r_p) and $\varphi_p + 180^\circ$ is the path angle of the apocenter (radius r_a). The position on the orbit where the vehicle passes its nearest distance to the celestial body is called pericenter.

Energy and angular momentum of conic orbits. Trajectories in the central gravitational field of a celestial body obey some “conservation laws”. We can write:

$$e = \text{constant} = \frac{1}{2}(u^2 + v^2) - \frac{\gamma}{r} \quad (4-50)$$

$$h = \text{constant} = r v \quad (4-51)$$

Gravity is the only force that acts on the vehicle. Since the gravity is a “conservative” force it does not dissipate mechanical energy; and the specific mechanical energy e of the orbit must be a constant quantity (compare equation 4-42). Since the gravity force aims everywhere inside the gravitational field exactly at the gravitational center it never causes a torque there; and the angular momentum h of the orbit with respect to the gravitational center must also be a constant quantity (compare equation 4-44).

Numerical eccentricity of conic orbits. When the specific energy e and angular momentum h of the conic orbit are well-known, we can calculate another quantity: it is the important numerical eccentricity ε of the orbit:

$$\varepsilon = \sqrt{2e \left(\frac{h}{\gamma}\right)^2 + 1} = \frac{r_a - r_p}{r_a + r_p} \quad (4-52)$$

$$\text{circular orbits :} \quad \varepsilon = 0; \quad e < 0$$

$$\text{elliptic orbits :} \quad \varepsilon < 1; \quad e < 0$$

$$\text{parabolic orbits :} \quad \varepsilon = 1; \quad e = 0$$

$$\text{hyperbolic orbits :} \quad \varepsilon > 1; \quad e > 0$$

The eccentricity ε is by definition always positive; the value of ε defines which type of orbit is under consideration (a circular, elliptic, parabolic or hyperbolic orbit).

Pericenter and apocenter. The orbit is characterized by the pericenter distance r_p and the apocenter distance r_a (for earth orbits: perigee and apogee; for solar orbits: perihelion and aphelion). Radius r_p is the nearest distance from the center of gravity, it is reached for $\varphi - \varphi_p = 0$ (equation 4-49). At this location the horizontal velocity v_p assumes its maximum value, while the vertical velocity vanishes, $u_p = 0$. On the opposite side of the celestial body there is apocenter position, for $\varphi - \varphi_p = 180^\circ$. The apocenter can just be reached on elliptic orbits ($\varepsilon < 1$); the horizontal velocity v_a assumes its minimum value there, and the vertical velocity vanishes again, $u_a = 0$. The apocenter is a hypothetical position for hyperbolic or parabolic orbits; it is impossible to reach this position when $\varepsilon \geq 1$. For circular orbits ($\varepsilon = 0$) we have $r_p = r_a = h/\sqrt{-2e}$; for parabolic orbits ($\varepsilon = 1$) we have $r_p = h^2/(2\gamma)$ and $r_a \rightarrow \infty$.

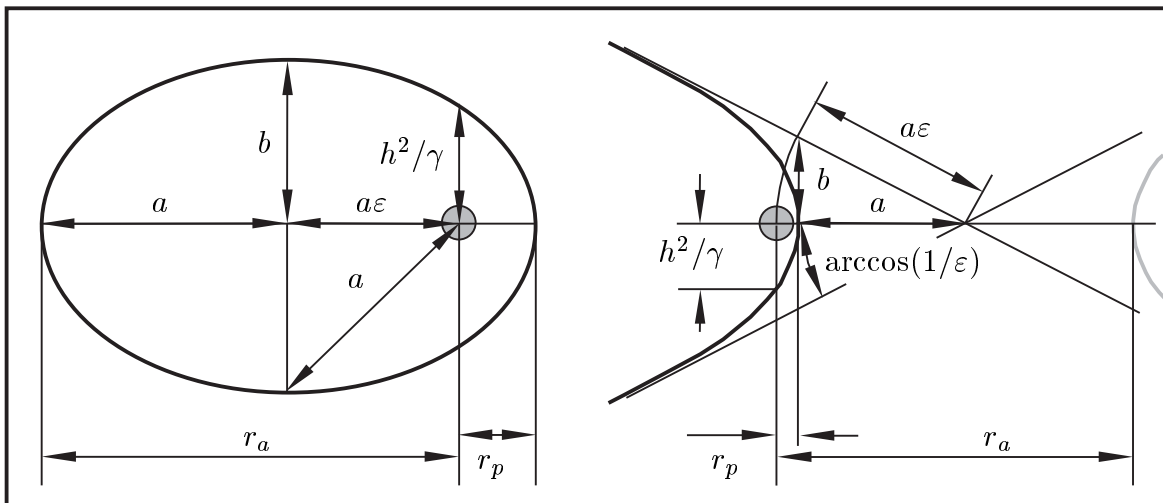
Geometry of conic orbits. Strictly speaking, circular or parabolic orbits do not exist in reality, because ε is never exactly zero or exactly one. The cases which are important in practice are elliptic orbits and hyperbolic orbits. The equations (4-53) show that the quantities a and b (semimajor axis and semiminor axis) are differently defined for elliptic or hyperbolic orbits. We may consider circular orbits as degenerated ellipses; the equations simplify to $a = b = r_p = r_a$ when $\varepsilon = 0$. We can see a parabolic orbit as a degenerated hyperbole; but a definition of the semiaxes a and b makes no sense when $\varepsilon = 1$. The following set of equations is valid:

elliptic orbits:

$$\begin{aligned} r_p &= \frac{h^2}{\gamma(1+\varepsilon)} = a(1-\varepsilon) \\ r_a &= \frac{h^2}{\gamma(1-\varepsilon)} = a(1+\varepsilon) \\ a &= \frac{r_a + r_p}{2} = \frac{h^2}{\gamma(1-\varepsilon^2)} = \frac{-\gamma}{2e} \\ b &= a\sqrt{1-\varepsilon^2} = h/\sqrt{-2e} = h\sqrt{a/\gamma} \end{aligned}$$

hyperbolic orbits:

$$\begin{aligned} r_p &= \frac{h^2}{\gamma(\varepsilon+1)} = a(\varepsilon-1) \\ r_a &= \frac{h^2}{\gamma(\varepsilon-1)} = a(\varepsilon+1) \\ a &= \frac{r_a - r_p}{2} = \frac{h^2}{\gamma(\varepsilon^2-1)} = \frac{\gamma}{2e} \\ b &= a\sqrt{\varepsilon^2-1} = h/\sqrt{2e} = h\sqrt{a/\gamma} \end{aligned} \quad (4-53)$$

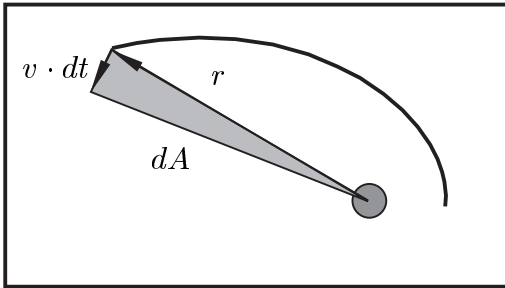


4.3.2 Flight Time on Conic Orbits

Kepler's laws. Modern sciences started at the beginning of the seventeenth century. The medieval geocentric cosmic system was displaced by the recognition that the earth is not the center of the universe. The astronomer Johannes Kepler (1571-1630) contributed essentially to the beginning of natural sciences with his studies on planets. Kepler formulated his famous three statements about the motion of planets:

1. Planets move on elliptic orbits, where the sun is in one of the focal points.
2. In a time interval Δt the connection line sun-planet (location vector \vec{r}) sweeps out the area ΔA ; for all planets ΔA is a linear function of the time interval Δt .
3. There exists a relationship between the orbital period T of the orbit of a planet and the semimajor axis a ; the ratio T^2/a^3 is the same for all planets.

Kepler's first law follows from the analytical integration of the equations of motion: the equation system (4-49) defines an elliptic orbit for $\varepsilon < 1$. Kepler's second law is nothing else than a geometrical interpretation of the conservation law for the angular momentum of the orbit (equation 4-51): remember that the expression $h = rv$ is constant on a conic orbit; r is the length of location vector \vec{r} , and v is the velocity component perpendicular to \vec{r} . The area dA of the infinitesimally small triangle formed in the time interval dt by a change of the location vector \vec{r} is a linear function of dt :



$$dA = \frac{1}{2} \cdot r \cdot v \, dt = \frac{h}{2} \, dt \quad (4-54)$$

It follows that dA/dt is a constant quantity.

To verify Kepler's third law we form the time integral of equation (4-54) for a complete orbital period T . The result of the integration of the left hand side is the area A ; integration of the right side yields the expression $h \cdot T/2$. The area A of the ellipse is also given by the equation $A = \pi ab$, when a and b are the semiaxes (the area of an ellipse is the area of a circle πa^2 , scaled by the ratio b/a of the semiaxes a and b):

$$T = \frac{2}{h} \cdot \pi \cdot ab = 2\pi \sqrt{a^3/\gamma} \quad \left(= \pi\gamma/\sqrt{|2e^3|} \right) \quad (4-55)$$

The semiminor axis b can be replaced by the expression $h\sqrt{a/\gamma}$, (equation 4-53). It is the essential statement of Kepler's third law, that the orbital period T of a conic orbit is exclusively a function of the value of the major semiaxis a (or the energy e).

Modern celestial mechanics can easily verify Kepler's laws. However, it is remarkable that even today we cannot formulate more "simple relationships" in planetary motion; and Kepler found his three laws by the accurate comparative analysis of observations.

Kepler's flight time equation. The equation system (4-49) determines $u(\varphi)$, $v(\varphi)$ and $r(\varphi)$; however, velocity and distance are given as functions of the coordinate φ , and not as functions of the time t . The time interval $\Delta t = t_2 - t_1$ elapses when the spacecraft coasts on a conic orbit from position φ_1 to position φ_2 . Therefore we have:

$$\dot{\varphi} = \frac{v}{r} = \frac{\gamma^2}{h^3} \cdot (1 + \varepsilon \cos(\varphi - \varphi_p))^2 \quad (4 - 56)$$

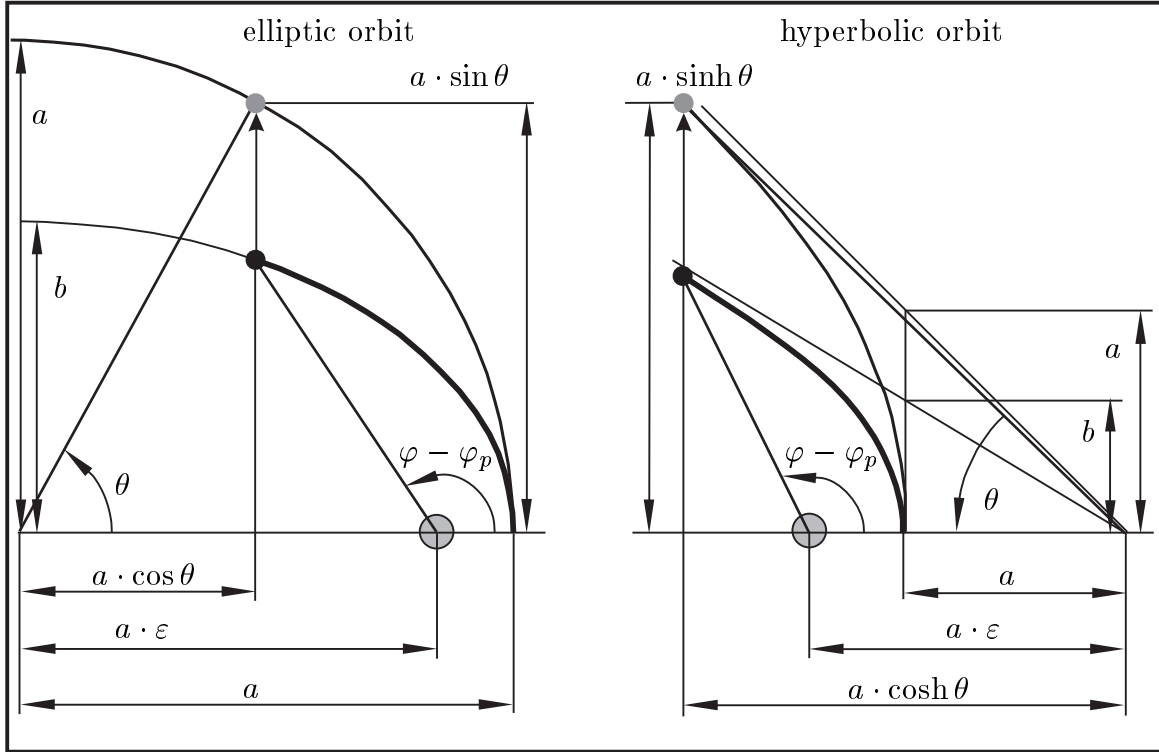
To compute the flight time Δt , we have to transform the equation above and integrate:

$$\Delta t = \frac{h^3}{\gamma^2} \int_1^2 \frac{d\varphi}{(1 + \varepsilon \cos(\varphi - \varphi_p))^2} \quad (4 - 57)$$

Unfortunately, this involves two problems. First, the integral takes quite different forms for $\varepsilon < 1$, $\varepsilon = 1$ or $\varepsilon > 1$ (for elliptic, parabolic, or hyperbolic orbits). Second, the results are transcendental equations (to solve them, numerical iteration processes have to be applied). We can compute directly the time interval Δt between two orbital positions with predefined true anomaly $\varphi - \varphi_p$, but we cannot transform the relationship into a reverse representation: the true anomaly $\varphi - \varphi_p$ for a predefined time interval Δt has to be computed iteratively by a numerical iteration scheme.

The problem has been known for some centuries already: for nautical navigation purposes it was of high practical importance to know the relationship between time and position of the planets. Many accurate and efficient numerical computation schemes have been developed to solve this problem. Today, position finding is done via navigation satellites, but actually this involves the same equations. With a computer it is not a problem anymore to find the accurate solution to a transcendental equation; however, a transparent representation of the flight time equation is helpful anyway.

Kepler has introduced the "eccentric anomaly" θ , besides the "true anomaly" $\varphi - \varphi_p$ another angle to describe the position on the orbit. In order to find this angle θ , the actual position has to be projected to a corresponding position on a reference orbit. The reference orbit is a circle for elliptic orbits around the same center point ($b = a$); in case of hyperbolic orbits it is a hyperbole with rectangular asymptotic lines ($b = a$). The "eccentric anomaly" θ is the angle between the line that connects the center of the reference figure with the orthogonal projection of the actual position to the reference figure (circle or hyperbole), and the major axis of the orbit. The flight time is the ratio of "mean anomaly" (for elliptic orbits the expression $\theta - \varepsilon \sin \theta$, for hyperbolic orbits the expression $\varepsilon \sinh \theta - \theta$) and "mean motion" (the expression $2\pi/T$ or $\sqrt{\gamma/a^3}$, it is the angular velocity on a circular orbit with exactly the same period T). Thus, to compute the flight time Δt between the two positions θ_1 and θ_2 , we exclusively have to know eccentricity ε and period T of the orbit. In case of an elliptic or circular orbit we must pay attention, there is the possibility of several orbital revolutions (number n) before the spacecraft reaches the final position (the eccentric anomaly θ is greater than π when the true anomaly $\varphi - \varphi_p$ is greater than π). A circular orbit is present for $\varepsilon = 0$, and the eccentric anomaly θ is identical with the true anomaly $\varphi - \varphi_p$. A parabolic orbit is present for $\varepsilon = 1$; with undefined period T a completely different equation has to be used for the flight time calculation.



elliptic orbit:

$$\sin \theta = \frac{\sqrt{1 - \varepsilon^2} \sin(\varphi - \varphi_p)}{1 + \varepsilon \cos(\varphi - \varphi_p)}$$

$$\cos \theta = \frac{\varepsilon + \cos(\varphi - \varphi_p)}{1 + \varepsilon \cos(\varphi - \varphi_p)}$$

$$\theta = 2 \arctan\left(\sqrt{\frac{1 - \varepsilon}{1 + \varepsilon}} \tan \frac{\varphi - \varphi_p}{2}\right)$$

$$\Delta t = \frac{T}{2\pi} [\theta - \varepsilon \sin \theta]_1^2 + nT$$

hyperbolic orbit:

$$\sinh \theta = \frac{\sqrt{\varepsilon^2 - 1} \sin(\varphi - \varphi_p)}{1 + \varepsilon \cos(\varphi - \varphi_p)}$$

$$\cosh \theta = \frac{\varepsilon + \cos(\varphi - \varphi_p)}{1 + \varepsilon \cos(\varphi - \varphi_p)}$$

$$\theta = 2 \operatorname{arctanh}\left(\sqrt{\frac{\varepsilon - 1}{\varepsilon + 1}} \tan \frac{\varphi - \varphi_p}{2}\right)$$

$$\Delta t = \frac{T}{2\pi} [\varepsilon \sinh \theta - \theta]_1^2 \quad (4 - 58)$$

circular orbit: $\Delta t = \frac{T}{2\pi} [\varphi]_1^2 + nT \quad (4 - 59)$

parabolic orbit: $\Delta t = \frac{h^3}{\gamma^2} \left[\frac{1}{2} \tan\left(\frac{\varphi - \varphi_p}{2}\right) + \frac{1}{6} \tan^3\left(\frac{\varphi - \varphi_p}{2}\right) \right]_1^2 \quad (4 - 60)$

For hyperbolic orbits the flight time calculation involves “hyperbolic functions”:

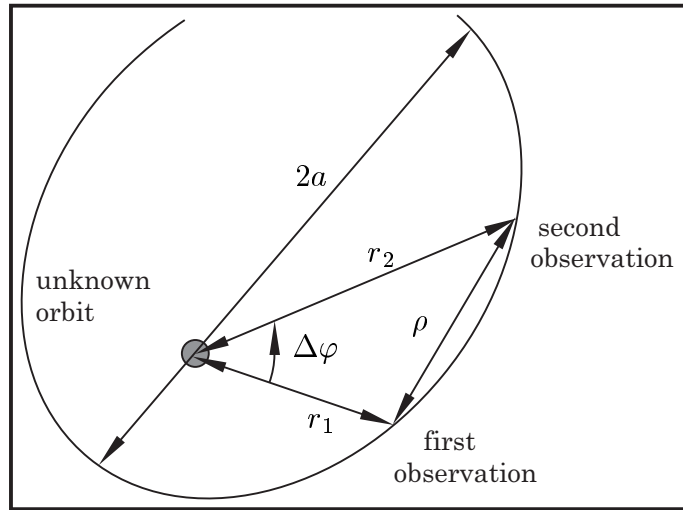
$$\sinh(x) = \frac{e^x - e^{-x}}{2}; \quad \cosh(x) = \frac{e^x + e^{-x}}{2}; \quad \operatorname{arctanh}(x) = \frac{1}{2} \ln \frac{1 + x}{1 - x}$$

The flight time equations (4-58) can be brought to a unique representation for elliptic and hyperbolic orbits when we use mathematics of complex numbers ($i = \sqrt{-1}$); but the interesting theory is for practical applications more confusing than helpful.

Lambert's theorem. Sometimes it is the problem to find a conic orbit which fits between two locations in space and satisfies a predefined transfer time. This problem is called "Lambert problem", named after the astronomer J.H.Lambert (1728-1777). It is a standard problem of celestial mechanics to compute the orbit of an object from observation data. Measured are usually at two moments the distances r_1 and r_2 , and the angle $\Delta\varphi = \varphi_2 - \varphi_1$ between the radial r_1 and the radial r_2 . The distance ρ between these two locations can be calculated using the theorem of the cosine. Thus:

$$\rho = \sqrt{r_1^2 + r_2^2 - 2r_1r_2 \cos(\varphi_2 - \varphi_1)} \quad (4 - 61)$$

The "Lambert theorem" serves now to calculate the elements of the conic orbit between the two locations. The theorem is nothing else than an elegant transformation of Kepler's flight time equation; it demonstrates that we can compute the flight time $\Delta t = t_2 - t_1$ on the conic section directly when we know just three values: the semimajor axis a , the sum $r_1 + r_2$ and the distance ρ :



elliptic orbit:
$$\Delta t = \sqrt{a^3/\gamma} \cdot ((\varsigma_1 - \varsigma_2) - (\sin \varsigma_1 - \sin \varsigma_2)) \quad (4 - 62)$$

$$\sin \varsigma_1 = \frac{1}{2} \sqrt{\frac{r_1 + r_2 + \rho}{a}} ; \sin \varsigma_2 = \frac{1}{2} \sqrt{\frac{r_1 + r_2 - \rho}{a}}$$

parabolic orbit:
$$\Delta t = \frac{1}{6\sqrt{\gamma}} \cdot (\sqrt{(r_1 + r_2 + \rho)^3} - \sqrt{(r_1 + r_2 - \rho)^3}) \quad (4 - 63)$$

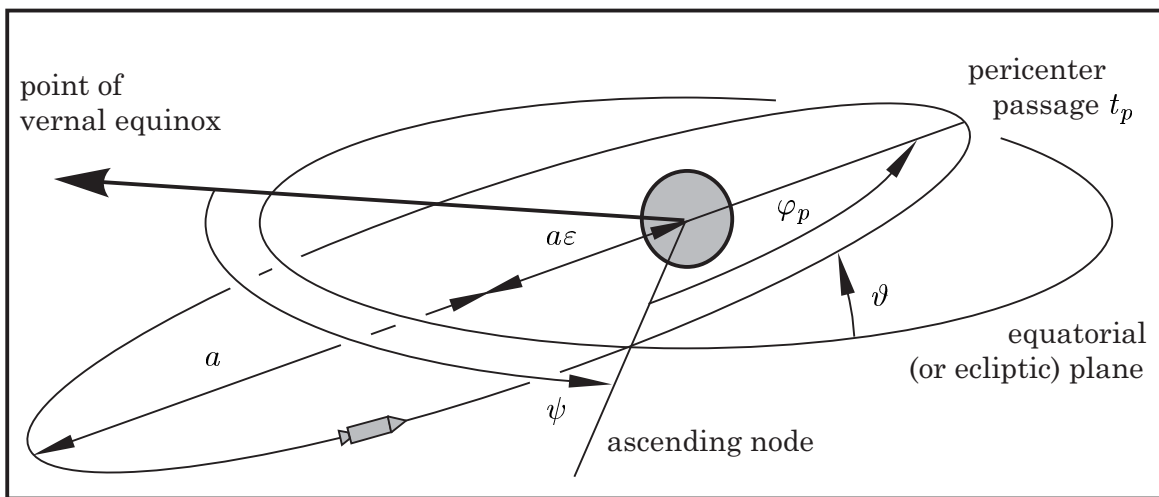
hyperbolic orbit:
$$\Delta t = \sqrt{a^3/\gamma} \cdot (-(\varsigma_1 - \varsigma_2) + (\sinh \varsigma_1 - \sinh \varsigma_2)) \quad (4 - 64)$$

$$\sinh \varsigma_1 = \frac{1}{2} \sqrt{\frac{r_1 + r_2 + \rho}{a}} ; \sinh \varsigma_2 = \frac{1}{2} \sqrt{\frac{r_1 + r_2 - \rho}{a}}$$

Lambert's theorem is a transcendental equation, exactly like Kepler's equation. An iterative computation process has to be applied anyway when it is the task to calculate the semimajor axis a of an orbit from observation data (r_1 , r_2 , $\Delta\varphi$ and Δt). Once a has been iterated, the orbital period T and the specific energy e of the orbit follow directly (equation 4-55); but the calculation of the other orbital elements remains cumbersome. To find the eccentricity ε , we have to substitute the expression $h^2/\gamma = a(1 - \varepsilon^2)$ in the equation system (4-49), then insert the distances $r_1(\varphi)$ and $r_2(\varphi)$, and finally eliminate the pericenter angle φ_p from the two nonlinear equations.

4.3.3 Orbital Elements

Definition of conic orbits. A material body which coasts in the gravitational field of a celestial body moves on a conic orbit (usually an ellipse or hyperbole, exceptionally a circle or parabole). The orbit is defined when the location vector \vec{r} and the velocity vector \vec{v} are known at any instant t . The location vector \vec{r} and the velocity vector \vec{v} change in time t (obeying six first-order differential equations 4-48). Therefore exactly six terms are necessary to define the orbit. The “time coordinate” is the seventh element; it determines the exact position of the moving object on the orbit. It is possible to specify the orbit by a set of six time-invariant elements. These data are called “orbital elements” of the object.



Semimajor axis and eccentricity. The size of the orbit is defined by its mechanical energy e and its angular momentum h (equations 4-50, 4-51). Remember that these two quantities are “conservation laws” for conic orbits. Once we know the values of e and h , we can readily compute the semi-axes a and b , or alternatively the pericenter radius r_p and the apocenter radius r_a (equations 4-53). It is a convention to use the major semi-axis a together with the eccentricity ε for the definition of the size of a conic orbit. The important orbital period T is a direct function of a (when the dimension of a is “astronomical units”, the orbital period T follows as $\sqrt{a^3}$ in the dimension of years). The eccentricity ε is a dimensionless quantity (equation 4-52), it gives directly information about which type of conic orbit is under consideration.

Node angle, inclination angle and pericenter angle. The attitude of the orbit is defined by three Euler angles. Strictly speaking, these angles determine the location of the pericenter with respect to inertial space. The angle ψ (called “ascending node”) is the angle between the “node axis” and a certain fixed direction in space (the line from the earth to the sun on March, 21st, called “point of vernal equinox”). The inclination angle ϑ is measured against the equator in case of earth orbits, or against the ecliptic in case of solar orbits. The pericenter angle φ_p is the angle between the direction of the “ascending node” and the radial of the pericenter.

Time of pericenter passage. We divide the angle of mean anomaly (in case of elliptic orbits $\theta - \varepsilon \sin \theta$, in case of hyperbolic orbits $\varepsilon \sinh \theta - \theta$) by the mean motion ($2\pi/T = \sqrt{\gamma/a^3}$) to get the time interval that elapsed since the last pericenter passage (equation 4-58). Then we subtract this time interval from the time coordinate to get the “time of pericenter passage” t_p . The quantity t_p is constant on a conic orbit:

$$t_p = t - \frac{T}{2\pi}(\theta - \varepsilon \sin \theta) \quad (\text{hyperbolic orbits : } t_p = t - \frac{T}{2\pi}(\varepsilon \sinh \theta - \theta)) \quad (4 - 65)$$

The orbit is entirely determined by a specification of semimajor axis a , eccentricity ε , node angle ψ , inclination angle ϑ , pericenter angle φ_p and pericenter passage time t_p . The problem is to find the actual position of the orbiting object at a given time t ; it requires the numerical solution to Kepler’s equation. The “eccentric anomaly” θ of the orbital position has to be computed iteratively (knowing $t - t_p$, $T = 2\pi\sqrt{a^3/\gamma}$ and ε). Once the true anomaly $\varphi - \varphi_p$ has been obtained (equation 4-58), the velocity components u and v , and the distance r follow immediately (equation system 4-49).

4.3.4 Perturbation of Low Earth Orbits

Lagrange’s equations for the perturbation of orbital elements. A satellite orbiting on a low altitude earth orbit experiences a remainder of atmospheric drag. Another major perturbation comes from the fact that the gravitational field of the earth is not perfectly spherical. Orbits are determined by a set of six data, the “orbital elements” (semimajor axis a , eccentricity ε , node angle ψ , inclination angle ϑ , pericenter angle φ_p and time of pericenter passage t_p). Strictly speaking, the orbital elements are not exactly time-invariant constants, since perturbations cause them to change slowly. To quantify these effects, we consider the thrust acceleration vector $\vec{s} = (s_1, s_2, s_3)$ as a “perturbation”. The “powered motion” of the spacecraft obeys the equation system (4-48). Now thrust is applied, and the orbital elements become functions of the time t . Differentiation leads us to a system of differential equations:

$$\begin{aligned} \dot{a} &= -\frac{a}{e}(us_1 + vs_2) && (\text{corresponding to: } \dot{e} = us_1 + vs_2) \\ \dot{\varepsilon} &= \frac{h}{\varepsilon\gamma^2}(uhs_1 + (2er + vh)s_2) && (\text{corresponding to: } \dot{h} = rs_2) \\ \dot{\psi} &= \frac{s_3 \sin \varphi}{v \sin \vartheta} \\ \dot{\vartheta} &= \frac{s_3}{v} \cos \varphi \\ \dot{\varphi}_p &= \frac{h}{(\gamma\varepsilon)^2}((\gamma - vh)s_1 + u(h + \frac{\gamma}{v})s_2) - \frac{s_3}{v} \sin \varphi \cot \vartheta && (4 - 66) \\ \dot{t}_p &= \frac{1}{2e}(2rs_1 - 3(t + t_p)(us_1 + vs_2)) + \frac{h^2}{2e(\gamma\varepsilon)^2}((\gamma - vh)s_1 + u(h + \frac{\gamma}{v})s_2) \end{aligned}$$

The equation system above is called “Lagrange equations” for orbital perturbations.

The equations (4-66) can be seen as replacement for the equations of motion (4-48); with the velocity components u , v and the radius r according to the equations (4-49). The energy $e = \gamma/(2a)$ and the angular momentum $h = rv = \sqrt{a\gamma|1 - \varepsilon^2|}$ are now functions of the time t , like all the orbital elements. For numerical integration it is probably not advisable to replace the equations of the motion (they are better conditioned for an application on a computer); but the effect of perturbations on the orbit can better be examined using the transformed system of “Lagrange equations” (named after the great mathematician Joseph de Lagrange, 1736-1813).

We can observe that the change of the orbital elements is a linear function of the magnitude s of the perturbation; but it is a nonlinear function of the direction of \vec{s} . Perturbations in 1-direction ($\vec{s} = (s_1, 0, 0)$) have obviously no effect on the inclination angle ϑ and the node angle ψ . Constant perturbations in 1-direction have also no immediate effect on the semimajor axis a and on the eccentricity ε : the energy e changes periodically ($\dot{e} = us_1$), and the angular momentum h is not affected at all ($\dot{h} = 0$). A shift of the pericenter $\dot{\varphi}_p$ is the only long-term effect of perturbations which act predominantly in radial direction. In contrast, perturbations which act in 2-direction ($\vec{s} = (0, s_2, 0)$) do change the energy and the angular momentum of the orbit; but as expected, they have no influence on the inclination angle and on the node angle. The only perturbation which can change the inclination angle and the node angle is a perturbation that acts in 3-direction ($\vec{s} = (0, 0, s_3)$); but on the other hand, a perturbation in 3-directions does not change the energy and the angular momentum.

Effect of rest atmosphere. We can use the method of orbital perturbations to compute the approximate life time of a satellite. When the satellite flies at a low altitude, the remainder of the atmosphere will slow down the satellite and bring it to a reentry. The vertical velocity u of the satellite is negligibly small until the final reentry phase, and the horizontal velocity v is just a function of the actual flight altitude, $v = \sqrt{\gamma/r}$. The drag deceleration d_2 (equation 4-24) acts as perturbation in 2-direction. We can approximate the atmospheric density ϱ by an exponentially declining function of the altitude (with the ground level density ϱ_0 and the scale height h_{scale} , equation 4-23). On orbit the drag deceleration d_2 is proportional to the square of satellite velocity v . Since the orbit remains more or less circular, we may take the first equation of the system (4-66) and replace the semimajor axis a of the orbit by the expression r . When we consider that $e = -0.5\gamma/r$ (equation 4-53) we can find the following relationship:

$$\dot{r} = -\sqrt{\gamma r} \cdot (C_D A/M) \varrho_0 e^{-(r - R_{earth})/h_{scale}} \quad (4 - 67)$$

or:
$$\Delta t = \frac{-1}{\sqrt{\gamma} (C_D A/M) \varrho_0} \cdot \int_{r_1}^{r_2} \frac{e^{(r - R_{earth})/h_{scale}}}{\sqrt{r}} dr \quad (4 - 68)$$

Term $(C_D A/M)$ is the “ballistic factor” of the satellite (the drag coefficient C_D multiplied with the reference area A and divided by the satellite mass M). Equation (4-68) defines the time interval $\Delta t = t_2 - t_1$ that elapses while the satellite descends from the higher altitude (radius r_1) to the lower altitude (radius r_2). The satellite is faster at the lower altitude; paradoxically, the drag deceleration speeds-up the satellite.

Let $\Delta r = r - R_{earth}$ be the altitude of the circular satellite orbit. The expression $1/\sqrt{\Delta r + R_{earth}}$ in the integral (4-68) can be expanded in series; the result is:

$$(\Delta r + R_{earth})^{-1/2} = R_{earth}^{-1/2} - \frac{1}{2} \cdot \Delta r \cdot R_{earth}^{-3/2} + \frac{3}{8} \cdot \Delta r^2 \cdot R_{earth}^{-5/2} \dots$$

When a satellite moves on a very low orbit, the scale height h_{scale} and the altitude Δr of the orbit are both much smaller than the radius of the earth ($R_{earth} = 6378$ km). We can integrate equation (4-68) using the approximation $\sqrt{\Delta r + R_{earth}} \approx \sqrt{R_{earth}}$. Then the descent time interval $\Delta t = t_2 - t_1$ follows as:

$$\Delta t \approx \frac{h_{scale} (e^{(r_1 - R_{earth})/h_{scale}} - e^{(r_2 - R_{earth})/h_{scale}})}{\sqrt{\gamma R_{earth}} (C_D A/M) \rho_0} \quad (4-69)$$

The scale height h_{scale} is not exactly constant but a function of the flight altitude (between 100 and 120 km we have approximately a scale height of 6 to 10 km). Equation (4-69) shows that the life time of a satellite grows exponentially with the initial altitude $r_1 - R_{earth}$ of the low earth orbit. When this initial altitude is increased just by one scale height, the life time of the satellite will be extended by the factor e (Euler's number $e=2.718281\dots$); and this means that a satellite will live about three times longer when we put it on an orbit which is just about 8 km higher.

Unfortunately, the equation (4-69) provides just limited help to predict the lifetime Δt of a satellite that orbits at a low altitude. In practice, the actual life time Δt can deviate by a factor 0.5 to 2.0 from the estimate at the beginning of the operation phase. Such long-term life-time predictions are difficult because of our insufficient knowledge about future solar activity. The solar activity influences the actual value of scale height h_{scale} considerably. Since the scale height h_{scale} is an argument of the exponential function in the equation (4-69), a small error in h_{scale} will have a strong impact on the life time of the satellite. Even with refined methods, in practice the actual day of a satellite decay is not predictable until about a week in advance.

Regression of the node. Rotating celestial bodies are not perfectly spherical. The diameter of earth at the equator exceeds its diameter at the poles by almost 40 km; the shape of earth is more an ellipsoid than a sphere. Consequently, the potential field of earth gravity $e_{potential}$ is not just a function of the distance r from the gravitational center, it depends also slightly on the latitude (the spherical angle between actual position and equator, Cardan angle Θ , where $\sin \Theta = \sin \vartheta \sin \varphi$). More accurate than equation (4-18) is the following relationship:

$$e_{potential} = -\frac{\gamma}{r} \cdot \left[1 - J_2 \cdot \left(\frac{R_{earth}}{r} \right)^2 \cdot \left(\frac{3}{2} \sin^2 \vartheta \sin^2 \varphi - \frac{1}{2} \right) \right] \quad (4-70)$$

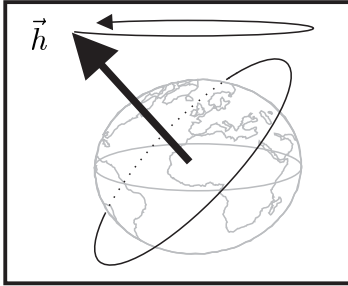
The dimensionless term $J_2 = 0.001082$ characterizes the oblateness of the earth. In order to find the gravitational perturbation vector ($\vec{s} = (s_1, s_2, s_3)$), we have to form the gradient vector of the potential of gravity (compare equation 4-19). The three direction coordinates are characterized by the vector $\vec{x} = (r, r\varphi, r\vartheta \sin \varphi)$.

Coordinate differentiation yields the three components of gravitational perturbation:

$$\begin{aligned} s_1 &= -\frac{\partial e_{potential}}{\partial r} + \frac{\gamma}{r^2} = \frac{3J_2\gamma R_{earth}^2}{2r^4} (3 \sin^2 \vartheta \sin^2 \varphi - 1) \\ s_2 &= -\frac{\partial e_{potential}}{r \partial \varphi} = -\frac{J_2\gamma R_{earth}^2}{r^4} (3 \sin^2 \vartheta \sin \varphi \cos \varphi) \\ s_3 &= -\frac{\partial e_{potential}}{r \sin \varphi \partial \vartheta} = -\frac{J_2\gamma R_{earth}^2}{r^4} (3 \sin \vartheta \cos \vartheta \sin \varphi) \end{aligned} \quad (4-71)$$

Term $-\gamma/r^2$ is the main gravitational acceleration. To consider just perturbations, we have to disregard this term and subtract it from the first equation of system (4-71).

The oblateness of the earth is responsible for the “regression of the node”, a constant slow motion of the node angle ψ . When the orbit is inclined against the equator, the attraction by the extra mass on the equator causes this gyroscopic effect on the orbit (the moment of the additional force lets the angular momentum vector \vec{h} “precess”).



To calculate the magnitude of this precessional movement, we insert the gravitational perturbation s_3 into the third equation of the system (4-66). Then we get:

$$\dot{\psi} = \frac{s_3 \sin \varphi}{v \sin \vartheta} = -\frac{3J_2\gamma R_{earth}^2 \cos \vartheta}{h^2} \cdot \frac{\sin^2 \varphi}{r(\varphi)} \cdot \dot{\varphi} \quad (4-72)$$

The angular momentum h , the eccentricity ε and the pericenter angle φ_p are constant during the short time interval of one orbital revolution, at least approximately. As the next step we insert $r(\varphi)$ (equation 4-49) and integrate one complete revolution:

$$\begin{aligned} \Delta\psi &= -\frac{3J_2\gamma R_{earth}^2 \cos \vartheta}{h^2} \cdot \frac{\gamma}{h^2} \cdot \int_0^{2\pi} (1 + \varepsilon \cos(\varphi - \varphi_p)) \sin^2 \varphi d\varphi \\ &= -\frac{3\pi J_2\gamma^2 R_{earth}^2 \cos \vartheta}{h^4} \end{aligned} \quad (4-73)$$

The orbital period T is the time interval of one revolution. We use the expression $a\gamma(1 - \varepsilon^2)$ instead of h^2 and find the velocity of the regression of the node as:

$$\dot{\psi}_{(regression\ of\ the\ node)} = \frac{\Delta\psi}{T} = -\frac{3\pi J_2 R_{earth}^2 \cos \vartheta}{T a^2 (1 - \varepsilon^2)^2} \quad (4-74)$$

For satellites on circular low earth orbit, we have $\varepsilon = 0$ and $a \approx R_{earth}$. One complete revolution of the node axis ψ is performed after n orbital revolutions of the satellite:

$$-2\pi = n \cdot \Delta\psi \approx -n \cdot 3\pi J_2 \cos \vartheta \quad (4-75)$$

So-called “sun-synchronous” orbits use the effect of this gravitational perturbation. When the satellite revolves in 2 hours, the node angle ψ revolves approximately in $4/(3J_2 \cos \vartheta)$ hours or $51.3/\cos \vartheta$ days. For $\vartheta \approx 82^\circ$ (high inclined polar orbits), the “regression of the node” needs exactly one year to complete one revolution.

Advance of the perigee. The oblateness of the earth is responsible for another significant long-term effect on orbital elements of earth satellites: the perigee of an eccentric orbit moves slowly. In contrast to the motion called “regression of the node” which is caused just by a perturbation in 3-direction, the “advance of the perigee” is effected by all the three perturbation components (s_1 , s_2 and s_3). The calculation of this effect is similar but requires more complicated transformations. We have to insert the perturbations (4-71) into the fifth equation of system (4-66) and integrate a complete orbital revolution ($0 \leq \varphi \leq 2\pi$), considering the orbital elements a , ε , φ_p as constant. During the comparatively short time interval of one orbital period, the oblateness of the earth causes the perigee of an inclined orbit to change by $\Delta\varphi_p$:

$$\Delta\varphi_p = -\frac{3\pi J_2 \gamma^2 R_{earth}^2}{h^4} \cdot \left(\frac{5}{2} \sin^2 \vartheta - 2\right) \quad (4-76)$$

The angular velocity of the “advance of the perigee” becomes:

$$\dot{\varphi}_p \text{ (advance of the perigee)} = \frac{\Delta\varphi_p}{T} = -\frac{3\pi J_2 R_{earth}^2 (2.5 \sin^2 \vartheta - 2)}{T a^2 (1 - \varepsilon^2)^2} \quad (4-77)$$

It depends on the inclination ϑ of the orbit, whether the perigee moves in positive or in negative direction. Of practical importance are orbital inclinations where the “advance of the perigee” vanishes: the perigee will not advance for $(2.5 \sin^2 \vartheta - 2) = 0$ or $\vartheta = 63.435^\circ$. This inclination is interesting for orbits with a high eccentricity.

4.3.5 Destination Orbits

Circular low earth orbit LEO. Many space missions go to low earth orbit, abbreviated “LEO”. Sometimes LEO is the final destination orbit; sometimes LEO is just used as “parking orbit”, to continue the mission later to a more distant orbit. When we disregard suborbital flights, LEO is the easiest accessible space destination. A launcher that departs in easterly direction has a particularly high payload capacity when it attains a circular equatorial LEO. LEOs are more or less circular orbits, because the altitudes of perigee and apogee are just slightly higher than the upper limits of the atmosphere. Typically, altitudes for the perigee or the apogee of a LEO are between 200 km to 500 km above the surface of the earth ($R_{earth} = 6378$ km).

The velocity of a satellite on a circular LEO depends on the altitude of the orbit. The lower the orbit, the faster the satellite moves on it. We can readily find the orbital velocity on a circular LEO when we equate the centrifugal acceleration γ/r^2 and the gravitational acceleration v^2/r . The result is the following condition:

$$v_{circular} = \sqrt{\gamma/r} \quad (4-78)$$

The value of the gravitational constant γ of the earth is $3.9866 \cdot 10^{14} \text{m}^3/\text{s}^2$. A satellite that orbits at an altitude of 200 km ($r=6578$ km) has a velocity of 7.785 km/s; the satellite moves a little slower at 300 km altitude (7.726 km/s for $r=6678$ km).

The orbital period T (time interval required for one revolution) on a circular LEO is:

$$T = 2\pi \cdot \sqrt{r^3/\gamma} \quad (4 - 79)$$

The period of a 200 km circular LEO is with 5309 seconds roughly one and a half hours. The orbital period of a satellite that flies on a hypothetical orbit at zero altitude is called “Schuler period”, $T = 84$ minutes when $r = R_{earth} = 6378$ km. The Schuler period takes similar values for celestial bodies with similar composition (the quotient r^3/γ accepts the same value for spherical bodies with the same density).

Sun synchronous orbit SSO. Highly inclined “polar” low earth orbits can take advantage of the gravitational perturbation effect that is called “regression of the node”. We have seen that it is possible to select the inclination ϑ and the altitude r of a polar orbit in a way that the node angle ψ performs one (retro-grade) revolution exactly in one year (equation 4-74). The consequence is that now the orbital plane is stabilized with respect to the direction of the sunlight. While the earth rotates, the satellite flies over every region on earth surface always at the same local daytime (for example, a satellite launched at 12:00 local time into a sun-synchronous polar orbit will pass all regions around midday at the sun illuminated side of the earth). The great advantage is that observation satellites can scan the surface of the earth at a constant solar illumination angle. Often the “ground-track” of polar orbit satellites is the same for every day, because it is possible to adjust the altitude of the orbit in a way that the satellite returns in exactly 24 hours to the same point above the earth.

Geostationary earth orbit GEO and geostationary transfer orbit GTO.

Particularly interesting for some commercial applications like telecommunication and broadcasting is the geostationary or geosynchronous earth orbit GEO. At an altitude of $r = 42250$ km (or 35872 km above the surface of the earth), the circular velocity of a satellite relative to a non-rotating earth is $v = 3.072$ km/s: the satellite performs one orbital revolution exactly in $T = 24$ hours. For an observer on the surface of the earth, the satellite appears to be always in the same position above the equator. It is the great advantage of GEO satellites that ground stations can use fixed antennae. Usually, space launchers do not deliver their payload directly into GEO position, but into a geostationary transfer orbit GTO (circularization is performed by the satellite itself). Perigee and apogee of the transfer orbit are located above the equator (with an apogee altitude at GEO position and a perigee altitude just on top of the atmosphere). The following table lists some representative data for a typical GTO:

$r_a = 42250$ km	$e = -\gamma/(2a) = -8.172$ km ² /s ²
$r_p = 6578$ km	$h = b \cdot \sqrt{\gamma/a} = 6737$ km ² /s
$\varepsilon = (r_a - r_p)/(r_a + r_p) = 0.730564$	$T = \pi \sqrt{a^3/\gamma} = 18980$ seconds
$a = (r_a + r_p)/2 = 24414$ km	$v_p = h/r_p = 10241$ m/s
$b = \sqrt{r_a r_p} = 16671$ km	$v_a = h/r_a = 1595$ m/s

The satellite needs $T/2 = 5.27$ hours to coast from the perigee to the apogee position.

Molniya orbits. Obviously, the geostationary orbit is not visible for ground stations in polar regions. Substantial parts of Russia are located far north, to ensure communication link to ground stations near the north pole, Russia uses so-called Molniya communication satellites. These Molniya satellites fly on high-eccentric orbits ($r_p = 6830$ km, $r_a = 43280$ km); with an orbital period of $T = 12$ hours the satellites are actually not “geostationary satellites”, but at least they remain during their apogee passage (twice every day) for an extended time interval approximately over the same (north) region on earth. To avoid that the apogee position of a Molniya satellite migrates slowly, the orbit has to use the critical inclination where the “advance of perigee” vanishes. Remember that this gravitational perturbation effect vanishes exactly for the inclination $\vartheta = 63.5^\circ$.

Parabolic earth escape trajectories. When a spacecraft accelerates until it reaches the escape velocity it will leave the gravitational field of the earth. A parabolic escape orbit requires the minimum mechanical energy for the escape maneuver. Thus:

$$e_{parabolic} = \frac{u^2 + v^2}{2} - \frac{\gamma}{r} = 0 \quad (4 - 81)$$

After an infinite flight time the spacecraft reaches an infinite distance where it has no velocity anymore. The parabolic escape velocity depends on the actual flight altitude:

$$v_{parabolic} = \sqrt{u^2 + v^2} = \sqrt{2\gamma/r} \quad (4 - 82)$$

The parabolic escape velocity $v_{parabolic}$ is exactly by the factor $\sqrt{2}$ higher than the circular velocity (equation 4-78), independent of the actual flight altitude (radius r). For example, at an altitude of 200 km the circular velocity amounts 7.785 km/s; and we can immediately calculate the minimum escape velocity at this altitude as $v_{parabolic} = \sqrt{2} \cdot 7.785$ km/s = 11.010 km/s.

Hyperbolic earth escape trajectories. From the practical point of view, hyperbolic escape orbits are more important than parabolic escape orbits. The spacecraft escapes faster; and the spacecraft has still a velocity with respect to the earth in interplanetary space. The mechanical energy of the hyperbolic escape orbit is positive:

$$e_{hyperbolic} = \frac{u^2 + v^2}{2} - \frac{\gamma}{r} = \frac{v_\infty^2}{2} \quad (> 0) \quad (4 - 83)$$

In interplanetary space, far away from the earth, the spacecraft has the velocity v_∞ :

$$v_\infty = \sqrt{u^2 + v^2 - 2\gamma/r} = \sqrt{v_{hyperbolic}^2 - v_{parabolic}^2} \quad (4 - 84)$$

$$v_\infty > v_{hyperbolic} - v_{parabolic}$$

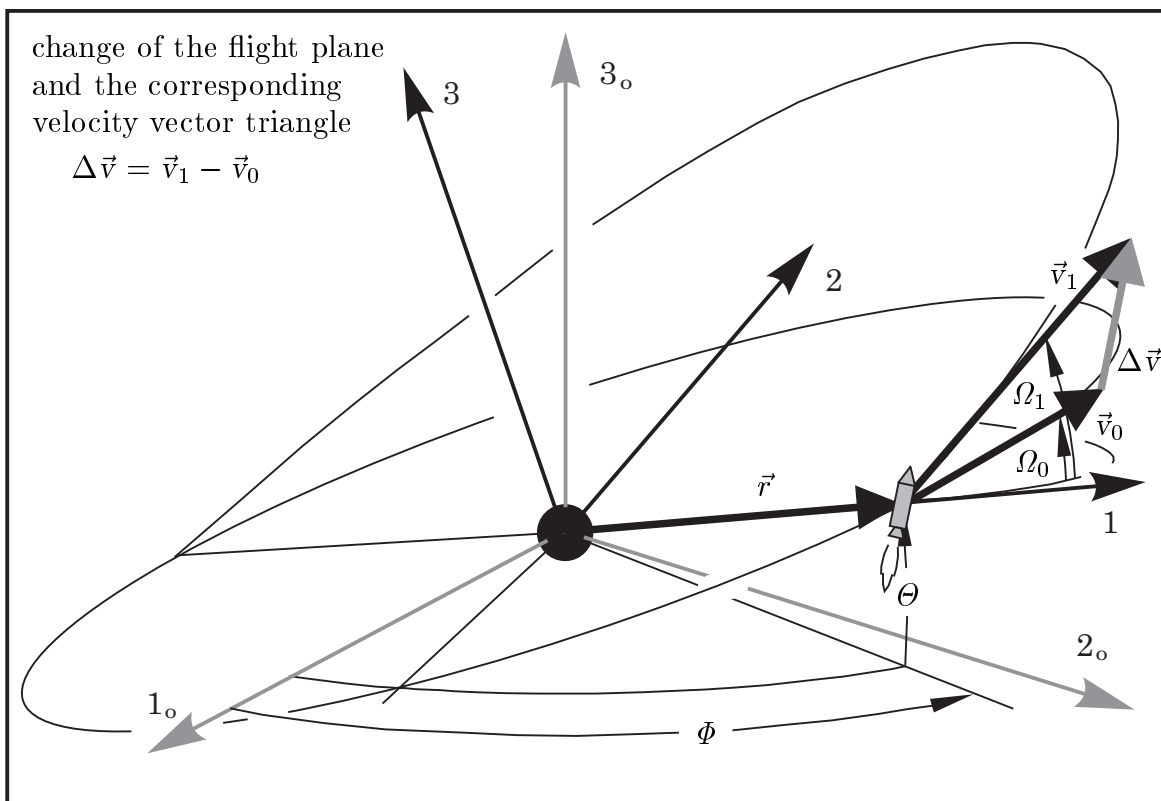
It is important to realize that the relative velocity v_∞ is much higher than the difference of the hyperbolic velocity and the parabolic velocity. For example, when the spacecraft is injected from a 200 km parking orbit with a velocity of 11510 m/s (just 500 m/s faster than the parabolic escape velocity), it will finally depart from the earth with the remarkable relative velocity of 3.35 km/s.

4.4. Impulsive Transfers

The spacecraft uses a transfer trajectory to travel from one orbit to another one. These transfer trajectories are composed of propulsive maneuvers (“thrust arcs”) separated by sections of conic orbits (“coast arcs” with switched-off engine). Often it is the case that we consider a transfer trajectory where the duration of the thrust arcs is much shorter than the duration of the coast arcs. When the thrust arcs are very short in comparison with the entire transfer time, we are allowed to approximate finite thrust maneuvers by impulsive maneuvers; and consider the transfer trajectory as an “impulsive transfer orbit”. Impulsive thrust maneuvers with infinitely short duration change the velocity of the spacecraft, but they do not change its location in space. We assume that only coast arcs change the location of the spacecraft, and that only coast arcs consume the flight time of the impulsive transfer trajectory.

4.4.1 Impulsive Thrusts

Changing of the velocity. Consider that the spacecraft moves on a conic orbit. Its position is defined by the location vector \vec{r} (length r , Cardan angles Φ and Θ). When an impulsive thrust maneuver Δv is applied, the velocity vector \vec{v} changes its direction and its magnitude instantaneously; but the vector \vec{r} remains constant during the whole impulsive maneuver. The actual flight plane can assume a different orientation after the maneuver, because the Cardan angle Ω can change its value:



The velocity vector \vec{v} changes its length and its direction during the infinitely short time interval of the impulsive maneuver. The index “0” means “immediately before the impulse” and the index “1” means “immediately after the impulse”. Therefore:

$$\Delta \vec{v} = \vec{v}_1 - \vec{v}_0 \quad (4-85)$$

As the next step we have to resolve this vector equation into its component notation. Therefore it is necessary that all vectors refer to the same coordinate system; however, the moving system changes its attitude while the impulse is applied. We take the attitude of the system immediately before the impulse as a basis, because the angles δ_0 and ϕ_0 of the impulsive thrust maneuver refer to the actual position of the moving system before the maneuver. The vector of the impulse $\Delta \vec{v}$ can be expressed using its magnitude Δv and the definition of the thrust direction with the angles δ_0 and ϕ_0 (equation 4-33). The velocity vector \vec{v}_0 takes the simple form of $\vec{v}_0 = (u_0, v_0, 0)$. Note that velocity vector \vec{v}_1 does not take the form of $\vec{v}_1 = (u_1, v_1, 0)$, this notation would refer to the attitude of the coordinate system immediately after the impulse. We have to apply matrix multiplication in order to transform the vector \vec{v}_1 in a way that its component notation refers to the attitude of the system before the impulse:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\Omega_1 - \Omega_0) & -\sin(\Omega_1 - \Omega_0) \\ 0 & \sin(\Omega_1 - \Omega_0) & \cos(\Omega_1 - \Omega_0) \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ v_1 \\ 0 \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \cos(\Omega_1 - \Omega_0) \\ v_1 \sin(\Omega_1 - \Omega_0) \end{pmatrix} \quad (4-86)$$

The rotation angle is $\Delta\Omega = \Omega_1 - \Omega_0$. Now we resolve the vector equation (4-85) as:

$$\Delta v \cdot \begin{pmatrix} \sin \phi_0 \cos \delta_0 \\ \cos \phi_0 \cos \delta_0 \\ \sin \delta_0 \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \cos(\Omega_1 - \Omega_0) \\ v_1 \sin(\Omega_1 - \Omega_0) \end{pmatrix} - \begin{pmatrix} u_0 \\ v_0 \\ 0 \end{pmatrix} \quad (4-87)$$

Finally, the equation (4-87) provides sufficient information to calculate the velocity of the spacecraft and the attitude of the coordinate system after the impulsive maneuver:

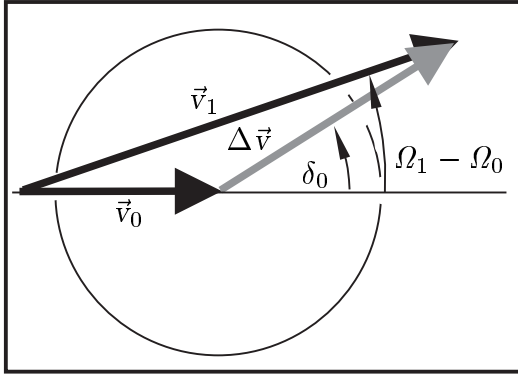
$$\begin{aligned} u_1 &= u_0 + \Delta v \sin \phi_0 \cos \delta_0 \\ v_1 &= \sqrt{(\Delta v \cos \phi_0 \cos \delta_0 + v_0)^2 + (\Delta v \sin \delta_0)^2} \\ \sin(\Omega_1 - \Omega_0) &= (\Delta v \sin \delta_0) / v_1 \\ \cos(\Omega_1 - \Omega_0) &= (\Delta v \cos \phi_0 \cos \delta_0 + v_0) / v_1 \end{aligned} \quad (4-88)$$

Plane impulsive maneuvers. The equation system (4-88) simplifies when the impulsive thrust vector $\Delta \vec{v}$ is located completely inside the actual flight plane ($\delta_0 = 0$):

$$\begin{aligned} u_1 &= u_0 + \Delta v \sin \phi_0 \\ v_1 &= v_0 + \Delta v \cos \phi_0 \\ \Omega_1 &= \Omega_0 \end{aligned} \quad (4-89)$$

Such plane impulsive maneuvers are important special cases in practice, because very often the final orbit is located exactly inside the same plane as the initial orbit. Then the transfer orbit is a plane trajectory, called “coplanar transfer orbit”.

Impulsive inclination change maneuvers. Another important special case is the case of “horizontal thrusting”, when a horizontal impulse is applied in the apse of a transfer trajectory. The inclination of the flight plane changes when $\phi_0 = 0, \delta_0 \neq 0$. When we look down at the maneuver from the top we see a velocity vector triangle:



$$\begin{aligned}
 u_1 &= u_0 \\
 \Delta v \cos \delta_0 &= v_1 \cos(\Omega_1 - \Omega_0) - v_0 \\
 \Delta v \sin \delta_0 &= v_1 \sin(\Omega_1 - \Omega_0) \quad (4 - 90)
 \end{aligned}$$

We can eliminate the thrust angle δ_0 to get:

$$\Delta v^2 = v_1^2 + v_0^2 - 2v_1v_0 \cos(\Omega_1 - \Omega_0) \quad (4 - 91)$$

The theorem of the cosine brings us to the same conclusion. When the change of the inclination is small, we have with $\cos(\Omega_1 - \Omega_0) \approx 1$ approximately $\Delta v \approx v_1 - v_0$. This means that changing the inclination is “cheap” (requires just a little propulsion) when the maneuver changes the inclination just slightly but the velocity considerably (the case: $\Omega_1 - \Omega_0 \ll 1, v_1 \gg v_0$). However, when the velocity is just changed slightly (the case: $v_1 \approx v_0, \cos(\Omega_1 - \Omega_0) \approx 1 - (\Omega_1 - \Omega_0)^2/2$), we can approximate the equation (4-91) by the expression $\Delta v = v_0 \cdot (\Omega_1 - \Omega_0)$; and this means that also a small inclination change maneuver is very expensive when the velocity v_0 is high.

Thrust direction before and after the impulse. During an impulsive maneuver the thrust direction remains constant with respect to inertial space, but not with respect to the moving coordinate system. The coordinate system turns instantaneously at the moment when the impulse is applied, and the thrust angles δ and ϕ assume different values before and after the impulse. The following relationship is valid:

$$\begin{pmatrix} \sin \phi_1 \cos \delta_1 \\ \cos \phi_1 \cos \delta_1 \\ \sin \delta_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\Omega_1 - \Omega_0) & \sin(\Omega_1 - \Omega_0) \\ 0 & -\sin(\Omega_1 - \Omega_0) & \cos(\Omega_1 - \Omega_0) \end{pmatrix} \cdot \begin{pmatrix} \sin \phi_0 \cos \delta_0 \\ \cos \phi_0 \cos \delta_0 \\ \sin \delta_0 \end{pmatrix} \quad (4 - 92)$$

The thrust direction immediately before the impulse (defined by thrust angles δ_0 and ϕ_0) has to be transformed to the component notation of the coordinate system immediately after the impulse. Again a rotation matrix serves for the transformation (compare equation 4-86, but this time it is the inverse matrix). We may conclude:

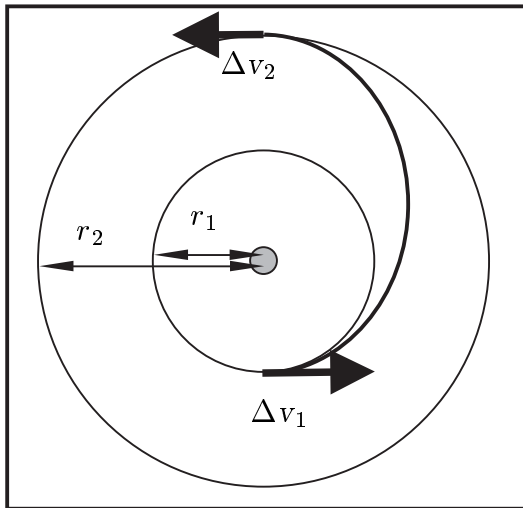
$$\begin{aligned}
 \sin \phi_1 \cos \delta_1 &= \sin \phi_0 \cos \delta_0 \\
 v_1 \cos \phi_1 \cos \delta_1 &= \Delta v(1 - \sin^2 \phi_0 \cos^2 \delta_0) + v_0 \cos \phi_0 \cos \delta_0 \\
 v_1 \sin \delta_1 &= v_0 \sin \delta_0 \quad (4 - 93)
 \end{aligned}$$

Using the equation (4-88) we can replace the terms $\sin(\Omega_1 - \Omega_0)$ and $\cos(\Omega_1 - \Omega_0)$.

4.4.2 The Hohmann Transfer

Plane transfer between two circular orbits. In the year 1925, Walter Hohmann (1880-1945) wrote his famous book about the “attainability of celestial bodies” (Die Erreichbarkeit der Himmelskörper). In this book he described a way to transfer a spacecraft from one orbit to another coplanar orbit; the trajectory which was later named after him. Hohmann trajectories are the most important type of rocket powered transfer orbits. However, it must be mentioned that these Hohmann transfer trajectories do not define the limits for reaching celestial bodies (Hohmann was not aware that planetary flybys can be utilized). Even though most transfer trajectories are not exactly Hohmann trajectories (often the initial and the final orbit are not circular and not coplanar), many two-impulse transfer orbits can be treated in a similar manner and are therefore also called “Hohmann transfer trajectories”.

However, the real Hohmann transfer is a plane trajectory between two circular orbits. It consists of two horizontal thrust maneuvers separated by one coast arc (180° arc length). To move from the initial orbit (radius r_1) to the final orbit (radius r_2), the spacecraft utilizes an elliptic orbit with the eccentricity $\varepsilon = (r_2 - r_1)/(r_2 + r_1)$ and the semimajor axis $a = (r_2 + r_1)/2$. Transfer time Δt is half the orbital period T . The spacecraft raises its velocity twice: the first impulsive maneuver accelerates the vehicle from the circular velocity $v_1 = \sqrt{\gamma/r_1}$ at radius r_1 to the pericenter velocity $v_p = v_1 \cdot \sqrt{1 + \varepsilon}$; the second maneuver accelerates the vehicle from the apocenter velocity $v_a = v_2 \cdot \sqrt{1 - \varepsilon}$ to the circular velocity $v_2 = \sqrt{\gamma/r_2}$ at the radius r_2 . Thus:



$$\Delta v_1 = \sqrt{\frac{\gamma}{r_1}} \cdot \left(\sqrt{\frac{2r_2}{r_1 + r_2}} - 1 \right)$$

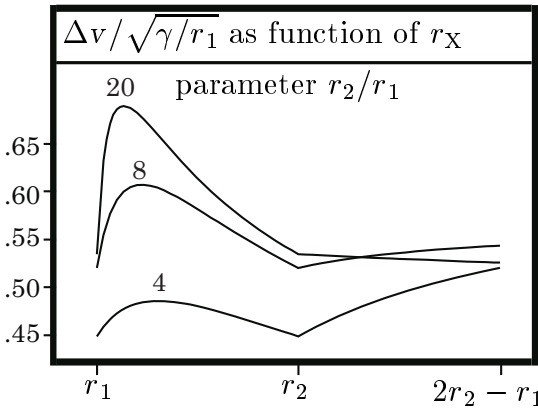
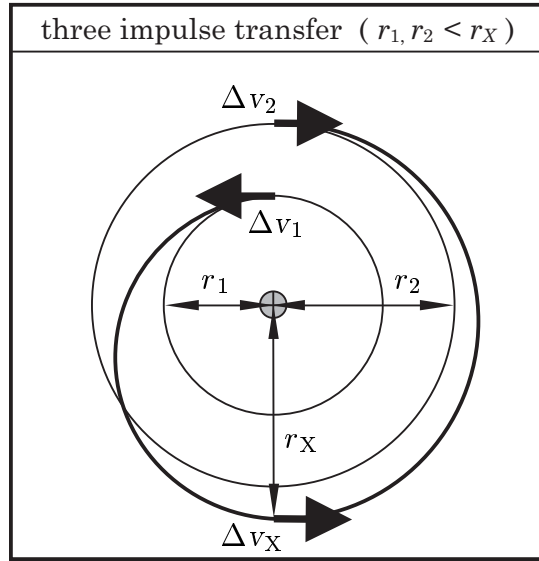
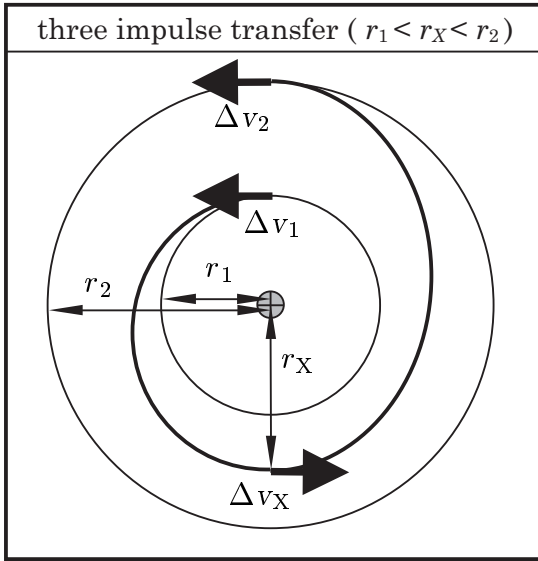
$$\Delta v_2 = \sqrt{\frac{\gamma}{r_2}} \cdot \left(1 - \sqrt{\frac{2r_1}{r_1 + r_2}} \right)$$

$$\Delta v = \Delta v_1 + \Delta v_2 \quad (4 - 94)$$

$$\Delta t = \pi \sqrt{\frac{(r_1 + r_2)^3}{8\gamma}}$$

A spacecraft which orbits on an equatorial LEO (altitude 200 km, $r_1 = 6578$ km) has a circular velocity of $v_1 = 7.785$ km/s. For changing to GTO ($r_2 = 42250$ km), the vehicle accelerates to perigee velocity $v_p = 10.241$ km/s. The spacecraft coasts 5.3 hours until it reaches apogee, where it has the smaller velocity $v_a = 1.595$ km/s. To depart for GEO, the spacecraft accelerates a second time until it reaches again the circular velocity, $v_2 = 3.072$ km/s at GEO altitude. The total velocity requirement for the maneuver amounts to $\Delta v = 2.456 + 1.477 = 3.933$ km/s.

Coplanar transfer with intermediate impulse. The Hohmann transfer with two horizontal thrust maneuvers is the simplest way to move from one circular orbit to another one. Usually, it is also the optimal way. It is easy to demonstrate that the overall velocity requirement Δv increases when we use a third impulse “on the way”: the three-impulse transfer utilizes an intermediate impulse (Δv_X , radius r_X) at the junction point of two 180° coast arcs. We have to consider two cases, since the radius r_X of the intermediate impulse can be smaller or greater than the final radius r_2 .



$$\Delta v_1 = \sqrt{\frac{\gamma}{r_1}} \cdot \left(\sqrt{\frac{2r_X}{r_1 + r_X}} - 1 \right)$$

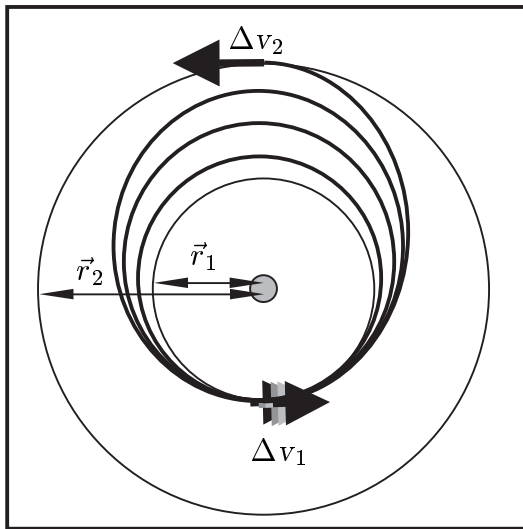
$$\Delta v_X = \sqrt{\frac{\gamma}{r_X}} \cdot \left(\sqrt{\frac{2r_2}{r_2 + r_X}} - \sqrt{\frac{2r_1}{r_1 + r_X}} \right)$$

$$\Delta v_2 = \sqrt{\frac{\gamma}{r_2}} \cdot \left(1 - \sqrt{\frac{2r_X}{r_X + r_2}} \right)$$

$$\Delta v = \Delta v_1 + \Delta v_X + |\Delta v_2| \quad (4 - 95)$$

The figure shows the velocity requirement Δv (related to the velocity $v_1 = \sqrt{\gamma/r_1}$ on the initial orbit) for several ratios r_2/r_1 as function of r_X . When $r_X = r_1$, the first impulse vanishes and we actually consider a two-impulse Hohmann transfer. When $r_X = r_2$, the last impulse vanishes and again the Hohmann transfer is present. Anyway the overall velocity requirement Δv increases when r_X lies between r_1 and r_2 . It is remarkable that for high ratios of r_2/r_1 the Hohmann transfer is not the optimum anymore; for $r_2/r_1 > 15.6$ and $r_X > r_2$ the three-impulse transfer saves indeed propulsion. The first impulse Δv_1 accelerates the spacecraft nearly to escape velocity; a small intermediate impulse Δv_X raises the pericenter to radius r_2 ; and finally a retro-impulse Δv_2 performs circularization. The three-impulse maneuver is of little importance in practice, because the transfer time is very long and the saving is small.

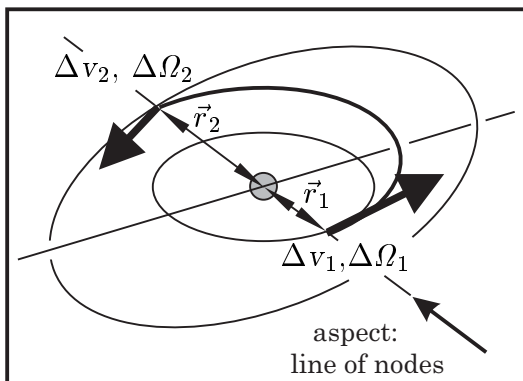
Transfer with several pericenter or apocenter burn maneuvers. Considered is the case of a usual Hohmann transfer trajectory. After the first impulsive thrust maneuver Δv_1 the spacecraft must have sufficient velocity to reach the altitude of the target orbit. When the burnout velocity is smaller, the apocenter altitude of the coast arc will remain below the altitude of the final orbit; but the spacecraft has the option to return on an elliptic orbit back to exactly the same location on orbit where it has performed the initial maneuver. Burning a second time it can complete the maneuver Δv_1 . In comparison to the two-impulse Hohmann transfer, the modified maneuver takes longer; but actually its overall velocity requirement Δv is the same.



The modified Hohmann transfer with several pericenter impulses (separated by elliptic coast arcs) requires all-together the same Δv_1 capacity at pericenter position as the usual two-impulse Hohmann transfer. Only when sufficient velocity has been attained, the spacecraft coasts to apocenter position, where the apocenter impulse Δv_2 accelerates it to circular velocity (also Δv_2 can be distributed among several burn maneuvers). A Hohmann transfer trajectory with several pericenter burn maneuvers needs a comparatively long transfer time, but sometimes a long flight time is desired.

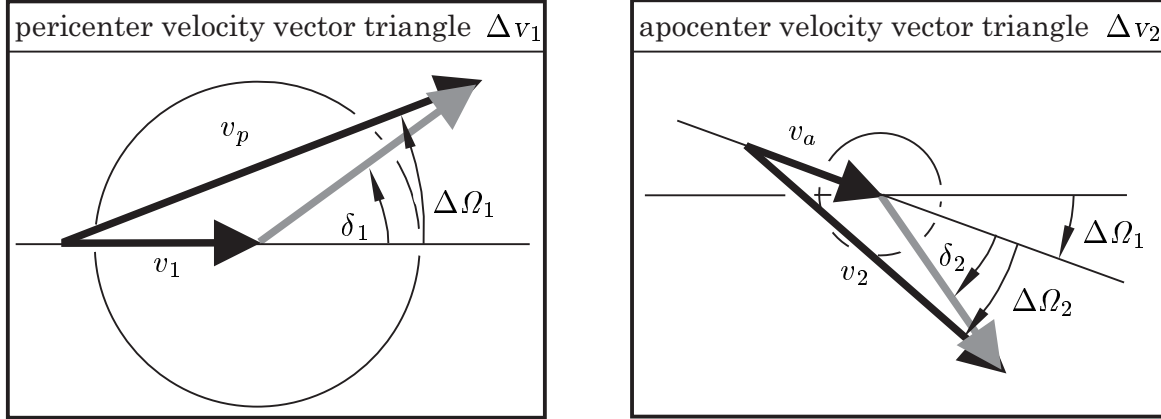
Transfers with several pericenter burn periods are important for low thrust vehicles, where the duration at pericenter position is too short for the complete maneuver.

Hohmann transfer with inclination change. Often initial and final orbit are not in the same plane: the transfer maneuver has not only to change the altitude but also the inclination of the orbit. Usually it is not a good idea to perform the inclination change by a separate maneuver. Remember that an inclination change is just “cheap” when the velocity is small, or when at least also the velocity increases considerably during the maneuver. The best option is to perform the inclination change while the motor burns anyway. Impulses Δv_1 and Δv_2 must be located on the intersection line between the planes of initial and final orbit (“line of nodes”).



The essential part of the inclination change $\Delta \Omega_2$ is performed during the apocenter maneuver Δv_2 , since the spacecraft is comparatively slow immediately before it reaches the apocenter. Also the pericenter maneuver Δv_1 contributes with a smaller part $\Delta \Omega_1$ to the entire inclination change, since we get small inclination changes at the pericenter nearly for free (when $\cos \Delta \Omega_1 \approx 1$, equation 4-91).

Pericenter maneuver Δv_1 and apocenter maneuver Δv_2 are both horizontal impulses (thrust angle $\phi = 0$). To analyse the relationship between the thrust angles δ_1 , δ_2 and the total inclination change $\Delta\Omega$, we can regard the velocity vector triangles for the pericenter impulse and for the apocenter impulse. We observe the two maneuvers from certain positions in space, located on the “line of nodes”. We always look down; and, when we see the velocity vector triangle, the celestial body lies behind it. The plane of the initial orbit is in both cases just a horizontal line when we look at it from above. The velocity vectors form the following triangles:



The angle δ_1 represents the thrust direction immediately before the pericenter impulse Δv_1 ; and the angle δ_2 is the thrust angle immediately before the apocenter impulse Δv_2 . The pericenter impulse accelerates the spacecraft from circular velocity v_1 to pericenter velocity v_p ; the apocenter impulse accelerates it from apocenter velocity v_a to circular velocity v_2 . Note that the values of the velocities v_p and v_a are not functions of the inclination change maneuver; however, v_p and v_a are entirely determined by initial radius r_1 and final radius r_2 (just like the values v_1 and v_2). The inclination angle $\Delta\Omega_1$ is free for optimization. The angle $\Delta\Omega_1$ remains constant while the spacecraft coasts from the pericenter position to the apocenter position. According to the equation (4-91), the transfer maneuver needs a total Δv capacity of:

$$\Delta v = \Delta v_1 + \Delta v_2 = \sqrt{v_p^2 + v_1^2 - 2v_p v_1 \cos \Delta\Omega_1} + \sqrt{v_a^2 + v_2^2 - 2v_a v_2 \cos \Delta\Omega_2} \quad (4-96)$$

All terms in the equation above are defined, except for the optimal distribution $\Delta\Omega = \Delta\Omega_1 + \Delta\Omega_2$; interesting is the distribution that requires the minimum Δv capacity for a given angle $\Delta\Omega$. To find it we may replace $\Delta\Omega_2$ by the expression $\Delta\Omega - \Delta\Omega_1$ and form $\partial\Delta v/\partial\Delta\Omega_1 = 0$. Finally, differentiation and transformation leads us to:

$$v_1 \sin \delta_1 = v_a \sin \delta_2 \quad (4-97)$$

The product of the initial velocity and the sine of the thrust angle δ is the same for both maneuvers; it defines the best distribution of the inclination change $\Delta\Omega$ among $\Delta\Omega_1$ and $\Delta\Omega_2$. For example, for a Hohmann transfer from a low earth orbit ($r_1 = 6578$ km, $v_1 = 7785$ m/s) to geostationary position ($r_2 = 42250$ km, $v_a = 1595$ m/s) the optimal distribution is determined by the ratio $\sin \delta_2/\sin \delta_1 = 4.88$.

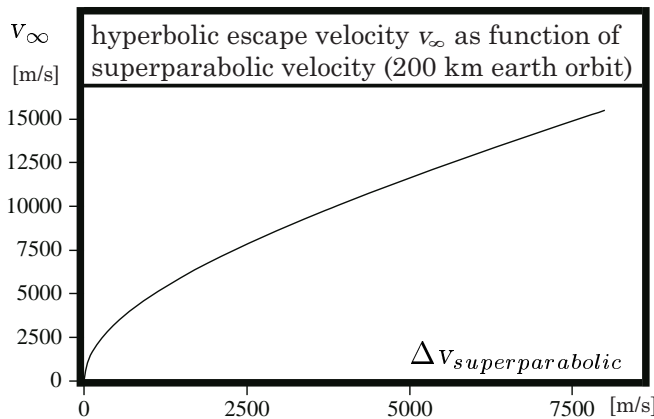
4.4.3 Impulsive Escape Maneuvers

Hyperbolic escape from circular low earth orbit. Remember that the parabolic escape velocity $v_{parabolic}$ is by the factor $\sqrt{2}$ higher than the circular velocity $v_{circular}$. Like the circular velocity, the parabolic velocity depends on the radius r of the orbit. From the practical point of view, hyperbolic escape maneuvers are more important than parabolic escape maneuvers: a spacecraft escapes faster with hyperbolic speed and has at an infinite distance still a relative velocity v_{∞} . We can consider a parabolic escape maneuver as a special case of hyperbolic maneuvers. The term $\Delta v_{parabolic}$ describes the minimum Δv capacity which is at least necessary to let the spacecraft escape from the gravitational field of the celestial body. We call the Δv capacity which exceeds the parabolic speed $\Delta v_{superparabolic}$. Thus:

$$\begin{aligned} v_{hyperbolic} &= v_{parabolic} + \Delta v_{superparabolic} \\ v_{hyperbolic} &= v_{circular} + \Delta v_{parabolic} + \Delta v_{superparabolic} \quad (4-98) \\ &\left(\text{with } v_{circular} = \sqrt{\gamma/r} ; \Delta v_{parabolic} = (\sqrt{2} - 1) \cdot \sqrt{\gamma/r} \right) \end{aligned}$$

Only the velocity increase $\Delta v_{superparabolic}$ gives at an infinite distance the relative velocity v_{∞} . To find v_{∞} we can use the energy conservation law of conic orbits (4-83), $e = (u^2 + v^2)/2 - \gamma/r = constant = v_{\infty}^2/2$. The result is the following relationship:

$$\begin{aligned} \Delta v_{superparabolic} &= \sqrt{v_{\infty}^2 + 2\gamma/r} - \sqrt{2\gamma/r} \quad (4-99) \\ \Delta v_{superparabolic} &< v_{\infty} \end{aligned}$$

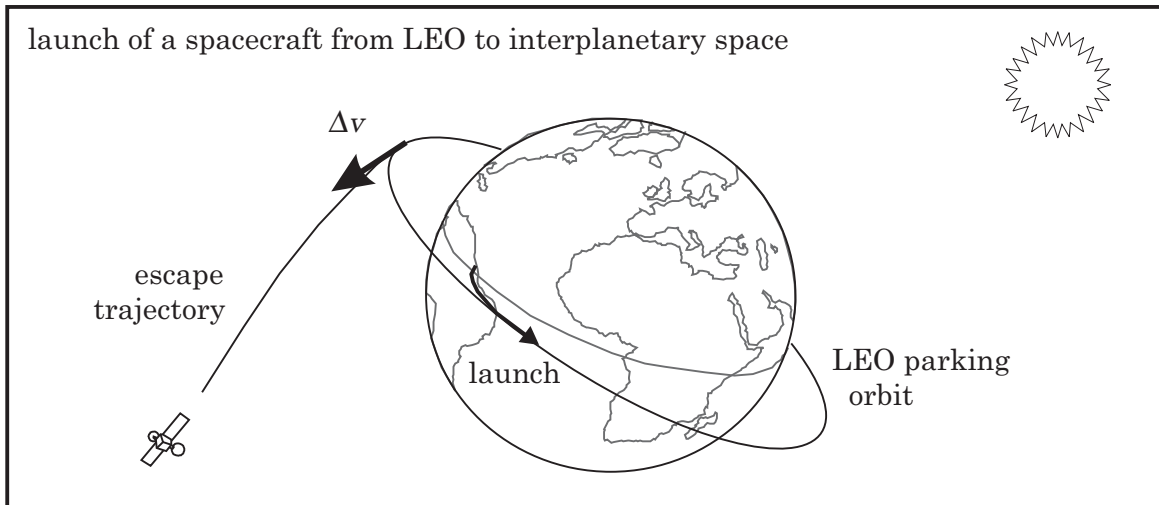


$\Delta v_{superparabolic}$	v_{∞}
10 m/s	469 m/s
50 m/s	1050 m/s
100 m/s	1487 m/s
250 m/s	2360 m/s
500 m/s	3356 m/s
750 m/s	4132 m/s
1000 m/s	4798 m/s
2500 m/s	7829 m/s
5000 m/s	11623 m/s
7500 m/s	14880 m/s

The capacity $\Delta v_{parabolic} = 3.225$ km/s is the minimum that is required to escape from a 200 km circular low earth orbit ($r = 6578$ km); this Δv requirement is smaller than the Δv requirement for a coplanar transfer into geostationary position ($\Delta v_{LEO \rightarrow GEO} = 3.933$ km/s). When the motor of the spacecraft burns longer, the spacecraft reaches superparabolic speed; and, after having left the gravitational field of the earth, the spacecraft escapes with the velocity v_{∞} relative to the earth. Note that the final relative velocity v_{∞} is much higher than $\Delta v_{superparabolic}$. Once the spacecraft has attained escape speed, the gravity of the celestial body helps.

Interplanetary escape from LEO. After the spacecraft has left the gravitational field of earth, it moves towards the target planet on a heliocentric transfer orbit. For this it is necessary that the hyperbolic earth escape trajectory “matches” the orbital elements of the heliocentric orbit. Escape direction, escape velocity and injection time have to fit. When the mission starts from a circular LEO, the right moment for engine ignition has to be calculated. We must consider that inclination changes are just “cheap” when they are really small. For launching an interplanetary spacecraft from a space-station, the orbit of the station must have the appropriate inclination.

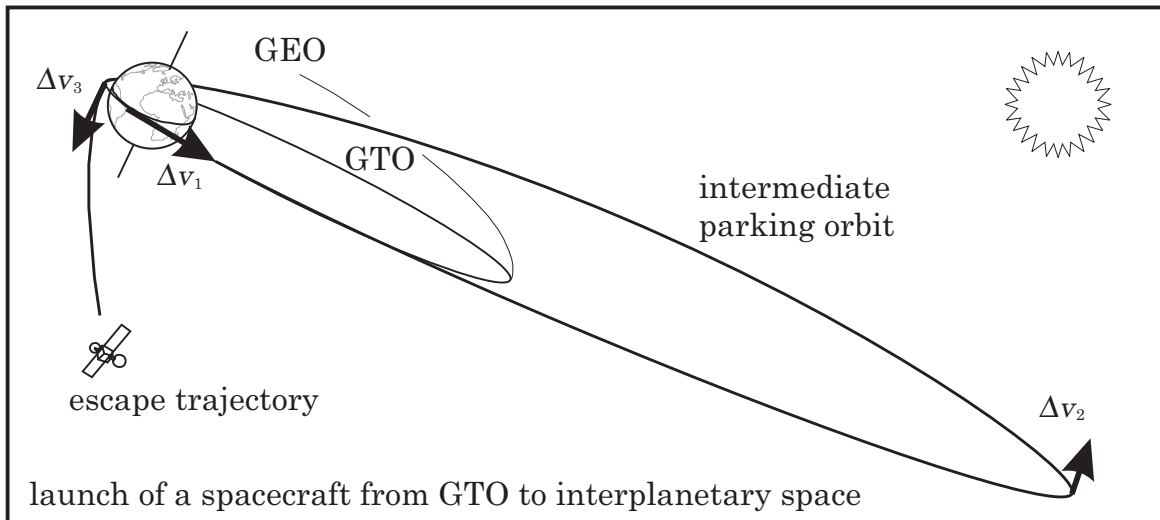
When the mission begins with a launch from the surface, the inclination of the orbit is provided by the launcher. Launch cannot occur any time, just once a day (the appropriate launch daytime has to be calculated). Utilizing an intermediate LEO (with appropriate inclination) avoids this disadvantage. The “parking orbit” brings the launcher’s upper stage more often in a suitable position for the planetary injection.



Interplanetary escape from GTO. The situation is more complicated when the mission starts from GTO. Some launchers are capable to launch several satellites at the same time (so-called “multiple launching”, a method to save some launch costs). All satellites must use on-board propulsion to maneuver from the terminal of the launcher (GTO) to their final destinations (GEO). A scientific deep space probe has to escape from the terminal of the launch vehicle GTO to its planetary destination.

GTOs have an inclination with respect to the earth equator, depending on the geographical latitude of the launch site. The heliocentric orbit lies usually more or less inside the ecliptic plane (the ecliptic is inclined against the equator, the angle is 23.4°). Often an inclination change is necessary, because it is improbable that the inclination of the GTO is suitable for a direct injection into interplanetary space. Since an inclination change at perigee speed is not a practicable option (for example, with a speed of 10 km/s, a change of the inclination by 30° requires a Δv of more than 5 km/s), a separate maneuver is required at low speed and high altitude. The place of this maneuver is preferably the apogee position of a high-eccentric intermediate orbit.

Thus, a high-eccentric intermediate orbit must be used to change the inclination. Three burn maneuvers are required to escape from GTO to the interplanetary target. First, the maneuver Δv_1 is executed in order to raise the apogee of the orbit. Therefore the engine of the spacecraft burns either immediately after the burnout of the upper-stage of the launcher, or at the first perigee passage. The Δv_1 for the first maneuver is slightly lower than what is required for a parabolic escape maneuver, $\Delta v_1 < 770$ m/s. Then the spacecraft coasts to the apogee of the intermediate orbit, where it performs the inclination change maneuver Δv_2 by thrusting perpendicular to the flight plane. The velocity requirement for the second maneuver depends on inclination angle and on the apogee altitude. The higher the apogee altitude the lower the velocity requirement Δv_2 (down to the theoretical value zero at infinite altitude), but also the flight time is longer on a highly eccentric parking orbit. When the apogee altitude is 385000 km (distance earth-moon), the coasting time is 5 days and the apogee velocity is 187 m/s; at a million kilometers the coasting time is 20 days and the apogee velocity 72 m/s. A small impulse Δv_2 suffices even for a substantial change of the orbital inclination. After having performed this maneuver, the spacecraft falls back to the perigee position where its propulsion system is re-ignited. The tangential impulse Δv_3 accelerates the spacecraft to hyperbolic escape velocity.



The spacecraft spends mission time coasting on the highly eccentric intermediate “parking orbit”. This waiting period makes it possible that the launch may occur on a date when the GEO satellites are ready. The launch may occur any day in a period of a few months before the actual interplanetary injection of the spacecraft. The earlier the launch occurs, the less propellant has to be spent for the inclination change maneuver; and the planetary mission itself takes much longer anyway.

Thus, the “date of launch” is in some limits a free variable, when we consider the interplanetary injection; but the “daytime of launch” is not a free variable. GTOs have their perigee and their apogee always above the equator. Since the perigee orientation with respect to interplanetary space has to fit for the interplanetary injection, it is essential that a suitable daytime is used for launching. To the other passengers, the simultaneously launched geostationary satellites, it makes no other difference.

5. Orbit Optimization

The optimization of transfer trajectories is a very important topic in the design and the preparation of space missions. A spacecraft must transport its payload on the optimal trajectory to the destination orbit; and, particularly when the payload is very small in comparison with the initial mass of the spacecraft, only small deviations from the optimal trajectory can deteriorate the capacity of the spacecraft badly. The problem is to find the best way through space where the mission is optimally performed (the way that allows the maximal payload capacity or the minimal flight time). We can solve the problem mathematically by an application of the Hamilton-Lagrange theory (also called “calculus of variations”). This method transforms the nonlinear trajectory optimization problem into a so-called “two point boundary value problem” (it is the problem of finding a solution to a system of differential equations which is determined by values at the initial and at the final state). Iterative search strategies have to be applied to solve such a “two point boundary value” problem numerically.

This chapter demonstrates how the Hamilton-Lagrange theory can solve trajectory optimization problems; but you will not find a prove of this mathematical method. Rigorous proofs you can find, for example, in the books of L.S.Pontryagin [“The Mathematical Theory of Optimal Processes”, Interscience, 1962], G.Leitmann [“Foundations of Optimal Control Theory”, McGraw Hill, New York, 1966]. Numerical solution strategies are treated in the books of J.P.Marek [“Optimal Space Trajectories”, Elsevier, Amsterdam, 1979] and A.E.Bryson/Y.C.Ho [“Applied Optimal Control”, Hemisphere Publishing Corporation, New York, 1975]. However, the fundamental books which treat the analytical theory of optimal transfer trajectories were written by D.F.Lawden [“Optimal Trajectories for Space Navigation”, Butterworth, London, 1963 and “Analytical Methods of Optimization”, Scottish Academic Press, Edinburgh and London, 1975]. Lawden introduced the “primer vector” and the “switch function”; he worked on trajectory optimization problems already at a time when no-one else imagined that optimal trajectories would become important one day.

5.1. Optimal Transfer Trajectories

The analytical optimization theory was established by Leonhard Euler (1707-1783), Joseph de Lagrange (1736-1813), Sir William Hamilton (1805-1865), and other famous mathematicians. Following this theory we consider the differential equations of the powered orbital motion of a spacecraft and form the so-called “Hamilton function” or “Hamiltonian”. Then, as the next step, we introduce the “Lagrange multipliers” and derive the “Lagrange equations” (an “adjoined” system of differential equations that determines the Lagrange multipliers). The behaviour of the thrust vector control functions is defined by “Pontryagin’s principle”: it requests for an “optimum control” a maximization of the Hamiltonian at every instant of the entire flight time. Pontryagin’s principle constitutes only a necessary condition for an optimal control.

We get a continuous time function for the optimal control of the thrust angles; but for the optimal control of the thrust magnitude we get a switch function that determines the right moments for switching between maximum thrust and zero thrust. However, it is important that we adjust the initial values of the control functions correctly: the steering of the thrust vector in direction and magnitude must bring the spacecraft to the predefined final destination orbit. When the trajectory obeys the Lagrange equations and the switch function, the mass of the spacecraft assumes a (relative) maximum at the final instant of the transfer trajectory. The search for these initial conditions is the numerical problem in spacecraft trajectory optimization. Solutions have to be found on a computer by the application of an iteration algorithm.

5.1.1 Application of the Hamilton-Lagrange theory

Optimization of functionals. Sometimes we have to find the optimal solution to a mathematical problem that is the function of just one parameter. Then the technique works as follows: we differentiate the objective function with respect to this parameter and find out where the differential quotient vanishes. When the problem is a continuous function $z(x)$ of the parameter x , the quotient dz/dx vanishes at the optimum (which could be a minimum or a maximum). The situation is similar when we intend to adjust two parameters: we can imagine the function $z(x, y)$ as a terrain with the altitude z over the $x-y$ -plane. The gradient vector $(\partial z/\partial x, \partial z/\partial y)$ vanishes on top of mountains, but also in valleys and on so-called “saddle points”. Second order derivatives can help us to indicate whether a solution is a maximum, a minimum or a saddle point. It is a “necessary condition” that the gradient vector vanishes at the top of a mountain; but unfortunately it is not a “sufficient condition” because we have no guarantee that there is just one mountain: a solution can be the global maximum but also just a local maximum. The problem becomes even more complicated when the objective function depends on many parameters: a parameter optimization strategy will probably not find the solution, the desired global optimum. The search will stop at the first local optimum where the gradient vector vanishes. When the problem is a function of many parameters, the danger increases that the algorithm stops at a local optimum, where actually the solution is badly non-optimal.

“Maximization of payload capacity” is usually the objective in spacecraft trajectory optimization (sometimes the objective is “minimization of flight time”). The problem is to optimize a so-called “functional”; a function of the final state. The value of this functional is controllable by three time functions for the thrust vector: two functions steer the thrust direction and one function steers the thrust magnitude (sometimes the thrust magnitude is a predetermined function of time, not open for optimization). In any case we do not have the problem to adjust a finite number of parameters: some of these thrust vector control functions are continuous functions of time. We could intersect the flight time and get a set of finite time intervals, with the intention to transform the problem into a parameter optimization problem. Then we could use the values of the thrust vector at the “corner times” as optimization parameters, approximating the continuous thrust angle control functions by linear polygonal lines.

The technique of parameter optimization is not commendable to solve a spacecraft trajectory optimization problem: this approach involves two severe disadvantages. First, the solution will be anyway only an approximation; and second, there is the danger that the parameter optimization strategy finds a badly non-optimal solution.

The mathematical formalism to solve trajectory optimization problems.

Fortunately there is a much better way to optimize spacecraft trajectories: it is the Hamilton-Lagrange theory, also called calculus of variations or Pontryagin's principle. Assume the motion is represented by a system of n first order differential equations:

$$\dot{x}_i = f_i(\vec{x}, \vec{q}, t) \quad ; \quad 1 \leq i \leq n \quad (5-1)$$

Term \vec{x} is the vector of the state variables, controllable by the control variables \vec{q} . The time derivatives of the state variables dx/dt are nonlinear functions $f(\vec{x}, \vec{q}, t)$. Then we introduce the "Lagrange multipliers" λ and construct the Hamiltonian \mathcal{H} as:

$$\mathcal{H} = \sum_{i=1}^n \lambda_i f_i \quad (5-2)$$

These Lagrange multipliers λ_i do not really have a physical meaning, they are just dimensionless mathematical quantities (elements of vector $\vec{\lambda}$). Their time functions are determined by the so-called "Lagrange equations", which here take the form of:

$$\dot{\lambda}_i = - \frac{\partial \mathcal{H}}{\partial x_i} \quad ; \quad 1 \leq i \leq n \quad (5-3)$$

Now it is a necessary condition for an optimum control that the vector \vec{q} maximizes the Hamiltonian H at every instant of time t ; however, the optimum control of vector \vec{q} must obey the limits of allowed control. Any other control of the vector \vec{q} in the allowed limits is non-optimal. This law is called "maximum principle of Pontryagin":

$$\mathcal{H}(\vec{\lambda}, \vec{x}, \vec{q}, t) \stackrel{\perp}{=} \text{maximum} \quad (5-4)$$

For continuous components of the control vector \vec{q} the maximum principle demands:

$$\frac{\partial \mathcal{H}}{\partial q} = 0 \quad (5-5)$$

The total time derivative and the partial time derivative of the Hamiltonian are equal:

$$\frac{d\mathcal{H}}{dt} = \frac{\partial \mathcal{H}}{\partial t} \quad (5-6)$$

$$\text{since:} \quad \dot{x}_i = \frac{\partial \mathcal{H}}{\partial \lambda_i} \quad ; \quad 1 \leq i \leq n \quad (5-7)$$

Therefore the Hamiltonian remains constant on the whole optimum trajectory when the equations of motion do not depend explicitly on the time t ; and the Hamiltonian constitutes then a first integral of the system of Lagrange equations (5-3).

The optimization objective is defined by the final conditions only; optimized is a functional of the final state and the final time. The state variables must coincide with the desired final state after the integration of the trajectory; therefore the initial values of the Lagrange multipliers λ have to be adjusted appropriately. The trajectory must be integrated obeying the maximum principle of Pontryagin, which constitutes a necessary (but not a sufficient) condition for an optimum control. “The trajectory is optimum” means it is a relative extreme in sense of the Hamilton-Lagrange theory.

Sometimes it makes sense to use another representation of the optimal control theory, using relationships which are apparently different from the equations (5-1) to (5-7). However, then it is just another formalism; the essential results are always the same. Our representation of the Hamilton-Lagrange theory is concentrated on the laws which we need for spacecraft trajectory optimization. We can always transform a trajectory optimization problem in a way that our representation of the method fits.

Restrictions or constraints. It can be the case that additional conditions restrict the state variables or the control variables. For examples, a parabolic antenna that aims at the earth can restrict the thrust direction of a planetary probe; or the earth atmosphere can prohibit that a space launcher upper stage flies below a certain altitude. Pontryagin’s principle is also valid for these cases. It states that the Hamiltonian \mathcal{H} must be maximized at every instant of time; however, just possible control is involved. When a control is not allowed because a restriction or constrain is valid, then another control has to be found which is allowed and which maximizes the Hamiltonian. This principle holds also for discontinuous control functions. Pontryagin proved that the Lagrange multipliers λ_i and the Hamiltonian \mathcal{H} remain continuous time functions even in case of discontinuous control, but their time derivatives $d\lambda/dt$ and dH/dt may be discontinuous functions at these “corners”. The behaviour of the control equations at discontinuities is also called “Weierstrass-Erdmann corner conditions”.

Restrictions can occur in form of equations or in form of inequalities. Let us first consider restriction in form of equations. Actually, the system of equations corresponds now to a system of reduced order. We can use every constrain to eliminate one control variable. Sometimes such a procedure is not possible or at least complicated (it depends on the form of the equations). The formalism of the Hamilton–Lagrange theory accepts a slightly different form when a set of additional equations is valid:

$$g_j(\vec{x}, \vec{q}, t) = 0 \quad ; \quad 1 \leq j \leq k \quad (5 - 8)$$

Naturally, the number of the restricting equations k must be smaller than the number of the control variables, otherwise the control system would be “over-defined”: we can eliminate one control variable for every additional equation. However, now we want to use the control theory in a different form and modify the Hamiltonian as follows:

$$\mathcal{H} = \sum_{i=1}^n \lambda_i f_i + \sum_{j=1}^k \mu_j g_j \quad (5 - 9)$$

Therefore we introduce new dimensionless multipliers μ for the additional equations.

Pontryagin's principle, determined by equation (5-4) or equation (5-5) respectively, is still valid in the same form. Also the Lagrange equation system (5-3) that defines the behaviour of the multipliers λ remains unchanged. However, we must not derive some more differential equations for the new multipliers μ_j . Condition (5-5), which is still valid for all continuous control variables, must be used indirectly to define the time functions for the new multipliers μ_j . Remember that every additional equation reduces the order of free control; and it is not possible to devise a control function for a control variable which has already been determined by an additional equation.

Now let us examine restrictions in form of inequalities, for example:

$$g(\vec{x}, \vec{q}, t) \leq 0 \quad (5-10)$$

Obviously, the restriction is either not violated on the optimal trajectory and can therefore be ignored; or it is violated and must therefore be regarded as a restriction in equation form. Remember the example of a spacecraft that has to obey a certain minimum flight altitude. It can be the case that the restriction is not effective, since anyway the optimal trajectory keeps the spacecraft always above the minimum altitude. However, it can also be the case that the unrestricted optimal trajectory would guide the spacecraft in some flight phases below this minimum altitude. Then the restriction becomes effective and keeps the spacecraft exactly at the minimum flight altitude. Thus, the optimal flight path consists of two different trajectory types: phases where the restriction is ignored and phases where the restriction is valid as an equation. The moments when the operation changes from one phase to another phase are called "corner times". At these corner times the control may take a discontinuous course; but Hamiltonian and Lagrange multipliers remain continuous functions of time.

Transversality conditions. The optimization objective is usually a nonlinear function of the state variables x_{end} at final time t_{end} . We can express the objective \mathcal{G}_1 as:

$$\mathcal{G}_1(\vec{x}_{end}, t_{end}) \stackrel{!}{=} \text{extreme} \quad (5-11)$$

Sometimes there are also restrictions for the final state variables in form of equations:

$$\mathcal{G}_j(\vec{x}_{end}, t_{end}) = 0 \quad ; 2 \leq j \leq k \quad (5-12)$$

These restrictions for the final state variables can be transformed into restrictions for the final values of the Lagrange multipliers, so-called "transversality conditions". Therefore the Hamilton-Lagrange theory uses the following mathematical formalism. When k equations restrict the final conditions, then the functional \mathcal{J} is established as:

$$\mathcal{J} = \sum_{j=1}^k \mu_j \mathcal{G}_j \quad (5-13)$$

Again new multipliers μ are introduced. Since their values are unknown, these multipliers have to be eliminated later (after differentiation) from the equation system.

Then the Lagrange multipliers λ_i and Hamiltonian \mathcal{H} must satisfy the following “transversality conditions” at the final instant of the optimal transfer trajectory:

$$\lambda_{i,end} = \frac{\partial \mathcal{J}}{\partial x_{i,end}} \quad ; \quad 1 \leq i \leq n \quad (5-14)$$

$$\mathcal{H}_{end} = - \frac{\partial \mathcal{J}}{\partial t_{end}} \quad (5-15)$$

In case a final state variable x_i is not restricted (when x_i is free at t_{end}), it follows:

$$\lambda_{i,end} = 0 \quad (5-16)$$

When the time t_{end} is unrestricted (free), we get a final condition for the Hamiltonian:

$$\mathcal{H}_{end} = 0 \quad (5-17)$$

When additionally the equations of motion (5-1) do not depend explicitly on time t , condition (5-6) is valid and the Hamiltonian vanishes on the whole optimal trajectory.

Numerical solution to a trajectory optimization problem. Let us now consider that we want to find a numerical solution to a system of differential equations. The integration procedure is comparatively simple when the initial values are known. Several fast integration routines are provided by numerical mathematics (for example “Runge-Kutta-Fehlberg” methods or “Runge-Kutta-Merson” methods). However, the procedure is much more complicated when the set of the initial conditions is not completely defined; for example, when the differential equation system is defined by conditions at the initial time and at the final time. Such a numerical problem is called a “two point boundary value problem”; and the optimization of trajectories involves usually this kind of problem. We have to find the initial values of the control variables which bring us the desired final state variables after trajectory integration. The adjustment of the initial values must be done by an iteration method: the algorithm estimates initial conditions, integrates the trajectory, and tries to find improved estimates for the initial conditions by an analysis of the deviations in the final state.

The strategy can be denoted as follows: we consider a system of differential equations:

$$\dot{x} = f(x, q, t) \quad ; \quad x(t_0) = x_0 \quad (5-18)$$

$$\dot{q} = g(x, q, t) \quad ; \quad x(t_1) = x_1 \quad (5-19)$$

Equation (5-18) determines the state functions; equation (5-19) the control functions. The vector of the state variables x is controlled by the vector of the control variables q (to simplify the notation we omit vector arrows). The boundaries of the state variables $x(t)$ are completely defined at the initial time t_0 and the final time t_1 ; however, the boundaries of the control variables $q(t)$ are unknown at the initial and final time. The Hamilton-Lagrange theory allows us always to transform a trajectory optimization problem into the form of a two point boundary value problem: the actual problem is then how we can find the solution to a nonlinear system of differential equations.

Functions $f(x, q, t)$ and $g(x, q, t)$ have to be integrated simultaneously. We insert the control functions $q(t)$ into the state functions f and get the final state variables x_1 as:

$$\begin{aligned} x_1 &= x_0 + \int_{t_0}^{t_1} f(x, q, t) dt \quad ; \quad q(t) = q_0 + \int_{t_0}^t g(x, q, t) dt \\ x_1 &= x_0 + \int_{t_0}^{t_1} f\left(x, q_0 + \int_{t_0}^t g(x, q, t) dt, t\right) dt = \mathcal{I}(q_0) \end{aligned} \quad (5-20)$$

All quantities of the right hand side of equation (5-20) are defined, except for the initial values q_0 . In a “condensed” notation, the vector function \mathcal{I} represents the numerical integrations. We have the problem of solving a nonlinear equation system: equation (5-20) must contain exactly the same number of unknown initial values q_0 as it contains final values x_1 , otherwise the equation system would have no solution.

We can compute the function \mathcal{I} also for estimated initial values \tilde{q}_0 ; however, this will not lead us to the desired final values x_1 . When \tilde{q}_0 is a good estimate, it will bring us approximately the desired final values x_1 . The next step is the linearization of $\mathcal{I}(q_0)$ in the vicinity of this estimate (using numerical differentiation techniques):

$$\mathcal{I}(q_0) = \mathcal{I}(\tilde{q}_0) + \left(\frac{\partial \mathcal{I}}{\partial q_0}\right)_{\tilde{q}_0} \cdot (q_0 - \tilde{q}_0) \quad (5-21)$$

The matrix $\partial \mathcal{I} / \partial q_0$ has to be transformed numerically to resolve the linear equation system above (the exponent -1 is actually wrong here, it means “inverted matrix”):

$$q_0 = \left[\left(\frac{\partial \mathcal{I}}{\partial q_0}\right)_{\tilde{q}_0} \right]^{-1} \cdot (x_1 - \mathcal{I}(\tilde{q}_0)) + \tilde{q}_0 \quad (5-22)$$

Hopefully, the solution is an improved estimate \tilde{q}_0 for the nonlinear system. We can start the method again with the new estimate (Newton’s method). Iteratively we can repeat the method, until finally the boundaries are satisfied with sufficient accuracy.

The Newton method needs an initial estimate which is at least “sufficiently good”. In practice, this requirement can involve serious numerical problems. For bad guesses the integration process can fail (“crash”), or the algorithm can be divergent. Sometimes a known solution for a similar problem can provide such a good initial estimate. A known solution can be transformed step by step into the desired solution, modifying in every step parameters and improving the initial estimate (“homotopy method”).

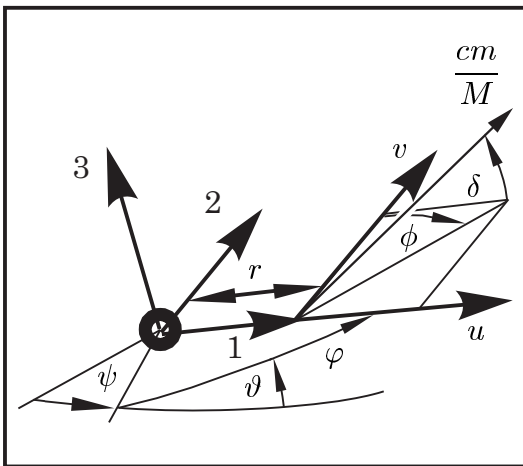
As an example, let us consider the generation of a “low-thrust” transfer trajectory. It is much easier to find an impulsive trajectory than a real low-thrust trajectory; but impulsive trajectories are good approximations for trajectories with high thrust levels. An impulsive reference trajectory can be used to generate a high thrust trajectory (with only small modifications of the control functions). A real thrust trajectory with high thrust level is a good approximation for a real thrust trajectory where the thrust level is a little lower. Again, data of a well-known trajectory are used as estimates for the generation of unknown trajectory control functions. Iteratively the thrust level can be diminished until finally the desired low thrust trajectory has been found.

5.1.2 Optimum Control of Thrust Arcs

Equations of motion. The “equations of motion” of a mechanical system are second order differential equations; they describe the influence of forces and moments on the motion of the system. The vector form of the equations of motion is independent of a coordinate system (like every vector equation); but it depends on the orientation of the coordinate system whether the component notation of the equations of motion takes a more complicated or a less complicated form. A suitable coordinate system can considerably simplify the component notation of a vector equation.

We could base the equations of motion on an inertial coordinate system; however, this would involve the disadvantage that the description of orbits is circumstantial. For example, consider a spacecraft that moves on a conic orbit: the equations of motion take a quite simple form when we use inertial coordinates (their vector form is $d^2\vec{r}/dt^2 = -\gamma \vec{r}/|\vec{r}|^3$), but the analytical solutions are complicated three-dimensional transcendental functions of time. The analytical solutions are much more transparent when we base them on polar coordinates, where the distance and the velocity are functions of the path angle: we get the well-known equations for conic orbits. Thus, a moving coordinate system that is carried along with the motion of the spacecraft seems to be much better conditioned for the interpretation of numerical results.

You can find a detailed description of this moving coordinate system in chapter four of this book. The center of the system coincides with the center of gravitation; and the orientation of the axes is defined as follows: every time the spacecraft moves on the 1-axis; and every time the 3-axis is perpendicular to the location vector \vec{r} and the velocity vector \vec{v} . Thus, the velocity vector is always in the 1-2-plane; and the location vector has just one component in 1-direction. The equation system (5-23) describes the trajectory of a spacecraft in the three-dimensional space near a celestial body completely by a system of seven first order differential equations. We can write:



$$\begin{aligned}
 \dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} + \frac{c m}{M} \sin \phi \cos \delta \\
 \dot{v} &= -\frac{u v}{r} + \frac{c m}{M} \cos \phi \cos \delta \\
 \dot{r} &= u \\
 \dot{\phi} &= \frac{v}{r} - \frac{c m}{M} \frac{\sin \delta}{v} \sin \varphi \cot \vartheta \\
 \dot{\vartheta} &= \frac{c m}{M} \frac{\sin \delta}{v} \cos \varphi \\
 \dot{\psi} &= \frac{c m}{M} \frac{\sin \delta}{v} \cdot \frac{\sin \varphi}{\sin \vartheta} \\
 \dot{M} &= -m
 \end{aligned} \tag{5-23}$$

The components of the velocity vector are termed u (vertical) and v (horizontal). Symbol r is the distance of the spacecraft from the gravitational center. The orientation of the coordinate system is defined by three Euler angles, named ψ , ϑ and φ .

The actual spacecraft mass M is the difference between initial mass and consumed propellant mass (the time integral of the mass flow rate m is exactly the mass of expelled propellant). Mass flow rate m , multiplied with the effective exhaust velocity c , constitutes the thrust of the engine. The thrust direction is defined by the thrust angles ϕ (in flight plane, measured against the horizontal line) and δ (out of actual flight plane, measured against the flight plane). The gravitational constant is termed as γ .

Now look at the system of equations from the mathematical point of view: $u, v, r, \varphi, \vartheta, \psi$ and M are the state variables; δ, ϕ and m are the control variables. Problem specific constants are c and γ . We control the spacecraft trajectory by steering the thrust vector in direction (ϕ and δ) and magnitude (m). By this we want to find a trajectory which brings the spacecraft from its well-known initial state (a position on orbit) to the desired final state (a position on another orbit). Additionally, as optimization objective, the transfer should require a minimum propellant consumption (synonymous with maximum final mass M). Transfer time may be predefined or free.

For liquid rocket engines operating in space the effective exhaust velocity c is constant. The value of c depends on the propellant which is used and on the construction of the motor. Most rocket engines work also with constant maximum mass flow rate m . Often the motor can be switched-off and switched-on again. The mass flow rate m becomes a control variable with a discontinuous course. Some rocket motors offer the option of throttling the thrust in some limits (down to 50%, by a reduction of the mass flow rate). In this case we may regard the thrust magnitude as a continuous control function, controllable in the limits of maximum and minimum thrust. A lot of theoretical effort has been invested in the analysis of such “intermediate thrust arcs” (sometimes trajectories with this type of thrusting are named “Lawden’s spirals”). In some cases complete analytical solutions are possible. However, it can be proved that these arcs are non-optimal; and that the real optimum trajectories consist of “thrust arcs” with full-operative engine and “coast arcs” with switched-off engine. Intermediate thrust arcs are ignored in this book due to missing practical importance.

When we consider solid rocket motors the thrust is a predefined function of time (often not constant). Solid motors cannot be switched-off and later be switched-on again. If two (or more) burning periods are required to complete the transfer trajectory, we will also need two (or more) solid rocket stages. Solid rocket motors can easily generate a high thrust force; often high thrust is desired but sometimes other restrictions prohibit high thrust. The effective exhaust velocity c of solid rocket motors is constant when the motor operates in the vacuum of space (at least approximately).

In the analysis of electric-propelled missions both c and m can be treated as control variables, however, existing ion engines work with constant exhaust velocity c (but it would also be possible to construct ion engines with variable exhaust velocity when this is desired). When the energy source is the sun (solar electric propulsion), the available energy diminishes with the distance r of the spacecraft from the sun. Then the mass flow rate m is a nonlinear function of r , and the exhaust velocity c is constant (the mass flow rate m of existing ion engines can be throttled down to maybe 50%).

Lagrange equations. We follow the Hamilton-Lagrange theory to derive optimal control laws. According to the equation (5-2) we construct the Hamiltonian as:

$$\begin{aligned}
\mathcal{H} = & \lambda_u \left(\frac{v^2}{r} - \frac{\gamma}{r^2} + \frac{c m}{M} \sin \phi \cos \delta \right) \\
& + \lambda_v \left(-\frac{u v}{r} + \frac{c m}{M} \cos \phi \cos \delta \right) \\
& + \lambda_r (u) \\
& + \lambda_\varphi \left(\frac{v}{r} - \frac{c m}{M} \frac{\sin \delta}{v} \sin \varphi \cot \vartheta \right) \\
& + \lambda_\vartheta \left(\frac{c m}{M} \frac{\sin \delta}{v} \cos \varphi \right) \\
& + \lambda_\psi \left(\frac{c m}{M} \frac{\sin \delta}{v} \cdot \frac{\sin \varphi}{\sin \vartheta} \right) \\
& + \lambda_M (-m)
\end{aligned} \tag{5-24}$$

The Hamiltonian \mathcal{H} is a linear function of m . It is important whether an expression in \mathcal{H} contains the mass flow rate m or not. Therefore we re-arrange \mathcal{H} correspondingly:

$$\begin{aligned}
\mathcal{H} = & \left\{ \left(\frac{v^2}{r} - \frac{\gamma}{r^2} \right) \lambda_u - \frac{u v}{r} \lambda_v + u \lambda_r + \frac{v}{r} \lambda_\varphi \right\} \\
& + m \left[\frac{c}{M} \{ \cos \delta (\lambda_u \sin \phi + \lambda_v \cos \phi) \right. \\
& \quad \left. + \frac{\sin \delta}{v} (-\lambda_\varphi \sin \varphi \cos \vartheta + \lambda_\vartheta \cos \varphi + \lambda_\psi \frac{\sin \varphi}{\sin \vartheta}) \right] - \lambda_M
\end{aligned} \tag{5-25}$$

The next step is the derivation of the Lagrange equations, as demanded by (5-3):

$$\begin{aligned}
\dot{\lambda}_u &= -\frac{\partial \mathcal{H}}{\partial u} = \frac{v}{r} \lambda_v - \lambda_r \\
\dot{\lambda}_v &= -\frac{\partial \mathcal{H}}{\partial v} = -2\frac{v}{r} \lambda_u + \frac{u}{r} \lambda_v - \frac{1}{r} \lambda_\varphi + \frac{c m}{M} \frac{\sin \delta}{v^2} (\dots) \\
\dot{\lambda}_r &= -\frac{\partial \mathcal{H}}{\partial r} = \left(\frac{v^2}{r^2} - 2\frac{\gamma}{r^3} \right) \lambda_u - \frac{u v}{r^2} \lambda_v + \frac{v}{r^2} \lambda_\varphi \left(-\frac{\partial m}{\partial r} [\dots] \right) \\
\dot{\lambda}_\varphi &= -\frac{\partial \mathcal{H}}{\partial \varphi} = \frac{c m}{M} \frac{\sin \delta}{v} (\lambda_\vartheta \sin \varphi + \lambda_\varphi \cos \varphi \cot \vartheta - \lambda_\psi \frac{\cos \varphi}{\sin \vartheta}) \\
\dot{\lambda}_\vartheta &= -\frac{\partial \mathcal{H}}{\partial \vartheta} = \frac{c m}{M} \frac{\sin \delta}{v} \frac{\sin \varphi}{\sin^2 \vartheta} (\lambda_\psi \cos \vartheta - \lambda_\varphi) \\
\dot{\lambda}_\psi &= -\frac{\partial \mathcal{H}}{\partial \psi} = 0 \\
\dot{\lambda}_M &= -\frac{\partial \mathcal{H}}{\partial M} = \frac{c m}{M^2} \{ \dots \}
\end{aligned} \tag{5-26}$$

We have omitted some expressions in parenthesis to make the notation of system (5-26) less complex; the contents of (...), [...] and {...} are defined by equation (5-25).

When the mass flow rate m is a function of r (solar electric propulsion in interplanetary space), then we have to consider this dependency and form the derivative $\lambda_r = -\partial\mathcal{H}/\partial r$ for the third Lagrange equation of the system (5-26).

According to Pontryagin's maximum principle (5-4), the Hamiltonian (5-25) must be maximized with respect to all possible controls. This means for the thrust angle ϕ :

$$\lambda_u \sin \phi + \lambda_v \cos \phi \stackrel{\perp}{=} \text{maximum} \quad (5-27)$$

The condition (5-5) states the same, since ϕ is a continuous function of time. Thus:

$$\frac{\partial\mathcal{H}}{\partial\phi} = \lambda_u \cos \phi - \lambda_v \sin \phi = 0 \quad \rightarrow \quad \tan \phi = \frac{\lambda_u}{\lambda_v} \quad (5-28)$$

Analogously, we use the condition (5-5) to obtain a control law for the thrust angle δ :

$$\frac{\partial\mathcal{H}}{\partial\delta} = 0 \quad \rightarrow \quad \tan \delta = \frac{\lambda_\vartheta \cos \varphi + \frac{\sin \varphi}{\sin \vartheta} (\lambda_\psi - \lambda_\varphi \cos \vartheta)}{v (\lambda_u \sin \phi + \lambda_v \cos \phi)} \quad (5-29)$$

We have to resist the temptation of using the condition (5-5) again and control the mass flow rate according to $\partial\mathcal{H}/\partial m = 0$. This would lead us to the case of a so-called "singular" thrust arc (\mathcal{H} is a linear function of m , so equation $\partial\mathcal{H}/\partial m = 0$ does not contain m anymore: therefore "singular"). No-one really doubts that these singular thrust arcs are non-optimal (but some mathematicians are interested in singular thrust arcs, because in some special cases it is possible to find analytical solutions). However, the actual optimum control of the thrust magnitude brings us trajectories which consists exclusively of "thrust arcs" and "coast arcs". The mass flow rate must be controlled discontinuously between maximum thrust and zero thrust. Pontryagin's principle (5-4) is still valid, but it is not possible anymore to conclude the validity of condition (5-5). Instead of this we summarize all terms in equation (5-25) which depend explicitly on the mass flow rate m ; and abbreviate the result using the name \mathcal{K} :

$$\mathcal{K} = \left[\frac{c}{M} \{ \cos \delta (\lambda_u \sin \phi + \lambda_v \cos \phi) + \frac{\sin \delta}{v} (-\lambda_\varphi \sin \varphi \cos \vartheta + \lambda_\vartheta \cos \varphi + \lambda_\psi \frac{\sin \varphi}{\sin \vartheta}) \} - \lambda_M \right] \quad (5-30)$$

It is the statement of Pontryagin's maximum principle that, for an optimum control, \mathcal{H} has to be maximized at every time; and this means with respect to equation (5-25):

$$\mathcal{H} = \{ \dots \} + m \cdot \mathcal{K} \stackrel{\perp}{=} \text{maximum} \quad (5-31)$$

\mathcal{K} is the switch function. In thrust periods its value is positive; in coast periods its value is negative. The motor must be switched-off when the switch function changes its sign from positive to negative. At the moment when the switch function becomes positive, the motor must be switched-on again. The instantaneous change of the control variable m will not lead to a discontinuous behaviour of the Hamiltonian \mathcal{H} .

Analysis of the control equations. The equations (5-24) to (5-31) determine the control of an optimal transfer trajectory; however, quite complicated nonlinear differential equations are involved. For a simplification of the notation we introduce now two abbreviations. The first important term is called P (“primer”), defined as:

$$P = \lambda_u \sin \phi + \lambda_v \cos \phi \quad (= \sqrt{\lambda_u^2 + \lambda_v^2}) \quad (5-32)$$

The other abbreviation is the rotation velocity of the actual flight plane, called ω :

$$\omega = \frac{c m}{M} \cdot \frac{\sin \delta}{v} \quad (5-33)$$

The notations for the Hamiltonian \mathcal{H} and the switch function \mathcal{K} become less complex when we use these abbreviations together with the control laws (5-28) and (5-29):

$$\begin{aligned} \mathcal{H} &= \left\{ \left(\frac{v^2}{r} - \frac{\gamma}{r^2} \right) \lambda_u - \frac{u v}{r} \lambda_v + u \lambda_r + \frac{v}{r} \lambda_\varphi \right\} + m \cdot \mathcal{K} \\ \mathcal{K} &= \left[\frac{c}{M} \frac{P}{\cos \delta} - \lambda_M \right] \end{aligned} \quad (5-34)$$

The Hamiltonian \mathcal{H} is not an explicit function of the time t ; therefore \mathcal{H} constitutes an integral of the Lagrange equations. We can re-write the Lagrange equations as:

$$\begin{aligned} \dot{\lambda}_u &= \frac{v}{r} \lambda_v - \lambda_r \\ \dot{\lambda}_v &= -2 \frac{v}{r} \lambda_u + \frac{u}{r} \lambda_v - \frac{1}{r} \lambda_\varphi + \omega P \tan \delta \\ \dot{\lambda}_r &= \left(\frac{v^2}{r^2} - 2 \frac{\gamma}{r^3} \right) \lambda_u - \frac{u v}{r^2} \lambda_v + \frac{v}{r^2} \lambda_\varphi \left(-\frac{\partial m}{\partial r} \cdot \mathcal{K} \right) \\ \dot{\lambda}_M &= \frac{c m}{M^2} \frac{P}{\cos \delta} \end{aligned} \quad (5-35)$$

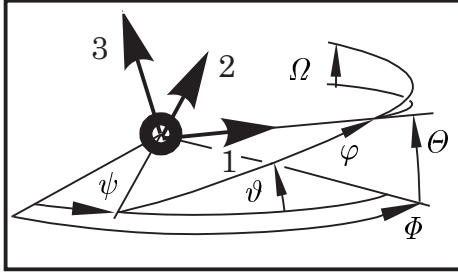
Fortunately, we can integrate three of the seven Lagrange equations analytically:

$$\begin{aligned} \lambda_\varphi &= (C_A \sin \psi - C_B \cos \psi) \sin \vartheta + C_C \cos \vartheta \\ \lambda_\vartheta &= C_A \cos \psi + C_B \sin \psi \\ \lambda_\psi &= C_C \end{aligned} \quad (5-36)$$

It is not easy to find these relationships, but you can easily differentiate them to verify their validity. The terms C_A , C_B and C_C are the integration constants. Now we are able to re-write the thrust angle control law (5-29) in a more transparent form:

$$\begin{aligned} \tan \delta &= \frac{1}{v P} \left(C_A \{ \cos \varphi \cos \psi - \sin \varphi \sin \psi \cos \vartheta \} \right. \\ &\quad \left. + C_B \{ \cos \varphi \sin \psi + \sin \varphi \cos \psi \cos \vartheta \} \right. \\ &\quad \left. + C_C \{ \sin \varphi \sin \vartheta \} \right) \end{aligned} \quad (5-37)$$

The Euler angles φ , ϑ and ψ define the position of the spacecraft and the orientation of the actual flight plane; the thrust angle δ controls the “rotation of the flight plane”.



Alternatively Cardan angles (Φ, Θ, Ω) can be used to define the control law for thrust angle δ :

$$\tan \delta = \frac{1}{v P} (C_A \cos \Phi \cos \Theta + C_B \sin \Phi \cos \Theta + C_C \sin \Theta) \quad (5 - 38)$$

The thrust angle δ can be controlled between $\pm 90^\circ$; and the control law (5-37) (or 5-38) will never become singular. However, the thrust angle ϕ can be controlled in any direction. For numerical calculations it is better to replace equation (5-28) by:

$$\sin \phi = \frac{\lambda_u}{P} \quad ; \quad \cos \phi = \frac{\lambda_v}{P} \quad (5 - 39)$$

This allows a control of ϕ between $\pm 180^\circ$ and avoids the singularity in $\tan \phi$ at $\pm 90^\circ$.

5.1.3 Control Equations for Impulsive Transfers

Impulsive transfer trajectories. Let us now analyze the behaviour of the control equations for the important case of “impulsive transfer trajectories”: when it is assumed that the burning periods are very short in comparison with the total flight time, then it is possible to consider in a simplification that the trajectory is composed of two arc types: impulsive thrust arcs and coast arcs. These impulsive trajectories constitute very often good approximations for real trajectories, particularly when the thrust of the spacecraft is high and the transfer time is long. Fortunately, for impulsive transfer trajectories it is possible to integrate the control equations analytically.

Impulsive thrust arcs. It is assumed that the spacecraft has an infinite mass flow rate ($m \rightarrow \infty$) during an infinitely short time interval ($\Delta t \rightarrow 0$) and as a consequence, that the spacecraft changes instantaneously its velocity and its mass, while it keeps its position in space. The spacecraft velocity after the impulse can be calculated by vector addition of the velocity vector before the impulse and the velocity change $\Delta \vec{v}$. We can use the famous Ciolcovskij equation to establish the relationship between the change of the mass and the change of the velocity; actually we can calculate the time integral of the thrust acceleration ($s = cm/M$) for a certain time interval Δt : M_0 is the mass of the spacecraft immediately before the impulse and M_1 is the mass immediately after the impulse, and the velocity change becomes $\Delta v = -c \ln(M_1/M_0)$ (compare chapter two of this book). Term c is the effective exhaust velocity.

Consider now the equations of motion (5-23) with an infinite mass flow rate ($m \rightarrow \infty$): we can ignore all terms which do not depend directly or indirectly on m . We can conclude that the velocity components u and v change during an impulsive thrust arc while the radius r is constant. Even when the spacecraft position is constant in space, the Euler angles alter because the attitude of the flight plane turns instantaneously. When we use Cardan angles to determine the position, we can see that the angles Φ and Θ are constant and the angle Ω changes according to the integral $\Omega = \int \omega dt$.

The Cardan angles Φ and Θ define a position in space which does not change when an impulse is applied. However, the Cardan angle Ω will alter its value when $\delta \neq 0$ (the thrust angle δ “out of the flight plane” turns the flight plane instantaneously). All the three Euler angles assume different values after the impulse when a rotation of Ω shifts the “line of nodes”; but the left hand sides of the following equations take immediately after the impulse the same values as immediately before the impulse:

$$\begin{aligned} \cos \psi \cos \varphi - \sin \psi \sin \varphi \cos \vartheta &= \text{constant} & (= \cos \Phi \cos \Theta) \\ \sin \psi \cos \varphi + \cos \psi \sin \varphi \cos \vartheta &= \text{constant} & (= \sin \Phi \cos \Theta) \\ \sin \varphi \sin \vartheta &= \text{constant} & (= \sin \Theta) \end{aligned} \quad (5-40)$$

It follows from these relationships (and equation 5-37) that the expression $(vP \tan \delta)$ is constant during an impulsive thrust arc. However, the thrust angle δ changes its values when the coordinate system turns instantaneously (the thrust angles δ and ϕ refer to the moving coordinate system, they can change their values even though the inertial thrust direction is conserved at the moments of impulsive thrusting). Chapter four of this book analyzes the geometry of the maneuver. We can calculate the thrust direction immediately after the impulse (index “1”) from the well-known thrust direction before the impulse (index “0”). The following equations are valid:

$$\begin{aligned} \sin \phi_1 \cos \delta_1 &= \sin \phi_0 \cos \delta_0 \\ v_1 \cos \phi_1 \cos \delta_1 &= \Delta v (1 - \sin^2 \phi_0 \cos^2 \delta_0) + v_0 \cos \phi_0 \cos \delta_0 \\ v_1 \sin \delta_1 &= v_0 \sin \delta_0 \end{aligned} \quad (5-41)$$

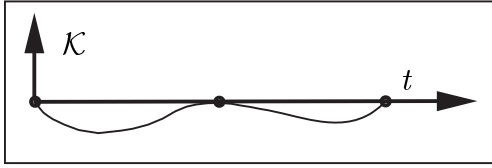
Let us now consider the Lagrange equations (5-35): the control equations degrade (like the equations of motion) when all terms are ignored if they are independent of m . We can see that multipliers λ_u and λ_r do not change during impulsive thrust arcs. Since the expressions $(vP \tan \delta)$ and $(v \sin \delta)$ remain constant during this time interval, we can follow that also the important expression $(P/\cos \delta)$ remains constant. Knowing this, we can calculate the instantaneous change of multipliers λ_v and λ_M :

$$\begin{aligned} \dot{\lambda}_u &= 0 & \lambda_{u1} &= \lambda_{u0} \\ \dot{\lambda}_v &= \omega P \tan \delta & \lambda_{v1} &= \lambda_{v0} (\cos \delta_0 \cos \phi_0) / (\cos \delta_1 \cos \phi_1) \\ \dot{\lambda}_r &= 0 & \lambda_{r1} &= \lambda_{r0} \\ \dot{\lambda}_M &= \frac{c m}{M^2} \frac{P}{\cos \delta} & \lambda_{M1} &= \lambda_{M0} \frac{M_0}{M_1} \end{aligned} \quad (5-42)$$

The equations (5-42) represent the integrals of system (5-35) for impulsive thrust arcs.

The behaviour of the switch function for impulsive transfers. It follows from the theory that the Hamiltonian \mathcal{H} and also the switch function \mathcal{K} (5-34) are always continuous functions of time, also in case of discontinuous control. As a consequence, switch function \mathcal{K} must vanish at moments of infinite thrusting ($\mathcal{K} = 0$ for $m \rightarrow \infty$). Since the expression $(P/\cos \delta)$ remains constant, we can differentiate equation (5-34) at once to form the time derivative of the switch function. The Lagrange equation system (5-42) shows us that also the derivative $\dot{\mathcal{K}}$ vanishes when an impulse is applied.

At moments of impulsive thrust the switch function must accept a convex behaviour: \mathcal{K} and its first time derivative $\dot{\mathcal{K}}$ are zero. Everywhere else the function \mathcal{K} is negative.



$$\begin{aligned}\mathcal{K}_{impulse} &= \frac{c}{M} \frac{P}{\cos \delta} - \lambda_M = 0 \\ \dot{\mathcal{K}}_{midcourse\ impulse} &= 0\end{aligned}\quad (5-43)$$

However, the situation can be different at the initial or at the final instant of the transfer trajectory: we cannot conclude anymore that the expression $(P/\cos \delta)$ is constant for the initial or final impulse; because the time before the initial impulse and the time after the final impulse are not necessarily parts of the optimal transfer trajectory. Therefore the derivative of the switch function can be different from zero ($\dot{\mathcal{K}}$ can be smaller zero at the initial impulse and greater zero at the final impulse); but in any case, at impulses the value of the switch function \mathcal{K} is always equal zero.

Coast arcs. When the motor is switched off ($m = 0$) the spacecraft moves on a conic orbit; and the motion is independent of the behaviour of the control variables. In any case the orbit is a circle, an ellipse, a parabola or a hyperbola. Obviously, the spacecraft mass M is constant for coast flight periods. Now the Euler angles are better conditioned for a description of the motion than the Cardan angles: the node angle ψ and the inclination angle ϑ remain constant; just the path angle φ is a transcendental function of time (determined by Kepler's equation). The radius r , the velocity components u and v can readily be written as functions of the path angle φ . You can find a detailed description of these equations in chapter four of this book.

Obviously, a steering of the thrust direction cannot influence the motion of a spacecraft on a coast arc. However, we cannot simply ignore the control functions during coast arcs: it is necessary to know the behaviour of the Lagrange multipliers (5-35) during these flight intervals, because they define the thrust angles ϕ and δ at the beginning of the next thrust period, and they define the course of the switch function \mathcal{K} . Fortunately, when $m = 0$ we can integrate the Lagrange equations (5-35) analytically:

$$\begin{aligned}\lambda_u &= \mathcal{H} \frac{2r - 3tu}{2e} + C_u u + C_\varphi (\gamma - hv) \\ \lambda_v &= \mathcal{H} \frac{-3tv}{2e} + C_u v + C_v r + C_\varphi (uh + \frac{u\gamma}{v}) \\ \lambda_r &= \mathcal{H} \frac{u - 3t(\gamma/r^2)}{2e} + C_u \frac{\gamma}{r^2} + C_v v + C_\varphi \frac{\gamma u}{r} \\ \lambda_\varphi &= \mathcal{H} \frac{-h}{2e} + C_\varphi \frac{2eh^2 + \gamma^2}{h} \\ \lambda_M &= \text{constant}\end{aligned}\quad (5-44)$$

The terms \mathcal{H} , C_u , C_v and C_φ are the integration constants. \mathcal{H} is the constant value of the Hamiltonian (5-34); and C_φ is equivalent to the constant value given by the first equation of system (5-36). The specific mechanical energy $e = (u^2 + v^2)/2 - \gamma/r$ and the specific angular momentum $h = rv$ are also constant during coast flight periods.

Control equations for the flight on circular orbits. The use of equation system (5-44) is not possible when the spacecraft moves on an orbit with an extremely low eccentricity ($u \rightarrow 0$, $hv \rightarrow \gamma$). We can observe that the equation system is singular for circular orbits (we would have $C_u \rightarrow \infty$, $C_v \rightarrow \infty$, and $C_\varphi \rightarrow \infty$). Therefore a different solution to the system (5-26) has to be used for the motion on circular orbits:

$$\begin{aligned}\lambda_u &= \mathcal{H} \frac{r}{e} + C_1 \sin(\dot{\varphi}t) + C_2 \cos(\dot{\varphi}t) \\ \lambda_v &= \mathcal{H} \frac{-3tv}{2e} + 2C_1 \cos(\dot{\varphi}t) - 2C_2 \sin(\dot{\varphi}t) + C_3 \\ \lambda_r &= \mathcal{H} \frac{-3t(\gamma/r^2)}{2e} + \dot{\varphi}C_1 \cos(\dot{\varphi}t) - \dot{\varphi}C_2 \sin(\dot{\varphi}t) + \dot{\varphi}C_3 \\ \lambda_\varphi &= \mathcal{H} \frac{-h}{2e} \\ \lambda_M &= \text{constant}\end{aligned}\tag{5-45}$$

The equations above are just valid for circular orbits; and, because of this, the solution is of little importance (in reality the spacecraft moves on an elliptic orbit because the eccentricity is never exactly zero). \mathcal{H} , C_1 , C_2 and C_3 are the integration constants. We can see that equation system (5-45) describes a harmonic oscillation for the Lagrange multipliers, with $\dot{\varphi} = v/r$ as the angular velocity of the circular orbit.

5.1.4 Final Conditions for Optimal Transfer Trajectories

Classification of transfer maneuvers. Every transfer trajectory is characterized by its final conditions. The following maneuvers are of practical importance:

Rendezvous: The spacecraft approaches a target object that moves on a predefined target orbit. Finally, at the end of the maneuver, position and velocity of the spacecraft coincide with position and velocity of the target object. The final path angle (the location where the spacecraft encounters the object) is predefined or optimized.

Positioning: The only difference to the rendezvous maneuver is that now the spacecraft enters the target orbit at any time (the transfer time is free). Position and velocity coincide with the target orbit at the final instant; but the target object is somewhere else on the same orbit. Again, the final path angle is predefined or free.

Flyby: Now the spacecraft approaches the target object on a way where the location coincides at the final instant, but not the velocity. The spacecraft does not enter the target orbit, but it remains on the transfer orbit where it has a velocity different from the velocity of the target object. Usually, the spacecraft velocity is free at the final time; while the encounter location on the target orbit is either predefined or free.

Escape: An escape maneuver will give the spacecraft sufficient energy to depart from the gravitational field of the celestial body. The final condition for an escape maneuver is that the spacecraft enters a predetermined hyperbolic (or sometimes parabolic) orbit. The transfer time and the final path angle are free for optimization.

Optimization of the final path angle. Let us now consider a rendezvous or positioning maneuver with a “free” final path angle φ_{end} (the location where the spacecraft enters the target orbit is free for optimization). Then the final values of the state variables (u_{end} , v_{end} , r_{end} , and in case of a rendezvous maneuver also t_{end}) are explicit functions of φ_{end} . The well-known equations for a conic orbit are valid:

$$\begin{aligned} u_{end} &= \frac{\gamma\varepsilon}{h} \sin(\varphi_{end} - \varphi_p) \\ v_{end} &= \frac{\gamma}{h} (1 + \varepsilon \cos(\varphi_{end} - \varphi_p)) \\ r_{end} &= \frac{h^2}{\gamma(1 + \varepsilon \cos(\varphi_{end} - \varphi_p))} \\ t_{end} &= t(\varphi_{end}) \end{aligned} \quad (5-46)$$

The values h (specific angular momentum), ε (eccentricity) and φ_p (pericenter angle) are elements of the predefined target orbit; and, when φ_{end} was given, we could use the equations (5-46) to compute the final state variables. In case of a rendezvous maneuver also the transfer time is a function of φ_{end} (but not quoted here). The formalism of the Hamilton-Lagrange theory allows us to transform these relationships into a final condition for the Lagrange multipliers. The method is called “derivation of transversality conditions”. First we construct the functional \mathcal{J} (5-13):

$$\begin{aligned} \mathcal{J} &= \mu_u (u_{end} - u(\varphi_{end})) + \mu_v (v_{end} - v(\varphi_{end})) + \mu_r (r_{end} - r(\varphi_{end})) \\ &\quad + \mu_t (t_{end} - t(\varphi_{end})) \end{aligned} \quad (5-47)$$

Then we differentiate this functional \mathcal{J} according to the conditions (5-14) and (5-15):

$$\begin{aligned} \lambda_{u,end} &= \frac{\partial \mathcal{J}}{\partial u_{end}} = \mu_u ; & \lambda_{v,end} &= \frac{\partial \mathcal{J}}{\partial v_{end}} = \mu_v ; & \lambda_{r,end} &= \frac{\partial \mathcal{J}}{\partial r_{end}} = \mu_r \\ \lambda_{\varphi,end} &= \frac{\partial \mathcal{J}}{\partial \varphi_{end}} = -\mu_u \frac{\partial u(\varphi_{end})}{\partial \varphi_{end}} - \mu_v \frac{\partial v(\varphi_{end})}{\partial \varphi_{end}} - \mu_r \frac{\partial r(\varphi_{end})}{\partial \varphi_{end}} - \mu_t \frac{\partial t(\varphi_{end})}{\partial \varphi_{end}} \\ \mathcal{H}_{end} &= - \frac{\partial \mathcal{J}}{\partial t_{end}} = -\mu_t \end{aligned}$$

It is easy to calculate the derivatives of the functions (5-46) with respect to the final path angle φ_{end} when we consider that $\partial t(\varphi_{end})/\partial \varphi_{end} = 1/\dot{\varphi}_{end}$ and $h = r_{end}v_{end}$. After having calculated the derivatives we eliminate the multipliers μ_u , μ_v , μ_r , μ_t from the five equations above. The result is one final condition at the final time t_{end} :

$$\mathcal{H} = \left(\frac{v^2}{r} - \frac{\gamma}{r^2} \right) \lambda_u - \frac{u v}{r} \lambda_v + u \lambda_r + \frac{v}{r} \lambda_\varphi \quad (5-48)$$

This condition is also valid on coast arcs: when the final path angle φ_{end} is free, the switch function \mathcal{K} is zero at the final time t_{end} , and the final orbit is a part of the transfer trajectory. However, the condition (5-48) is not valid when φ_{end} is restricted; then the final \mathcal{K} is positive and the final orbit is not a part of the transfer trajectory.

Optimization of the flight time. It is the important point in question whether the equations of motion contain quantities which depend explicitly on the time t . We can observe that there are no explicit time dependencies in the equation system (5-23), and we can follow for the Hamiltonian (5-24) that $\partial\mathcal{H}/\partial t = 0$. Then condition (5-6) states that $d\mathcal{H}/dt = 0$: the Hamiltonian has the same constant value \mathcal{H} on the whole optimal trajectory (but the value of \mathcal{H} is not constant when the thrusting is a pre-determined time function; for example when c or m are explicit functions of the time.)

Additionally, it follows from condition (5-17) that $\mathcal{H} = 0$ at the final instant of a maneuver with a free transfer time (unrestricted or optimized). We may summarize:

$$\begin{aligned}\mathcal{H}_{final\ time\ optimized} &= 0 \\ \mathcal{H}_{final\ time\ predefined} &= constant\end{aligned}\quad (5-49)$$

For positioning maneuvers the value of \mathcal{H} vanishes during the complete transfer time; for rendezvous maneuvers \mathcal{H} is not zero, usually, but \mathcal{H} is another constant value.

Final conditions for flyby maneuvers. We talk about a flyby maneuver when the spacecraft encounters the target object with a relative velocity. The location where the flyby takes place is a position on the final target orbit; this place is either entirely predetermined or a function of the final path angle φ_{end} of the target object.

The flyby velocity is usually free for optimization. When the velocity components u_{end} and v_{end} are unrestricted, we can consider the condition (5-16) and conclude:

$$\lambda_{u,end} = 0 \quad ; \quad \lambda_{v,end} = 0 \quad (5-50)$$

This means that the control law (5-28) for thrust angle ϕ becomes an expression of the type 0/0, and ϕ is undefined at the final instant. The plausible result is that the thrust angle ϕ can be directed arbitrarily at the moment of the encounter.

We may expect that also the control law for the thrust angle δ is an expression of the type 0/0 at the final instant of a flyby trajectory. The transversality conditions can be used to demonstrate that this is the case: the location of the encounter is predetermined for a flyby maneuver and the encounter velocity is free (in contrast to a rendezvous maneuver where both, location and velocity, have finally to coincide with the corresponding values of the target orbit). A location in space is entirely determined by three values (for example: radius r and the Cardan angles Φ and Θ). When the location of the flyby maneuver is given, then the expressions $(\sin \Phi \cos \Theta)$ and $(\sin \Theta)$ are predetermined at the final time, and we can write for the instant t_{end} :

$$\begin{aligned}\cos \psi \cos \varphi - \sin \psi \sin \varphi \cos \vartheta &= \sin \Phi \cos \Theta \quad (= \text{predetermined}) \\ \sin \varphi \sin \vartheta &= \sin \Theta \quad (= \text{predetermined})\end{aligned}\quad (5-51)$$

The Euler angle ψ defines the orientation of the “line of nodes”. When we change ψ (shift the line of nodes) and keep the flyby location constant, then the Euler angles ϑ and φ change their values; but the left sides of the equations (5-51) remain constant.

The equations (5-51) are “restrictions in equation form” at the final time of the flyby trajectory. We construct the functional \mathcal{J} (5-13) for the Lagrange multipliers (5-14):

$$\begin{aligned}\mathcal{J} &= \mu_1(\cos \psi \cos \varphi - \sin \psi \sin \varphi \cos \vartheta - \sin \Phi \cos \Theta) + \mu_2(\sin \varphi \sin \vartheta - \sin \Theta) \\ \lambda_\varphi &= \partial J / \partial \varphi = \mu_1(-\sin \varphi \cos \psi - \sin \psi \cos \varphi \cos \vartheta) + \mu_2(\sin \vartheta \cos \varphi) \\ \lambda_\vartheta &= \partial J / \partial \vartheta = \mu_1(\sin \psi \sin \varphi \sin \vartheta) + \mu_2(\sin \varphi \cos \vartheta) \\ \lambda_\psi &= \partial J / \partial \psi = \mu_1(-\sin \psi \cos \varphi - \cos \psi \sin \varphi \cos \vartheta)\end{aligned}\quad (5-52)$$

After elimination of the unknown multipliers μ_1 and μ_2 we get the final condition:

$$\lambda_\vartheta \cos \varphi + (\lambda_\psi - \lambda_\varphi \cos \vartheta) \cdot \frac{\sin \varphi}{\sin \vartheta} = 0 \quad (5-53)$$

When we compare this result with the equation (5-29) and consider that the multipliers λ_u and λ_v equal zero at the final instant, then we see that also the control law for the thrust angle δ becomes an expressions of type 0/0 at the end of a flyby trajectory. As expected, the thrust direction is not defined at the moment of the encounter.

Final conditions for escape maneuvers. “Maximization of mechanical energy” is the optimization objective for escape maneuvers. We can write for the energy e :

$$e = \frac{u^2 + v^2}{2} - \frac{\gamma}{r} \stackrel{!}{=} \text{maximum} \quad (5-54)$$

Therefore we establish the functional \mathcal{J} (5-13) and differentiate (according to 5-14):

$$\mathcal{J} = \mu \left(\frac{u^2 + v^2}{2} - \frac{\gamma}{r} - e \right) \quad (5-55)$$

$$\lambda_u = \frac{\partial \mathcal{J}}{\partial u} = \mu u \quad ; \quad \lambda_v = \frac{\partial \mathcal{J}}{\partial v} = \mu v \quad ; \quad \lambda_r = \frac{\partial \mathcal{J}}{\partial r} = \mu \gamma / r^2$$

The thrust angle ϕ is always controlled as $\tan \phi = \lambda_u / \lambda_v$ (compare equation 5-28). We can conclude that the final condition for escape maneuvers is “tangential thrust”:

$$\tan \phi_{end} = \frac{u}{v} \quad (5-56)$$

To get the “mechanical work” of a force that acts along a way we have to integrate the dot vector product of the force vector and the infinitesimal distance vector. Analogously, the time derivative of the specific mechanical energy e is equivalent to the dot vector product of thrust acceleration vector \vec{s} and velocity vector \vec{v} . Therefore:

$$\dot{e} = \vec{s} \cdot \vec{v} \quad (5-57)$$

Obviously, the thrust vector must be aligned with the velocity vector for the fastest increase of energy. This statement is true for any instant of the flight time. We might expect that “a maximization of the final energy” requests having tangential thrust all the time. Surprisingly this is not the case: energy maximization does not request tangential thrusting; but, at least, it requests tangential thrusting at the final instant.

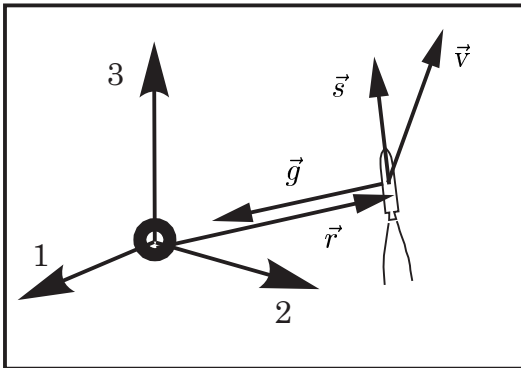
5.2. Behaviour of Lagrange Multipliers and Switch Function

We have seen that spacecraft trajectory optimization involves “two point boundary value problems”, where we actually have to integrate differential equations and additionally find the solutions to nonlinear equation systems. When we integrate the equations of motion and the Lagrange equations simultaneously, we must observe that the switch function is valid on the optimal trajectory; it defines whether a thrust arc or a coast arc is present. The numerical integration of the trajectory must satisfy special final conditions. The theory sounds easy, but put into practice is difficult. There is the danger that a trajectory is non-optimal because it is based on wrong assumptions (switch function must be correct and final conditions must be satisfied). This justifies a closer look at the same problem, using different approaches.

5.2.1 Vector Representation of the Optimal Trajectories

Equations of motion for a transfer trajectory in vector form. The equations of motion are vector equations. We can write down a very compact representation for optimal trajectories when we use the vector form of an inertial system, and such a form is quite transparent to draw some general conclusions (but this vector form is not advisable for a numerical integration: the specification of orbits is complicated; and anyway the vectors have to be resolved into their component notations).

Let \vec{r} be the location vector and \vec{v} be the velocity vector. The thrust acceleration vector \vec{s} and the gravitational acceleration vector \vec{g} act on the spacecraft (mass M):



$$\begin{aligned}\dot{\vec{v}} &= \vec{s} + \vec{g} \\ \dot{\vec{r}} &= \vec{v} \\ \dot{M} &= -m\end{aligned}\quad (5-58)$$

$$\left(|\vec{s}| = \frac{c \cdot m}{M} ; \quad \vec{g} = -\frac{\gamma \vec{r}}{|\vec{r}|^3} \right)$$

The mass M diminishes in time t ; and the amount \dot{M} (the time derivative of M) is equivalent to the rate of expenditure of propellant m . The expression $c \cdot m$ represents the thrust of the engine; we assume a constant exhaust velocity c . The equation system (5-58) represents seven first order differential equations (like the system 5-23): we have three equations for the location vector and three equations for the velocity vector; and one equation to determine the actual mass. Using inertial components, the vectors can be written as $\vec{r} = (r_1, r_2, r_3)$, $\vec{v} = (v_1, v_2, v_3)$ and $\vec{g} = (g_1, g_2, g_3)$.

We do not necessarily consider a central gravitational field coincident with the center of the coordinate system (in this case we may write for the gravitational acceleration $\vec{g} = -\gamma \vec{r} / |\vec{r}|^3$). The gravitational field may also be different: it may be the asymmetric field of an oblate celestial body, a uniform field with parallel gravity or even the field of several celestial bodies. The only assumption is that the vector of gravitational attraction \vec{g} depends on nothing else than on the position \vec{r} of the spacecraft.

The thrust acceleration vector \vec{s} is subjected to the following set of constraints:

$$\begin{aligned} |\vec{s}| &= \sqrt{s_1^2 + s_2^2 + s_3^2} \quad (= cm/M) \\ 0 &\leq m \leq \text{constant maximum} \end{aligned} \tag{5-59}$$

We assume a discontinuous control of the mass flow rate m between $m = 0$ (coast arcs) and $m = \text{constant maximum}$ (thrust arcs). The thrust direction is not restricted.

Control of the thrust direction. The Hamilton-Lagrange theory accepts a compact form when we use vectors for the representation of the optimal thrust direction. Following the equation (5-9) we establish the Hamiltonian \mathcal{H} as:

$$\mathcal{H} = \vec{\lambda}_v \cdot (\vec{s} + \vec{g}) + \vec{\lambda}_r \cdot \vec{v} - \lambda_M \cdot m + \mu \cdot \left(\frac{cm}{M} - |\vec{s}| \right) \tag{5-60}$$

Remember that \mathcal{H} has to be maximized for an optimal control (maximum principle). We can observe that for an optimal control the thrust direction \vec{s} must be aligned with the vector $\vec{\lambda}_v$ (the dot product takes its maximum when both vectors are aligned). The vector $\vec{\lambda}_v$ is an important quantity, it is renamed and called \vec{P} “primer vector”. Every time the thrust vector \vec{s} must be directed into the same direction as this primer vector \vec{P} . The time function of the primer vector is defined by a second order differential equation; we find it when we form the Lagrange equations (5-3). Thus:

$$\left. \begin{aligned} \dot{\vec{\lambda}}_v &= -\vec{\lambda}_r \\ \dot{\vec{\lambda}}_r &= -\nabla(\vec{\lambda}_v \cdot \vec{g}) \end{aligned} \right\} \ddot{\vec{P}} = \nabla(\vec{P} \cdot \vec{g}) \tag{5-61}$$

The equation (5-61) determines the primer vector (operator ∇ is called “Nabla”). When we want to integrate this second order differential vector equation numerically, we need to know five initial values for the primer vector and its time derivative (the length of \vec{P} is not important to the direction of \vec{s}). Everywhere on the optimal trajectory, the primer vector and its first time derivative are continuous time functions.

Term μ (equation 5-60) is a quantity with a physical interpretation: μ is equivalent to the length of the primer vector. Since the thrust direction (vector \vec{s}) is a continuous function of time, we may apply condition (5-5) to every component of \vec{s} and follow:

$$\left. \vec{\lambda}_v = \mu \cdot \frac{\vec{s}}{|\vec{s}|} \right\} \mu = |\vec{P}| = \sqrt{P_1^2 + P_2^2 + P_3^2} \tag{5-62}$$

The length of the primer vector \vec{P} is important to find the optimal thrust magnitude.

Optimal thrust magnitude control. The next step is the analysis of the behaviour of the switch function \mathcal{K} . Hamiltonian \mathcal{H} (equation 5-60) is rearranged to find \mathcal{K} as:

$$\mathcal{H} = \left\{ \vec{\lambda}_v \cdot (\vec{s} + \vec{g}) + \vec{\lambda}_r \cdot \vec{v} - \mu \cdot |\vec{s}| \right\} + m \cdot \left[\mu \cdot \frac{c}{M} - \lambda_M \right] \quad (5-63)$$

$$\mathcal{K} = \left[\mu \cdot \frac{c}{M} - \lambda_M \right] \quad (5-64)$$

The expression in curly braces is the value of Hamiltonian \mathcal{H} on coast arcs ($m = 0$). The expression in square brackets is switch function \mathcal{K} . The motor must be switched discontinuously between maximum thrust and zero thrust: the switch function is positive for thrust arcs and negative for coast arcs. \mathcal{K} is a function of two multipliers: μ (the length of the primer vector) and λ_M (adjoined to the equation for mass M). According to the condition (5-3) we form the Lagrange equation for multiplier λ_M :

$$\dot{\lambda}_M = \mu \cdot \frac{cm}{M^2} \quad (= |\vec{P}| \cdot \frac{cm}{M^2}) \quad (5-65)$$

The time derivative $\dot{\mathcal{K}}$ of the switch function can also be denoted in a simple form. We differentiate \mathcal{K} (equation 5-64) and insert the relationship (5-65). Finally we get:

$$\dot{\mathcal{K}} = \dot{\mu} \cdot \frac{c}{M} \quad (= |\dot{\vec{P}}| \cdot \frac{c}{M}) \quad (5-66)$$

Obviously, with $m = 0$ both, spacecraft mass M (equation 5-58) and multiplier λ_M (equation 5-65), are constant if the motor is switched-off. During the time interval of a coast arc the switch function is a direct function of the length μ of primer vector \vec{P} .

Differential vector equations for optimal maneuvers. Finally, we can write down the conditions for optimal trajectory control in their most compact vector form:

$$\ddot{\vec{r}} = \vec{s} + \vec{g} \quad (5-67)$$

$$\ddot{\vec{P}} = \nabla(\vec{P} \cdot \vec{g}) \quad (5-68)$$

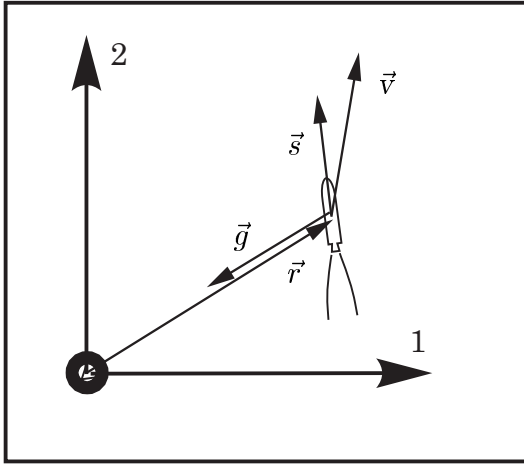
$$\dot{M} = -m \quad (5-69)$$

$$\dot{\mathcal{K}} = |\dot{\vec{P}}| \cdot c/M \quad (5-70)$$

We use exclusively location vector \vec{r} , thrust vector \vec{s} , gravity vector \vec{g} and primer vector \vec{P} . Every time vector \vec{P} is aligned with vector \vec{s} . Spacecraft mass M , mass flow rate m , constant exhaust velocity c and switch function \mathcal{K} are not vector quantities. When $K > 0$ we have $m = \text{constant maximum}$; when $K < 0$ we have $m = 0$. Seven initial state variables are given: the predetermined initial orbit allows us to calculate the vector \vec{r} and its time derivative $\dot{\vec{r}}$, and the initial mass M is also well-known. Usually, nothing is known about the initial values of the control variables. We can choose any value for the initial length of \vec{P} (for example $|\vec{P}| = 1$), and then we have to find six more initial values (five for \vec{P} , $d\vec{P}/dt$ and one for \mathcal{K}) to meet six predetermined state variables \vec{r} and $\dot{\vec{r}}$ at final time. The final mass M is maximized.

5.2.2 Plane Optimal Trajectories

Optimal control using inertial coordinates. Unfortunately, analytical solutions to the behaviour of the primer vector in a central gravitational field are not available. Anyway, we depend on numerical integration procedures. When we use an inertial representation of the equations of motion, we must resolve all vectors to get their component form. In case of a plane motion we just have two-dimensional vectors:



$$\begin{aligned} \ddot{r}_1 &= \frac{cm}{M} \cdot \frac{s_1}{s} - g \frac{r_1}{r} \\ \ddot{r}_2 &= \frac{cm}{M} \cdot \frac{s_2}{s} - g \frac{r_2}{r} \\ \dot{M} &= -m \end{aligned} \tag{5-71}$$

$$\begin{aligned} \text{with : } r &= \sqrt{r_1^2 + r_2^2} \\ s &= \sqrt{s_1^2 + s_2^2} = \frac{cm}{M} \\ g &= \sqrt{g_1^2 + g_2^2} = \frac{\gamma}{r^2} \end{aligned}$$

The equation (5-68) determines the primer vector P ; but we must resolve this vector representation and form the component notation (evaluating the “Nabla” operator).

Considering that $g_1 = -\gamma r_1 / \sqrt{r_1^2 + r_2^2}^3$ and $g_2 = -\gamma r_2 / \sqrt{r_1^2 + r_2^2}^3$ we can denote:

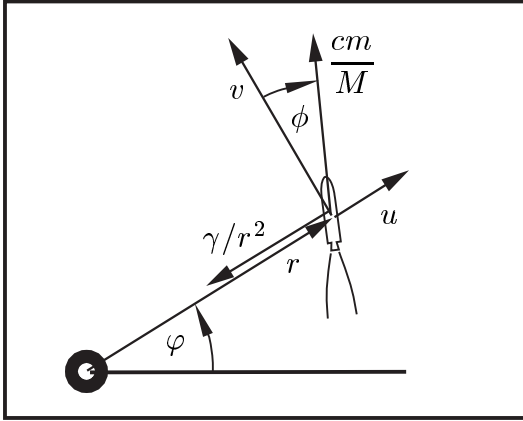
$$\begin{aligned} \ddot{P}_1 &= P_1 \cdot \frac{\partial g_1}{\partial r_1} + P_2 \cdot \frac{\partial g_2}{\partial r_1} = \gamma P_1 \left(\frac{3r_1^2}{r^5} - \frac{1}{r^3} \right) + \gamma P_2 \left(\frac{3r_1 r_2}{r^5} \right) \\ \ddot{P}_2 &= P_1 \cdot \frac{\partial g_1}{\partial r_2} + P_2 \cdot \frac{\partial g_2}{\partial r_2} = \gamma P_1 \left(\frac{3r_1 r_2}{r^5} \right) + \gamma P_2 \left(\frac{3r_2^2}{r^5} - \frac{1}{r^3} \right) \end{aligned} \tag{5-72}$$

Systems (5-71) and (5-72) have to be integrated simultaneously. Therefore we have to transform these second order differential equations into first order representations. For the time derivatives of the components of location vector \vec{r} we can simply use the velocity components (for example $\dot{r}_1 = v_1$ and $\dot{r}_2 = v_2$); for the time derivatives of the components of the primer vector \vec{P} we may write \dot{P}_1 and \dot{P}_2 . We need the time derivatives of the components of the primer vector anyway for a determination of the switch function \mathcal{K} . According to the equation (5-70) we can conclude:

$$\dot{\mathcal{K}} = \frac{c}{M} \cdot (P_1 \dot{P}_1 + P_2 \dot{P}_2) / \sqrt{P_1^2 + P_2^2}$$

The numerical integration of the trajectory is still troublesome: we have to guess four initial control variables to make an agreement with four final state variables (or four transversality conditions). The fifth state variable (or a functional of the final state and the final time) represents the optimization objective. Since the orbit definition is circumstantial anyway, it cannot be recommended to use the inertial coordinate representation for the numerical optimization of transfer trajectories.

Optimal control using polar coordinates. Polar coordinates are better conditioned for the description of orbital motion (5-23). Considered is a spacecraft that moves in the central gravitational field, surrounding a celestial body. Again we examine a plane motion where the thrust angle δ vanishes. The position of the spacecraft is determined by r (radius) and φ (path angle), with pole at the center of attraction. Using velocity components $u = \dot{r}$ (vertical) and $v = r\dot{\varphi}$ (horizontal), the motion can be formulated by a system of first-order differential equations. Symbol γ represents the gravitational constant; and again the effective exhaust velocity c is assumed to be constant. M is the mass of the spacecraft and m its propellant consumption rate.



The plane equations of motion are then:

$$\begin{aligned}\dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} + \frac{c m}{M} \sin \phi \\ \dot{v} &= -\frac{u v}{r} + \frac{c m}{M} \cos \phi \\ \dot{r} &= u \\ \dot{\varphi} &= \frac{v}{r} \\ \dot{M} &= -m\end{aligned}\quad (5-73)$$

The powered motion should be optimally controlled when the spacecraft transfers from one orbit to another coplanar orbit. To find optimal time functions for thrust angle ϕ and mass flow rate m we follow the optimization formalism of the Hamilton Lagrange theory, equations (5-1) to (5-7). The Hamiltonian \mathcal{H} is constructed as:

$$\begin{aligned}\mathcal{H} &= \lambda_u \left(\frac{v^2}{r} - \frac{\gamma}{r^2} + \frac{c m}{M} \sin \phi \right) \\ &+ \lambda_v \left(-\frac{u v}{r} + \frac{c m}{M} \cos \phi \right) + \lambda_r u + \lambda_\varphi \left(\frac{v}{r} \right) - \lambda_M m\end{aligned}\quad (5-74)$$

The behaviour of the Lagrange multipliers is defined by the Lagrange equations (5-3):

$$\begin{aligned}\dot{\lambda}_u &= -\frac{\partial \mathcal{H}}{\partial u} = \frac{v}{r} \lambda_v - \lambda_r \\ \dot{\lambda}_v &= -\frac{\partial \mathcal{H}}{\partial v} = -2\frac{v}{r} \lambda_u + \frac{u}{r} \lambda_v - \frac{1}{r} \lambda_\varphi \\ \dot{\lambda}_r &= -\frac{\partial \mathcal{H}}{\partial r} = \left(\frac{v^2}{r^2} - 2\frac{\gamma}{r^3} \right) \lambda_u - \frac{u v}{r^2} \lambda_v + \frac{v}{r^2} \lambda_\varphi \\ \dot{\lambda}_\varphi &= -\frac{\partial \mathcal{H}}{\partial \varphi} = 0 \\ \dot{\lambda}_M &= -\frac{\partial \mathcal{H}}{\partial M} = \frac{c m}{M^2} (\lambda_u \sin \phi + \lambda_v \cos \phi)\end{aligned}\quad (5-75)$$

Note that multipliers λ in these differential equations are continuous time functions.

Then, the optimal control law for thrust angle ϕ is determined by the equation (5-5):

$$\frac{\partial \mathcal{H}}{\partial \phi} = \lambda_u \cos \phi - \lambda_v \sin \phi = 0 \quad ; \quad \tan \phi = \frac{\lambda_u}{\lambda_v} \quad (5-76)$$

Besides the direction of thrust, the spacecraft can be controlled by its thrust level. Usually, the thrust of a rocket engine (given by the product of fuel consumption m and effective exhaust velocity c) is constant. Such is often the case for rocket engines using chemical propellant. A decrease in the rate of propellant expenditure m will throttle the motor, but this raises technical problems and trajectories flown with throttled engines (“singular thrust arcs”) are non-optimal. Optimal trajectories consist of flight periods using maximum thrust and flight periods with the engine off.

According to Pontryagin’s principle (equation 5-4), Hamiltonian \mathcal{H} and Lagrange multipliers λ must be continuous at moments of discontinuous control. If the multipliers do not vanish simultaneously, it follows from this condition that thrust angle ϕ and its first time derivative $\dot{\phi}$ are continuous along the whole optimal trajectory.

The course of mass flow rate m , however, is discontinuous at moments when the engine is switched. We may conclude that a continuous course of Hamilton function and Lagrange multipliers demands the following conditions at the switching times:

$$\begin{aligned} \mathcal{H} &= \left(\frac{v^2}{r} - \frac{\gamma}{r^2} \right) \lambda_u - \frac{u v}{r} \lambda_v + u \lambda_r + \frac{v}{r} \lambda_\varphi \\ \mathcal{K} &= \frac{c}{M} (\lambda_u \sin \phi + \lambda_v \cos \phi) - \lambda_M = 0 \end{aligned} \quad (5-77)$$

The final values of the Hamiltonian \mathcal{H} and the Lagrange multipliers λ depend on the restrictions at the final time, called “transversality conditions”. The Hamiltonian \mathcal{H} is a constant since it does not depend explicitly on the time t . In case the flight time is not predetermined, \mathcal{H} vanishes at the final time and therefore along the whole optimal trajectory. Of course, flight-time-unrestricted trajectories demand less fuel consumption than trajectories with predetermined time of arrival.

Often, we also have no restrictions for the final path angle φ . Thus, the concerning multiplier λ_φ will accept the value zero at the final instant; and, since λ_φ is a constant, it will also vanish along the whole optimal trajectory. Therefore we may write for transfer trajectories with unrestricted final time and unrestricted final path angle:

$$\lambda_\varphi = 0 \quad (5-78)$$

$$\mathcal{H} = 0 \quad (5-79)$$

The problem now is to calculate the initial values of the Lagrange equations in such a way, that the spacecraft reaches a desired terminal orbit after having flown its transfer trajectory. A flight path controlled by these equations maximizes the final value of the rocket mass M . Analytical solutions can be derived for flight periods with switched-off engine; but under all circumstances trajectory integration has to be done on a computer, because analytical solutions to thrust periods are not available.

Optimal control using energy and angular momentum as coordinates. We can use alternatively inertial coordinates or polar coordinates to formulate the equations of motion, expecting that the conditions for optimal control are substantially affected by the form we represent the motion with. The actual trajectory in physical sense is always the same (always it is just a different representation of the same flight path through space). Use of quite different approaches to find equivalent laws can make a complicated problem more transparent: as another completely different alternative we can use energy e and angular momentum h to represent the plane motion of a rocket powered spacecraft in the central gravitational field of a celestial body:

$$\begin{aligned} e &= \frac{1}{2}(u^2 + v^2) - \frac{\gamma}{r} \\ h &= r v \end{aligned} \quad (5-80)$$

Differentiation of the equations above enables us to replace the system (5-73) by:

$$\begin{aligned} \dot{e} &= \frac{c m}{M}(u \sin \phi + v \cos \phi) \\ \dot{h} &= \frac{c m}{M}(r \cos \phi) \\ \dot{r} &= u \\ \dot{\phi} &= \frac{v}{r} \\ \dot{M} &= -m \end{aligned} \quad (5-81)$$

Specific energy e and specific angular momentum h of the orbit are “state variables”, together with polar coordinates r and ϕ and rocket mass M . The velocity components u (vertical) and v (horizontal) are no state variables anymore; they are treated as if they were control variables; u and v are not free for optimization but subjected to constraints, as represented by the equations (5-80). The only real control variables are ϕ and m . Thrust angle ϕ is a continuous function of time; however, mass flow rate m is a discontinuous function of time (motor switched-on or switched-off). Like always, we consider a thrust acceleration cm/M with constant maximum mass flow rate and constant exhaust velocity c . The gravitational constant is called γ .

Since we have restrictions in form of equations, we must now use scheme (5-9) to establish the Hamiltonian. The multipliers adjoined to the differential equations are termed λ ; the multipliers adjoined to the constraints are termed μ . We denote \mathcal{H} as:

$$\begin{aligned} \mathcal{H} &= \lambda_e \frac{cm}{M}(u \sin \phi + v \cos \phi) + \lambda_h \frac{cm}{M}(r \cos \phi) + \lambda_r u + \lambda_\phi \frac{v}{r} - \lambda_M m \\ &+ \mu_e \left(\frac{1}{2}(u^2 + v^2) - \frac{\gamma}{r} - e \right) + \mu_h (rv - h) \end{aligned} \quad (5-82)$$

Under certain circumstances it is possible that we consider a simplified problem: \mathcal{H} becomes constantly zero when we regard problems with unrestricted final time; and if we impose no restriction concerning the final path angle ϕ , we may even ignore the equation $\dot{\phi} = v/r$ (since \mathcal{H} is not an explicit function of ϕ we have $\lambda_\phi = 0$).

Application of law (5-3) brings us the adjointed system of Lagrange equations. Thus:

$$\begin{aligned}
 \dot{\lambda}_e &= -\frac{\partial \mathcal{H}}{\partial e} = \mu_e \\
 \dot{\lambda}_h &= -\frac{\partial \mathcal{H}}{\partial h} = \mu_h \\
 \dot{\lambda}_r &= -\frac{\partial \mathcal{H}}{\partial r} = -\lambda_h \frac{cm}{M} \cos \phi + \lambda_\varphi \frac{v}{r^2} - \mu_e \frac{\gamma}{r^2} - \mu_h v \\
 \dot{\lambda}_\varphi &= -\frac{\partial \mathcal{H}}{\partial \varphi} = 0 \\
 \dot{\lambda}_M &= -\frac{\partial \mathcal{H}}{\partial M} = \frac{cm}{M^2} (\lambda_e (u \sin \phi + v \cos \phi) + \lambda_h r \cos \phi) \quad (5-83)
 \end{aligned}$$

The velocity components u and v are considered as control variables, subjected to constraints. We can use law (5-5) twice to find the courses of multipliers μ_e and μ_h :

$$\begin{aligned}
 0 &= \frac{\partial \mathcal{H}}{\partial u} = \lambda_e \frac{cm}{M} \sin \phi + \lambda_r + \mu_e u \\
 0 &= \frac{\partial \mathcal{H}}{\partial v} = \lambda_e \frac{cm}{M} \cos \phi + \lambda_\varphi \frac{1}{r} + \mu_e v + \mu_h r \quad (5-84)
 \end{aligned}$$

The “control equations” (5-84) are just helpful to eliminate the unknown multipliers μ_e and μ_h from the system of Lagrange equations (5-83) (the procedure is straightforward and not quoted here). Finally, control law (5-5) for thrust angle ϕ becomes:

$$0 = \frac{\partial \mathcal{H}}{\partial \phi} = \frac{cm}{M} (\lambda_e (u \cos \phi - v \sin \phi) - \lambda_h r \sin \phi) \quad (5-85)$$

The switch conditions for the optimal control of mass flow rate m can be stated as:

$$\mathcal{H} = \lambda_r u + \lambda_\varphi \frac{v}{r} \quad (5-86)$$

$$\mathcal{K} = \frac{c}{M} (\lambda_e (u \sin \phi + v \cos \phi) + \lambda_h (r \cos \phi)) - \lambda_M = 0 \quad (5-87)$$

We can observe that the representation of motion and control using orbital energy and angular momentum as coordinates is helpful to understand the behaviour of the control functions for the motion on coast arcs (with $m = 0$ the state variables e , h , M and the multiplier λ_M remain constant). Actually, it is possible to find the complete analytical solution (5-44) of system (5-35) for coast arcs by a similar method (using Lagrange’s equation for the perturbation of orbital elements as a replacement for the equations of motion). For the motion on thrust arcs the representation in such a form is not very helpful. The set of equations which we have found is equivalent to sets provided by analysis of motion using inertial coordinates or polar coordinates (system 5-71 or system 5-73): we do not get relationships which can bring us to new conclusions. However, we can see that quite different approaches are possible for the same problem when we use different coordinates or different equations of motion. It depends on the point of view, which representation is the most transparent.

5.2.3 Elimination of the Lagrange Multipliers

Plane thrust angle control. As we have seen, so-called Lagrange multipliers determine the optimal control of the thrust vector (its direction and magnitude). These multipliers do not really have a physical interpretation; they are given as functions of time in the form of differential equations, where the initial conditions are missing. The combination of all this creates problems: when you do not have the analytical solution to a system of differential equations, and when you also do not have the initial conditions for these equations, then the analysis is usually difficult. In some cases it is possible to replace the Lagrange multipliers by notations with the thrust angles and their time derivatives. Again, the replacement does not bring new relationships; but thrust angles have a physical meaning, and therefore a notation with the thrust angles and their time derivatives can be more transparent.

We make use of the gravitational acceleration $g = \gamma/r^2$ and the thrust acceleration $s = cm/M$ and transform the system (5-73) to represent it by two second-order differential equations. Radius r and path angle φ determine the position of the spacecraft:

$$\begin{aligned} \text{vertically :} \quad & \ddot{r} - r\dot{\varphi}^2 = s \sin \phi - g \\ \text{horizontally :} \quad & r\ddot{\varphi} + 2\dot{r}\dot{\varphi} = s \cos \phi \end{aligned} \quad (5-88)$$

In the equation system above, intentional use of $\dot{r} = u$ and $\dot{\varphi}r = v$ is made for the components of the velocity vector: we want to minimize the number of different terms in the equations. The thrust angle ϕ should be controlled optimally. The first three Lagrange equations (5-75) together with control equation (5-76) provide a set of four equations to calculate ϕ (assuming $\lambda_\varphi = 0$). We use analytical differentiation and eliminate the Lagrange multipliers λ_u , λ_v and λ_r . The differentiation allows us to replace the set of four control equations by one second-order differential equation:

$$r(\ddot{\phi} - \ddot{\varphi}) + 2\dot{r}(\dot{\phi} - \dot{\varphi}) - 3g \sin \phi \cos \phi + 2r(\dot{\phi}^2 - 3\dot{\phi}\dot{\varphi} + 2\dot{\varphi}^2) \tan \phi = 0 \quad (5-89)$$

Control of thrust angle ϕ according to the equation above gives the final value of the spacecraft mass M a (relative) maximum. The differential equation (5-89) constitutes a necessary condition for a time optimal or fuel-consumption optimal control. The equation is valid for optimal or non-optimal functions of thrust acceleration s (even if s is a discontinuous time function, for example in case of multi-stage rockets).

Equivalent (but better conditioned for application on computers) is the first-order differential equation form of this control law. We introduce the variable λ and write:

$$\begin{aligned} \dot{\phi} &= \lambda (\cos \phi / r)^2 + \dot{\varphi} \\ \dot{\lambda} &= (4\lambda\dot{\varphi} + 3g r) \tan \phi \end{aligned} \quad (5-90)$$

Like the Lagrange multipliers, the “help-variable” λ has no physical interpretation; λ is just a nonlinear function of the thrust angle ϕ and its first time derivative $\dot{\phi}$.

Thrust magnitude control. During thrust-flight periods the rocket motor operates with maximum thrust; and the spacecraft expels fuel mass. Consequently the thrust acceleration s increases in time. The working rocket engine must be switched-off at the instant when the following switch-off condition has been reached:

$$\frac{d}{dt} \left(\frac{\dot{r}}{r \tan \phi} - \dot{\phi} \right) = 0 \quad (5-91)$$

This condition is equivalent to the first equation of system (5-77), using the Lagrange equations (5-75) and putting $\mathcal{H} = 0$ (we consider a maneuver with free arrival time). Condition (5-91) is valid as an equation during the coast arc that follows and, during that time period, it defines the thrust angle ϕ as an integral of equation (5-89):

$$\tan \phi = \frac{\dot{r}}{r \dot{\phi} + C_v r} \quad (5-92)$$

Term C_v represents the integration constant (remember that we avoid the use of $u = \dot{r}$ and $v = r\dot{\phi}$ to minimize the number of different terms in the equations). Of course, there is no necessity to control the thrust angle when the motor is not operative. In any case the spacecraft moves along a conic orbit (often an elliptic orbit); but we can calculate the optimal switch-on instant by using equation (5-92) as a control law. Beginning and final time of the coast arc are connected by equation system (5-77). After eliminating the Lagrange multipliers, the second condition takes the form of:

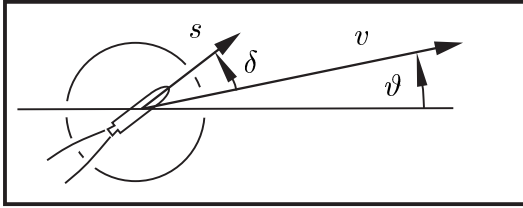
$$\frac{\dot{r}_1}{\sin \phi_1} = \frac{\dot{r}_2}{\sin \phi_2} \quad (5-93)$$

In this notation, index “1” refers to the beginning instant and index “2” to the final instant of the coast arc. A determination of this switch-on condition demands the computation of thrust angle ϕ during the coasting time; ϕ and its first time derivative $\dot{\phi}$ are continuous everywhere. Since we considered $\mathcal{H} = 0$ (equation 5-77), the flight time of a trajectory controlled by these switch conditions is fuel-consumption optimal.

The equations of motion (5-88) determine the spacecraft trajectory as a function of its control. Thrust angle control law (5-89) and switch conditions (5-91) and (5-93) determine the course of the control. Consequently, the terminal orbit reached by the spacecraft depends on nothing else than on the initial values of the control equations. If initial orbit and initial time are given, the whole transfer trajectory of the spacecraft (and, thus, the final destination orbit) depends upon only the initial values ϕ and $\dot{\phi}$ of control equation (5-89). We have just two values to control the destination orbit.

Unfortunately, there is no possibility to analyze the initial conditions analytically as a function of the destination orbit, for there is no analytical solution to the motion during thrust arcs. The final trajectory calculation must be done numerically on a computer, which takes initial trial values and integrates the trajectory. From deviations in the final values of the differential equations the initial values can be adjusted. The integration process must be repeated until the final values are satisfied.

Change of the inclination angle. When the destination orbit is not coplanar with the initial orbit, the spacecraft has to change the inclination angle while performing its orbital transfer. By inclining the thrust acceleration (vector \vec{s}) out of the trajectory plane, the thrust acceleration component in the trajectory plane is diminished by $\cos \delta$ (calling the angle between the thrust vector and the trajectory plane δ). Component $s \sin \delta$ of thrust acceleration vector \vec{s} creates a rotation of the flight plane. We consider a spacecraft that moves horizontally in the apse of a transfer ellipse ($u = 0$, $\phi = 0$ and $\varphi = 0$, equation 5-23). There it performs a high-thrust inclination change maneuver, characterized by the following set of two differential equations:



$$\begin{aligned} \dot{v} &= s \cos \delta \\ \dot{\vartheta} &= \frac{s \sin \delta}{v} \end{aligned} \quad (5-94)$$

Rotation velocity $\dot{\vartheta}$ of the trajectory plane is simply the component $s \sin \delta$ of the thrust acceleration vector divided by the horizontal component of the spacecraft velocity v . The value of v can be evaluated by equating the acceleration \dot{v} with the thrust acceleration $s \cos \delta$ that lies in the trajectory plane (radial components of thrust and velocity can be neglected in the apse of a transfer ellipse). The problem is to find the optimal control function for the thrust angle δ . Hamiltonian (5-2), Lagrange equations (5-3) and maximum principle (5-5) determine the optimal course of δ . Therefore:

$$\begin{aligned} \mathcal{H} &= \lambda_v \cdot s \cos \delta + \lambda_{\vartheta} \frac{s \sin \delta}{v} \\ \dot{\lambda}_v &= - \frac{\partial \mathcal{H}}{\partial v} = \lambda_{\vartheta} \frac{s \sin \delta}{v^2} \\ \dot{\lambda}_{\vartheta} &= - \frac{\partial \mathcal{H}}{\partial \vartheta} = 0 \\ \frac{\partial \mathcal{H}}{\partial \delta} &= -\lambda_v s \sin \delta + \lambda_{\vartheta} \frac{s \cos \delta}{v} = 0 \end{aligned} \quad (5-95)$$

Next step is the elimination of the Lagrange multipliers λ_{ϑ} and λ_v . Differentiation and transformation brings us to the following control equation for the thrust angle δ :

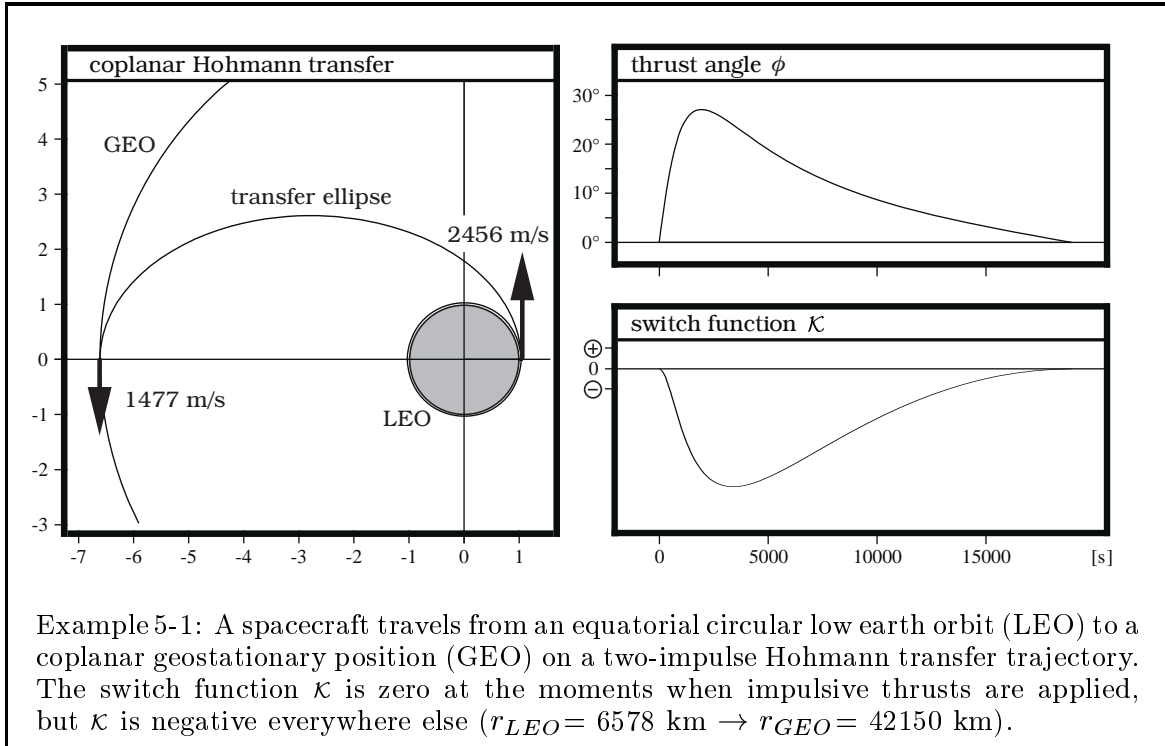
$$\dot{\delta} + \frac{s}{v} \sin \delta = \dot{\delta} + \dot{\vartheta} = 0 \quad (5-96)$$

As the inclination ϑ increases, the thrust angle δ must be diminished. The plausible and well-known result is a thrust direction constant with respect to inertial space.

At the end of the maneuver the desired inclination change $\Delta\vartheta$ has to be realized. We have to adjust the initial value of the differential equation (5-96) appropriately to satisfy this final condition. The analytical solution to this problem is easy: we can readily find the thrust angle δ at the beginning of thrusting when we form the velocity triangle of the maneuver (initial velocity, final velocity and velocity change).

5.2.4 The Course of the Switch Function

The switch function for a Hohmann transfer. The coplanar Hohmann transfer trajectory consists of two impulsive burn maneuvers with horizontal thrusting, separated by a 180° coast arc. The thrust angle ϕ is zero at both apses of the transfer ellipse; and the thrust angle δ is zero everywhere since no inclination change is made. The Hohmann transfer is an optimal trajectory considering its velocity requirement.



In case of a plane trajectory we may use instead of system (5-23) the simplified system (5-73) to define the equations of motion. Then the optimal control functions are given by the equations (5-74) to (5-79). The transfer is a “free positioning maneuver”: we have $\mathcal{H} = 0$ and $\lambda_\varphi = 0$, because final time and final path angle φ are unrestricted (compare equations 5-78 and 5-79). Except for λ_M which behaves discontinuously at the moments of impulsive thrusting, all the other Lagrange multipliers are continuous functions of time (also λ_v in this case since $\delta = 0$, compare equation system 5-42): these multipliers assume the same values immediately after and immediately before the impulse. Since the multipliers λ_u and λ_v are continuous function of time, equation (5-92) is valid to determine the optimal control law for thrust angle ϕ :

$$\tan \phi = \frac{\lambda_u}{\lambda_v} = \frac{u}{v + C_v \cdot r} \tag{5-97}$$

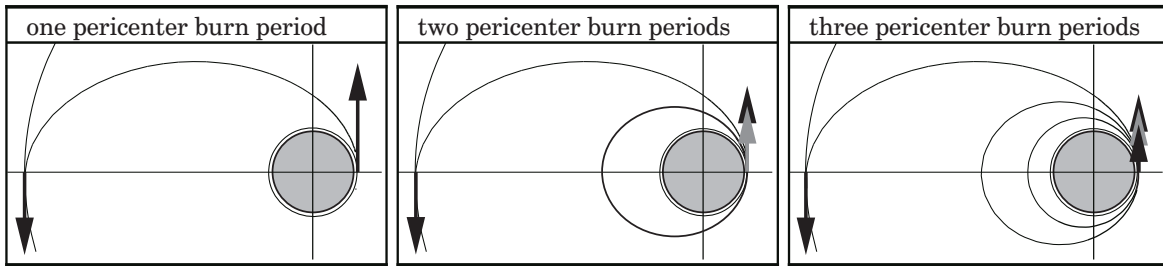
Disregarding the value of constant C_v , control law (5-97) requests horizontal thrusting at pericenter and apocenter position of the Hohmann transfer ellipse (where $u = 0$). We need to know the value of C_v just to calculate the course of the switch function.

We know that the switch function \mathcal{K} vanishes at moments when impulses are applied (condition 5-43). It follows from the switch condition (5-77) that the multiplier λ_v accepts at both impulses the same value (because the spacecraft mass M and the multiplier λ_M are constant on coast arcs). The multiplier λ_u is zero where the vertical velocity u vanishes (on the apses of the coast arcs). Then the constant C_v becomes:

$$C_v = \frac{v_{apocenter} - v_{pericenter}}{r_{pericenter} - r_{apocenter}} \quad (5 - 98)$$

Term v is the horizontal component of the velocity vector, term r is the distance of the spacecraft from the center of gravitation. Knowing constant C_v we can plot the time functions of thrust angle ϕ and switch function \mathcal{K} . As expected, the Hohmann transfer orbit is an optimal trajectory in the sense of the Hamilton-Lagrange theory. We can observe that \mathcal{K} and $\dot{\mathcal{K}}$ vanish at moments when impulsive thrusts are applied.

Trajectories with several burn periods. The real-thrust Hohmann transfer is not the only optimal way to reach the destination orbit: other trajectories exist which are also free positioning maneuvers, subjected to the same conditions. We can split the time interval of the pericenter burn maneuver into two (or several) burning periods. Always a nearly 360° coast arc brings the spacecraft back to pericenter position, where the next burning phase is executed (until finally the maneuver is completed).

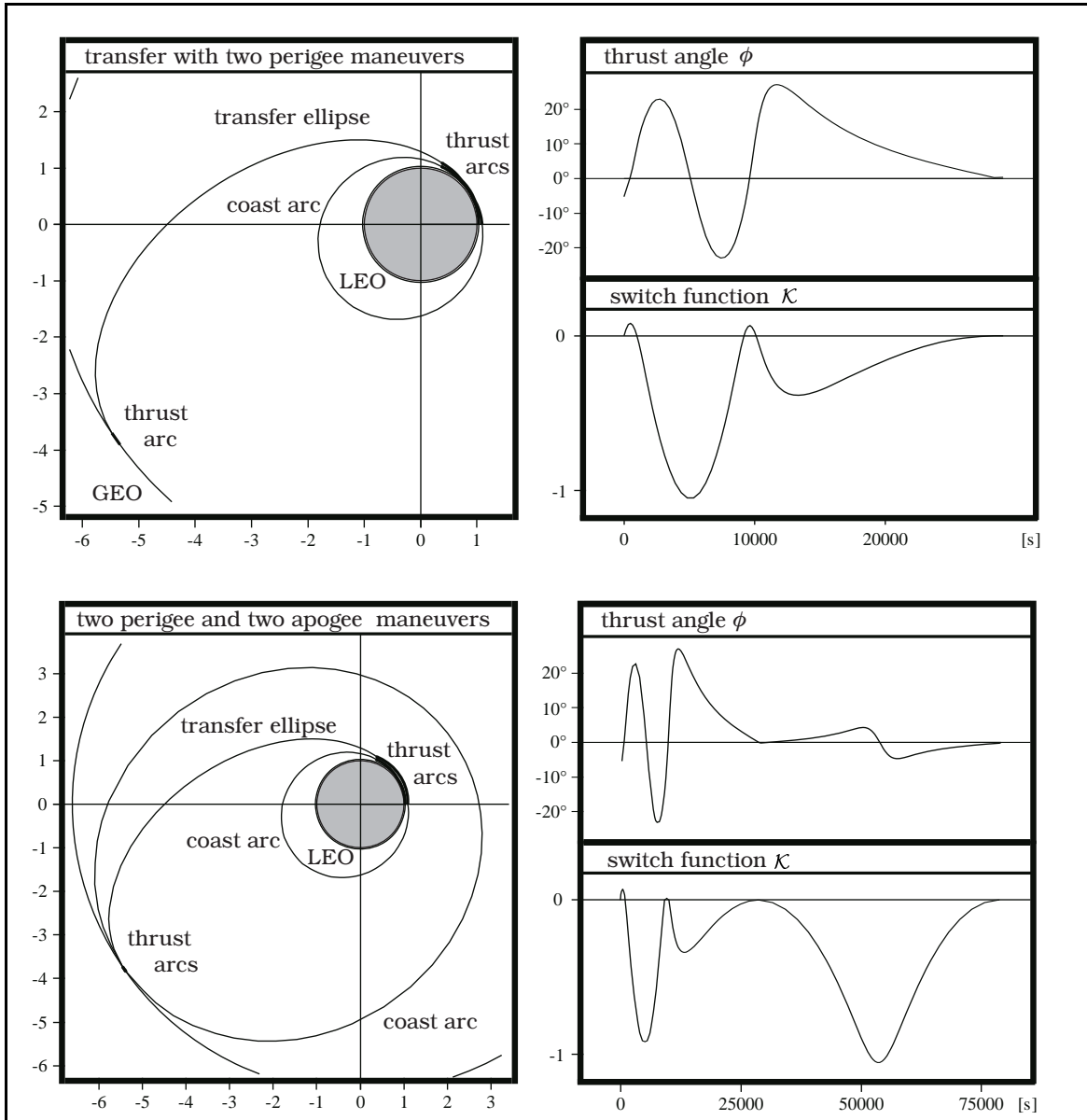


These nearly 360° coast arcs begin with cutoff of the engine, they end with restart of the engine immediately before the next pericenter passage. The equation (5-92) determines the corner conditions, together with the relationship (5-93). We can denote:

$$\begin{aligned} r(\text{end of pericenter passage}) &= +r(\text{beginning of next pericenter passage}) \\ u(\text{end of pericenter passage}) &= -u(\text{beginning of next pericenter passage}) \\ v(\text{end of pericenter passage}) &= +v(\text{beginning of next pericenter passage}) \\ \phi(\text{end of pericenter passage}) &= -\phi(\text{beginning of next pericenter passage}) \end{aligned} \quad (5 - 99)$$

The same splitting procedure can be done with the burn period at apocenter position. Several burn phases prolong the total transfer time; but, due to lower gravity losses in case of low thrust spacecraft, several burn periods require always less propellant.

All trajectories are positioning maneuvers with free time of arrival, they are all “local optima” in the sense of the Hamilton-Lagrange theory. “Global optimum” is just the transfer trajectory with an infinite number of pericenter and apocenter burn periods (of infinitely short duration). It will bring the spacecraft to target orbit after infinitely long time, requiring exactly the same velocity Δv as the impulsive Hohmann transfer.

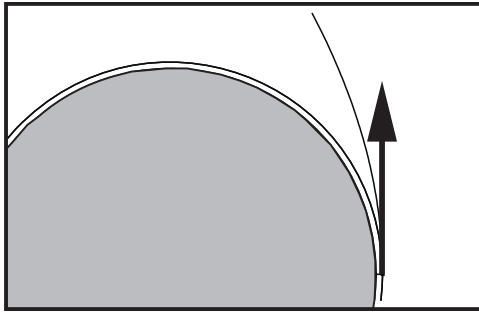


Example 5-2: The satellite transportation from a space station orbit to a destination orbit serves as an example for an optimal positioning maneuver with free time of arrival. It is necessary to raise the velocity at the perigee and the apogee of the transfer ellipse. Several optimal solutions exist because the velocity increase can be distributed among several burning phases. The figures show thrust angle control function and switch function for transfers with two perigee and two apogee thrust periods; considered are multi-burn positioning maneuvers which bring a spacecraft from 200 km circular low earth orbit to coplanar geostationary orbit (initial thrust acceleration $s = 1 \text{ m/s}^2$; exhaust velocity $c = 4330 \text{ m/s}$). The splitting of the perigee burn phase reduces slightly the propulsion requirement but increases considerably the transfer time:

- 1 pericenter maneuver: $\Delta v = 4058 \text{ m/s}$ in 5.76 hours;
- 2 pericenter maneuvers: $\Delta v = 3967 \text{ m/s}$ in 8.06 hours;
- 3 pericenter maneuvers: $\Delta v = 3947 \text{ m/s}$ in 10.9 hours;
- 4 pericenter maneuvers: $\Delta v = 3940 \text{ m/s}$ in 14.0 hours.

All trajectories constitute optimal solutions in sense of the Hamilton-Lagrange theory.

The control of impulsive escape trajectories. The optimization objective for escape maneuvers is a maximization of the total mechanical energy at burnout (5-54). Equation (5-56) shows that the optimization objective demands tangential thrusting at the final time (the thrust angle δ vanishes when no inclination change is made). Since the inertial thrust direction remains constant during an impulsive maneuver, we can conclude that the Lagrange multipliers λ_u , λ_v and λ_r are continuous everywhere. It follows that the control law for the thrust angle ϕ is a quite simple relationship:



$$\tan \phi = \frac{u}{v} \tag{5 - 100}$$

with: $\lambda_u = \mu \cdot u$; $\lambda_v = \mu \cdot v$

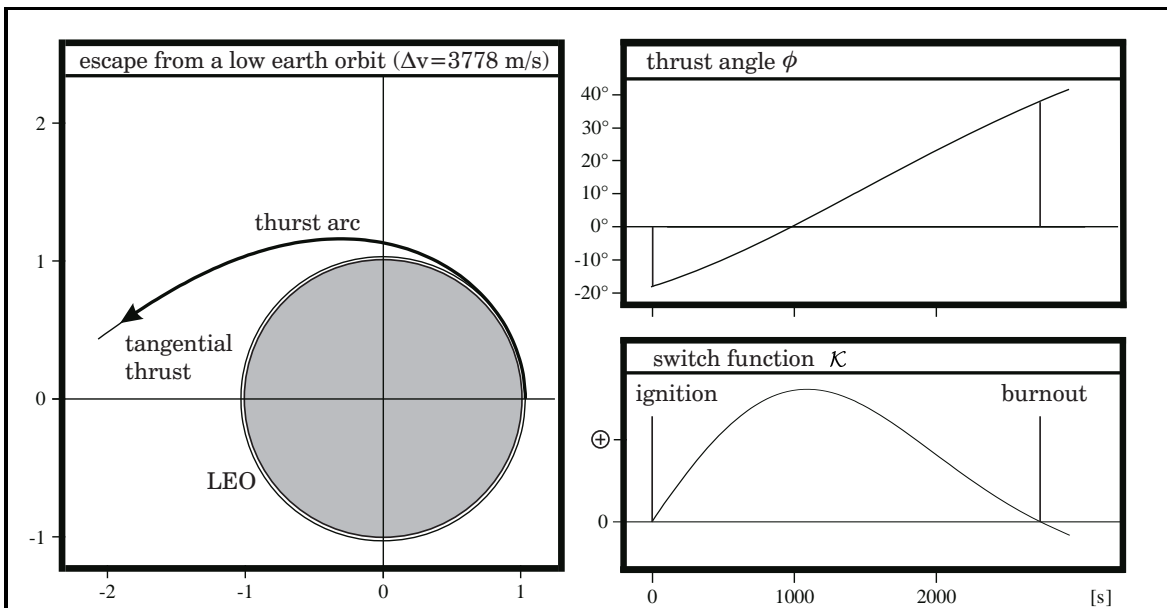
$$\lambda_r = \mu \cdot \gamma / r^2$$

and: $\mathcal{H} = 0$ (free final time)

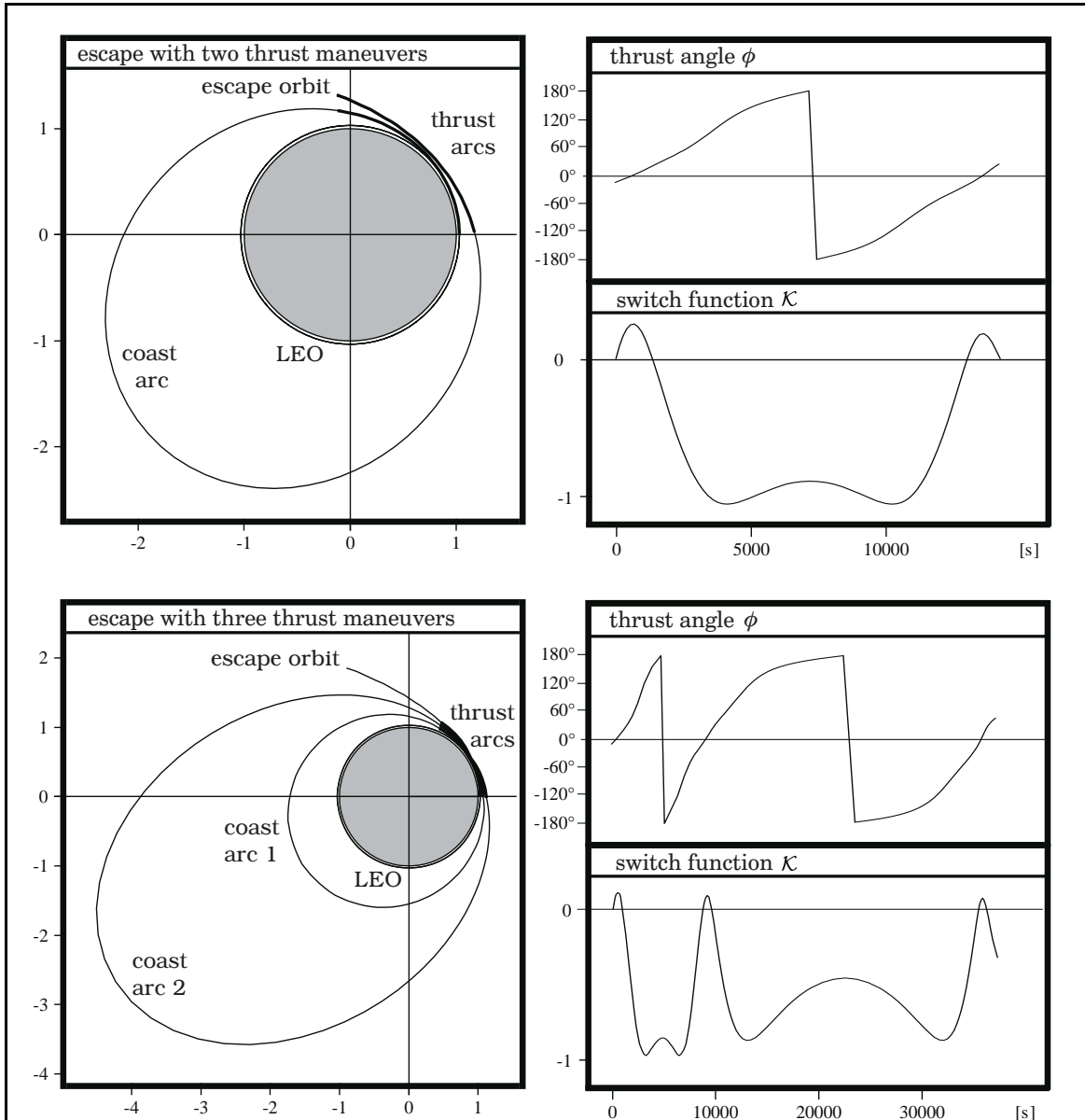
$$\lambda_\phi = 0$$
 (free final path angle)

Condition (5-77) demands that the impulsive escape maneuver requires a free final time and a free final path angle: when $\tan \phi = u/v$ immediately before thrusting, then the switch function \mathcal{K} is a linear function of the velocity $v = \sqrt{u^2 + v^2}$ on the conic orbit; and it follows that the spacecraft should ignite the rocket engine exactly when it passes the pericenter position of the initial orbit.

Low thrust escape trajectories. The calculation of low thrust escape trajectories has to be done numerically. Now we have tangential thrust only at the final instant.



Example 5-3a: A “low-thrust” spacecraft escapes from a circular low earth orbit (LEO), with tangential thrust only at the final time. During the maneuver the switch function \mathcal{K} is positive. ($r_{LEO} = 6578 \text{ km} \rightarrow u_\infty = 3536 \text{ m/s}$, $s = 1 \text{ m/s}^2$, $c = 4330 \text{ m/s}$)



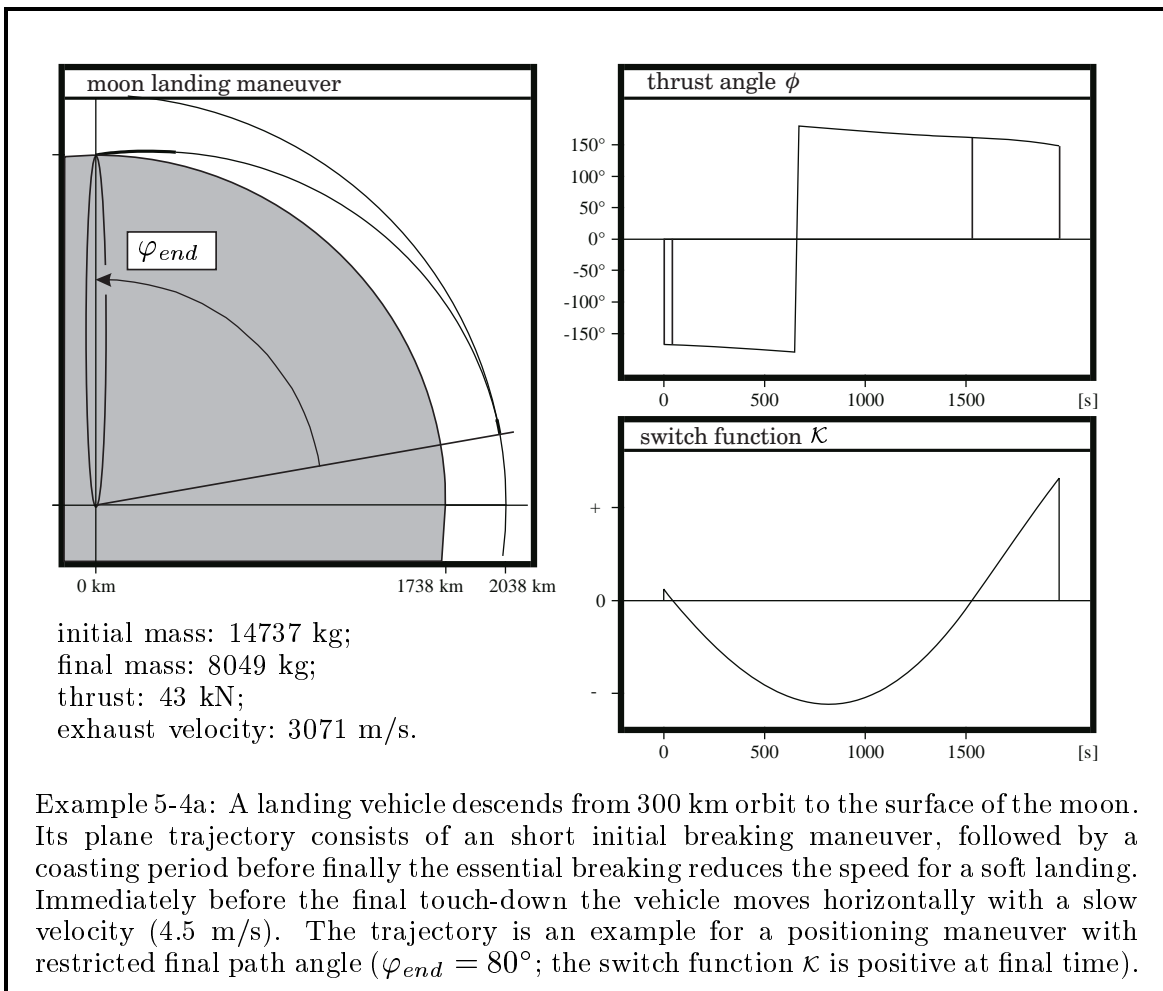
Example 5-3b: Low thrust earth escape maneuvers can be performed via several burn periods at pericenter passage, with the advantage of reduced gravity losses and the disadvantage of extended flight time. The following table compares four different ways to accelerate from a low earth orbit to hyperbolic velocity ($s = 1 \text{ m/s}^2$, $c = 4330 \text{ m/s}$, $r_{LEO} = 6578 \text{ km} \rightarrow u_{\infty} = 3536 \text{ m/s}$, impulsive thrust requires $\Delta v = 3778 \text{ m/s}$):

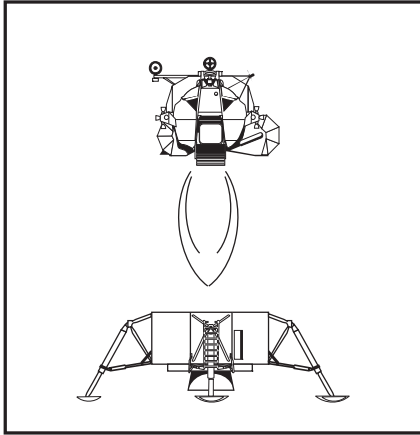
- 1 pericenter maneuver: $\Delta v = 4326 \text{ m/s}$ in 0.76 hours;
- 2 pericenter maneuvers: $\Delta v = 3949 \text{ m/s}$ in 3.96 hours;
- 3 pericenter maneuvers: $\Delta v = 3957 \text{ m/s}$ in 10.1 hours;
- 4 pericenter maneuvers: $\Delta v = 3822 \text{ m/s}$ in 20.1 hours;

The behaviour of the switch function is interesting: during coast arcs κ remains negative indicating that the engine must be switched off. At apogee passage the thrust direction is inverted ($\phi = \pm 180^\circ$), and κ accepts a local maximum near zero; but when the apogee altitude is very high, κ becomes positive indicating that then an apogee retro-thrust is optimal in order to reduce the perigee altitude for a more efficient final burn period (the atmosphere of earth makes such a mission unworkable).

Moon landing maneuvers. All the preceding examples treated trajectories with unrestricted final path angle. Then the condition (5-48) is valid and the target orbit can be considered as a part of the optimal transfer trajectory. As a consequence the switch function \mathcal{K} must vanish at final time (and in case of impulsive trajectories also the time derivative of the switch function $d\mathcal{K}/dt$). We will use trajectories of a moon landing vehicle to demonstrate examples where \mathcal{K} is positive at the end of a maneuver.

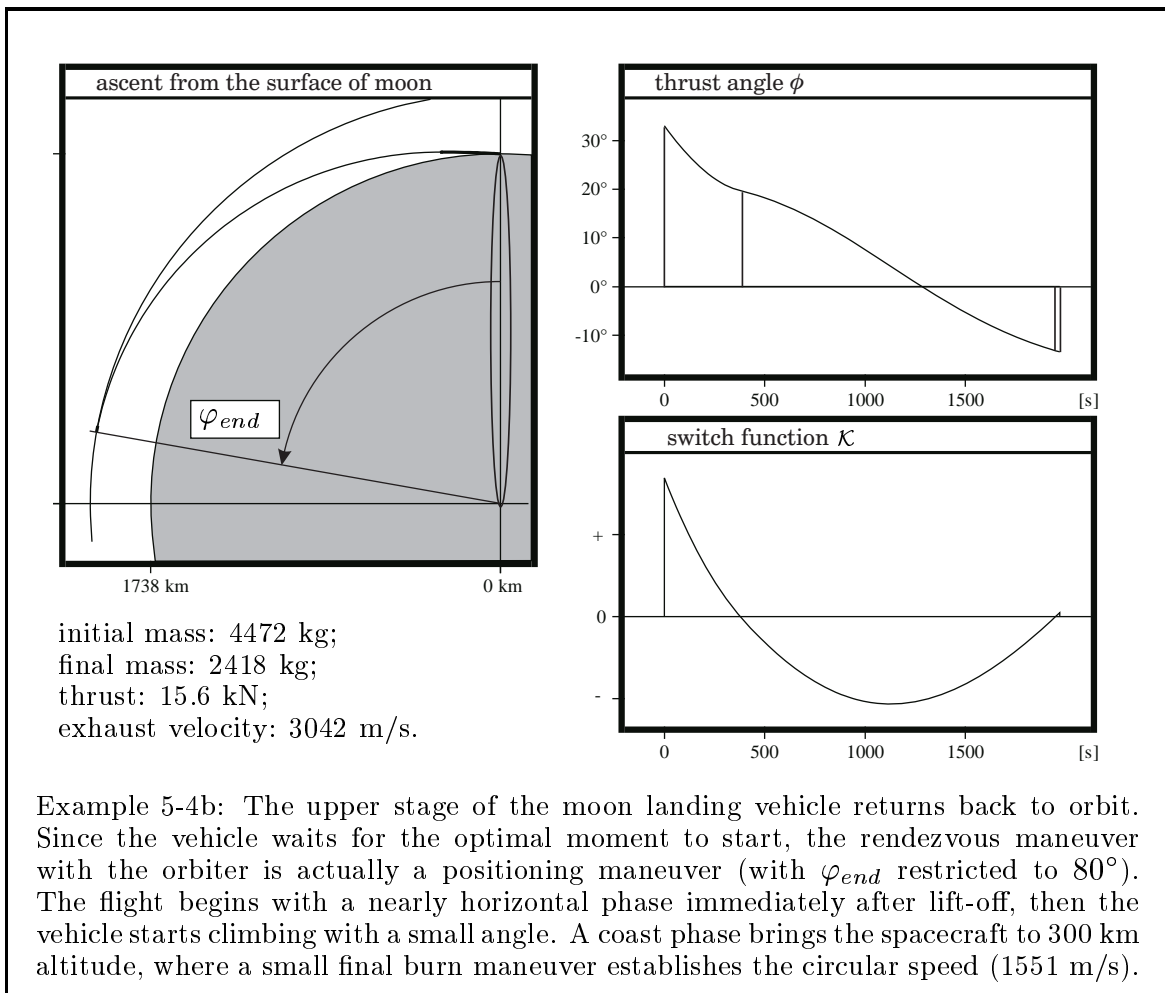
The monthly revolution of the moon (gravitational constant $\gamma = 4.89804 \cdot 10^{12} \text{ m}^3/\text{s}^2$, radius = 1738 km) causes a slow motion of its surface, about 4.5 m/s at the equator. We assume that a highly eccentric elliptic orbit exists inside the moon; this “orbit” has the eccentricity $\varepsilon = 0.999999984$ and touches the lunar surface with its apocenter; and the apocenter of this hypothetical orbit is supposed to be the “terminal position” for a lunar landing craft. Now the final path angle φ_{end} is a predetermined trajectory parameter: we may assume the final orbit anywhere inside the moon, but the vehicle has to land on the apocenter of this final orbit and not anywhere else on the same orbit. The final path angle φ_{end} is restricted for a descent trajectory of a lunar lander, and consequently the multiplier λ_φ is not zero anymore. We can compare trajectories for different values of the parameter φ_{end} (performing parametric trajectory analysis).





Calculation of lunar ascent trajectories is similar. When the landing vehicle does not proceed directly to the earth, it will perform a rendezvous maneuver with a moon orbiter. Since the vehicle can wait for the optimal moment to lift-off (it waits for the best “launch window”), the transfer time is not restricted. We can conclude that the Hamiltonian vanishes on the whole ascent trajectory ($\mathcal{H} = 0$); and actually we do not consider a rendezvous maneuver but a positioning maneuver. The flight time Δt and the velocity requirement Δv are functions of the predetermined final path angle φ_{end} .

The propellant consumption accepts its minimum when φ_{end} is approximately 180° ; and then multiplier λ_φ is zero and the switch function \mathcal{K} vanishes at the final instant. The climb rate gets better and the flight time gets shorter when we reduce φ_{end} (immediately after lift-off the lunar landing vehicle moves nearly horizontally). When we reduce the angle φ_{end} the switch function \mathcal{K} becomes positive at the final instant (this indicates that the restricted trajectory consumes more propellant).



Rendezvous and flyby. We call the trajectory of a spacecraft which encounters a target object on a target orbit a “rendezvous maneuver”: the spacecraft enters the target orbit and meets the object there without relative velocity. We talk about a “flyby maneuver” when the encounter takes place with a relative velocity at final time. The flight time is restricted for both maneuvers; and we examine the case $\mathcal{H} \neq 0$: the flight time is predetermined by the initial and final position of the target object.

First we have to solve the “Lambert problem” numerically in order to calculate such a trajectory: we have to find the conic orbit which connects the initial with the final location and additionally satisfies the predetermined transfer time (a description how to solve the Lambert problem is given in the chapter eight of this book). The next step is that we establish the corresponding two-impulse reference trajectory: the geometry of the trajectory is entirely determined without any optimization, and we can readily calculate thrust direction and velocity requirement for the impulsive maneuver. The Lagrange multipliers on a coast arc are defined by the equation system (5-44). We know that the value of the switch function \mathcal{K} is zero at both corners of the coast arc, and consequently the expression $(P/\cos \delta)$ accepts at the end of the coast arc the same value as at the beginning (we have $P^2 = \lambda_u^2 + \lambda_v^2$, with $\lambda_u = P \sin \phi$ and $\lambda_v = P \cos \phi$, compare the equations 5-32 and 5-39). Therefore we can solve a linear equation system to calculate the four integration constants \mathcal{H} , C_u , C_v and C_φ . Thus:

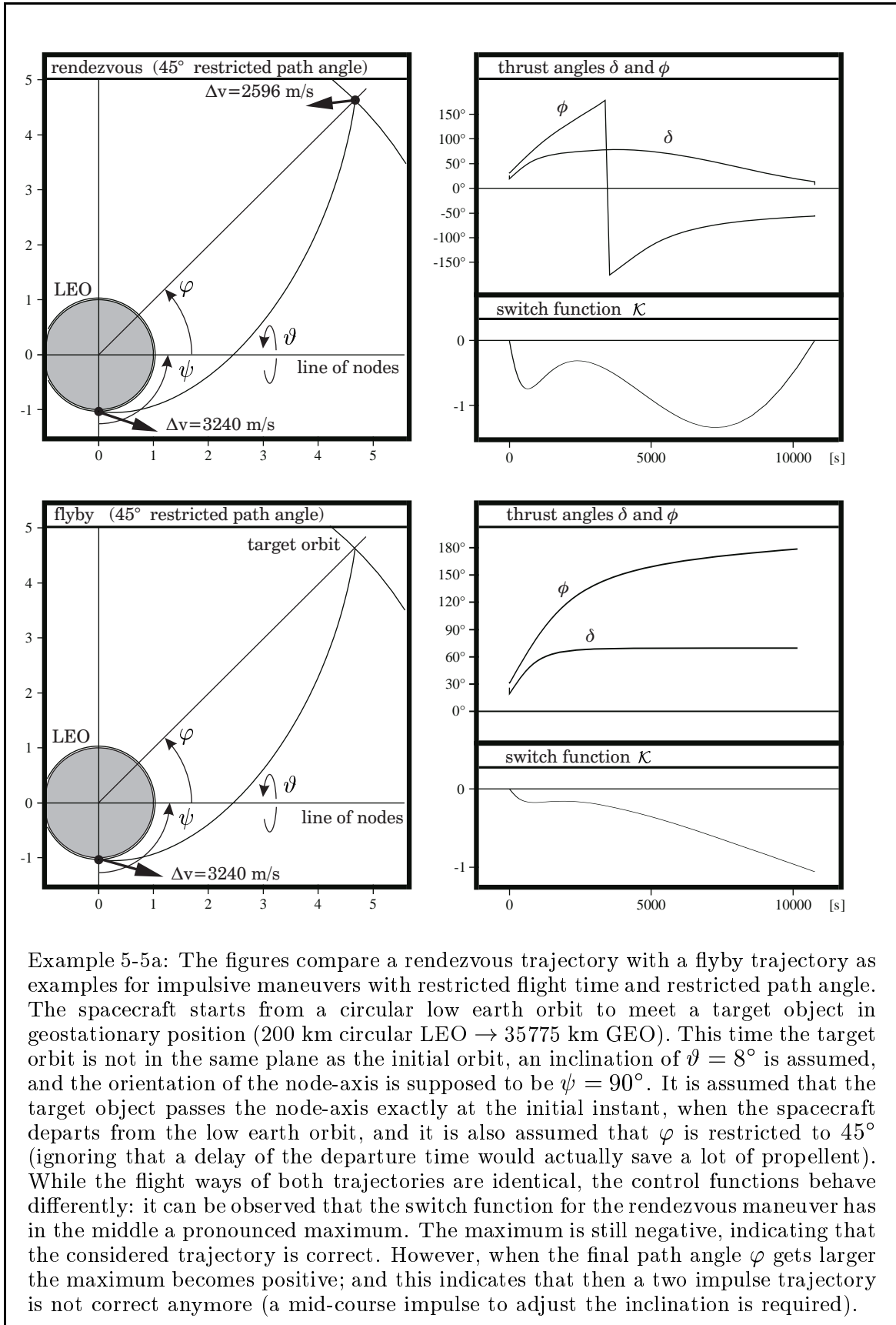
$$\begin{aligned}\sin \phi_1 \cos \delta_1 &= \mathcal{H} \frac{r_1}{e} + C_u u_1 + C_\varphi (\gamma - h v_1) \\ \cos \phi_1 \cos \delta_1 &= C_u v_1 + C_v r_1 + C_\varphi u_1 (h + \frac{\gamma}{v_1}) \\ \sin \phi_2 \cos \delta_2 &= \mathcal{H} \frac{2r_2 - 3tu_2}{2e} + C_u u_2 + C_\varphi (\gamma - h v_2) \\ \cos \phi_2 \cos \delta_2 &= \mathcal{H} \frac{-3tv_2}{2e} + C_u v_2 + C_v r_2 + C_\varphi u_2 (h + \frac{\gamma}{v_2})\end{aligned}\quad (5 - 101)$$

Since we have already solved the Lambert problem we know the thrust direction during the impulsive maneuvers. Knowing δ (equation 5-37) at both corners of the coast arc we find the constants C_A and C_B by solving another equation system ($C_C = \lambda_\varphi$):

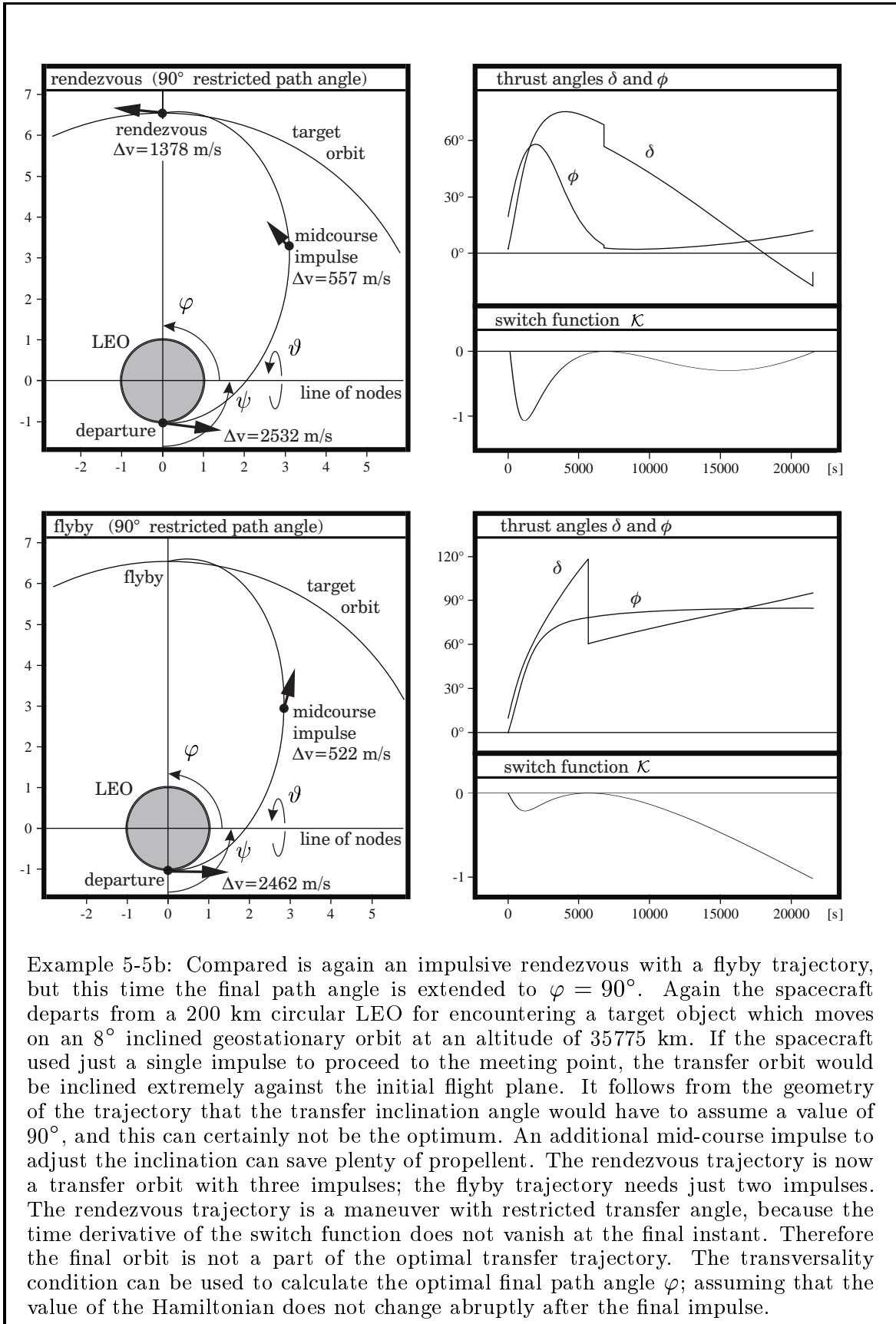
$$\begin{aligned}\sin \delta_1 &= \frac{1}{v_1} C_A \\ \sin \delta_2 &= \frac{1}{v_2} (C_A \cos \varphi_2 + C_B \sin \varphi_2 \cos \vartheta_2 + C_C \sin \varphi_2 \sin \vartheta_2)\end{aligned}\quad (5 - 102)$$

The “index 1” refers to “immediately after the initial impulse” and the “index 2” to “immediately before the final impulse”. As the last step we find the initial values of the Lagrange multipliers before the first impulse by an application of equation (5-42).

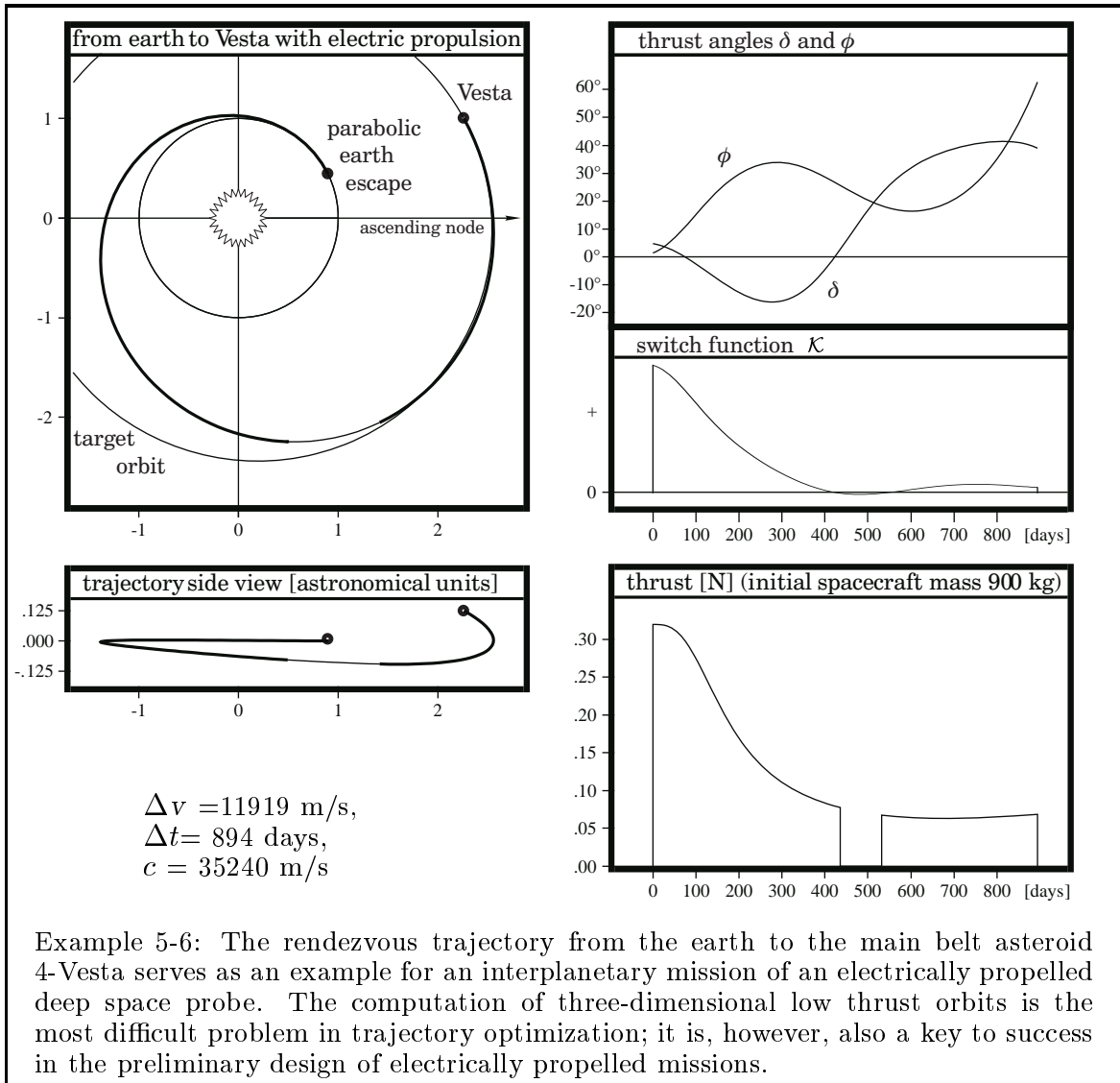
The geometry of the trajectory is exactly the same for a flyby maneuver; but the control functions are different then. When the velocity is free at the final instant, then conditions (5-50) and (5-53) are valid and we have an undefined thrust direction at the moment of the encounter ($\tan \phi = 0/0$, $\tan \delta = 0/0$). To calculate the control functions of a flyby maneuver we may also use the equations (5-101) and (5-102), but then the three expressions $(\sin \phi_2 \cos \delta_2)$, $(\cos \phi_2 \cos \delta_2)$ and $(\sin \delta_2)$ equal zero.



Example 5-5a: The figures compare a rendezvous trajectory with a flyby trajectory as examples for impulsive maneuvers with restricted flight time and restricted path angle. The spacecraft starts from a circular low earth orbit to meet a target object in geostationary position (200 km circular LEO \rightarrow 35775 km GEO). This time the target orbit is not in the same plane as the initial orbit, an inclination of $\vartheta = 8^\circ$ is assumed, and the orientation of the node-axis is supposed to be $\psi = 90^\circ$. It is assumed that the target object passes the node-axis exactly at the initial instant, when the spacecraft departs from the low earth orbit, and it is also assumed that ϕ is restricted to 45° (ignoring that a delay of the departure time would actually save a lot of propellant). While the flight ways of both trajectories are identical, the control functions behave differently: it can be observed that the switch function for the rendezvous maneuver has in the middle a pronounced maximum. The maximum is still negative, indicating that the considered trajectory is correct. However, when the final path angle ϕ gets larger the maximum becomes positive; and this indicates that then a two impulse trajectory is not correct anymore (a mid-course impulse to adjust the inclination is required).



Trajectories of electrically propelled spacecraft. Ion engines promise the feasibility of deep space missions which are impossible for conventional propulsion or at least extremely difficult; considered are particularly a “comet nucleus sample return” mission and a “multi asteroid rendezvous and flyby” mission. In the preliminary design of such a mission, trajectory computation techniques play a key role. The source of energy for propulsion is the sun: the “solar electric propulsion spacecraft” is equipped with large solar arrays which provide the electric power for the engines. However, the thrust declines as the available energy diminishes with distance from the sun r , and the mass flow rate m is proportional to the factor $r^{-1.7}$ because the solar cells generate more power when it gets colder (but they fail completely when it is colder than a certain temperature). Even though the thrust is extremely low, the exhaust velocity of the engines is by a factor of typically 10 better than the exhaust velocity of chemical engines (for example $c=30$ km/s to 60 km/s); and the long operation time interval (years) gives the probe a high Δv capacity. The chapter nine of this book gives a description of electric propulsion system and some mission examples.



The optimization of three-dimensional trajectories is complicated, since it usually involves an unknown number of thrust periods. Often three thrust periods are required for three-dimensional rendezvous trajectories: the first maneuver is necessary to leave the initial orbit, the second maneuver is used to change the inclination of the transfer orbit, and the third maneuver is used to enter the target orbit. However, the inclination change is also performed during the first and during the last maneuver; and it is also possible that a mid-course maneuver is not necessary. Sometimes even more than three thrust periods are required, particularly when the transfer time is long.

We have to look at the switch function when we want to find out whether a trajectory is correct or not: the trajectory is certainly not the optimum when the switch function is positive anywhere during a coasting period or negative during a thrusting period. When the switch function of a two impulse trajectory is positive we apply a small mid-course maneuver at the moment when the switch function passes its maximum. The additional maneuver deviates the final values of the trajectory, and we correct the final values by an adjustment of the initial values of the Lagrange multipliers. By this we increase the mid-course impulse iteratively until finally the switch function is correct. When the switch function is negative everywhere, then we have generated a three impulse trajectory that is optimal in the sense of the Hamilton-Lagrange theory.

Docking. In practice a rendezvous maneuver can be dangerous. When a manned shuttle is trying to approach a space station on low earth orbit, it always involves the risk that the shuttle collides with the station. To reduce the risk of collision, the shuttle must first approach a position in the vicinity of the station. From this position the shuttle starts with precise navigation for the final rendezvous and docking maneuver. A convenient point to start the final approach is located exactly on the target orbit, on a position where the path angle φ is some degrees smaller than the path angle of the space station. When the shuttle has reached this point, it executes a small retro-thrust maneuver to reduce the eccentricity of its orbit; and then the relative position of the shuttle performs a cycloid motion with respect to the space station (the relative radius r and the relative path angle φ oscillate with the angular velocity of the orbit, where a slow forward drift of the path angle is super-imposed). Precise adjustment of the retro-thrust maneuver is necessary: after one (or several) orbital revolutions the shuttle arrives at the space station with a slow relative velocity.

Satellite position stabilization. Gravitation of sun and moon cause geostationary satellites to drift away from their nominal position. To keep a broadcasting satellite in position (a cube in space with edges of about 100 km length), rendezvous maneuvers with the nominal position have to be performed every now and then (actually in time intervals of about two weeks). For practical reasons it is not desired that the satellite changes its attitude when a station keeping maneuver is executed. Usually, control thrusters have fixed-canted nozzles. The thrust direction is not changeable, and the satellite needs at least six control thrusters to be able to maneuver in three different directions in space. Inclination, altitude and eccentricity of the orbit are adjusted by individual burn maneuvers. In this case the rendezvous maneuver with the nominal position is not an optimal trajectory in the sense of the Hamilton-Lagrange theory.

6. Attitude Stabilization

A spacecraft may be regarded as a mass particle without size when its translational motion is under consideration; but every spacecraft is an object with a real physical size and thus it cannot only translate but also rotate. In spaceflight dynamics, the rotational motion can be treated separately from the translational motion, since usually the size of a spacecraft is very small compared to the size of its orbit. Orbiting in space the spacecraft rotates about its center of mass, while its center of mass is displaced simultaneously. The rotational motion of spacecraft is the content of this section.

Nearly all satellites have to attain a special predetermined attitude when they are operative in space. Solar arrays, parabolic antennae, sensors and other equipment must point into defined directions. Without control or stabilization, the attitude of a space vehicle would soon drift away from the desired orientation. The subject of spacecraft attitude dynamics is to calculate the rotational behaviour of satellites and to find appropriate means to establish and keep orientation in space.

Gyro dynamics is the appropriate tool for the analysis of the rotational behaviour of space vehicles. Like particle dynamics, gyro dynamics is an old (well-established) subject of technical mechanics. A gyro can execute two different types of motion: a nutation as the result of special initial conditions, and a precession as the result of an external torque. The behaviour of a gyro is fascinating and seemingly paradox, but naturally the movement of a gyro obeys the rules of technical mechanics.

The book of K.Magnus [“Kreisel”, Springer Verlag, Berlin-Heidelberg-New York, 1971] gives a detailed introduction to gyro dynamics. Special emphasis on the attitude dynamics of spacecraft is made in the books of W.T.Thomson [“Introduction to Space Dynamics”, Dover Publications Inc., New York, 1986], M.H.Kaplan [“Modern Spacecraft Dynamics and Control”, John Wiley and Sons, New York, 1976] and M.J.Sidi [“Spacecraft Dynamics and Control”, Cambridge University Press, 1997].

6.1. The Angular Motion of Satellites

The rotational behaviour of spinning satellites is described by “equations of motion”. These equations for the rotational behaviour of satellites are second order differential equations, like the equations of motion for the translational behaviour; but now the first time derivatives refer to the angular velocity, and the second time derivatives refer to the angular acceleration. In astronautics, trajectory dynamics and attitude dynamics are nearly perfectly decoupled: the rotation has no influence on the orbit of the spacecraft, and usually the orbit has no influence on the attitude of the spacecraft (a large gravity stabilized space station is an example where this is not the case). The equations of motion for the rotational behaviour of a satellite consider the influence of external moments (torques) which act on the satellite. We can establish these equations by an application of the “angular momentum law”.

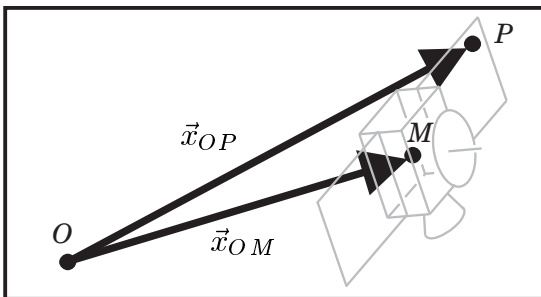
6.1.1 Newton's Law for an Object with Physical Size

Definitions. Before we start with calculations we have to define some expressions (otherwise there is a high possibility of confusion, particularly when the analysis is very difficult, it is necessary to determine uniquely the applicability of formulae). Any material object with infinitely small size we will call a “mass particle”. This includes mass particles of infinitely small mass dM . If an object has a physical size it is termed “extended object”. Thus, extended objects are also gases, liquids or elastic bodies. Extended objects consist of a (usually infinite) number of mass particles. Sometimes the location of all mass particles is invariant in relation to the other body elements: then the object is termed a “rigid body”. Often, satellites are not perfectly rigid bodies, as they may have flexible antennae or tanks which store liquid propellant.

We will call every coordinate system “inertial coordinate system” if no forces or moments are introduced which come originally from a motion of the system. Thus, a non-rotating system that moves with constant speed is an inertial system. When we change the definition for the location vector, the origin of the coordinate system does not matter any more, and every non-rotating system becomes an inertial system. The origin (or center) of the coordinate system implies that location vectors start at this point; but when by definition location vectors locate positions just in relation to other well-defined positions, the origin of the coordinate system becomes obsolete. It is always advisable to indicate clearly where a location vector starts. In this case a coordinate system is just useful to resolve vector equations into component notations. For vector resolution it is not necessary to specify the origin of the coordinate system.

When a rigid body is under consideration, a coordinate system that is fixed with respect to the body is called a “body frame of reference”. Its attitude is always coincident with the orientation of the rigid body. Usually, the center of the body frame of reference is coincident with the center of mass of the rigid body.

The center of mass of an extended object. Consider an extended mass M which consists of an infinite number of mass particles dM (we have $M = \int_M dM$). The point of application of the gravity attraction vector is called “center of mass”. It is the same point for any orientation of the mass in a parallel gravitational field:



$$\vec{x}_{OM} = \frac{1}{M} \cdot \int_M \vec{x}_{OP} dM \quad (6-1)$$

$$\vec{v}_M = \frac{1}{M} \cdot \int_M \vec{v}_P dM \quad (6-2)$$

$$\left(\text{since : } \dot{\vec{x}}_{OM} = \frac{1}{M} \cdot \int_M \dot{\vec{x}}_{OP} dM \right)$$

Location vector \vec{x}_{OM} locates the center of mass M in relation to reference point O . Velocity vector \vec{v}_M defines the relative velocity of the center of mass.

Linear momentum and angular momentum of a mass particle. We make use of an inertially fixed reference point O and consider the motion of a mass particle dM . The mass particle dM has the velocity \vec{v}_P in relation to inertial space. According to their equations of condition, linear momentum $d\vec{V}$ and angular momentum $d\vec{H}_O$ are:

$$d\vec{V} = \vec{v}_P \cdot dM \quad (6-3)$$

$$d\vec{H}_O = \vec{x}_{OP} \times d\vec{V} \quad (6-4)$$

Both, linear momentum and angular momentum, are vectors. The linear momentum vector has always the same direction as the velocity vector, because it is simply the product of mass and velocity. The angular momentum vector is perpendicular to the location vector and perpendicular to the linear momentum vector, because it is the vector cross product. While the linear momentum is independent of the reference point, the angular momentum is usually different if another reference point is used.

Linear momentum and angular momentum of an extended object. Analogously we find linear and angular momentum of an extended object by integration:

$$\vec{V} = \int_M \vec{v}_P \cdot dM = M \cdot \vec{v}_M \quad (6-5)$$

$$\vec{H}_O = \int_M (\vec{x}_{OP} \times \vec{v}_P) dM \quad (6-6)$$

Again, vector \vec{H}_O depends on point O ; and vector \vec{V} does not. The linear momentum \vec{V} of an extended object is simply its mass multiplied with the absolute velocity of the center of mass. The calculation of the angular momentum \vec{H}_O is more complicated.

Newton's law. In classical mechanics, many relationships follow directly from a geometrical interpretation of Newton's three laws (as announced in his "Principia"). Newton recognized that the absolute time derivative of the linear momentum \vec{V} of an object is caused by the external force \vec{F} that acts on the object. Analogously, the absolute time derivative of the angular momentum \vec{H} of the object is caused by the external moment (or torque) \vec{M} , provided that an inertially fixed reference point is taken as a basis. These relationships are used to establish the equations of motion. Usually, the first form of "Newton's law" is used to analyze the translational motion, while the second form is used for the rotational motion:

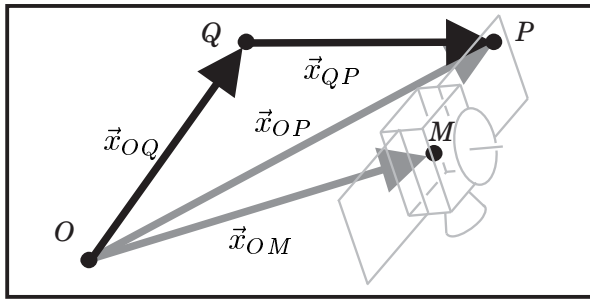
$$d\vec{V} / dt = \vec{F} \quad (6-7)$$

$$d\vec{H}_O / dt = \vec{M}_O \quad (6-8)$$

Often, equation (6-7) is interpreted as "force is mass multiplied with acceleration"; but in this form the statement is not precise and thus not always correct. We should always use a more precise formulation: here the word "force" means the result of all external forces that act on the extended object when it is completely liberated from its surrounding. The "mass" of the object must be constant, the mass may not be a function of time or location; and finally the word "acceleration" means the acceleration of the location of the center of mass with respect to inertial space.

6.1.2 The Angular Momentum Law for another Reference Point

Transformation of the moment vector. Equation (6-8) states that the change in time of the angular momentum is the same physical quantity as the external torque vector which acts on the extended object. In satellite attitude dynamics, the angular momentum law in the form of equation (6-8) is not appropriate for many problems. Even though the equation (6-8) is valid, it cannot be applied directly in this form. Usually we need the representation of the law in a moving coordinate system. The transformation is a comparatively complicated procedure: as a first step we



have to transfer the external moment vector from the inertial fixed point O to another reference point Q. Usually point Q is not fixed with respect to inertial space, and Q is not necessarily coincident with the center of mass M:

$$\vec{M}_O = \vec{M}_Q + \vec{x}_{OQ} \times \vec{F} \quad (6 - 9)$$

Forces with a line of action distant from point O generate in O the torque vector \vec{M}_O . When we change the reference point from O to Q, we shift the torque vector directly. Additionally we have to consider that also the external force vector \vec{F} acts on point O, an additional moment $\vec{x}_{OQ} \times \vec{F}$ (or $-\vec{x}_{QO} \times \vec{F}$) must be included in the relationship.

Transformation of the angular momentum vector. The next step is the transformation of equation (6-6), which is the conditional equation of the absolute angular momentum \vec{H}_O . The objective is to find a representation which involves the relative angular momentum \vec{H}_Q . We can constitute \vec{H}_Q when we know the relative distance \vec{x}_{QP} and relative velocity $\vec{v}_{P,relative}$ of a particle dM. For any point Q we have:

$$\vec{x}_{OP} = \vec{x}_{OQ} + \vec{x}_{QP} \quad (6 - 10)$$

$$\vec{v}_{P,inertial} = \vec{v}_Q + \vec{v}_{P,relative} \quad (6 - 11)$$

Now we insert these relationships into equation (6-6), to find:

$$\begin{aligned} \vec{H}_O &= \int_M [(\vec{x}_{OQ} + \vec{x}_{QP}) \times (\vec{v}_Q + \vec{v}_{P,relative})] dM \\ &= \int_M (\vec{x}_{OQ} \times \vec{v}_{P,inertial}) dM + \int_M (\vec{x}_{QP} \times \vec{v}_Q) dM + \int_M (\vec{x}_{QP} \times \vec{v}_{P,relative}) dM \\ &= \vec{x}_{OQ} \times \vec{v}_{M,inertial} \cdot M + \vec{x}_{QM} \cdot M \times \vec{v}_Q + \vec{H}_Q \quad (6 - 12) \end{aligned}$$

This is the general relationship between the inertial angular momentum \vec{H}_O and the relative angular momentum \vec{H}_Q . Reference point Q can be any moving point. Equation (6-12) accepts a simpler form under some important special conditions: point Q can also be fixed with respect to inertial space (then we have $\vec{v}_Q = 0$); Q can coincide with the center of mass (then we have $\vec{x}_{QM} = 0$); and Q can move parallel to the center of mass (therefore we have $\vec{v}_Q = \vec{v}_M$ and $\vec{x}_{OM} = \vec{x}_{OQ} + \vec{x}_{QM}$).

Of practical interest are three special cases where simpler relationships between the inertial angular momentum \vec{H}_O and the relative angular momentum \vec{H}_Q are valid:

$$Q \text{ is inertially fixed:} \quad \vec{H}_O = \vec{x}_{OQ} \times \vec{V} + \vec{H}_Q \quad (6-13)$$

$$Q \text{ is center of mass:} \quad \vec{H}_O = \vec{x}_{OM} \times \vec{V} + \vec{H}_M \quad (6-14)$$

$$\vec{v}_Q = \vec{v}_M \text{ and } \vec{x}_{OM} = \vec{x}_{OQ} + \vec{x}_{QM} : \quad \vec{H}_O = \vec{x}_{OM} \times \vec{V} + \vec{H}_Q \quad (6-15)$$

The first transformation is important in gyro dynamics when instead of O another fixed reference point Q is used as basis for the calculation of the angular momentum vector. The second transformation is often required to transform the angular momentum vector from the center of mass M to a fixed point O or vice versa. Naturally, the center of mass can also be a point that is fixed with respect to inertial space. The third case occurs when we use a reference point Q that moves in the same way as the center of mass M . When we compare equation (6-14) with (6-15), we can observe that for this case the angular momentum \vec{H}_Q is equal to the angular momentum \vec{H}_M .

Transformation of the angular momentum law. Next step is the transformation of the angular momentum law. We insert (6-9) and (6-12) into law (6-8) to find:

$$\begin{aligned} & \frac{d(\vec{x}_{OQ} \times M\vec{v}_{M,inertial})}{dt} + \frac{d(\vec{x}_{QM} \times M\vec{v}_Q)}{dt} + \frac{d\vec{H}_Q}{dt} = \vec{M}_Q + \vec{x}_{OQ} \times \vec{F} \quad (6-16) \\ & = (\vec{v}_Q \times M\vec{v}_{M,inertial}) + (\vec{x}_{OQ} \times \frac{d\vec{V}}{dt}) + (\vec{v}_{M,relative} \times M\vec{v}_Q) + (\vec{x}_{QM} \times M\frac{d\vec{v}_Q}{dt}) + \frac{d\vec{H}_Q}{dt} \end{aligned}$$

Since $d\vec{V}/dt = \vec{F}$, the second terms of both sides of equation (6-16) cancel out. First and third term of the right hand side will also cancel out, because we have $\vec{v}_{M,inertial} = \vec{v}_Q + \vec{v}_{M,relative}$ and $\vec{v}_Q \times \vec{v}_Q = 0$. For reference point Q law (6-8) yields:

$$(\vec{x}_{QM} \times M\frac{d\vec{v}_Q}{dt}) + \frac{d\vec{H}_Q}{dt} = \vec{M}_Q \quad (6-17)$$

Very often (but not always) reference point Q is not accelerated (then $d\vec{v}_Q/dt = 0$). Also very often Q is the center of mass M of the extended mass (then $\vec{x}_{QM} = 0$). Thus, in both cases the angular momentum law accepts the very simple form of:

$$\frac{d\vec{H}_Q}{dt} = \vec{M}_Q \quad (6-18)$$

Remember that the angular momentum vector \vec{H}_Q is by definition the integral of the vector cross product of location vector \vec{x}_{QP} and infinitesimal small linear momentum vector $\vec{v}_{P,relative} \cdot dM$ (compare equation 6-12). “Reference point Q ” is always the location where the location vector \vec{x}_{QP} starts. We should emphasize that Q is not necessarily fixed with respect to the object, and that the object does not necessarily rotate around point Q (sometimes reference point Q is located outside the object). The “rotation axis” of a rigid body is a straight line, it connects all the elements which are not moving at a certain moment (but often the rotation axis itself is moving). We can just say that Q coincides with the “center of the rotation”, when Q is in rest with respect to the reference system and Q is fixed with respect to the rigid body.

6.1.3 Inertia of a Rigid Body

Moment of inertia matrix. Before we are able to resolve equation (6-18) into a component notation of a moving coordinate system, we have to learn more about the inertial behaviour of rigid bodies. Now we regard exclusively the relative angular momentum \vec{H}_M of a rigid body which rotates around reference point M . We can write:

$$\vec{H}_M = \int_M (\vec{x}_{MP} \times \vec{v}_{P,relative}) dM \quad (6-19)$$

Reference point M is the center of mass. Often M is inertially fixed, but this is not mandatory for equation of condition (6-19). Vector \vec{x}_{MP} locates mass element dM . Since we assumed that the rigid body rotates around point M , we can find an expression for the relative velocity $\vec{v}_{P,relative}$ of each mass element dM :

$$\vec{v}_{P,relative} = \vec{\omega} \times \vec{x}_{MP} \quad (6-20)$$

Equation (6-20) is the equation of condition for the rotation vector $\vec{\omega}$. As a vector equation, this relationship is independent of a coordinate system. Without dedication to a particular coordinate system, we write for location vector $\vec{x}_{MP} = (x_1, x_2, x_3)$ and for rotation vector $\vec{\omega} = (\omega_1, \omega_2, \omega_3)$. Naturally, the elements of \vec{x}_{MP} and $\vec{\omega}$ depend on the reference system that we take as a basis; but for every coordinate system we get:

$$\vec{H}_M = \int_M \begin{pmatrix} + (x_2^2 + x_3^2) \omega_1 - x_1 x_2 \omega_2 - x_1 x_3 \omega_3 \\ - x_1 x_2 \omega_1 + (x_1^2 + x_3^2) \omega_2 - x_2 x_3 \omega_3 \\ - x_1 x_3 \omega_1 - x_2 x_3 \omega_2 + (x_1^2 + x_2^2) \omega_3 \end{pmatrix} \cdot dM \quad (6-21)$$

Equation (6-21) is the component form of vector equation (6-19), valid for every coordinate system. Usually, the elements of vector $\vec{\omega}$ are not constant but functions of time. The elements of vector \vec{x}_{MP} locate mass element dM : it is very important to notice that vector \vec{x}_{MP} is constant if a body frame of reference is taken as a basis. Then we can express the angular momentum vector as product of a matrix and the rotation velocity vector $\vec{\omega}$. This matrix is called “moment of inertia matrix”. Thus:

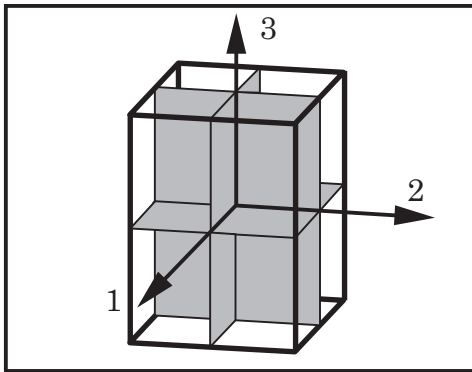
$$\vec{H}_M = \mathfrak{S}_M \cdot \vec{\omega} = \begin{pmatrix} \mathcal{A} & \mathcal{F} & \mathcal{E} \\ \mathcal{F} & \mathcal{B} & \mathcal{D} \\ \mathcal{E} & \mathcal{D} & \mathcal{C} \end{pmatrix} \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \quad (6-22)$$

with diagonal elements: $\mathcal{A} = \int_M (x_2^2 + x_3^2) dM$, $\mathcal{B} = \int_M (x_1^2 + x_3^2) dM$, $\mathcal{C} = \int_M (x_1^2 + x_2^2) dM$ and off-diagonal elements: $\mathcal{D} = - \int_M x_2 x_3 dM$, $\mathcal{E} = - \int_M x_1 x_3 dM$, $\mathcal{F} = - \int_M x_1 x_2 dM$

The moment of inertia matrix \mathfrak{S} is a symmetric matrix (referring to other particular properties \mathfrak{S} is also called “second range tensor” or “dyadic”). Matrix \mathfrak{S} contains all information on the mass distribution within a rigid body. When we use a body system of reference, the elements of the moment of inertia matrix \mathfrak{S} are constant. In this important case these elements are independent of the attitude of the system. Then we can calculate these elements disregarding the motion of the rigid body.

Principal body axis frame. Equation (6-22) shows that the angular momentum \vec{H} is a linear vector function of the angular velocity $\vec{\omega}$. Usually, \vec{H} and $\vec{\omega}$ are not aligned. Here we can observe a significant difference when we compare the properties of the angular momentum with the linear momentum: \vec{V} is always aligned with velocity \vec{v}_M .

For a better understanding of the inertial behaviour of a rigid body we will now regard a simple case: the body should be a homogeneous brick with three planes of symmetry. The coordinate system (“body frame of reference”) should be placed into the intersection lines of these planes of symmetry. Since the brick is symmetric with respect to the coordinate system, the three off-diagonal elements of the moment of inertia matrix \mathfrak{S} must vanish ($\mathcal{D} = 0$, $\mathcal{E} = 0$, $\mathcal{F} = 0$).



A body-fixed coordinate system where the off-diagonal elements of the moment of inertia matrix are zero is called a “principal body axis frame”. In the component notation of the principal body axis frame the angular momentum vector \vec{H}_M takes the form of:

$$\vec{H}_M = \begin{pmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{B} & 0 \\ 0 & 0 & \mathcal{C} \end{pmatrix} \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \quad (6-23)$$

When we assume that the length of the edges are all different, we have $\mathcal{A} \neq \mathcal{B} \neq \mathcal{C}$. In this case there are exactly three directions where the angular momentum and the rotation velocity are parallel: \vec{H}_M is aligned with $\vec{\omega}$ when the body rotates on one of the coordinate axes. When $\mathcal{A} = \mathcal{B}$, the vectors \vec{H}_M and $\vec{\omega}$ are aligned if $\vec{\omega} = (0, 0, \omega_3)$ or if $\vec{\omega} = (\omega_1, \omega_2, 0)$. All elements are equal when the body is a cube ($\mathcal{A} = \mathcal{B} = \mathcal{C}$), and then the angular momentum vector is always aligned with the rotation vector.

Rotations on the principal body axes. Let us consider the body rotates with the angular velocity ω_1 on the 1-axis. Torque \mathcal{M}_1 will increase (or decrease) its spin rate. The angular momentum is a linear function of the rotation velocity; both vectors are aligned and fixed with respect to inertial space. Therefore we can conclude:

$$\mathcal{M}_1 = \frac{dH_1}{dt} = \mathcal{A} \cdot \frac{d\omega_1}{dt}, \quad \text{or:} \quad \mathcal{A} = \frac{dH_1}{d\omega_1} = \frac{H_1}{\omega_1} \quad (6-24)$$

This follows from the angular momentum law (6-18). Thus, term $\mathcal{A} = \int_M (x_2^2 + x_3^2) dM$ is a scale for the “power of resistance” to torques. A ring around the 1-axis with mass M and a certain radius $R_{\mathcal{A}}$ has the same resistance to torques as the brick (expression $(x_2^2 + x_3^2)$ is the rectangular distance of particle dM from the spin axis). We can replace every diagonal element of the moment of inertia matrix by an expression which uses the mass multiplied by the square of a certain “radius of gyration”:

$$MR_{\mathcal{A}}^2 = \mathcal{A}, \quad MR_{\mathcal{B}}^2 = \mathcal{B}, \quad MR_{\mathcal{C}}^2 = \mathcal{C} \quad (6-25)$$

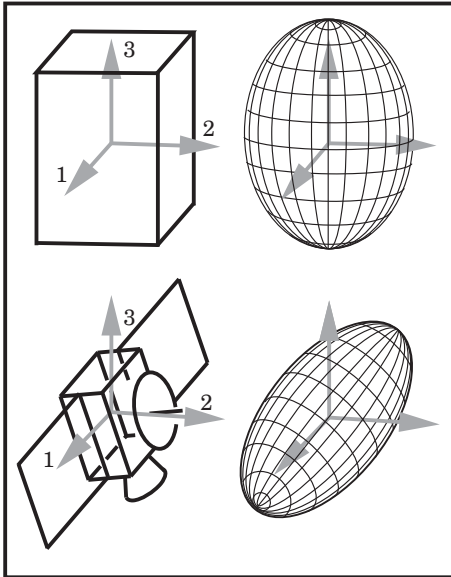
Sometimes notations with the “radius of gyration R ” are preferred in gyro-dynamics.

Ellipsoid of inertia. The next point in question is what happens if the rigid body rotates on another (arbitrarily directed) spin axis. The spin axis is determined by its unit direction vector, the rotation vector $\vec{\omega} = (\omega_1, \omega_2, \omega_3)$ divided by its length $|\vec{\omega}|$.

The angular momentum vector \vec{H}_M is not aligned with the rotation vector $\vec{\omega}$ anymore; but the angular momentum is still a linear vector function of the rotation velocity: an increase of the spin rate will let the angular momentum vector grow proportionally (the angular momentum law (6-18) states that torques are required to change the angular momentum vector with respect to inertial space). Using the direction of the spin axis, the angular momentum vector can be resolved into two rectangular components. The component that is aligned with the spin axis is important to find the moment of inertia (“power of resistance to torques”) for a rotation on this axis. Again, we use term $MR_{\mathfrak{S}}^2$ to characterize the moment of inertia (and compare it with the inertia of a ring around the rotation axis of mass M and radius $R_{\mathfrak{S}}$). Thus:

$$MR_{\mathfrak{S}}^2 = \frac{\vec{H}_M \cdot \vec{\omega} / |\vec{\omega}|}{|\vec{\omega}|} \quad (6-26)$$

We form the scalar (dot) vector product of angular momentum vector \vec{H}_M and the unit vector in direction of the spin axis. This yields the length of the component of the angular momentum vector in direction of the spin axis. Then we divide the result by the length of the rotation vector to find the moment of inertia (“power of resistance to torques”) for a rotation on the spin axis. We can observe that radius $R_{\mathfrak{S}}$ is just a function of the orientation of $\vec{\omega}$, but not a function of its length $|\vec{\omega}|$. The function $R_{\mathfrak{S}}(\vec{\omega}/|\vec{\omega}|)$ determines a three-dimensional geometrical figure.



Equation (6-26) applied to equation (6-23) gives (principal body axis frame):

$$MR_{\mathfrak{S}}^2 = \mathcal{A} \frac{\omega_1^2}{|\vec{\omega}|^2} + \mathcal{B} \frac{\omega_2^2}{|\vec{\omega}|^2} + \mathcal{C} \frac{\omega_3^2}{|\vec{\omega}|^2} \quad (6-27)$$

Equation (6-26) applied to equation (6-22) gives (the general case):

$$MR_{\mathfrak{S}}^2 = \mathcal{A} \frac{\omega_1^2}{|\vec{\omega}|^2} + \mathcal{B} \frac{\omega_2^2}{|\vec{\omega}|^2} + \mathcal{C} \frac{\omega_3^2}{|\vec{\omega}|^2} + 2\mathcal{D} \frac{\omega_2\omega_3}{|\vec{\omega}|^2} + 2\mathcal{E} \frac{\omega_1\omega_3}{|\vec{\omega}|^2} + 2\mathcal{F} \frac{\omega_1\omega_2}{|\vec{\omega}|^2} \quad (6-28)$$

Both geometrical figures represent ellipsoids (the inverse length $1/R_{\mathfrak{S}}$ determines the “ellipsoids of inertia”). The symmetry axes are aligned with the coordinate axes when a principal axis frame of reference is used (equation 6-27); but in the general case (6-28) the symmetry axes are not aligned with the axes of the coordinate system.

Direction of the principal axis. When we analyze the inert behaviour of a rigid body as function of the spin direction, we find the geometrical figure of an ellipsoid. This is also the case for the general moment of inertia matrix \mathfrak{S}_M (equation 6-22) where the reference system is arbitrarily directed. If two completely different rigid bodies have identical ellipsoids, these bodies behave identically concerning rotations. The rotational behaviour of a rigid body is entirely determined by the ellipsoid of inertia; and the ellipsoid of inertia is entirely determined when in the principal body frame of reference the three diagonal elements (terms \mathcal{A} , \mathcal{B} and \mathcal{C}) are well-known. The “ellipsoid of inertia” defined by the length $1/R_{\mathfrak{S}}$ resembles the shape of the body: an oblate disk has a lentiform ellipsoid, and a thin stick has an elongated ellipsoid.

The problem is to find the attitude of the principal axis frame of reference (and the elements \mathcal{A} , \mathcal{B} and \mathcal{C}) for a rigid body with a well-known matrix of inertia (6-28). When the body rotates on one of its principal axes, the angular momentum vector is aligned with the rotation velocity vector. Only then the following condition is valid:

$$\begin{pmatrix} \mathcal{A} & \mathcal{F} & \mathcal{E} \\ \mathcal{F} & \mathcal{B} & \mathcal{D} \\ \mathcal{E} & \mathcal{D} & \mathcal{C} \end{pmatrix} \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \lambda \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \quad (6-29)$$

Condition (6-29) has a non-trivial solution (with $\vec{\omega} \neq 0$) for exactly three directions of the rotation velocity vector $\vec{\omega}$: the three directions of the principal axes. Then multiplier λ is the moment of inertia for a rotation on the concerning principal axis. When we write down the component form of condition (6-29), we get a system of three equations. Since the length of $\vec{\omega}$ does not matter, the equation system must be singular when the body rotates on one of the principal axes. The direction of the principal axes can be found by the solution of a so-called “eigenvalue problem”. Non-trivial solutions of condition (6-29) are just possible if the determinate of the linear equation system vanishes:

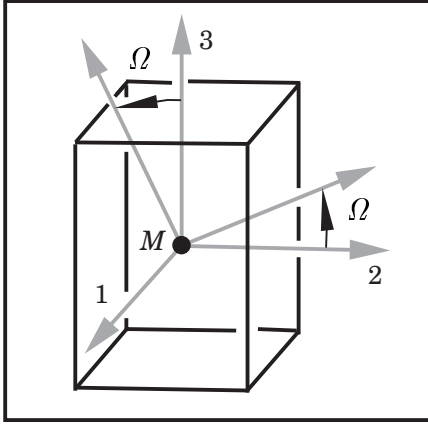
$$\begin{vmatrix} (\mathcal{A} - \lambda) & \mathcal{F} & \mathcal{E} \\ \mathcal{F} & (\mathcal{B} - \lambda) & \mathcal{D} \\ \mathcal{E} & \mathcal{D} & (\mathcal{C} - \lambda) \end{vmatrix} = 0 \quad (6-30)$$

$$\text{or: } (\mathcal{A} - \lambda)(\mathcal{B} - \lambda)(\mathcal{C} - \lambda) + 2\mathcal{F}\mathcal{D}\mathcal{E} - (\mathcal{A} - \lambda)\mathcal{D}^2 - (\mathcal{B} - \lambda)\mathcal{E}^2 - (\mathcal{C} - \lambda)\mathcal{F}^2 = 0 \quad (6-31)$$

Multipliers λ are called “eigenvalues” of the matrix, equation (6-31) is the “characteristic equation” of the problem. For rigid bodies with realistic mass distribution, equation (6-31) must have three solutions. These three solutions are the moments of inertia for rotations on the principal axes (the principal moments of inertia $\lambda_1 = \mathcal{A}$, $\lambda_2 = \mathcal{B}$, $\lambda_3 = \mathcal{C}$). Once we have found the principle moments of inertia, we can readily calculate the direction of the principal axes: we simply insert a solution into condition (6-29) to get a singular system of three equations (the length of $\vec{\omega}$ does not matter). We can choose one component of $\vec{\omega}$ arbitrarily and use two equations to calculate the other two components of $\vec{\omega}$. Then the direction of $\vec{\omega}$ shows us the direction of the concerning principal axis.

Angular displacement of the reference system. A body frame of reference exists for every rigid body, where the off-diagonal elements (the products of inertia \mathcal{D} , \mathcal{E} and \mathcal{F}) vanish and the diagonal elements (moments of inertia \mathcal{A} , \mathcal{B} and \mathcal{C}) assume relative extremes. In case the body has a homogeneous mass distribution and a simple symmetrical shape (a stick, brick, cylinder, disk, spherical segment or something like that), the orientation of the principal axes is evident and the principal moments of inertia can be found by analytical integration. Alternatively we can find the moments of inertia of many bodies with different shape tabulated in fundamental books on mechanics. When we want to calculate the inertia matrix of a satellite, we can compose (add up) this matrix if we know the moments of inertia of all elementary parts of the satellite. Therefore every individual inertia matrix has to refer to the same coordinate system. Using transformation matrices, we can change the representation of the inertia matrix from one coordinate system to another one.

First, let us consider rotational changes of the reference system. As we have seen, the rigid body is represented in its principal axis frame of reference by:



$$\vec{H}_M = \begin{pmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{B} & 0 \\ 0 & 0 & \mathcal{C} \end{pmatrix} \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \quad (6-32)$$

$$\text{with: } \int_M x_1^2 dM = (\mathcal{B} + \mathcal{C} - \mathcal{A})/2$$

$$\int_M x_2^2 dM = (\mathcal{A} + \mathcal{C} - \mathcal{B})/2$$

$$\int_M x_3^2 dM = (\mathcal{A} + \mathcal{B} - \mathcal{C})/2$$

$$\text{and: } \int_M x_1 x_2 dM = 0$$

$$\int_M x_2 x_3 dM = 0$$

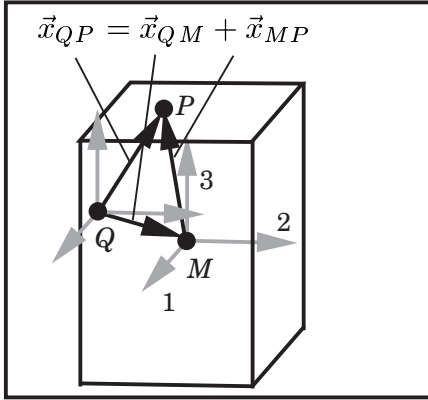
$$\int_M x_1 x_3 dM = 0$$

The vectors \vec{H}_M (the angular momentum) and $\vec{\omega}$ (the rotation velocity) are written in the component notation of the principal axis frame of reference. Now we represent the body in another system that is rotated on the 1-axis, with the twisting angle Ω :

$$\vec{H}_M = \begin{pmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{B} \cos^2 \Omega + \mathcal{C} \sin^2 \Omega & (\mathcal{C} - \mathcal{B}) \sin \Omega \cos \Omega \\ 0 & (\mathcal{C} - \mathcal{B}) \sin \Omega \cos \Omega & \mathcal{B} \sin^2 \Omega + \mathcal{C} \cos^2 \Omega \end{pmatrix} \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \quad (6-33)$$

Now vectors \vec{H}_M and $\vec{\omega}$ are written in the component notation of the rotated frame of reference. The new matrix of inertia can be expressed using the principal moments of inertia \mathcal{A} , \mathcal{B} and \mathcal{C} . We find the new elements by a simple coordinate transformation: in a component notation of the principal axis frame of reference, $\vec{x}_{MP} = (x_1, x_2, x_3)$ is the vector which locates mass element dM . When we take the rotated system, the same vector takes the form of $\vec{x}_{MP} = (x_1, x_2 \cos \Omega + x_3 \sin \Omega, -x_2 \sin \Omega + x_3 \cos \Omega)$. With these new components we must form the integrals of the individual elements of the matrix of inertia (6-22), considering the relationships (6-32). Equation (6-33) serves as an example for a single rotation of the coordinate system. To come to an arbitrary new orientation of the reference system, three rotations are required.

Parallel axis displacement. Equations (6-14) and (6-15) indicate that, when point Q and point M move parallel, the angular momentum \vec{H}_Q (with respect to Q) is the same as angular momentum \vec{H}_M (with respect to center of mass M). Then we have:



$$\vec{v}_{M,relative} = \frac{1}{M} \int_M \vec{v}_{P,relative} dM = 0 \quad (6-34)$$

Consequently:

$$\begin{aligned} \vec{H}_Q &= \int_M (\vec{x}_{QP} \times \vec{v}_{P,relative}) dM \quad (6-35) \\ &= \int_M ((\vec{x}_{QM} + \vec{x}_{MP}) \times \vec{v}_{P,relative}) dM \\ &= \int_M (\vec{x}_{MP} \times \vec{v}_{P,relative}) dM = \vec{H}_M \end{aligned}$$

Note that the velocity $\vec{v}_{P,relative}$ of particle dM relative to the reference point is the same for point Q and for point M : the velocity $\vec{v}_{M,relative}$ equals zero. However, point Q is not fixed with respect to a body frame of reference of the rotating body.

Now let us consider a different case, where Q is a fixed location with respect to the rigid body. Point Q does not move parallel with point M when the body rotates, and consequently $\vec{H}_Q \neq \vec{H}_M$. We can write for the relative angular momentum vector:

$$\vec{H}_Q = \int_M (\vec{x}_{QP} \times (\vec{\omega} \times \vec{x}_{QP})) dM \quad (6-36)$$

The body rotates about point Q , and $\vec{v}_{P,relative} = \vec{\omega} \times \vec{x}_{QP}$ replaces equation (6-20). We will assume that the elements of the moment of inertia matrix are well-known regarding a body frame of reference centered in point M (terms \mathcal{A} , \mathcal{B} , \mathcal{C} , \mathcal{D} , \mathcal{E} , \mathcal{F}). The question is what happens if the body frame of reference is parallel-displaced, maintaining the orientation of its axes.

Vector $\vec{x}_{QP} = \vec{x}_{QM} + \vec{x}_{MP}$ locates mass particle dM . We can use the elements of $\vec{x}_{MP} = (x_1, x_2, x_3)$ and $\vec{x}_{QM} = (\Delta x_1, \Delta x_2, \Delta x_3)$ for the composition of vector \vec{x}_{QP} ; $\vec{x}_{QP} = (x_1 + \Delta x_1, x_2 + \Delta x_2, x_3 + \Delta x_3)$. In a body frame of reference, the elements of \vec{x}_{QM} are constant. Since point M is the center of mass, we have $\int_M \vec{x}_{MP} dM = 0$. Evaluation of equation (6-36) gives finally (with \mathcal{A} , \mathcal{B} , \mathcal{C} , \mathcal{D} , \mathcal{E} , \mathcal{F} of equation 6-22):

$$\vec{H}_Q = \begin{pmatrix} \mathcal{A} + M(\Delta x_2^2 + \Delta x_3^2) & \mathcal{F} - M\Delta x_1\Delta x_2 & \mathcal{E} - M\Delta x_1\Delta x_3 \\ \mathcal{F} - M\Delta x_1\Delta x_2 & \mathcal{B} + M(\Delta x_1^2 + \Delta x_3^2) & \mathcal{D} - M\Delta x_2\Delta x_3 \\ \mathcal{E} - M\Delta x_1\Delta x_3 & \mathcal{D} - M\Delta x_2\Delta x_3 & \mathcal{C} + M(\Delta x_1^2 + \Delta x_2^2) \end{pmatrix} \cdot \vec{\omega} \quad (6-37)$$

Vector \vec{H}_Q is the angular momentum when the rigid body rotates with $\vec{\omega}$ around reference point Q . We can observe that anyway the diagonal elements of the matrix of inertia are always positive and bigger than for a rotation about the center of mass. Off-diagonal elements (\mathcal{D} , \mathcal{E} , \mathcal{F}) appear when we translate a principal axis body frame of reference that was originally centered in point M (provided that two or three elements of the displacement vector $\vec{x}_{QM} = (\Delta x_1, \Delta x_2, \Delta x_3)$ are not zero).

6.1.4 Kinetic Energy of a Rigid Body

Energy and work. A force that accelerates a mass produces work; this work is stored in the motion of the mass in form of “kinetic energy”. To calculate the kinetic energy of mass particle dM , we have to substitute the force by the time derivative of the linear momentum $dM \cdot d\vec{v}_P/dt$, consider $\vec{v}_P \cdot dt$ as coordinate and integrate (\vec{v}_P is the absolute velocity of dM). The kinetic energy of a mass particle becomes $1/2 \cdot \vec{v}_P^2 dM$. For the entire kinetic energy E of an extended mass M we can write:

$$E_{kinetic} = \frac{1}{2} \int_M |\vec{v}_P|^2 dM \quad (6-38)$$

Translational and rotational kinetic energy. Now we consider a rigid body that translates and rotates simultaneously. If we take the center of mass as reference point, we can split up the kinetic energy in a translational part and in a rotational part. The velocity \vec{v}_P of a mass particle can be replaced by $\vec{v}_P = \vec{v}_M + \vec{\omega} \times \vec{x}_{MP}$, where \vec{v}_M is the velocity of the center of mass of the rigid body and $\vec{\omega}$ its rotation vector:

$$\begin{aligned} E_{kinetic} &= \frac{1}{2} \int_M (\vec{v}_M + \vec{\omega} \times \vec{x}_{MP})^2 dM \\ &= \frac{1}{2} \int_M [\vec{v}_M^2 + 2\vec{v}_M \cdot (\vec{\omega} \times \vec{x}_{MP}) + (\vec{\omega} \times \vec{x}_{MP})^2] dM \end{aligned} \quad (6-39)$$

The first term inside the square brackets can be integrated immediately. The integral of the second term vanishes since the expression $\int_M \vec{x}_{MP} dM$ equals zero (it is the equation of condition for the center of mass). Also the integral of third term yields a simple expression, it can be written as $(\vec{\omega} \cdot \vec{H}_M)$. Finally we get:

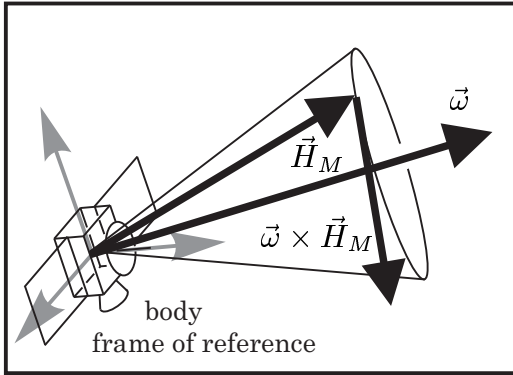
$$\begin{aligned} E_{kinetic} &= E_{kinetic,translational} + E_{kinetic,rotational} \\ &= \frac{1}{2} M \vec{v}_M^2 + \frac{1}{2} \vec{\omega} \cdot (\mathfrak{S}_M \cdot \vec{\omega}) \end{aligned} \quad (6-40)$$

Equation (6-40) determines the entire kinetic energy of a rigid body of mass M . \mathfrak{S}_M is the moment of inertia matrix with respect to the center of mass (point M). Term \vec{v}_M is the translation velocity vector and term $\vec{\omega}$ is the rotation velocity vector.

6.1.5 Equations of Motion of a Rotating Rigid Body

The angular momentum vector in a moving coordinate system. We return to the angular momentum law (equation 6-17). When we take the center of mass M as reference point (or any other point which is not accelerated), the angular momentum law accepts a form which is particularly simple (equation 6-18). The absolute change of the angular momentum vector in time is the change of this vector with respect to inertial space. Using a moving frame of reference, the absolute time derivative consists of two parts: a change of the vector with respect to the moving system, and a change of the vector introduced by the angular motion of the coordinate system.

Then we calculate the absolute time derivative of the angular momentum vector (expression $d\vec{H}/dt$ in equation 6-18) simply by differentiation of the components (forming the relative time derivative of \vec{H}), adding to the result the cross product of rotation velocity vector $\vec{\omega}_{system}$ and angular momentum vector \vec{H} . The rule is simply:



$$\vec{\mathcal{M}}_M = \left(\frac{d\vec{H}_M}{dt} \right)_{inertial} \quad (6-41)$$

$$\vec{\mathcal{M}}_M = \left[\frac{d\vec{H}_M}{dt} \right]_{relative} + \vec{\omega}_{system} \times \vec{H}_M$$

$$\vec{H}_M = \mathfrak{S}_M \cdot \vec{\omega}_{rigid\ body} \quad (6-42)$$

It must be emphasized that term $\vec{\omega}_{system}$ is the rotation velocity vector of the moving coordinate system; and term $\vec{\omega}_{rigid\ body}$ is the rotation velocity of the rigid body (sometimes it is sensible to use a special coordinate system that rotates with another angular velocity as the rigid body). However, when we consider a body frame of reference, these two rotation velocities are identical. Remember that the elements of the matrix of inertia \mathfrak{S}_M are time invariant if a body frame of reference is used. Then we insert the angular momentum vector (6-42) into law (6-41) and conclude:

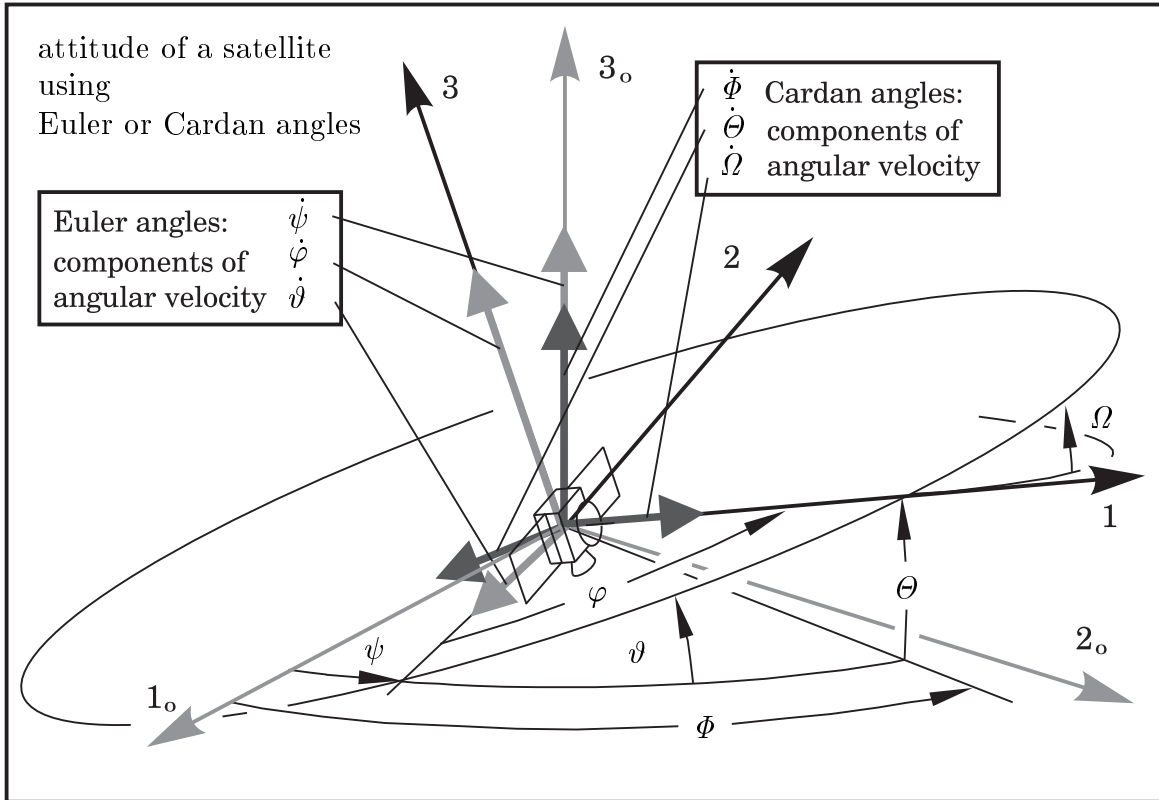
$$\vec{\mathcal{M}}_M = \mathfrak{S}_M \cdot \dot{\vec{\omega}} + \vec{\omega} \times (\mathfrak{S}_M \cdot \vec{\omega}) \quad (6-43)$$

Dynamical Euler equations. Equation (6-43) is of fundamental importance to gyro-dynamics (it is the vector form of the so-called “dynamical Euler equations”). We find this law by a transformation of the angular momentum law (6-18), taking as a basis a body frame of reference coincident with the center of mass of the rigid body. The elements of the matrix of inertia \mathfrak{S}_M are constant in time; and the rotation velocity of the rigid body is identical with the rotation velocity of the coordinate system. When equation (6-43) is resolved into its component notation, the use of a principal axis frame of reference is favorable: the off-diagonal elements (\mathcal{D} , \mathcal{E} , \mathcal{F}) of the matrix of inertia vanish, and just diagonal elements (\mathcal{A} , \mathcal{B} , \mathcal{C}) are involved. Thus:

$$\begin{aligned} \mathcal{M}_1 &= \mathcal{A} \dot{\omega}_1 + (\mathcal{C} - \mathcal{B}) \omega_2 \omega_3 \\ \mathcal{M}_2 &= \mathcal{B} \dot{\omega}_2 + (\mathcal{A} - \mathcal{C}) \omega_1 \omega_3 \\ \mathcal{M}_3 &= \mathcal{C} \dot{\omega}_3 + (\mathcal{B} - \mathcal{A}) \omega_1 \omega_2 \end{aligned} \quad (6-44)$$

Equation system (6-44) is the component form of the dynamical Euler equations. These differential equations allow us to calculate the rotation velocity $\vec{\omega}$ as a function of time when a satellite spins without attitude control ($\vec{\mathcal{M}} = 0$), or when a spinning satellite is controlled by application of thrust ($\vec{\mathcal{M}} \neq 0$). When $\vec{\omega}$ is a predetermined function of time, equation system (6-44) can be used to calculate torques ($-\vec{\mathcal{M}}$) caused by rotating parts of a satellite (gyros or spinning platforms inside a satellite).

Kinematical Euler equations. Finally we must know how to compute the attitude of a spinning satellite when the rotation velocity vector $\vec{\omega}$ is a well-known time function. Using equation system (6-44), the components of vector $\vec{\omega}$ refer to a moving coordinate system (the “body frame of reference”). We can use alternatively Euler angles $(\psi, \vartheta, \varphi)$ or Cardan angles (Φ, Θ, Ω) to determine the attitude of this moving system $(1, 2, 3)$ with respect to an inertial coordinate system $(1_o, 2_o, 3_o)$:



The time derivatives of the Euler angles $\dot{\psi}$, $\dot{\vartheta}$, $\dot{\varphi}$ or alternatively the time derivatives of the Cardan angles $\dot{\Phi}$, $\dot{\Theta}$, $\dot{\Omega}$ can be used to compose the rotation velocity vector $\vec{\omega}$:

$$\vec{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} \dot{\psi} \sin \vartheta \sin \varphi + \dot{\vartheta} \cos \varphi \\ \dot{\psi} \sin \vartheta \cos \varphi - \dot{\vartheta} \sin \varphi \\ \dot{\varphi} + \dot{\psi} \cos \vartheta \end{pmatrix} = \begin{pmatrix} \dot{\Phi} \sin \Theta + \dot{\Omega} \\ \dot{\Phi} \cos \Theta \sin \Omega - \dot{\Theta} \cos \Omega \\ \dot{\Phi} \cos \Theta \cos \Omega + \dot{\Theta} \sin \Omega \end{pmatrix} \quad (6-45)$$

To find the attitude of the satellite we have to transform vector equation (6-45), insert the components of $\vec{\omega}$ and integrate a first order differential equation system:

$$\begin{aligned} \dot{\psi} &= (\omega_1 \sin \varphi + \omega_2 \cos \varphi) / \sin \vartheta & \dot{\Phi} &= (\omega_2 \sin \Omega + \omega_3 \cos \Omega) / \cos \Theta \\ \dot{\vartheta} &= \omega_1 \cos \varphi - \omega_2 \sin \varphi & \dot{\Theta} &= -\omega_2 \cos \Omega + \omega_3 \sin \Omega \\ \dot{\varphi} &= \omega_3 - (\omega_1 \sin \varphi + \omega_2 \cos \varphi) \cdot \cot \vartheta & \dot{\Omega} &= \omega_1 - (\omega_2 \sin \Omega + \omega_3 \cos \Omega) \cdot \tan \Theta \end{aligned} \quad (6-46)$$

The nonlinear differential equations (6-46) are called “kinematical Euler equations”. They serve to obtain the attitude of a moving coordinate system as function of time.

6.2. Spin Stabilization

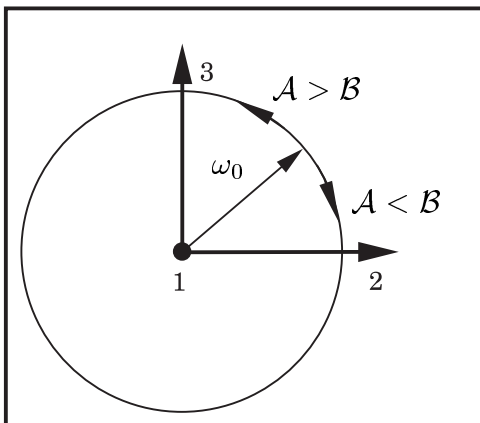
We have to integrate a system of three first-order differential equations, the so-called “dynamical Euler equations” (6-44), to calculate the rotation velocity vector of a satellite as a function of the external moment vector. Both vectors are given in a component notation of the body frame of reference. When we want to find the actual attitude of the satellite, we have to insert the components of the rotation velocity vector into the “kinematical Euler equations” (6-46), and integrate simultaneously another system of first-order differential equations. It is not advisable to integrate simply the components of the rotation velocity vector: integration of ω_1 , ω_2 and ω_3 yields attitude angles which are not appropriate for a determination of the vehicle’s attitude (these angles are nonholonomic coordinates). In case of three-dimensional rotations, the correct way to the attitude of the vehicle is the integration of the Euler equations.

6.2.1 Motion of a Torque-free Satellite

The axisymmetric spacecraft. In some special situations it is possible to find analytical solutions. We take the equation system (6-44) and consider a torque-free satellite, $\vec{M} = 0$. The solution to the general case ($\mathcal{A} \neq \mathcal{B} \neq \mathcal{C}$) involves so-called “elliptical integral functions”; and it depends on the point of view whether these functions are “analytical solutions”. Considerably less complex and much more important in practice is the motion of spacecraft with axisymmetric moment of inertia. A spinning satellite has usually an axisymmetric moment of inertia ($\mathcal{A} \neq \mathcal{B} = \mathcal{C}$). You can readily verify that for $\mathcal{B} = \mathcal{C}$ the following rotation vector $\vec{\omega}$ constitutes the general solution to the dynamical Euler equation system:

$$\vec{\omega} = \begin{pmatrix} \omega_1 \\ \omega_0 \cdot \sin((1 - \mathcal{A}/\mathcal{B}) \omega_1 (t - t_0)) \\ \omega_0 \cdot \cos((1 - \mathcal{A}/\mathcal{B}) \omega_1 (t - t_0)) \end{pmatrix} \quad (6 - 47)$$

When the axisymmetric satellite rotates on the 1-axis, \mathcal{A} is the moment of inertia. The moment of inertia is termed \mathcal{B} for every axis which is rectangular to the 1-axis.

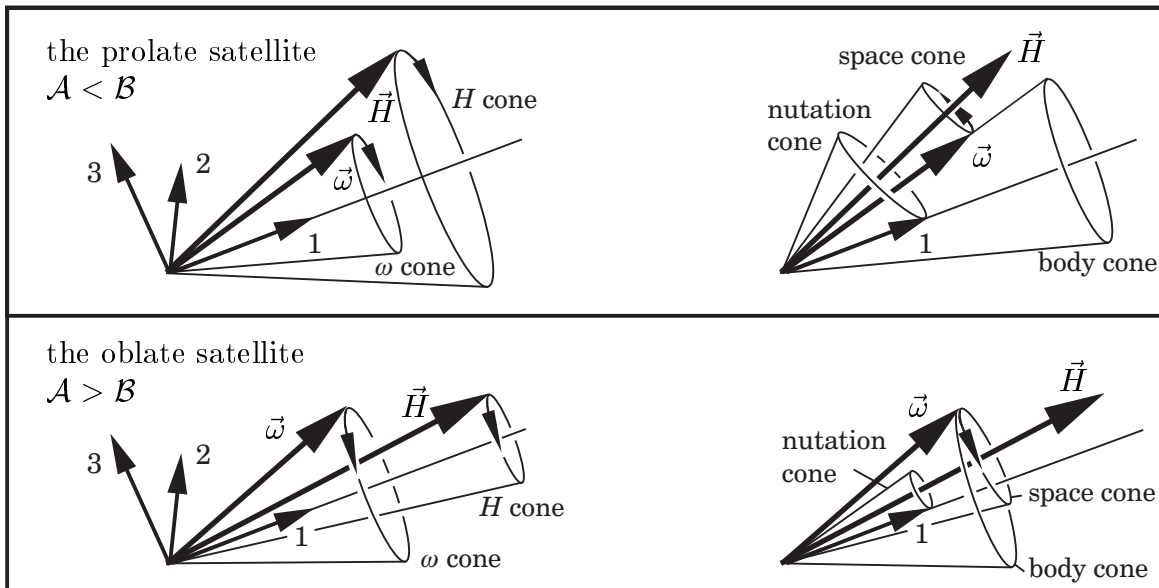


The integration constants are ω_1 (the constant component of the rotation velocity vector in the symmetry axis), ω_0 (the amplitude of the circular motion of the tip of vector $\vec{\omega}$) and t_0 (the initial instant). An observer in the body frame of reference sees that vector $\vec{\omega}$ moves on the surface of a cone. The rotation frequency of this motion is $\omega_1(\mathcal{B} - \mathcal{A})/\mathcal{B}$. The tip of the vector $\vec{\omega}$ rotates clockwise when the satellite is a prolate rotor ($\mathcal{A} < \mathcal{B}$), it rotates counterclockwise when the satellite is an oblate rotor ($\mathcal{A} > \mathcal{B}$).

Knowing the matrix of inertia and the rotation velocity vector, we can write down the vector of the angular momentum \vec{H} of the rotating satellite (equation 6-23):

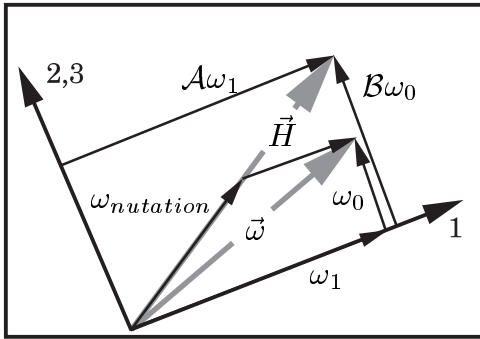
$$\vec{H} = \begin{pmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{B} & 0 \\ 0 & 0 & \mathcal{B} \end{pmatrix} \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} \mathcal{A} \omega_1 \\ \mathcal{B} \omega_2 \\ \mathcal{B} \omega_3 \end{pmatrix} \quad (6-48)$$

Since ω_1 is constant in the body frame of reference, the component of the angular momentum vector in 1-direction $\mathcal{A} \cdot \omega_1$ must also be constant. The two other components (in 2-direction and in 3-direction) define a circular motion, like the components of the rotation vector. Seen from the body frame of reference, both vectors execute a coning motion: vector \vec{H} rotates synchronously on the 1-axis with the same frequency as vector $\vec{\omega}$. When $\mathcal{A} < \mathcal{B}$ (a prolate satellite), \vec{H} is more distant from the 1-direction than vector $\vec{\omega}$. Then the angle of the H cone is larger than the angle of the ω cone. When the satellite is an oblate rotor, $\mathcal{A} > \mathcal{B}$ and the H cone is inside the ω cone.



Now we try to figure out the motion seen from an inertial frame of reference. The direction of the rotation velocity vector $\vec{\omega}$ is not constant in time, neither in the body frame of reference nor in the inertial system; but the absolute angular momentum must be a constant vector, since all external torques equal zero. The direction of \vec{H} is fixed with respect to inertial space; vector $\vec{\omega}$ and the unit vector in 1-direction rotate on the axis of \vec{H} . The motion can be regarded as the rolling of the "body cone" (fixed in the satellite) on the "space cone" (fixed in inertial space). The vector $\vec{\omega}$ determines the actual rotation axis, it is always located exactly there where the body cone touches the space cone. The unit vector in 1-direction determines the attitude of the satellite, this vector travels through the "nutration cone". Nutration cone and space cone have the same centerline (the axis of the angular momentum vector \vec{H}). Seen from this line, vector $\vec{\omega}$ and the body-fixed 1-direction are on the same side when $\mathcal{A} < \mathcal{B}$ (the prolate satellite). Then the body cone rolls outside on the space cone. When $\mathcal{A} > \mathcal{B}$ (the oblate satellite), $\vec{\omega}$ is on the other side as the 1-direction. Then the body cone rolls on the space cone, which is now inside the body cone.

Nutation frequency. The body-fixed 1-axis rotates counterclockwise on the space-fixed \vec{H} -direction. To find the rotation velocity $\omega_{nutation}$, we resolve vector $\vec{\omega}$ into two (usually not rectangular) components: one in the direction of the angular momentum vector \vec{H} , the other one in the direction of the body-fixed 1-axis. Therefore:



$$\frac{\omega_{nutation}}{\omega_0} = \frac{|\vec{H}|}{B \omega_0} = \frac{\sqrt{(A \omega_1)^2 + (B \omega_0)^2}}{B \omega_0}$$

$$\omega_{nutation} = \sqrt{\left(\frac{A}{B} \omega_1\right)^2 + \omega_0^2} \quad (6 - 49)$$

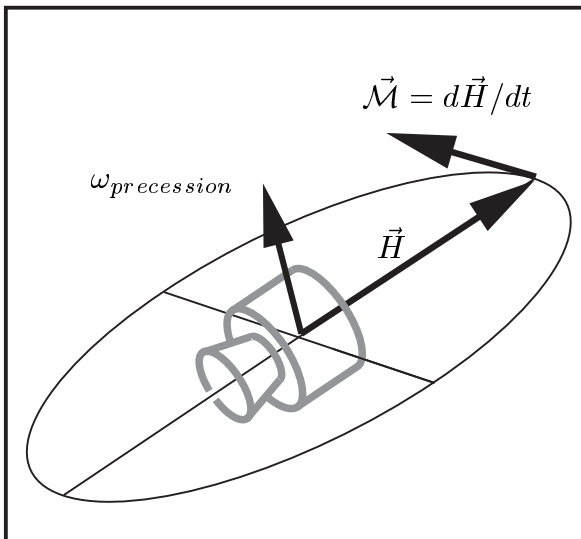
$$\approx \frac{A}{B} \omega_1$$

The nutation frequency $\omega_{nutation}$ is smaller than $|\vec{\omega}|$ for a prolate rotor ($A < B$) and larger than $|\vec{\omega}|$ for an oblate rotor ($A > B$). When the nutation amplitude is small ($\omega_0 \ll \omega_1$), the nutation frequency is approximately a linear function of $|\vec{\omega}|$.

6.2.2 Attitude Maneuver of a Spinning Satellite

Precession of a spinning satellite. When a satellite rotates on its symmetry axis, the direction of the spin axis is aligned with the direction of the angular momentum. This direction is stabilized with respect to inertial space: as mandated by the angular momentum law ($d\vec{H}/dt = \vec{M}$), changing the angular momentum with respect to inertial space $d\vec{H}/dt$ is just possible by the application of external torques \vec{M} .

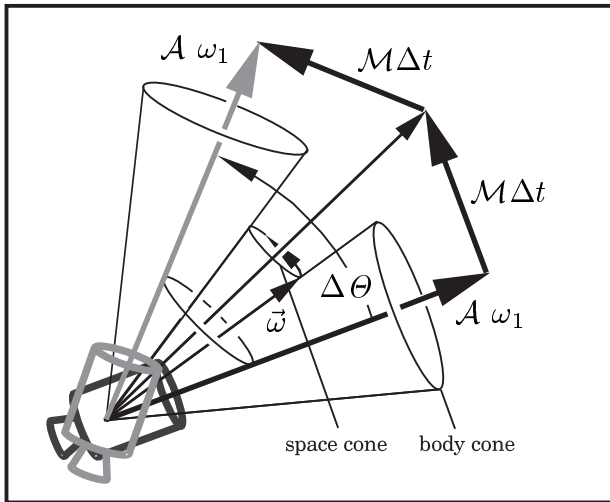
When the external torque vector \vec{M} is aligned with the angular momentum vector \vec{H} , it changes the spin rate. However, in this case the spin direction remains unchanged. For an attitude maneuver we need a torque vector that is perpendicular to



the angular momentum vector: when \vec{M} is perpendicular to \vec{H} , it conserves the spin rate while it pivots the spin direction towards the direction of the external torque vector. The torque vector should always be perpendicular to the spin axis, but it should not spin simultaneously with the satellite. A direction that is fixed with respect to inertial space rotates in the body frame of reference. The attitude control system of the satellite must generate a rotating torque vector (but the torque is just a rotating vector seen from the rotating body frame of reference).

Impulsive thrust attitude maneuver. The conventional way to generate an external moment for changing the attitude of a satellite is the use of control thrusters (orbiting satellites use sometimes magnetorquers, utilizing the earth's magnetic field). We consider a spinning satellite. It rotates without nutation on the body fixed 1-axis ($\omega_1 \neq 0, \omega_0 = 0$). The angular momentum \vec{H} is aligned with the spin axis and has the magnitude $\mathcal{A} \cdot \omega_1$. When the actuators are not mounted on a despun platform, their line of action runs periodically through the appropriate orientation (the thrust direction is determined with respect to inertial space). We can operate the attitude control thrusters several times, but always just for a very short time interval Δt . In order to change the direction of angular momentum vector \vec{H} , these thrusters execute impulsive burn maneuvers of the magnitude $\Delta \vec{H} = \vec{\mathcal{M}} \Delta t$. The direction of the impulsive torque $\vec{\mathcal{M}}$ should always be perpendicular to the body-fixed 1-axis.

At the right moment the control thrusters are fired. The burn maneuver changes the direction and also the magnitude of the angular momentum \vec{H} (it creates a rectangular component $\mathcal{M} \Delta t$). We can use equation (6-44) to verify that the burn maneuver alters also the rotation vector $\vec{\omega}$ (it creates the rectangular component $\mathcal{M} \Delta t / \mathcal{B}$). After the impulsive torque has been applied, $\vec{\omega}$ is not aligned with \vec{H} anymore:



the satellite executes a nutational motion now (provided that $\mathcal{A} \neq \mathcal{B}$). The body cone rolls on the space cone, and the body-fixed 1-direction rotates on the new direction of \vec{H} . When the body-fixed 1-direction arrives on the other side of the space cone (half a rotation period later), we execute the same burn maneuver once again. The second maneuver re-establishes the parallel orientation of $\vec{\omega}$ and \vec{H} . Now the satellite rotates again without nutational oscillation on the new spin axis.

Propellant is very expensive in space. We need propellant to operate the attitude control thrusters for the generation of impulsive torques. The change of the attitude angle $\Delta \theta$ requires the application of two impulsive burn maneuvers $\mathcal{M} \Delta t$, where:

$$\mathcal{M} \Delta t = \mathcal{A} \omega_1 \cdot \tan(\Delta \theta / 2) \quad (6 - 50)$$

Usually, the moment \mathcal{M} of the attitude control system is predetermined, and we can use the relationship above to calculate the “time interval of impulsive burning” Δt as a function of angle $\Delta \theta$. The tangent function in equation (6-50) is responsible for the fact that it saves propellant when we split a maneuver with large angle $\Delta \theta$ into several smaller maneuvers. An example is the 180° attitude turn maneuver of a spin stabilized upper stage: the spacecraft requires more than two burn maneuvers to reverse the thrust direction between the perigee and the apogee of a transfer ellipse.

6.2.3 Stability of a Spinning Satellite

Rotations on the principal axis of inertia. We have seen that the rotational motion of a satellite with symmetric moment of inertia can be interpreted as the rolling of the body cone on the space cone. When the satellite is a perfect rigid-body, the motion is stable. The word “stability” has many interpretations; here it means that the cone angles remain constant in time. The nutation amplitude will not grow.

Generally, the rotational motion of a torque-free rigid body is stable when the body rotates on the principal axis with the maximum moment of inertia, or when it rotates on the principal axis with the minimum moment of inertia. When the body rotates on the axis with the intermediate moment of inertia, the motion is unstable. Now we consider the general rigid body ($\mathcal{A} \neq \mathcal{B} \neq \mathcal{C}$) that rotates predominantly on the 1-axis of its principal body frame of reference. Equation (6-44) determines the motion of the rotation vector $\vec{\omega}$ (with $\vec{M} = 0$). We can observe that the first component of this vector ω_1 is not constant anymore, since now $\mathcal{B} \neq \mathcal{C}$. However, component ω_1 is still approximately constant when the components ω_2 and ω_3 are much smaller than ω_1 :

$$\begin{aligned} 0 &= \mathcal{A} \dot{\omega}_1 + (\mathcal{C} - \mathcal{B}) \omega_2 \omega_3 \quad , \quad \rightarrow \quad \omega_1 \approx \text{constant} \\ 0 &= \mathcal{B} \dot{\omega}_2 + (\mathcal{A} - \mathcal{C}) \omega_1 \omega_3 \\ 0 &= \mathcal{C} \dot{\omega}_3 + (\mathcal{B} - \mathcal{A}) \omega_1 \omega_2 \end{aligned} \quad (6 - 51)$$

The second and the third equation of the system above describe a linear oscillation, where the stability of the oscillation is determined by the “characteristic equation”. We find the characteristic equation of this differential equation system when we insert the solution $\omega_2 = C_1 e^{\lambda t}$, $\omega_3 = C_2 e^{\lambda t}$, and eliminate from the equations the amplitude coefficients C_1 and C_2 . Multiplier λ is the “characteristic value” of the problem:

$$\lambda^2 + \frac{\mathcal{A} - \mathcal{C}}{\mathcal{B}} \omega_1 \cdot \frac{\mathcal{A} - \mathcal{B}}{\mathcal{C}} \omega_1 = 0 \quad (6 - 52)$$

$$\lambda_{1,2} = \pm \sqrt{-\frac{(\mathcal{A} - \mathcal{C}) \cdot (\mathcal{A} - \mathcal{B})}{\mathcal{B}\mathcal{C}}} \omega_1^2 \quad (6 - 53)$$

The roots of the characteristic equation may not have positive real parts, otherwise the oscillation is unstable (an imaginary part corresponds to a periodic oscillation, a negative real part corresponds to an exponentially declining amplitude). We can observe that the motion is unstable when \mathcal{A} is intermediate in value to \mathcal{B} and \mathcal{C} . Then the positive real root of the characteristic equation indicates that a motion with exponentially growing amplitude is present. The oscillation is stable when \mathcal{A} is bigger than \mathcal{B} and bigger than \mathcal{C} (when the body rotates on the principal axis of the maximum moment of inertia). The motion seems also to be stable when the body rotates on the principal axis of its minimum moment of inertia (when \mathcal{A} is smaller than \mathcal{B} and smaller than \mathcal{C}). In satellite dynamics, however, the rotation of a satellite on the axis of the minimum moment of inertia is just stable for a short time interval. Unfortunately, for extended time periods of a day or longer, the motion is unstable.

Dissipation of energy. Energy dissipation is responsible for the instability of rotations on the axis of the minimum moment of inertia. Equation (6-40) determines the mechanical energy of a rigid body. When we use a principal body frame of reference and consider just the rotational part of the energy, we can write down the equation as:

$$2E = \mathcal{A} \omega_1^2 + \mathcal{B} \omega_2^2 + \mathcal{C} \omega_3^2 \tag{6 - 54}$$

Expression E is the entire rotational energy of the rigid body as a function of the rotation velocity vector $\vec{\omega}$. The notation resembles the equation for the square of the length of the angular momentum vector H^2 . Using equation (6-23), we can denote:

$$H^2 = \mathcal{A}^2 \omega_1^2 + \mathcal{B}^2 \omega_2^2 + \mathcal{C}^2 \omega_3^2 \tag{6 - 55}$$

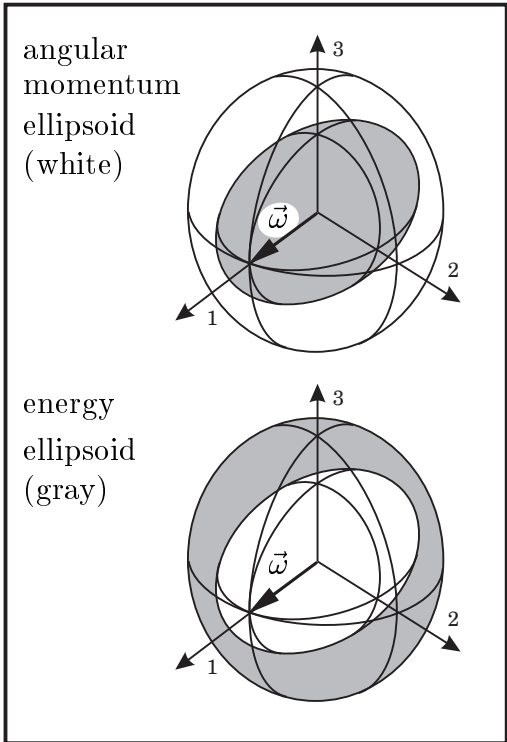
Both, energy E and angular momentum \vec{H} , are constant for a torque-free rigid body. The expressions for $2E$ and H^2 are similar but not identical functions of vector $\vec{\omega}$. They define different ellipsoids in the body frame of reference. We divide these equations by their left sides to find “energy ellipsoid” and “angular momentum ellipsoid”:

energy ellipsoid:
$$\frac{\mathcal{A}}{2E} \omega_1^2 + \frac{\mathcal{B}}{2E} \omega_2^2 + \frac{\mathcal{C}}{2E} \omega_3^2 = 1 \tag{6 - 56}$$

angular momentum ellipsoid:
$$\frac{\mathcal{A}^2}{H^2} \omega_1^2 + \frac{\mathcal{B}^2}{H^2} \omega_2^2 + \frac{\mathcal{C}^2}{H^2} \omega_3^2 = 1 \tag{6 - 57}$$

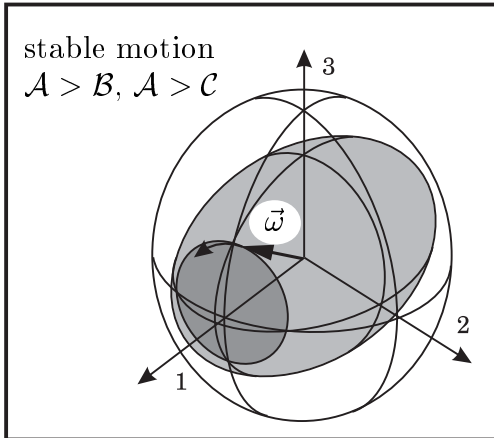
Note that the energy ellipsoid has axes of the length: $\sqrt{2E/\mathcal{A}}$, $\sqrt{2E/\mathcal{B}}$, $\sqrt{2E/\mathcal{C}}$; the angular momentum ellipsoid has axes of the length: H/\mathcal{A} , H/\mathcal{B} , H/\mathcal{C} .

The intersection line of the ellipsoids is a three-dimensional figure called “polhode”. $\vec{\omega}$ is a moving vector in the body frame of reference, its tip lies always on this curve.

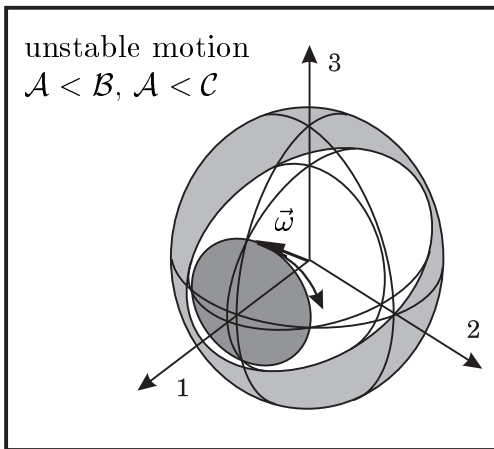


Let us first consider that \mathcal{A} is the maximum moment of inertia, and that the rigid body rotates exclusively on the 1-axis ($\omega_2 = 0$, $\omega_3 = 0$). We find that $2E = H^2/\mathcal{A}$, and the polhode is a single point on the 1-axis. The energy ellipsoid is smaller than the angular momentum ellipsoid, since $2E/\mathcal{B}$ is smaller than H^2/\mathcal{B}^2 , and $2E/\mathcal{C}$ is smaller than H^2/\mathcal{C}^2 . Then the energy ellipsoid lies entirely inside the angular momentum ellipsoid, but the two ellipsoids touch each other at the point of the polhode. Let us now consider that \mathcal{A} is the minimum moment of inertia. The two ellipsoids still touch each other at the point of the polhode, but now the energy ellipsoid is bigger than the angular momentum ellipsoid (since $2E/\mathcal{B} > H^2/\mathcal{B}^2$ and $2E/\mathcal{C} > H^2/\mathcal{C}^2$). In this case the energy ellipsoid lies entirely outside the angular momentum ellipsoid.

A torque-free rigid body cannot dissipate mechanical energy. However, no satellite is a perfectly rigid body, every satellite can dissipate mechanical energy (usually the satellite has flexible antennae, tanks with liquid propellant and so on). The satellite is a “semi-rigid body”, and the mechanical energy E is a declining function of time t .



When the satellite rotates on the 1-axis and \mathcal{A} is the maximum moment of inertia, the rotation is stable: dissipation of mechanical energy damps out any nutational motion. The energy ellipsoid shrinks while the angular momentum ellipsoid conserves its size. The polhode for the nutational oscillation is a curve which spirals slowly towards the 1-axis. Finally, when the polhode reaches the 1-axis, the satellite cannot dissipate energy anymore. The oblate satellite rotates stable.



However, when \mathcal{A} is the minimum moment of inertia, the motion is unstable. Energy dissipation excites a nutational motion with growing amplitude. Again, the energy ellipsoid shrinks slowly while the angular momentum ellipsoid conserves its size; but this time the polhode is a curve which spirals slowly away from the 1-axis. Even though the angular momentum vector of the motion is not altering its orientation and its length, the satellite tumbles wildly after some time. The prolate satellite rotates unstable.

6.2.4 The Single-Spin Satellite

The semi-rigid spacecraft. As we have seen, the rotation of a torque-free satellite on the axis of the maximum moment of inertia is always stable; but the rotation on the axis of the minimum moment of inertia is just stable for a short time period. Rotations on the intermediate axis of inertia are always unstable. Now we return to the axisymmetric satellite ($\mathcal{A} \neq \mathcal{B} = \mathcal{C}$). Equation (6-47) determines the nutational motion when the satellite rotates on the 1-axis of its body frame of reference. The motion conserves in any case the direction and length of the angular momentum \vec{H} . Only if the satellite were a perfectly rigid body the motion would also conserve the mechanical energy E ; but the satellite is never a perfectly rigid body, it is always a “semi-rigid” body. The energy of the rotation is a never increasing function of time.

When the mechanical energy E of the rotational motion is a well-known time function, then the growing (or shrinking) of the amplitude of the nutation is determined by:

$$2E(t) = \mathcal{A} \omega_1^2 + \mathcal{B} (\omega_2^2 + \omega_3^2) \quad (6-58)$$

$$H^2 = \mathcal{A}^2 \omega_1^2 + \mathcal{B}^2 (\omega_2^2 + \omega_3^2) \quad (6-59)$$

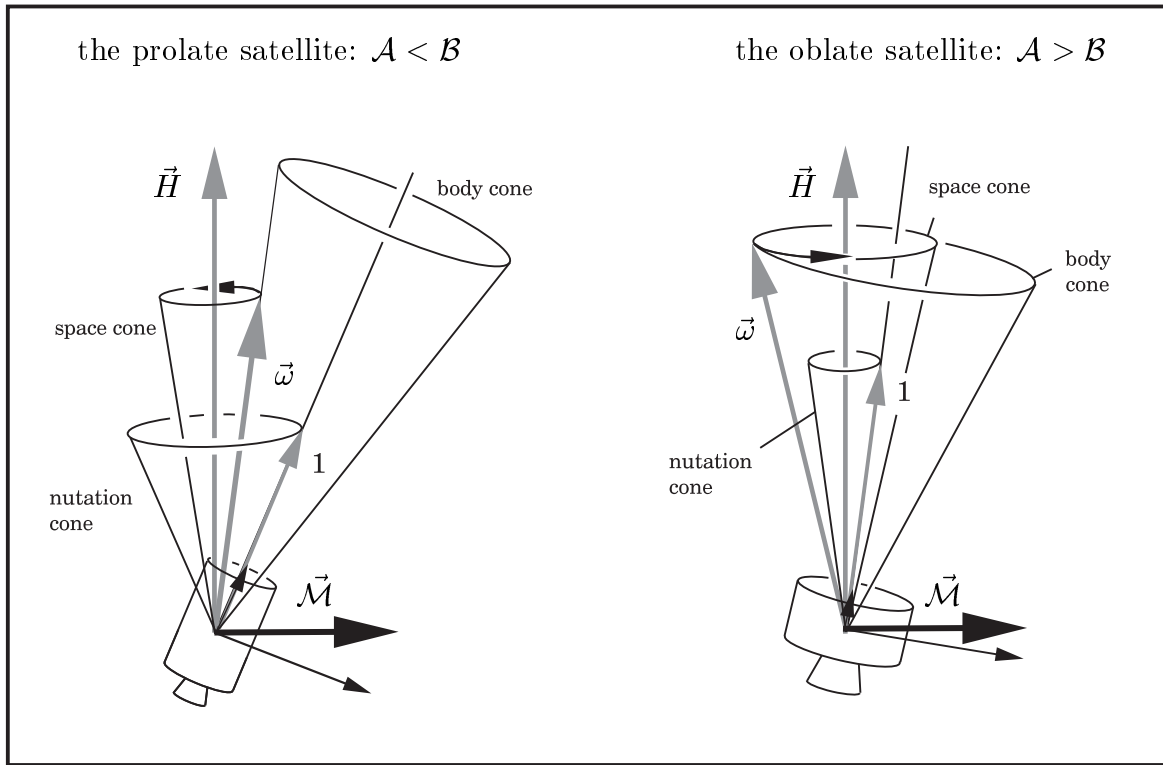
Expression $\omega_0 = \sqrt{\omega_2^2 + \omega_3^2}$ is the amplitude of the nutation; ω_1 is the component of the rotation vector that is aligned with the 1-axis of the body frame of reference. In case of a stable motion, ω_1 grows while ω_0 shrinks. The square of the length of the angular momentum vector (term H^2) is always constant. The equations (6-58) and (6-59) determine the behaviour of the nutation amplitude as function of time entirely.

The angular momentum law in form of the dynamical Euler equations (system 6-44) cannot be applied to calculate the stability of the motion of semi-rigid spacecraft. The equation of condition for the angular momentum vector ($\vec{H} = (\mathcal{A} \omega_1, \mathcal{B} \omega_2, \mathcal{B} \omega_3)$) is based on the assumption that the satellite is a perfectly rigid body; and this assumption is not valid any more for a semi-rigid spacecraft. If we used system (6-44), we would just find out that a rotating torque-free rigid body does not dissipate energy. Strictly speaking, also the definition for the energy (equation 6-40) uses the assumption of a rigid body. It is not obvious that the energy-sink approach is correct.

Nutation dampers. When we want to treat the satellite as a rigid body and apply the angular momentum law in form of equation (6-44), we must use another approach. Now we consider the loss of mechanical energy as the effect of nutation dampers. Nutation dampers absorb mechanical energy but conserve the angular momentum. The damper masses exert on the satellite the external damper moment $\vec{\mathcal{M}}_{damper}$. The point in question is orientation and magnitude of the damper moment vector.

$$\begin{aligned} \vec{\mathcal{M}}_{damper} &= \begin{pmatrix} \mathcal{A} \dot{\omega}_1 \\ \mathcal{B} \dot{\omega}_2 + (\mathcal{A} - \mathcal{B}) \omega_1 \omega_3 \\ \mathcal{B} \dot{\omega}_3 + (\mathcal{B} - \mathcal{A}) \omega_1 \omega_2 \end{pmatrix} \\ &= C_{damper} \cdot \begin{pmatrix} \mathcal{B} (\omega_2^2 + \omega_3^2) / (\mathcal{A} \omega_1) \\ -\omega_2 \\ -\omega_3 \end{pmatrix} \end{aligned} \quad (6-60)$$

You can readily verify that the model above complies with the following conditions: To ensure that the damper moment does not change the length of angular momentum, the damper moment must always be rectangular to the angular momentum vector (the vector dot product vanishes, $\vec{\mathcal{M}}_{damper} \cdot \vec{H} = 0$). Additionally, to ensure that the damper moment does not change the orientation of angular momentum vector, the damper moment must rotate ($\vec{\mathcal{M}}_{damper}$ is always in the same plane as $\vec{\omega}$ and \vec{H}). Finally, to ensure that the motion is stable, the damper moment lets ω_1^2 grow (it has a positive component in the body-fixed 1-direction). Term C_{damper} is a factor that provides information on how much time it takes to dissipate the mechanical energy.



The effect of the damper moment is that the angular momentum vector \vec{H} of the satellite executes a coning motion in inertial space (the cone angle is very small since the damper moment is small). On the long-term, however, the angular momentum vector of the satellite keeps its length and its direction. The reaction moment has a similar effect on the damper masses (the entire angular momentum of satellite and damper masses is a vector which is exactly constant with respect to inertial space); but the damper moment changes the rotation velocity vector and by this also the mechanical energy of the satellite:

$$\dot{E} = \vec{\omega} \cdot \vec{M}_{damper} = C_{damper}(\omega_2^2 + \omega_3^2) \cdot (\mathcal{B}/\mathcal{A} - 1) \quad (6 - 61)$$

Important is the sign of \dot{E} . When $\mathcal{A} < \mathcal{B}$ (the prolate satellite), $\dot{E} > 0$ and the satellite needs active nutation dampers. When $\mathcal{A} > \mathcal{B}$ (the oblate satellite), then $\dot{E} < 0$ and a satellite with passive nutation dampers rotates stable. We can observe that a satellite with a spherical ellipsoid of inertia ($\mathcal{A} = \mathcal{B}$) cannot dissipate energy. In this case the rotation vector $\vec{\omega}$ is always parallel with the angular momentum vector \vec{H} , and the satellite rotates stable without nutational oscillation (compare equation 6-44). Erroneously we might follow that such a construction is advisable; but in practice it is impossible to build a satellite with exactly spherical ellipsoid; the satellite with a nearly spherical ellipsoid of inertia would lose mechanical energy. The consequence is that the spherical satellite would not rotate about a defined axis (it would soon start tumbling wildly). We can conclude that the correct construction of single-spin satellite is an oblate rotor with a defined rotation axis.

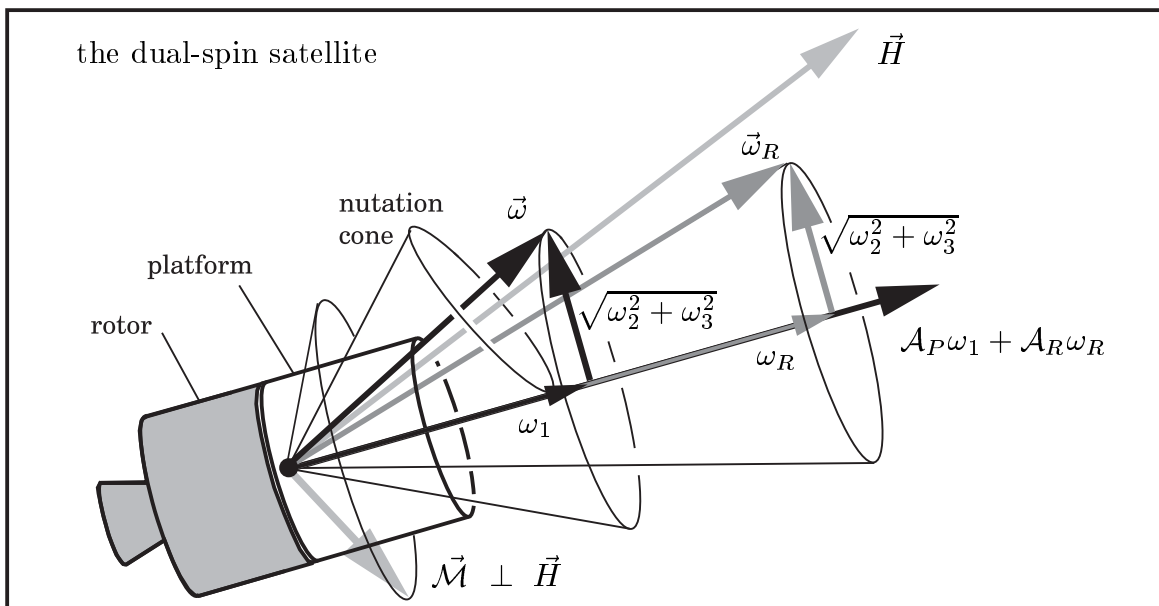
Time interval of stable rotation. Finally, it is interesting how long it takes to damp down a nutational motion, or how long a prolate spacecraft is stable. When we assume that the energy dissipation rate is proportional to the amplitude of the nutation, then the factor C_{damper} in equation (6-60) is constant in time. The vector equation (6-60) can be used to find a relationship that determines the nutation amplitude as a function of time. We can multiply the second component of this vector equation by ω_2 and the third component by ω_3 and add up the results, considering that $\omega_0^2 = \omega_2^2 + \omega_3^2$ and $d\omega_0^2/dt = 2\omega_2\dot{\omega}_2 + 2\omega_3\dot{\omega}_3$. Then we get:

$$\frac{d}{dt}\omega_0^2 = -\frac{2C_{damper}}{\mathcal{B}}\omega_0^2 \quad , \quad \text{or:} \quad \omega_0(t) = \omega_0(t_0) \cdot e^{-C_{damper} \cdot t/\mathcal{B}} \quad (6-62)$$

For an oblate satellite the amplitude of the nutation $\omega_0 = \sqrt{\omega_2^2 + \omega_3^2}$ is an exponentially declining function of time t . For a prolate satellite without active nutation dampers, however, C_{damper} is a negative constant. Any nutation amplitude ω_0 at the initial instant t_0 will grow exponentially. The time interval of “quasi stable motion” depends on the quality of the initial adjustment and on the rigidity of the vehicle (the value of C_{damper}). A solid upper stage can be kept stable for the duration of the mission, but a satellite with flexible antennae will tumble after less than a day.

6.2.5 The Dual-Spin Satellite

Shape of the satellite and the payload bay. As we have seen, stability of the rotational motion requires that the satellite is an oblate axisymmetric rotor. Often this requirement does not meet the dimensions of the payload bay of conventional satellite launchers, particularly in the case of a big satellite. To fit in the payload shroud, the satellite should usually be a prolate rotor. The solution to this problem is the construction of a so-called “dual-spin” satellite. Under certain conditions a dual-spin satellite can rotate stable on its axis, even when it is a prolate construction.



The angular momentum vector of a dual-spin satellite. To analyze the rotational behaviour of a dual-spin satellite, we use the “damper moment approach”: we consider that the spacecraft consists of two perfectly rigid bodies with axisymmetric ellipsoid of inertia, where nutation dampers are responsible for the loss of mechanical energy. One part of the satellite is called “platform” (index P), the other one is called “rotor” (index R). For the derivation of the equations of the motion we will use the body-fixed frame of reference of the platform: the platform rotates with ω_1 on the 1-axis, the rotor rotates with ω_R exactly on the same axis. We will assume that the rotor rotates faster than the platform ($\omega_R > \omega_1$), and that both components of the rotation are positive. Then the entire angular momentum \vec{H} is composed of the angular momentum of the platform and the angular momentum of the rotor:

$$\vec{H} = \begin{pmatrix} \mathcal{A}_P \omega_1 \\ \mathcal{B}_P \omega_2 \\ \mathcal{B}_P \omega_3 \end{pmatrix} + \begin{pmatrix} \mathcal{A}_R \omega_R \\ \mathcal{B}_R \omega_2 \\ \mathcal{B}_R \omega_3 \end{pmatrix} = \begin{pmatrix} \mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R \\ \mathcal{B} \omega_2 \\ \mathcal{B} \omega_3 \end{pmatrix} \quad (6-63)$$

The motions of platform and rotor in 2-direction and 3-direction are always combined. We may use their combined moment of inertia \mathcal{B} to describe the nutational oscillation. According to equation (6-60) for the single-spin satellite, the damper moment is:

$$\begin{pmatrix} \mathcal{M}_1 \\ \mathcal{M}_2 \\ \mathcal{M}_3 \end{pmatrix} = C_{damp} \begin{pmatrix} \frac{\mathcal{B} (\omega_2^2 + \omega_3^2)}{\mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R} \\ -\omega_2 \\ -\omega_3 \end{pmatrix} = \begin{pmatrix} \mathcal{A}_P \dot{\omega}_1 + \mathcal{A}_R \dot{\omega}_R \\ \mathcal{B} \dot{\omega}_2 + (\mathcal{A} - \mathcal{B}) \omega_1 \omega_3 + \mathcal{A}_R \omega_3 \omega_R \\ \mathcal{B} \dot{\omega}_3 + (\mathcal{B} - \mathcal{A}) \omega_1 \omega_2 - \mathcal{A}_R \omega_2 \omega_R \end{pmatrix} \quad (6-64)$$

The damper moment vector $\vec{\mathcal{M}}$ is always rectangular to the angular momentum \vec{H} ; and it is in the same plane as the rotation vector $\vec{\omega}$ and \vec{H} . The vector equation (6-64) describes the nutational motion of the dual-spin satellite, where a vector notation of the body frame of reference of the platform is used. Vector $\vec{\omega}$ determines the rotation velocity of the coordinate system. We have to use relationship (6-42) for the differentiation of the angular momentum vector, since the dynamical Euler equations (in the form of equation system 6-44) do not apply for the motion of the rotor.

Excited by the damper moment, the angular momentum \vec{H} performs a coning motion with small cone angle and short rotation period; but the length and the inertial direction of \vec{H} are kept constant when a long time interval is under consideration.

$$H^2 = (\mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R)^2 + \mathcal{B}^2 (\omega_2^2 + \omega_3^2) \quad (6-65)$$

$$C_{damp} = \frac{\frac{d}{dt}(\mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R)^2}{2 \mathcal{B} (\omega_2^2 + \omega_3^2)} \quad (6-66)$$

Term H^2 is constant in a long-time treatment, and we can conclude that for stability the expression $(\omega_2^2 + \omega_3^2)$ must decrease while the expression $(\mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R)^2$ grows. Transformed into equation (6-66), the first component equation of (6-64) shows that the motion is stable when the damping factor C_{damp} assumes a positive value.

Conditions for the stability of a dual-spin satellite. Stability analysis is more complex for a dual-spin satellite than for a single-spin satellite. It is not only important which component (platform or rotor) dissipates the energy, but also how fast the energy is lost. An internal moment between platform and rotor (friction in the bearings or an actuator motor) can exchange angular momentum between the two components of the satellite, but cannot change the angular momentum \vec{H} . The component of this vector in 1-direction (term $\mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R$) can only be changed by the first component of the damper moment \mathcal{M}_1 . We can write:

$$\begin{aligned}\mathcal{M}_1 &= \mathcal{A}_P \dot{\omega}_1 + \mathcal{A}_R \dot{\omega}_R = \mathcal{M}_P + \mathcal{M}_R \\ \mathcal{M}_P &= \mathcal{A}_P \dot{\omega}_1 = X \cdot \mathcal{M}_1\end{aligned}\quad (6-67)$$

$$\mathcal{M}_R = \mathcal{A}_R \dot{\omega}_R = (1 - X) \cdot \mathcal{M}_1 \quad (6-68)$$

We use factor X to split the damper moment into a part \mathcal{M}_P that acts on the platform and another part \mathcal{M}_R that acts on the rotor; X can assume the following values:

- $X < 0$ a motor moment accelerates the rotor and decelerates the platform;
- $X = 0$ the platform is not damped, or a motor keeps ω_1 constant;
- $0 < X < 1$ both parts of the satellite use nutation dampers, platform and rotor;
- $X = 1$ just the platform dissipates energy, the rotor is a rigid body;
- $X > 1$ bearing friction brakes the rotor and accelerates the platform.

The mechanical energy of the dual-spin satellite is given by the following equation:

$$2E = \mathcal{A}_P \omega_1^2 + \mathcal{A}_R \omega_R^2 + \mathcal{B} (\omega_2^2 + \omega_3^2) \quad (6-69)$$

For stability E must be a declining function of time, otherwise the satellite needs active nutation dampers. The time derivative of the energy \dot{E} is determined by:

$$\begin{aligned}\dot{E} &= \omega_1 \mathcal{M}_P + \omega_R \mathcal{M}_R + \omega_2 \mathcal{M}_2 + \omega_3 \mathcal{M}_3 \\ &= C_{damper} \cdot (\omega_2^2 + \omega_3^2) \cdot \left[\frac{\mathcal{B} (X \omega_1 + (1 - X) \omega_R)}{\mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R} - 1 \right]\end{aligned}\quad (6-70)$$

For stability with passive nutation dampers it is required that the expression in square brackets is smaller than zero, thus: $(\mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R) > \mathcal{B} \cdot (X \omega_1 + (1 - X) \omega_R)$.

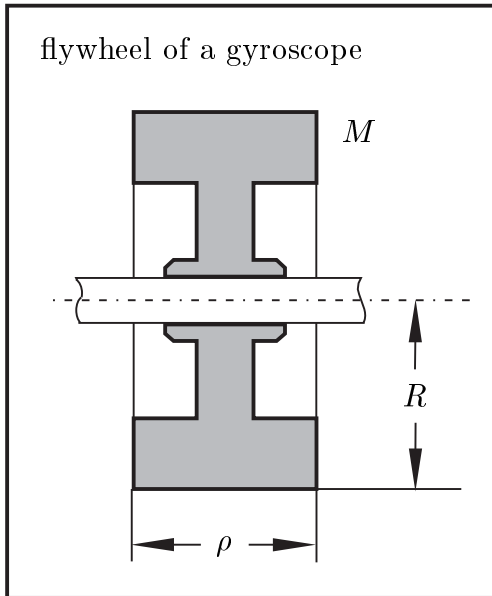
A satellite with $\mathcal{A}_P + \mathcal{A}_R > \mathcal{B}$ (the oblate dual-spin satellite) is probably stable. More important in practice is the prolate dual-spin satellite, where $\mathcal{A}_P + \mathcal{A}_R < \mathcal{B}$. When the nutation dampers are mounted on the platform, $X \approx 1$ and $\omega_R \approx \text{constant}$. It follows as the criterion for passive stability that $(\mathcal{A}_P \omega_1 + \mathcal{A}_R \omega_R) > (\mathcal{B} \omega_1)$. The satellite will be stable if the rotor is spinning much faster than the platform ($\omega_R \gg \omega_1$). If the nutation dampers were mounted on the rotor ($X \approx 0$), the satellite would be unstable. For passive stability of a prolate dual-spin satellite it is required that the rapidly spinning rotor dissipates the energy at a lower rate than the slowly spinning platform. The rotor of the dual-spin satellite should be a rigid body, and the platform should be equipped with nutation dampers.

6.3. Three-Axis Stabilization

When it is not desired that the satellite rotates, it needs another stabilization method. The attitude control and stabilization system of the satellite must be able to generate appropriate external control moments. In space, aerodynamic fins do nearly not work anymore; and the gravity force is not available on board of a space vehicle. To generate control moments, orbiting satellites use sometimes the interaction of the earth's magnetic field with on-board electro-magnets (so-called "magnetorquers"); but the usual method to exert external moments on the satellite is use of small thrusters, which involve the disadvantage of propellant consumption. To avoid that propellant is constantly expended while the control thrusters stabilize the attitude of a satellite, the control system uses additionally internal moments generated by fast rotating gyros (momentum wheels). These wheels reduce the demand for attitude stabilization propellant considerably (but they cannot eliminate it entirely). The method of attitude control by momentum wheels is called "three axis stabilization".

6.3.1 The Gyroscope Wheel

Moment of inertia. Momentum wheels in gyroscopic instruments are fast-rotating symmetrical disks. These wheels behave surprisingly: when we pivot the spin axis, the wheel reacts with a moment (or motion) in an unexpected direction.



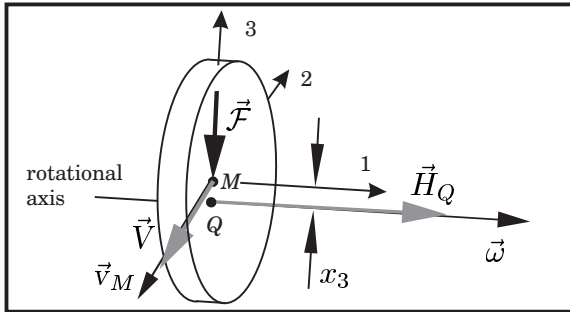
We can describe the power of inertia of a flywheel using a matrix notation of a body-fixed principal axis coordinate system (6-23). Assume that the flywheel has the mass M and the shape of a cylindrical disk, with the radius R and the thickness ρ . Then \mathcal{A} is the moment of inertia for rotations on the 1-axis (the axis of symmetry), and \mathcal{B} is the moment of inertia for rotations on the 2-axis and on the 3-axis. These terms can be written as:

$$\mathcal{A} = \frac{MR^2}{2} \quad (6-71)$$

$$\mathcal{B} = \frac{M}{4} \left(R^2 + \frac{\rho^2}{3} \right)$$

Mass elements which are distant from the spin axis contribute more to the moment of inertia than mass elements which are near to the spin axis. When weight saving is important, we will try to remove mass from the inner part of the wheel (for example, the white regions in the figure above). Then we have to correct the equations (6-71) appropriately (when removed parts of the wheel have also the shape of a cylindrical disk, we may use the same law and subtract the rotational inertia).

Balancing of wheels. Before we analyze the complicated behaviour of gyroscopic wheels in detail we have to study their shape and mass distribution. A fast rotating gyro wheel should not exert periodic forces or moments on the bearings. We say that a wheel is in an unbalanced state when its rotation about a fixed axis causes unwanted vibrations. A wheel is “statically unbalanced” when its center of mass is not located on the rotation axis. Consider the case that the wheel (mass M) rotates on an axis parallel to the 1-direction of the principal axis body frame of reference. The displacement is x_3 , and the moment of inertia \mathfrak{I} is bigger for rotations about Q than it is for rotations about the center of mass M . We can use the equations (6-5) and (6-37) to calculate the linear momentum \vec{V} and the angular momentum vector \vec{H} :

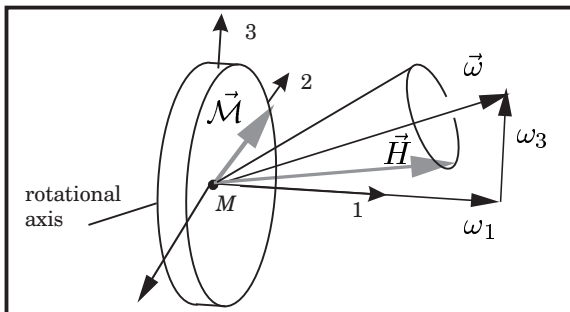


$$\vec{V} = M\vec{v}_M = \begin{pmatrix} 0 \\ M\omega_1 x_3 \\ 0 \end{pmatrix} \quad (6-72)$$

$$\vec{H}_Q = \mathfrak{I}_Q \cdot \vec{\omega} = \begin{pmatrix} (\mathcal{A} + Mx_3^2) \omega_1 \\ 0 \\ 0 \end{pmatrix}$$

Both, \vec{V} and \vec{H} , are constant vectors with respect to the body frame of reference; but while the angular momentum \vec{H} does not change with respect to inertial space, the linear momentum \vec{V} rotates. The change of the linear momentum is equivalent to a periodical bearing force \vec{F} that acts on the wheel. The wheel vibrates, it does not run concentrically. To avoid these oscillations, we have to “statically balance” the wheel and observe that the center of mass is located exactly on the rotation axis.

A wheel is “dynamically unbalanced” when the rotation axis is not parallel with the direction of a principal axis of its body frame of reference. Now we consider the case that the wheel rotates about the center of mass M , but the fixed rotation axis is inclined with respect to the 1-direction ($\vec{\omega} = (\omega_1, 0, \omega_3) = \text{constant}$, with $\omega_3 \ll \omega_1$):



$$\vec{H}_M = \mathfrak{I}_M \cdot \vec{\omega} = \begin{pmatrix} \mathcal{A} \omega_1 \\ 0 \\ \mathcal{C} \omega_3 \end{pmatrix} \quad (6-73)$$

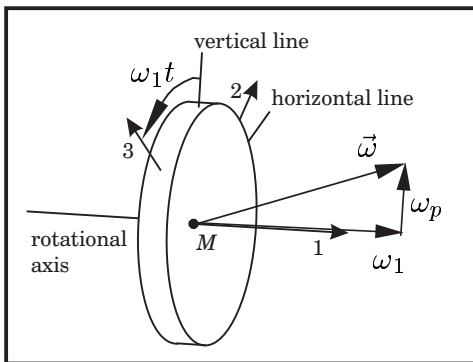
$$\vec{M}_M = \vec{\omega} \times \vec{H}_M = \begin{pmatrix} 0 \\ (\mathcal{A} - \mathcal{C}) \omega_1 \omega_3 \\ 0 \end{pmatrix}$$

The angular momentum \vec{H} is still a constant vector in the body frame of reference. However, \vec{H} changes with respect to inertial space. We may conclude that the wheel requires for the motion the moment \vec{M} (exerted from the bearings on the wheel). When $\mathcal{A} > \mathcal{C}$, the bearing moment \vec{M} aims into the positive 2-direction of the body frame of reference. The moment \vec{M} is zero when $\mathcal{A} = \mathcal{C}$ and negative when $\mathcal{A} < \mathcal{C}$.

The case $\mathcal{A} > \mathcal{C}$ corresponds to an oblate wheel (a disk), where the main share of the mass is located near the 2-3 plane. The case $\mathcal{A} < \mathcal{C}$, however, corresponds to a prolate wheel (a stick), where the main share of the mass is located near the 1-axis. Centrifugal forces act on all mass elements of the rotating wheel. These centrifugal forces cause a moment on the wheel when the spin axis is not exactly parallel with the 1-direction of the body frame of reference. Due to shaft flexibility and bearing slackness we can assume that the spin axis is never perfectly aligned with the 1-direction. Then the oblate rotor with $\mathcal{A} > \mathcal{C}$ is better conditioned for an application as gyro wheel than the prolate rotor: the centrifugal forces act in a way that they try to adjust the 1-direction with the spin axis. In case of a prolate rotor the centrifugal forces try to deviate the 1-direction, and the rotor runs unsteady causing vibrations.

The symmetrical gyroscopic wheel. We talk about a symmetrical rotor when two of its principal moments of inertia are the same, for example $\mathcal{A} \neq \mathcal{B} = \mathcal{C}$. This is the case when the body has a rotationally symmetrical mass distribution. Flywheels of gyroscopic instruments are nearly always oblate symmetrical rotors.

To understand why symmetrical flywheels have better true running qualities we consider the case of an asymmetrical rotor, with $\mathcal{A} \neq \mathcal{B} \neq \mathcal{C}$. The rotor rotates on its body-fixed 1-axis; additionally we pivot the rotation axis in a way that the rotor executes a precessional movement in the horizontal plane. The rotation velocity ω_1 of the rotor is a constant vector in the body frame of reference, while the precession velocity ω_p is a constant vector with respect to inertial space. Term t is the time coordinate; vector ω_p rotates with the angle $(\omega_1 t)$ in the body frame of reference. There the rotation vector takes the form of $\vec{\omega} = (\omega_1, \omega_p \sin(\omega_1 t), \omega_p \cos(\omega_1 t))$. Thus:



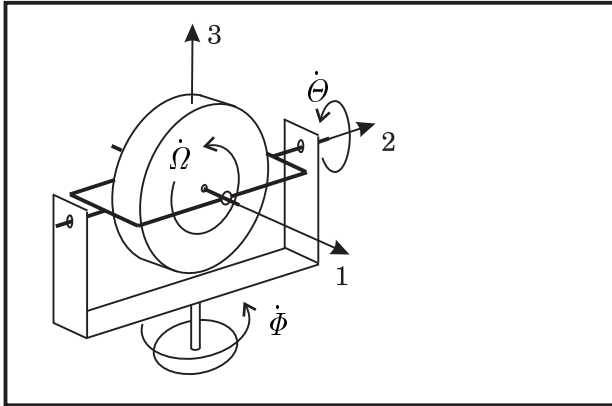
$$\vec{H} = \mathfrak{S} \cdot \vec{\omega} = \begin{pmatrix} \mathcal{A} \omega_1 \\ \mathcal{B} \omega_p \sin(\omega_1 t) \\ \mathcal{C} \omega_p \cos(\omega_1 t) \end{pmatrix} \quad (6-74)$$

$$\vec{\mathcal{M}} = \begin{pmatrix} (\mathcal{C} - \mathcal{B}) \omega_p^2 \sin(\omega_1 t) \cos(\omega_1 t) \\ (+\mathcal{A} - (\mathcal{C} - \mathcal{B})) \omega_p \omega_1 \cos(\omega_1 t) \\ (-\mathcal{A} - (\mathcal{C} - \mathcal{B})) \omega_p \omega_1 \sin(\omega_1 t) \end{pmatrix}$$

To find the bearing moment $\vec{\mathcal{M}}$, we have to differentiate the angular momentum vector \vec{H} with respect to inertial space. Since \vec{H} is not a constant vector in the body frame of reference, we have to consider both, relative changes and changes which come from the rotation of the coordinate system. The dynamical Euler equations (relationships 6-41) apply directly to the problem. We can observe that the notation of the moment $\vec{\mathcal{M}}$ simplifies substantially when a symmetrical rotor is used ($\mathcal{B} = \mathcal{C}$). Only then the motion of the rotor generates a non-vibrating horizontal moment \mathcal{M} ($\mathcal{M} = \mathcal{A} \omega_p \omega_1$ aims always in a horizontal direction, rectangular to the spin axis).

6.3.2 The Gyro as Sensor

The gimbal frame of reference. The “body frame of reference” is appropriate to analyze the motion of rigid bodies. However, it involves the disadvantage that some vectors, which are actually constant with respect to the bearings, rotate rapidly in the body-fixed system (compare equation 6-74). We can avoid this problem when we use another coordinate system. We will call it “gimbal frame of reference” since it is fixed with respect to the gimbal mounting of the gyroscope. Like the principal axis body frame of reference, its 1-direction is always aligned with the spin axis of the symmetrical rotor ($\mathcal{A} \neq \mathcal{B} = \mathcal{C}$); but in contrast to the body-fixed system, the 2-direction does not rotate. The 2-direction of the gimbal frame of reference remains horizontally. The wheel can be pivoted about the vertical axis (Cardan angle Φ) and about the horizontal axis (Cardan angle Θ , by turning the gimbal); it rotates on the 1-axis (Cardan angle Ω). Thus, the gimbal frame of reference assumes always the same orientation as the wheel; the only exception is that it does not execute the rotation on the 1-axis. Now we have to consider two different angular velocities, vector $\vec{\omega}_{gyro}$ of the wheel and vector $\vec{\omega}_{system}$ of the gimbal frame of reference:



$$\vec{\omega}_{gyro} = \begin{pmatrix} \dot{\Phi} \sin \Theta + \dot{\Omega} \\ -\dot{\Theta} \\ \dot{\Phi} \cos \Theta \end{pmatrix} \quad (6-75)$$

$$\vec{\omega}_{system} = \begin{pmatrix} \dot{\Phi} \sin \Theta \\ -\dot{\Theta} \\ \dot{\Phi} \cos \Theta \end{pmatrix} \quad (6-76)$$

(compare equation 6-45)

The dynamical Euler equations (6-44) do not apply anymore, since $\vec{\omega}_{gyro}$ and $\vec{\omega}_{system}$ are different vectors. The angular momentum \vec{H} (equation 6-42) takes the form of:

$$\vec{H} = \mathfrak{S} \cdot \vec{\omega}_{gyro} = \begin{pmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{B} & 0 \\ 0 & 0 & \mathcal{B} \end{pmatrix} \cdot \begin{pmatrix} \dot{\Phi} \sin \Theta + \dot{\Omega} \\ -\dot{\Theta} \\ \dot{\Phi} \cos \Theta \end{pmatrix} \quad (6-77)$$

Advantageously, the symmetrical wheel with $\mathcal{C} = \mathcal{B}$ allows us to use the gimbal frame of reference, since here the matrix of inertia \mathfrak{S} is invariant when the wheel rotates on the spin axis (compare equation 6-33). To find the moment vector $\vec{\mathcal{M}}$, we have to apply relationship (6-41) and differentiate \vec{H} with respect to inertial space. Therefore:

$$\vec{\mathcal{M}} = \frac{d}{dt} \begin{pmatrix} \mathcal{A}(\dot{\Phi} \sin \Theta + \dot{\Omega}) \\ -\mathcal{B}\dot{\Theta} \\ \mathcal{B}\dot{\Phi} \cos \Theta \end{pmatrix} + \begin{pmatrix} \dot{\Phi} \sin \Theta \\ -\dot{\Theta} \\ \dot{\Phi} \cos \Theta \end{pmatrix} \times \begin{pmatrix} \mathcal{A}(\dot{\Phi} \sin \Theta + \dot{\Omega}) \\ -\mathcal{B}\dot{\Theta} \\ \mathcal{B}\dot{\Phi} \cos \Theta \end{pmatrix} \quad (6-78)$$

This time the components of $\vec{\mathcal{M}}$ refer to a notation of the gimbal frame of reference.

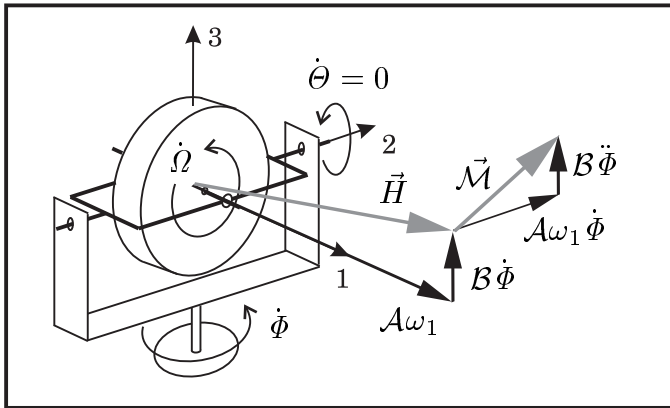
As final result we get the differential equations of motion for the symmetrical wheel:

$$\mathcal{M}_1 = \mathcal{A} \frac{d}{dt} (\dot{\Phi} \sin \Theta + \dot{\Omega}) \tag{6-79}$$

$$\mathcal{M}_2 = -\mathcal{B} \ddot{\Theta} + \mathcal{A} (\dot{\Phi} \sin \Theta + \dot{\Omega}) \dot{\Phi} \cos \Theta - \mathcal{B} \dot{\Phi}^2 \sin \Theta \cos \Theta \tag{6-80}$$

$$\mathcal{M}_3 = \mathcal{B} \ddot{\Phi} \cos \Theta - 2\mathcal{B} \dot{\Phi} \dot{\Theta} \sin \Theta + \mathcal{A} (\dot{\Phi} \sin \Theta + \dot{\Omega}) \dot{\Theta} \tag{6-81}$$

Moment of constant precession. The first case which we want to consider is the locked-up gimbal. The gimbal is not allowed to move with respect to the 2-axis, but it senses a moment $\vec{\mathcal{M}}$ (“output”) as function of the angular rate $\dot{\Phi}$ (“input”). Furthermore, we will assume that the spin rate $\omega_1 = \dot{\Omega}$ is constant. The notations for \vec{H} (equation 6-77) and for $\vec{\mathcal{M}}$ (equation 6-78) simplify considerably when Θ vanishes:



$$\vec{H} = \begin{pmatrix} \mathcal{A} \omega_1 \\ 0 \\ \mathcal{B} \dot{\Phi} \end{pmatrix} \tag{6-82}$$

$$\vec{\mathcal{M}} = \begin{pmatrix} 0 \\ \mathcal{A} \omega_1 \dot{\Phi} \\ \mathcal{B} \dot{\Phi} \end{pmatrix} \tag{6-83}$$

When the flywheel spins rapidly ($\omega_1 \gg \dot{\Phi}$), the input rate $\dot{\Phi}$ produces a proportional gyro moment ($\mathcal{M}_2 = \mathcal{A}\omega_1\dot{\Phi}$, exerted from the bearings on the rotor). The countervailing bearing moment acts in 2-direction, it prevents that the flywheel puts up its spin axis into upright position. The rotor tries to align its spin direction ω_1 with the direction of the forced motion ω_3 . It is remarkable that the input rate $\dot{\Phi}$ has no influence on the spin velocity ω_1 . Astonishingly, a constant rate $\dot{\Phi}$ requires a constant moment in 2-direction, but not a moment in 3-direction. Just when the input rate $\dot{\Phi}$ accelerates, a bearing moment appears also in 3-direction ($\mathcal{M}_3 = \mathcal{B}\ddot{\Phi}$).

The rate gyro. We can use the gyroscopic effect to build an instrument for measuring the angular rate $\dot{\Phi}$. The usual way to indicate the moment \mathcal{M}_2 is to use a spring-damper mechanism on the 2-axis. This allows rotational motions on the 2-axis, however, we will assume that the angle Θ remains small ($\sin \Theta \approx \Theta$, $\cos \Theta \approx 1$). We learn from equation (6-79) that the angular rate $\omega_1 = \dot{\Phi} \sin \Theta + \dot{\Omega}$ is constant (provided that no moment \mathcal{M}_1 accelerates or brakes the flywheel). When the flywheel spins rapidly, $\omega_1 \approx \dot{\Omega}$, and we can neglect term ($\dot{\Phi} \sin \Theta$) in equation (6-80):

$$\mathcal{M}_2 = C_{spring} \cdot \Theta + C_{damper} \cdot \dot{\Theta} = -\mathcal{B} \ddot{\Theta} + \mathcal{A} \omega_1 \dot{\Phi} \tag{6-84}$$

The bearing moment \mathcal{M}_2 is exerted by a spring (constant C_{spring}) proportional to the displacement Θ and by a damper (constant C_{damper}) proportional to the rate $\dot{\Theta}$.

Now our objective is to find appropriate values for the constants C_{spring} and C_{damper} . The spring influences the sensitivity of the rate gyro, the damper diminishes periodic oscillations. When the motion is excited by an instantaneous change of the rate $\dot{\Phi}$ (a step function or an exciting impulse), the response value Θ should reach its new equilibrium state as soon as possible (without overswinging or slow creep speed).

The angle Θ behaves like a harmonic oscillator, excited by the function $\mathcal{A} \omega_1 \dot{\Phi}$. We find the general solution to this type of differential equation (6-84) by the superposition of the solution to the homogenous form (without excitation function) with one particular solution to the inhomogenous form (with excitation function). The equilibrium state $\Theta_{equilibrium}$ is reached when $\dot{\Theta}$ and $\ddot{\Theta}$ vanish. Therefore:

$$C_{spring} \cdot \Theta_{equilibrium} = \mathcal{A} \omega_1 \dot{\Phi} \quad (6 - 85)$$

Mathematics shows that the solution to the homogenous form of equation (6-84) can accept three different forms: a periodic damped oscillation when the damping is low, and an exponentially declining curve when the damping is high. Both solutions are not appropriate for our purpose. What we want is the intermediate third form, a fast declining function without overswinging. For this case the damper constant is:

$$C_{damper} = \sqrt{2\mathcal{B} \cdot C_{spring}} \quad (6 - 86)$$

The problem is the correct adjustment of the constant C_{spring} . The gyro is sensitive when we use a soft spring, but then the output angle Θ exceeds the allowed limit for large input rates $\dot{\Phi}$ (the gyro is said to be “saturated”). When we use a strong spring, the instrument is less sensitive but can tolerate larger input rates.

The integrating rate gyro. As the next case we consider that spring and damper are removed from the bearings of the rate gyro ($\mathcal{M}_2 = 0$). Now the displacement Θ is not small any more, but $\omega_1 = (\dot{\Phi} \sin \Theta + \dot{\Omega})$ is still constant if \mathcal{M}_1 equals zero. We can observe that $\mathcal{M}_3 = \mathcal{A} \omega_1 \dot{\Theta} = constant$, $\dot{\Phi} = constant$ is a particular solution, describing a precession on the 2-axis with constant rate $\omega_{precession} = \mathcal{M}_3 / (\mathcal{A} \omega_1)$. Equation (6-80) serves to find the “output” rate $\dot{\Theta}$ as function of the “input” rate $\dot{\Phi}$, and equation (6-81) serves to find the moment \mathcal{M}_3 that is required for the motion. When $\dot{\Phi}$ is constant, we can readily integrate equation (6-80). We assume the initial state without any precessional movement ($\Theta = 0$, $\dot{\Theta} = 0$) and find:

$$\dot{\Theta} = \sqrt{2 (\mathcal{A}/\mathcal{B}) \omega_1 \dot{\Phi} \sin \Theta - \dot{\Phi}^2 \sin^2 \Theta} \quad (6 - 87)$$

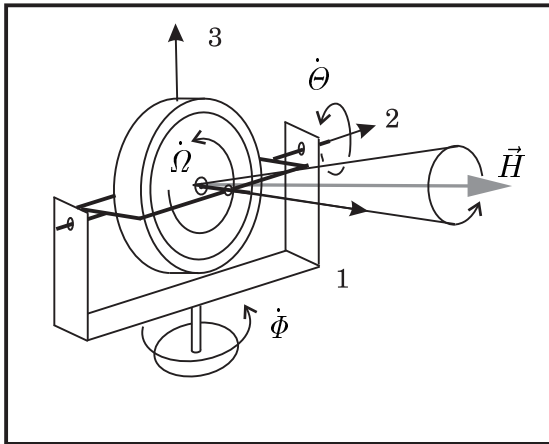
The result is a nonlinear rotation velocity $\dot{\Theta}$. When the flywheel rotates rapidly, the output rate $\dot{\Theta}$ can be much higher than the input rate $\dot{\Phi}$. Again, the rotor tries to align its spin direction ω_1 with the direction of the forced angular motion $\dot{\Phi}$. The instrument becomes a so-called “integrating rate gyro” when we consider the displacement angle Θ as the output. Angle Θ is now a direct function of angle Φ , provided that the gyro executes no precessional movement at the initial state. The range for measuring Φ is limited, since the gyro is saturated when $\Theta = \pm 90^\circ$. A damper on the 2-axis can reduce the sensitivity of the instrument.

The attitude reference gyro. Now we add a second (outer) gimbal to the gyro to enable it to rotate also about the vertical direction. We know that the gyro must conserve the length and direction of its actual angular momentum \vec{H} when no bearing moments \vec{M} act on the rotor. Equations (6-79), (6-80) and (6-81) confirm the trivial solution ($\dot{\Omega} = constant, \theta = constant, \dot{\Phi} = constant$) for a fast-rotating flywheel.

However, the gyro is also able to execute another movement that is called “nutation”. The preconditions for this motion are certain initial condition. When $\vec{M} = 0$, the rotation velocity vector ω_{gyro} accepts the following solution (compare equation 6-78):

$$\vec{\omega}_{gyro} = \begin{pmatrix} \dot{\Phi} \sin \theta + \dot{\Omega} \\ -\dot{\Theta} \\ \dot{\Phi} \cos \theta \end{pmatrix} = \begin{pmatrix} \omega_1 \\ -\dot{\Theta}_{max} \cdot \sin((\mathcal{A}/\mathcal{B} - 1)\omega_1 t + \Omega) \\ +\dot{\Theta}_{max} \cdot \cos((\mathcal{A}/\mathcal{B} - 1)\omega_1 t + \Omega) \end{pmatrix} \quad (6 - 88)$$

Term ω_1 is the constant component of the rotation vector in 1-direction; term $\dot{\Theta}_{max}$ is the amplitude of the oscillation (the maximum value of the angular velocity $\dot{\Theta}$). The nutation frequency of a fast-rotating wheel is approximately $\omega_{nutation} \approx \omega_1 \cdot \mathcal{A}/\mathcal{B}$ (assuming $\omega_1 t \approx \Omega$). The nutation velocity is proportional to the spin velocity ω_1 , in contrast to the precession velocity which is inversely proportional to ω_1 .



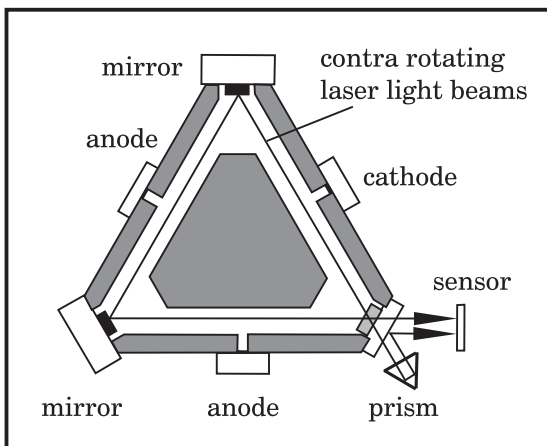
Now we consider the motion of the arrow head of the unit vector in the body-fixed 1-direction. Term $(-\dot{\Theta})$ is the velocity in 3-direction (“north-south”), term $(\dot{\Phi} \cos \theta)$ is the velocity in 2-direction (“east-west”). The velocity in 1-direction equals zero. The arrow head rotates about a certain axis fixed in inertial space, its motion is exactly a circle. The axis that is fixed in inertial space has the direction of the angular momentum vector \vec{H} .

In practice, the application of the double gimbaled gyro as a direction reference instrument involves serious problems. The direction of the angular momentum is not fixed with respect to inertial space, unfortunately, but it drifts slowly. First, in practice it is impossible to produce the gimbal bearings completely without any friction. A small rest bearing friction will cause the gyro to execute a precessional movement. Second, the center of mass of the rotor must be located exactly on the intersection point of the gimbal bearing axes, otherwise acceleration (or gravity) moments let the gyro drift. Third, also the mass of the gimbals cannot be neglected. When we consider the moments of inertia of the gimbals in the calculations, we find that now the nutation is not exactly a circular motion anymore. The reason for this drift effect is that, in the gimbal frame of reference, the moment of inertia of the outer gimbal, with respect to “east-west” movements $(\dot{\Phi} \cos \theta)$, depends on the actual “north” value θ . Even a very precise gyro can just be used as attitude reference for a short time period.

The inertial platform. The uncertain double-gimbaled attitude reference gyro is not used in astronautics. With precise rate gyros it is possible to measure on-board the components of the rotation velocity $\vec{\omega}$. The computer of the attitude control system finds the actual attitude of the vehicle (determined by Euler angles $(\psi, \vartheta, \varphi)$ or Cardan angles (Φ, Θ, Ω) alternatively) by the “real-time” numerical integration of the “kinematical Euler equations” (6-46). This method is called “strap-down inertial navigation”. The attitude must be well-known at the initial instant and has to be updated in certain time intervals. Since the output values of the rate gyros are never perfectly accurate, the attitude reference of the inertial navigation system will also drift. Update information is necessary particularly for long space missions (provided by “horizon crossing sensors”, “star trackers”, “sun acquisition sensors” and so on).

Strap-down navigation systems have now more or less replaced old-fashioned inertial platforms. The modern “strap-down” navigation system uses gyros which are fixed with respect to the vehicle; the attitude of inertial reference (“platform”) is calculated by a computer. The old-fashioned inertial navigation method uses real physical platforms. Such a gimbaled platform stabilizes its attitude by actuators in the gimbal bearings, using the input signals of precise rate gyros on the platform (the actuators control the attitude of the platform in a way that the angular velocities measured by the rate gyros are zero). The inertial platform is a complex technical instrument. Its only advantage, in comparison with modern strap-down systems, is that it does not need a digital computer: the inertial attitude is given directly by the platform.

The laser-gyro. The employment of rate gyros is not the only way to sense the angular motion of rotating objects: so-called “laser-gyros” use the constant velocity of light; “solid state gyros” use the influence of the Coriolis force on a rotating mechanical resonator. Particularly interesting for space applications are laser-gyros. They have actually nothing to do with mechanical gyros, except for the fact that they are also used to measure angular velocities. As an optic-electronic device the laser gyro is considered as more accurate and more reliable than the mechanical rate gyro. A light beam “rotates” inside the light path of a laser gyro, via a triangular tube with mirrors as corner reflectors (ring-laser gyros) or via an optic coil (fiber optic gyros).



Two contra rotating beams of coherent laser light are transmitted around the light path, both light beams travel in opposite direction with the velocity of light and complete one rotation in exactly the same time interval; but when the laser gyro rotates on an axis orthogonal to the light path, transmitter and sensor are displaced during the travel time and the travel distance is not the same anymore. The difference in the arrival time is extremely short, but it can be measured accurately using a laser light phase shift.

6.3.3 The Gyro as Actuator

Gyrostabilizers. Up to now we considered the gyro as a small sensor, small in comparison with the size of the spacecraft. The motion of the spacecraft had an effect on the gyro, but the inertia of the gyro did not influence the motion of the spacecraft. Especially when the gyro is big, this is not true anymore: then the inertia of the gyro influences the motion of the spacecraft. When the gyro runs concentrically on the spin axis, it has the angular momentum $\vec{H} = \mathcal{A} \cdot \vec{\omega}_{spin}$, where term \mathcal{A} is the moment of inertia. The angular momentum \vec{H} is aligned with the spin axis $\vec{\omega}_{spin}$. We have seen that the gyro reacts differently on torques: when the torque $\mathcal{M}_{aligned}$ acts in the direction of the spin axis, it accelerates (or decelerates) the spin rate; but when the torque $\mathcal{M}_{orthogonal}$ acts in a direction orthogonal to the spin axis, the gyro reacts with a slow precessional rotation. The motion $\omega_{precession}$ happens in a way that the rotating gyro tries to align its spin axis with the direction of the external torque. Using mathematical relationships, we can formulate this typical behaviour as follows:

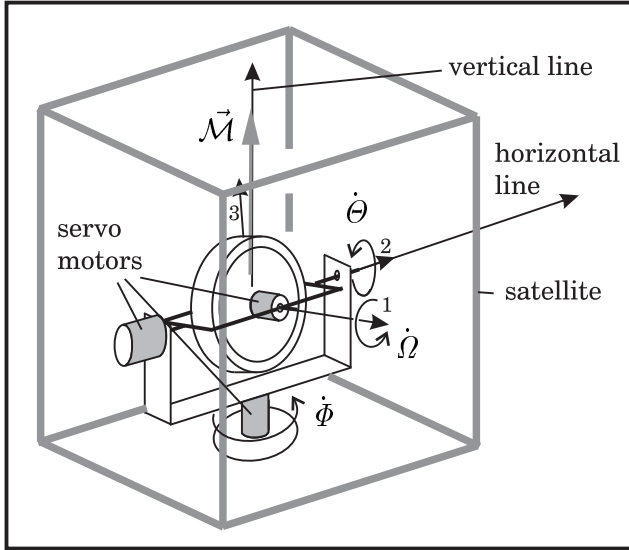
$$\omega(t)_{spin} = \int_0^t \frac{\mathcal{M}_{aligned}}{\mathcal{A}} dt + \omega(0)_{spin} \quad (6-89)$$

$$\omega(t)_{precession} = \frac{\mathcal{M}_{orthogonal}}{\mathcal{A} \cdot \omega_{spin}} \quad (6-90)$$

When a satellite is equipped with a fast-rotating momentum wheel, it “inherits” this typical gyroscopic behaviour. On orbit the satellite is always exposed to small external moments (a remainder of drag, light pressure, irregular gravity and so on). Without gyro the original attitude of the spacecraft would soon drift away, and the satellite would start tumbling; but a gyro inside the satellite stabilizes the attitude. When the moment acts in a direction orthogonal to the spin axis it causes a drift of the attitude. However, the attitude drift is very small and the displacement angle is proportional to the time interval of the perturbation. When the moment acts in direction aligned with the spin axis it causes a rotation of the satellite. However, it is easy to exchange angular momentum between the flywheel and the satellite. A controlled drive motor can stop the rotation and pivot the satellite about the spin axis of the gyro to any desired orientation. The method of controlling the attitude of a spacecraft by a single momentum wheel is called “one axis stabilization”.

The gimbaled actuator gyro. A fully gimbaled gyro can control the satellite’s attitude in all directions, provided that the bearings are equipped with servomotors. The principle is easy: the angular momentum conservation law tells us that the entire angular momentum of satellite and gyro together is constant in time when no external moments act on the satellite. The servomotors can exchange angular momentum between satellite and wheel and damp down nutational motion. So, assuming that the satellite stops moving at the initial and at the final time of an attitude changing maneuver, direction and length of the angular momentum vector of the wheel are exactly the same before and after the maneuver. When the maneuver is executed slowly, the angular momentum of the satellite can be neglected. Then the wheel keeps the length and the direction of its angular momentum during the entire maneuver.

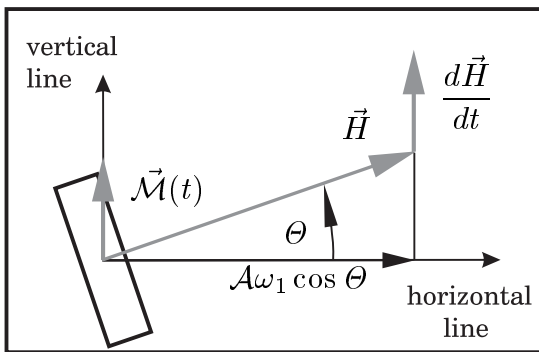
Satellite attitude maneuver. A satellite will change its attitude when a gyro is turned inside the satellite. Assume that the satellite is in rest, the working machine is a fast rotating fully gimballed gyro with servomotors in all gimbal bearings. Our objective is to generate a rotation of the satellite about the vertical direction. Conveniently we use a “gimbal frame of reference” for the analysis: its attitude



is determined by the Cardan angles Φ (not with respect to the satellite but with respect to inertial space) and Θ (with respect to the horizon). The symmetrical flywheel has the moments of inertia \mathcal{A} for rotations on the symmetry axis (1-direction) and \mathcal{B} for rotations orthogonal to the symmetry axis. To initiate the motion, the gyro must generate a moment \mathcal{M} about the vertical axis:

$$\vec{\mathcal{M}} = \begin{pmatrix} \mathcal{M}(t) \sin \Theta \\ 0 \\ \mathcal{M}(t) \cos \Theta \end{pmatrix} \quad (6-91)$$

Vector (6-91) determines the torque exerted from the servomotors on the flywheel. We consider that the torque vector is aiming upwards: the satellite will start to rotate into “west direction” (with a rotation velocity vector aiming downwards). Obviously, just the servomotors for the 1-axis and for the 3-axis must drive the wheel, the horizontal moment vanishes. Equations (6-79) to (6-81) apply to the problem. When the wheel spins rapidly, the spin rate ω_1 is much higher than the rates $\dot{\Phi}$ and $\dot{\Theta}$, and the terms $\mathcal{B}\dot{\Phi}$ and $\mathcal{B}\dot{\Theta}$ can be neglected in comparison with the term $\mathcal{A}\omega_1$. Then a simplified equation system follows from the equations (6-79) and (6-81):



$$\mathcal{M}(t) \sin \Theta = \mathcal{A} \cdot \dot{\omega}_1 \quad (6-92)$$

$$\mathcal{M}(t) \cos \Theta = \mathcal{A} \cdot \omega_1 \cdot \dot{\Theta} \quad (6-93)$$

The solution to the equation system is:

$$\mathcal{A} \omega_1 \cos \Theta = \text{constant} \quad (6-94)$$

$$\mathcal{M}(t) = \mathcal{A} \omega_1 \cos \Theta \, d(\tan \Theta)/dt \quad (6-95)$$

Term $\mathcal{A}\omega_1 \cos \Theta$ is the horizontal component of the angular momentum vector \vec{H} ; term $\mathcal{M}(t) = d\vec{H}/dt$ is the variation of the angular momentum vector in time. At the beginning of the motion ($\Theta = 0$) the gyro reacts on the torque with a comparatively slow precession. It puts its spin axis upwards, without changing its spin rate. When the gyro comes to its saturation point ($\Theta = 90^\circ$) it stops the precessional movement. Then the gyro must speed up its spin rate to exert a moment about the vertical axis.

However, our objective was to turn the satellite, not to exert a moment about the vertical direction for an infinite time interval. Therefore we just have to generate a vertical torque for a short time period. We can stop the precession of the gyro early before it reaches the saturation point. Then the gyro keeps its orientation while the satellite rotates ($\dot{\Phi} = 0$). To stop the rotation of the satellite, we simply have to turn the gyro back to its original orientation ($\Theta = 0$). The “gimbal frame of reference” accepts the same orientation with respect to inertial space before and after the maneuver.

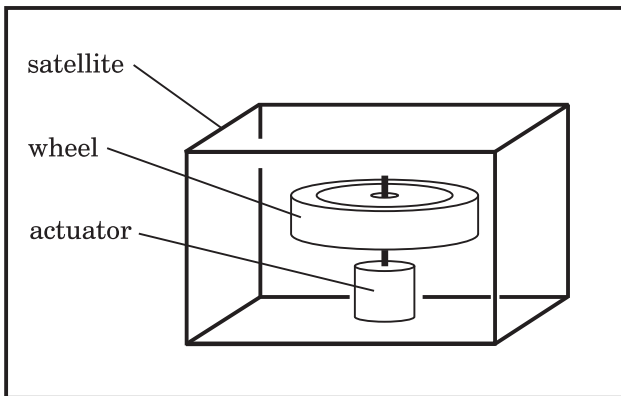
Finally, we want to evaluate equation (6-80) to find the relationship between the angles Φ and Θ . When Φ is constant and $\dot{\Phi}$ is zero, then the precession rate $\dot{\Theta}$ is also constant and the motion is exactly determined by the equations (6-94) and (6-95). However, the servomotors can control the gyro in a way that it also oscillates slightly about the vertical line. Since term $\mathcal{A}\omega_1 \cos \Theta$ is constant and term $\mathcal{B}\dot{\Phi}^2 \sin \Theta \cos \Theta$ is negligibly small, we can integrate equation (6-80) and conclude in an approximation:

$$\mathcal{B} \cdot \dot{\Theta} = (\mathcal{A}\omega_1) \cdot \dot{\Phi} \quad (6-96)$$

Term $\mathcal{A}\omega_1$ is the length of the angular momentum vector at the beginning of the maneuver, when the gyro spun concentrically on the horizontal line ($\Theta = 0$ and $\dot{\Theta} = 0$). Equation (6-96) confirms that the angle Φ remains small during the whole maneuver.

The reaction wheel. The application of gimballed bearings for rapidly rotating gyros is not unproblematic in space (an attitude control system that uses gimballed gyros is not very reliable). Gimbals can be avoided when several flywheels are used.

The working principle of reaction wheels is quite simple. Regard a satellite and consider that the reaction wheel mounted inside the satellite is initially in rest. When a servo-drive starts rotating the wheel, the wheel builds up an angular momentum.

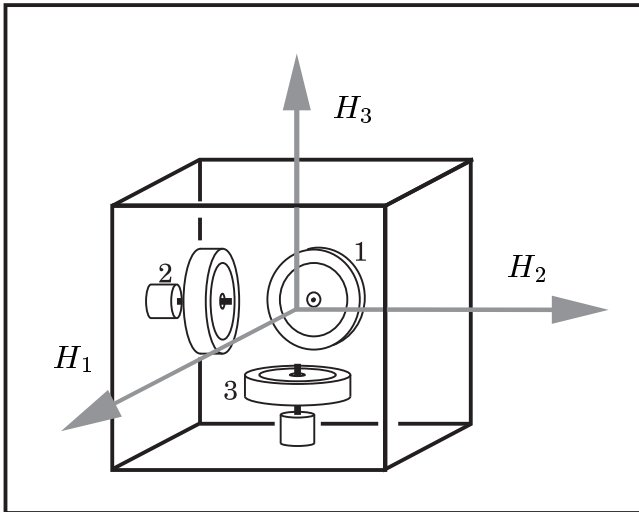


The satellite must begin to rotate on the same axis into the opposite direction, to cancel out with its angular momentum the angular momentum of the wheel. When the attitude change maneuver is completed, the servomotor must brake down the rotation of the wheel, and the satellite stops simultaneously. Finally, satellite and reaction wheel are in rest again.

At least three reaction wheels are necessary to provide control about all three axes. It is not necessary that the three operative rotation axes are mutually perpendicular. Often, the attitude control system of a satellite employs more than three wheels to increase the reliability by redundancy. For example, four wheels can be mounted in such a way that their axes form a pyramid-like configuration. When one of the four wheels fails, the operative wheels are still capable to control the satellite completely. Problematic lubrication limits the life time of reaction wheels to ten years, typically.

The momentum wheel. A rapidly rotating reaction wheel becomes a momentum wheel. Like the reaction wheel, the momentum wheel can be used to generate a control torque about its rotation axis by a servo-motor which accelerates or decelerates the wheel. Additionally, an angular momentum vector is associated with the rapidly rotating wheel; it stabilizes the direction of the rotation axis. When a satellite is controlled by a single momentum wheel (“one axis stabilization”), it behaves similar to a dual-spin satellite: one axis is stable with respect to inertial space, and an actuator can turn the satellite about this axis utilizing the inertia of the wheel.

The situation is similar when a satellite is controlled by several momentum wheels. Consider the case that three wheels are mounted on mutually perpendicular axes. We find the entire angular momentum \vec{H} when we add the three individual angular momentum vectors of the wheels and the angular momentum vector of the satellite (reference point should be the center of mass of the satellite). When the external torque that acts on the satellite equals zero, the angular momentum law states $d\vec{H}/dt = 0$. It tells us that the entire angular momentum vector keeps its direction and length with respect to inertial space. Assume that the satellite is in rest while the wheels are rotating. The wheels stabilize the axis of the angular momentum vector. When we want to turn the satellite on the 3-axis, we have to alter the spin rates of the wheels on the 2-axis and on the 1-axis, observing that the length of their combined angular momentum remains constant all the time. For example, a turn of the satellite on the vertical 3-axis into “west-direction” requires a reduction of the spin rate of wheel 2 and simultaneously a raising of the spin rate of wheel 1. A slight reduction of the rate of wheel 3 is also necessary during the maneuver, when the satellite turns rapidly and builds up its own angular momentum.



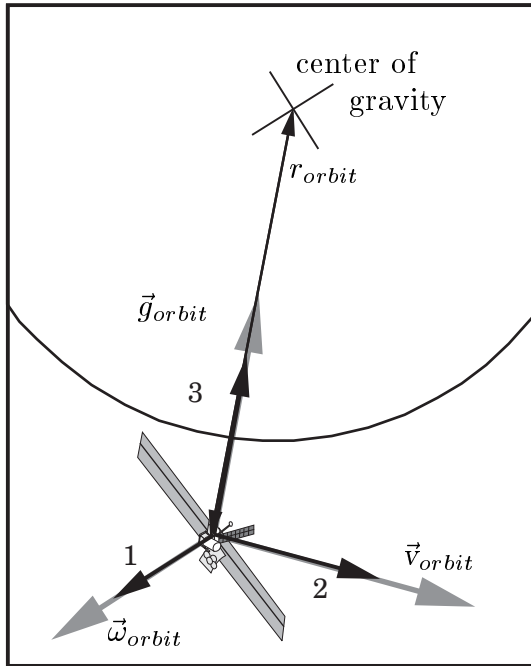
Unloading of momentum wheels. Reaction/momentum wheels are sensitive actuators. In contrast to attitude control thrusters they have not the unwanted characteristic of a permanent propellant consumption. Unfortunately, they cannot eliminate the need of propellant consumption entirely: wheels become faster and faster when they neutralize a perturbation torque that acts all the time in one predominant direction. When the wheels arrive at their saturation rate, they have to be “unloaded”. During a “momentum wheel unloading maneuver” the spin rate of the gyros is reduced while control thrusters hold the attitude of the satellite. In earth orbit, electro magnets are able to generate small torques. When a three-axis stabilized satellite is orbiting on a low earth orbit, magnet-torquers can perform the unloading of the wheels. Then it is not at all necessary to consume propellant for attitude control.

6.4. Gravity Gradient Stabilization

Rotation period and orbital period are identical for some natural satellites in the solar system (for example our moon or planet Mercury). The rotational motion is coupled with the translational motion in a way that an observer on the surface of the earth sees always the same side of the moon. Up to now we have treated the attitude dynamics of satellites as perfectly independent of their orbital mechanics. However, particularly when a satellite is very big and has not a spherical shape, there is a small interaction between the orbit and the attitude: the orbital period is not exactly the same as the orbital period of a mass particle; and the attitude is influenced by a small torque which comes from the gradient of the central gravitational field. The rotational motion of big satellites is coupled with the orbital motion. Even though the interaction is extremely small, it can be used for the attitude stabilization of large space stations. This method of attitude control is called “gravity gradient stabilization”.

6.4.1 The Motion of a Gravity Gradient Satellite

Attraction in the central gravitational field. Consider a satellite that moves on a circular orbit in the central gravitational field of a celestial body. The gravitational acceleration is inversely proportional to the square of the distance to the center of the attraction. When we use a coordinate system where the 2-axis is aligned with the velocity vector and the 3-axis aims towards the gravitational center, we can write:

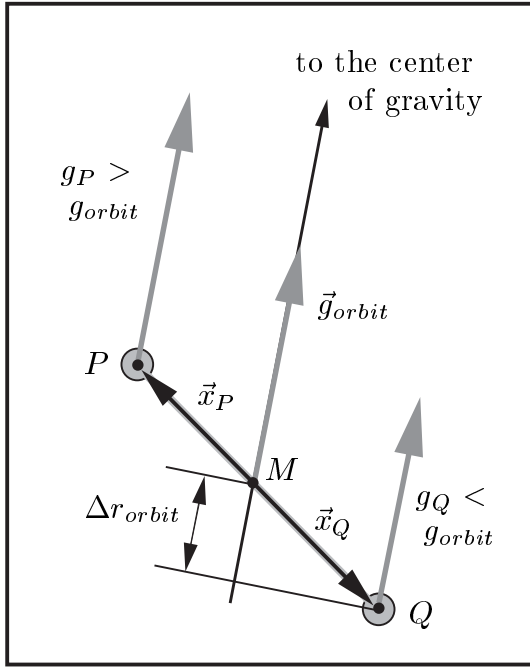


$$\vec{\omega}_{orbit} = \begin{pmatrix} \sqrt{\gamma/r_{orbit}^3} \\ 0 \\ 0 \end{pmatrix} \quad (6-97)$$

$$\vec{v}_{orbit} = \begin{pmatrix} 0 \\ \sqrt{\gamma/r_{orbit}} \\ 0 \end{pmatrix} \quad (6-98)$$

$$\vec{g}_{orbit} = \begin{pmatrix} 0 \\ 0 \\ \gamma/r_{orbit}^2 \end{pmatrix} \quad (6-99)$$

Term r_{orbit} is the distance to the gravitational center, term γ is the gravity constant. Rotation velocity $\vec{\omega}_{orbit}$, translation velocity \vec{v}_{orbit} and gravitational acceleration \vec{g}_{orbit} form a “right-hand” frame system (the vectors are perpendicular to each other). The expression $\vec{g}_{orbit}/g_{orbit}$ is the unit vector that aims at the gravitational center.



Let us now consider what happens to a satellite that has the form of the masses which weight-lifters use. When the satellite consists of two equal masses P and Q , its center of mass M (in the middle of the straight line between the two weights) moves on the circular orbit. Weight P (nearer to the gravitational center) is a little more attracted than weight Q ; and weight P experiences a slightly lower centrifugal force than weight Q . The result is a torque that tries to adjust the orientation of the elongated satellite with the direction towards the gravitational center. Our next objective is to calculate the torque on the satellite that comes from the small variation in the gravity field.

The gravity gradient vector. We find the gradient of a vector field when we differentiate the field with respect to directions coordinates; the gradient determines with its length and its direction the maximum increase of the field. The gradient of the gravitational acceleration γ/r^2 aims at the gravitational center and has the length $2\gamma/r^3$ or $2\omega_{orbit}^2$. However, the spacecraft moves on a circular orbit where it experiences the centrifugal acceleration v^2/r . The gradient of the corresponding attraction has the length v_{orbit}^2/r^2 or ω_{orbit}^2 and aims also at the gravitational center.

The gravity gradient torque. The difference in the attraction between point P and the center of mass M (or between point Q and M respectively) can be calculated using the magnitude of the gravity gradient vector. The difference Δg_P is approximately the distance Δr_{orbit} (added or subtracted) multiplied with the length of the attraction gradient vector $3\omega_{orbit}^2$. We find the distance Δr_{orbit} of point P when we form the scalar vector product of the vector \vec{x}_P (the vector which locates P with respect to M) and the unit vector in direction of the gravitational center (remember that the scalar “dot” vector product yields the product of the lengths of the two vectors, multiplied by the cosine of the angle between them). Therefore we can write:

$$\Delta g_P \approx 3\omega_{orbit}^2 \cdot \Delta r_{orbit} \quad (6-100)$$

$$\Delta r_{orbit} \approx \vec{x}_P \cdot \vec{g}_{orbit} / g_{orbit} \quad (6-101)$$

Let us now treat the variation of the attraction as a vector $\Delta \vec{g}_P = \Delta g_P \cdot \vec{g}_{orbit} / g_{orbit}$ that aims at the gravitational center (or away from it). The concerning torque is:

$$\vec{M} = \int_M (\vec{x}_P \times \Delta \vec{g}_P) dM = \frac{3\omega_{orbit}^2}{g_{orbit}} \int_M (\vec{x}_P \times \vec{g}_{orbit} \cdot (\vec{x}_P \cdot \vec{g}_{orbit})) dM \quad (6-102)$$

We find the entire torque \mathcal{M} that acts on the satellite (mass M) by integration, considering the torques $\vec{x} \times \Delta \vec{g}$ of all the mass elements dM of the whole satellite.

Equation (6-102) involves the rotation velocity ω_{orbit} of the satellite orbit and the unit vector $\vec{g}_{orbit}/g_{orbit}$ that aims at the gravitational center. The integral of all mass elements dM (located by vector \vec{x}_P) can be solved when we consider the satellite as a rigid body and use the conditional equations for the elements of the matrix of inertia:

$$\mathfrak{S} = \begin{pmatrix} \mathcal{A} & \mathcal{F} & \mathcal{E} \\ \mathcal{F} & \mathcal{B} & \mathcal{D} \\ \mathcal{E} & \mathcal{D} & \mathcal{C} \end{pmatrix} \quad (6-103)$$

The diagonal elements: $\mathcal{A} = \int_M (x_2^2 + x_3^2) dM$, $\mathcal{B} = \int_M (x_1^2 + x_3^2) dM$, $\mathcal{C} = \int_M (x_1^2 + x_2^2) dM$ and off-diagonal elements: $\mathcal{D} = -\int_M x_2 x_3 dM$, $\mathcal{E} = -\int_M x_1 x_3 dM$, $\mathcal{F} = -\int_M x_1 x_2 dM$ make use of the location vector $\vec{x}_P = (x_1, x_2, x_3)$ to locate a mass element dM at P .

When we write the gravity acceleration vector as $\vec{g}_{orbit} = (g_1, g_2, g_3)$, then we have $\vec{x}_P \times \vec{g}_{orbit} = (x_2 g_3 - x_3 g_2, -x_1 g_3 + x_3 g_1, x_1 g_2 - x_2 g_1)$; $\vec{x}_P \cdot \vec{g}_{orbit} = x_1 g_1 + x_2 g_2 + x_3 g_3$. The vector form of equation (6-102) for the gravity gradient torque becomes finally:

$$\begin{aligned} \vec{\mathcal{M}} &= \frac{3 \omega_{orbit}^2}{g_{orbit}^2} \int_M \begin{pmatrix} g_2 g_3 (x_2^2 - x_3^2) + x_2 x_3 (g_3^2 - g_2^2) + x_1 x_2 g_1 g_3 - x_1 x_3 g_1 g_2 \\ g_1 g_3 (x_3^2 - x_1^2) + x_1 x_3 (g_1^2 - g_3^2) - x_1 x_2 g_2 g_3 + x_2 x_3 g_1 g_2 \\ g_1 g_2 (x_1^2 - x_2^2) + x_1 x_2 (g_2^2 - g_1^2) + x_1 x_3 g_2 g_3 - x_2 x_3 g_1 g_3 \end{pmatrix} dM \\ &= \frac{3 \omega_{orbit}^2}{g_{orbit}^2} \begin{pmatrix} g_2 g_3 (\mathcal{C} - \mathcal{B}) - \mathcal{D} (g_3^2 - g_2^2) - \mathcal{F} g_1 g_3 + \mathcal{E} g_1 g_2 \\ g_1 g_3 (\mathcal{A} - \mathcal{C}) - \mathcal{E} (g_1^2 - g_3^2) + \mathcal{F} g_2 g_3 - \mathcal{D} g_1 g_2 \\ g_1 g_2 (\mathcal{B} - \mathcal{A}) - \mathcal{F} (g_2^2 - g_1^2) - \mathcal{E} g_2 g_3 + \mathcal{D} g_1 g_3 \end{pmatrix} \end{aligned} \quad (6-104)$$

As a vector equation, the definition (6-102) is valid in every coordinate system; but when we resolve the vector equation to get the component notation (6-104), we have to decide us for a certain coordinate system. When we use a body frame of reference, the elements of the matrix of inertia (6-103) are constant in time. When we use a principal axes body frame of reference, the notation for the gravity gradient torque vector becomes particularly simple. Nearly always a principal axis body frame of reference is used for the analysis of gravity gradient satellites (remember that in a principal axes body frame of reference the diagonal elements \mathcal{A} , \mathcal{B} and \mathcal{C} are extremes while the off-diagonal elements \mathcal{D} , \mathcal{E} and \mathcal{F} are zero). Then we can write:

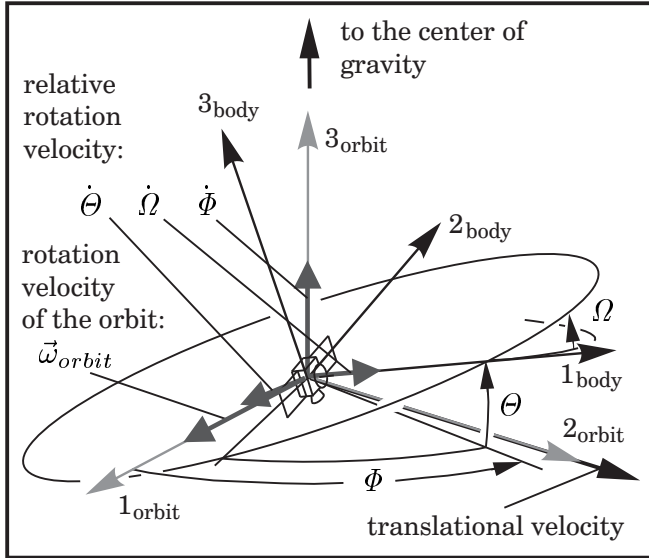
$$\vec{\mathcal{M}} = \frac{3 \omega_{orbit}^2}{g_{orbit}^2} \begin{pmatrix} g_2 g_3 (\mathcal{C} - \mathcal{B}) \\ g_1 g_3 (\mathcal{A} - \mathcal{C}) \\ g_1 g_2 (\mathcal{B} - \mathcal{A}) \end{pmatrix} \quad (6-105)$$

The order of magnitude of the gravity gradient torque can be compared with the gyroscopic torque $\vec{\omega} \times (\mathfrak{S} \cdot \vec{\omega})$, assuming that the satellite rotates slowly with a rotation velocity ω that is similar to rotation velocity of the orbit ω_{orbit} . The gravity gradient torque cannot be neglected for earth orbiting satellites. It can mean a burden for the active control system of a satellite that has to accept a certain defined attitude with respect to inertial space (for example a large telescope or a space station with fixed solar arrays). On the other hand, the gravity gradient torque is useful to stabilize the attitude of a space station with respect to the direction towards earth.

Equations of motion for the gravity gradient satellite. When we insert the gravity gradient vector (6-105) into the dynamical Euler equations (6-44), we find:

$$\begin{aligned} \mathcal{A} \dot{\omega}_1 + (\mathcal{C} - \mathcal{B}) \omega_2 \omega_3 &= 3 \omega_{orbit}^2 (\mathcal{C} - \mathcal{B}) g_2 g_3 / g_{orbit}^2 \\ \mathcal{B} \dot{\omega}_2 + (\mathcal{A} - \mathcal{C}) \omega_1 \omega_3 &= 3 \omega_{orbit}^2 (\mathcal{A} - \mathcal{C}) g_1 g_3 / g_{orbit}^2 \\ \mathcal{C} \dot{\omega}_3 + (\mathcal{B} - \mathcal{A}) \omega_1 \omega_2 &= 3 \omega_{orbit}^2 (\mathcal{B} - \mathcal{A}) g_1 g_2 / g_{orbit}^2 \end{aligned} \quad (6 - 106)$$

This first order differential equation system involves two vectors: $\vec{\omega}$ (the rotation velocity vector of the satellite) and $\vec{g}_{orbit}/g_{orbit}$ (the unit vector that aims at the gravitational center). In order to use these vectors in the equation system (6-106), we have to resolve them into a component notation of the body frame of reference. In the body-fixed coordinate system, however, the direction of the gravitational attraction depends on the actual attitude of the satellite. Therefore it is necessary that we determine the actual attitude of the satellite with respect to the orientation of another coordinate system, where the direction of the gravity vector is well-known.



We will call this other coordinate system “orbit frame of reference”. Its 3-axis aims always at the center of gravity; its 2-axis is always aligned with the translational velocity of the satellite. The “orbit frame of reference” is not an inertial system, since it rotates slowly on the 1-axis (the rotation velocity of the system is exactly ω_{orbit}). We will use the Cardan angles Φ , Θ and Ω to specify the attitude of the “body frame of reference” with respect to the well-known attitude of the “orbit frame of reference”.

The component notation of a vector in the “orbit frame of reference” \vec{x}_{orbit} can be transformed to the corresponding component notation in the “body frame of reference” \vec{x} by multiplication with the following transformation matrix. We can find this matrix by a geometrical analysis, or we can take it from books on gyro dynamics (a description of coordinate transformations is given in the chapter four of this book):

$$\vec{x} = \begin{pmatrix} \cos \Phi \cos \Theta & \sin \Phi \cos \Theta & \sin \Theta \\ -\sin \Phi \cos \Omega - \cos \Phi \sin \Theta \sin \Omega & \cos \Phi \cos \Omega - \sin \Phi \sin \Theta \sin \Omega & \cos \Theta \sin \Omega \\ \sin \Phi \sin \Omega - \cos \Phi \sin \Theta \cos \Omega & -\cos \Phi \sin \Omega - \sin \Phi \sin \Theta \cos \Omega & \cos \Theta \cos \Omega \end{pmatrix} \cdot \vec{x}_{orbit} \quad (6-107)$$

The unit vector in the 3-direction of the orbit frame of reference determines the direction to the center of gravity. To use this vector as $\vec{g}_{orbit}/g_{orbit}$ in the equation system (6-106), we have to transform it to a notation of the body frame of reference.

The result of the transformation is the third column of the transformation matrix:

$$\frac{\vec{g}_{orbit}}{g_{orbit}} = \begin{pmatrix} \sin \Theta \\ \cos \Theta \sin \Omega \\ \cos \Theta \cos \Omega \end{pmatrix} \quad (6-108)$$

The next step is to find the component notation of the rotation vector $\vec{\omega}$ of the satellite in the body frame of reference. The rotation vector $\vec{\omega}$ consists of two parts: $\vec{\omega}_{relative}$ (the rotation velocity of the satellite relative to the orbit frame of reference) and $\vec{\omega}_{orbit}$ (the rotation velocity of the orbit frame of reference). The components of the relative rotation are functions of the Cardan angles and their time derivatives, they can be found by a geometrical analysis (the “kinematical Euler equations”, described in detail in the chapter four of this book). The rotation velocity of the orbit frame of reference is the term ω_{orbit} multiplied by the unit vector in 1-direction. The transformation of this vector into a component notation of the body frame of reference is again a comparatively simple procedure. The rotation vector $\vec{\omega}$ becomes:

$$\begin{aligned} \vec{\omega} &= \vec{\omega}_{relative} + \vec{\omega}_{orbit} & (6-109) \\ &= \begin{pmatrix} \dot{\Phi} \sin \Theta + \dot{\Omega} \\ \dot{\Phi} \cos \Theta \sin \Omega - \dot{\Theta} \cos \Omega \\ \dot{\Phi} \cos \Theta \cos \Omega + \dot{\Theta} \sin \Omega \end{pmatrix} + \omega_{orbit} \begin{pmatrix} \cos \Phi \cos \Theta \\ -\sin \Phi \cos \Omega - \cos \Phi \sin \Theta \sin \Omega \\ \sin \Phi \sin \Omega - \cos \Phi \sin \Theta \cos \Omega \end{pmatrix} \end{aligned}$$

As the last step we have to insert the expressions for the unit vector in direction of the center of gravity ($\vec{g}_{orbit}/g_{orbit}$, equation 6-108) and the rotation vector of the satellite ($\vec{\omega}$, equation 6-109) into the Euler equations (system 6-106). Finally we get:

$$\begin{aligned} &\mathcal{A} \cdot (\ddot{\Phi} \sin \Theta + \dot{\Phi} \dot{\Theta} \cos \Theta + \ddot{\Omega} + \omega_{orbit}(-\dot{\Phi} \sin \Phi \cos \Theta - \dot{\Theta} \cos \Phi \sin \Theta)) + \\ &(\mathcal{C} - \mathcal{B}) \cdot (\dot{\Phi} \cos \Theta \sin \Omega - \dot{\Theta} \cos \Omega + \omega_{orbit}(-\sin \Phi \cos \Omega - \cos \Phi \sin \Theta \sin \Omega)) \cdot \\ &\quad (\dot{\Phi} \cos \Theta \cos \Omega + \dot{\Theta} \sin \Omega + \omega_{orbit}(\sin \Phi \sin \Omega - \cos \Phi \sin \Theta \cos \Omega)) \\ &\quad = 3 \omega_{orbit}^2 (\mathcal{C} - \mathcal{B}) \cos^2 \Theta \sin \Omega \cos \Omega \\ &\mathcal{B} \cdot (\ddot{\Phi} \cos \Theta \sin \Omega - \dot{\Phi} \dot{\Theta} \sin \Theta \sin \Omega + \dot{\Phi} \dot{\Omega} \cos \Theta \cos \Omega - \ddot{\Theta} \cos \Omega + \\ &\quad \dot{\Theta} \dot{\Omega} \sin \Omega + \omega_{orbit}(-\dot{\Phi} \cos \Phi \cos \Omega + \dot{\Omega} \sin \Phi \sin \Omega + \\ &\quad \dot{\Phi} \sin \Phi \sin \Theta \sin \Omega - \dot{\Theta} \cos \Phi \cos \Theta \sin \Omega - \dot{\Omega} \cos \Phi \sin \Theta \cos \Omega)) + \\ &(\mathcal{A} - \mathcal{C}) \cdot (\dot{\Phi} \sin \Theta + \dot{\Omega} + \omega_{orbit} \cos \Phi \cos \Theta) \cdot \\ &\quad (\dot{\Phi} \cos \Theta \cos \Omega + \dot{\Theta} \sin \Omega + \omega_{orbit}(\sin \Phi \sin \Omega - \cos \Phi \sin \Theta \cos \Omega)) \\ &\quad = 3 \omega_{orbit}^2 (\mathcal{A} - \mathcal{C}) \sin \Theta \cos \Theta \cos \Omega \\ &\mathcal{C} \cdot (\ddot{\Phi} \cos \Theta \cos \Omega - \dot{\Phi} \dot{\Theta} \sin \Theta \cos \Omega - \dot{\Phi} \dot{\Omega} \cos \Theta \sin \Omega + \ddot{\Theta} \sin \Omega + \\ &\quad \dot{\Theta} \dot{\Omega} \cos \Omega + \omega_{orbit}(\dot{\Phi} \cos \Phi \sin \Omega + \dot{\Omega} \sin \Phi \cos \Omega + \\ &\quad \dot{\Phi} \sin \Phi \sin \Theta \cos \Omega - \dot{\Theta} \cos \Phi \cos \Theta \cos \Omega + \dot{\Omega} \cos \Phi \sin \Theta \sin \Omega)) + \\ &(\mathcal{B} - \mathcal{A}) \cdot (\dot{\Phi} \sin \Theta + \dot{\Omega} + \omega_{orbit} \cos \Phi \cos \Theta) \cdot \\ &\quad (\dot{\Phi} \cos \Theta \sin \Omega - \dot{\Theta} \cos \Omega + \omega_{orbit}(-\sin \Phi \cos \Omega - \cos \Phi \sin \Theta \sin \Omega)) \\ &\quad = 3 \omega_{orbit}^2 (\mathcal{B} - \mathcal{A}) \sin \Theta \cos \Theta \sin \Omega \quad (6-110) \end{aligned}$$

The differentiation of $\vec{\omega}$ leads to comparatively complicated equations of the motion.

6.4.2 Stability of the Gravity Gradient Satellite

Stationary attitude. The equations of motion for the gravity gradient satellite (6-110) are a complicated system of nonlinear second-order differential equations. However, we can observe that the equation system has a particular solution where the satellite is in rest with respect to the orbit frame of reference. This is the case when the three Cardan angles (Φ , Θ and Ω) equal zero (then the rotation velocity vector is $\vec{\omega} = (\omega_{orbit}, 0, 0)$ and the unit vector to the center of gravity is $\vec{g}_{orbit}/g_{orbit} = (0, 0, 1)$). One principal axis of the satellite points towards the central body, another one is aligned with the orbit normal, and the third principal axis is aligned with the translational velocity vector. The case is important in practice, since usually it is desired that earth orbiting satellites face the earth always with the same side. The point in question is under which conditions would the earth-stationary orientation be stable.

Equations of motion for small perturbations. We will use a conventional way to analyze the stability of the gravity gradient satellite. The first step is the linearization of the equation system (6-110) in the vicinity of the equilibrium solution (consider the Cardan angles Φ , Θ and Ω and their time derivatives as small; replace the sine function of a small angle by the angle itself; replace the cosine function of a small angle by the value 1; and neglect all terms which involve products of small values). The result of the procedure is the following linear equation system:

$$A \ddot{\Omega} - 3(C - B) \Omega \omega_{orbit}^2 = 0 \quad (6-111)$$

$$B \ddot{\Theta} - (A - B - C) \dot{\Phi} \omega_{orbit} + 4(A - C) \Theta \omega_{orbit}^2 = 0 \quad (6-112)$$

$$C \ddot{\Phi} + (A - B - C) \dot{\Theta} \omega_{orbit} - (B - A) \Phi \omega_{orbit}^2 = 0 \quad (6-113)$$

The equation system above determines a linear oscillation (term ω_{orbit} is constant).

The characteristic equation. As the next step we have to insert the trial solutions: $\Omega = C_1 \cdot e^{\omega_{orbit} \lambda t}$, $\Theta = C_2 \cdot e^{\omega_{orbit} \lambda t}$, $\Phi = C_3 \cdot e^{\omega_{orbit} \lambda t}$, into the linear equation system (6-111 to 6-113) and eliminate the amplitude coefficients C_1 , C_2 and C_3 . Since the first equation (6-111) is not coupled with the second and the third equation, the elimination of C_1 from equation (6-111) is easy. We find the following condition:

$$\lambda^2 - 3 \frac{C - B}{A} = 0 \quad (6-114)$$

Elimination of C_2 and C_3 from the equations (6-112) and (6-113) is more complicated, because the two equations are coupled. We find after some transformations:

$$\lambda^4 + \left[1 + 3 \frac{A - C}{B} + \frac{A - C}{B} \cdot \frac{A - B}{C} \right] \cdot \lambda^2 + 4 \frac{A - C}{B} \cdot \frac{A - B}{C} = 0 \quad (6-115)$$

The roots of the characteristic equation determine the solution to the system. When a root is a complex numbers, its imaginary parts corresponds to a harmonic oscillation and its real part corresponds to an increasing or decreasing amplitude.

For stability no root of the characteristic equation may have a positive real part (otherwise an unstable solution with exponentially increasing amplitude is present).

Stability analysis of the characteristic equation (6-114) for the motion Ω is easy. The condition is a simple quadratic equation; we get as the first stability criterion:

$$\mathcal{B} > \mathcal{C} \quad (6-116)$$

The characteristic equation (6-115) for the coupled motion Θ and Φ is a quadratic equation of the type $\lambda^4 + p\lambda^2 + q = 0$, or transformed $(\lambda^2)_{1,2} = -\frac{1}{2}p \pm \frac{1}{2}\sqrt{p^2 - 4q}$. For stability we do not want to have positive real parts of any one of the four roots λ . Using mathematics of complex numbers, we can readily verify that stability is present when the value of λ^2 is a purely negative real number (when λ^2 is negative and has no imaginary part). We can conclude that stability is present when $p > 0$ as well as $q > 0$ (otherwise one solution for λ^2 would be a positive real number), and $p^2 - 4q > 0$ (otherwise λ^2 would be a complex number). These apparently three conditions are actually just two conditions, namely $q > 0$ and $p > 2\sqrt{q}$. When we evaluate the characteristic equation (6-115) using the condition $q > 0$, we find the second stability criterion for the gravity gradient satellite:

$$(\mathcal{A} - \mathcal{C}) \cdot (\mathcal{A} - \mathcal{B}) > 0 \quad (6-117)$$

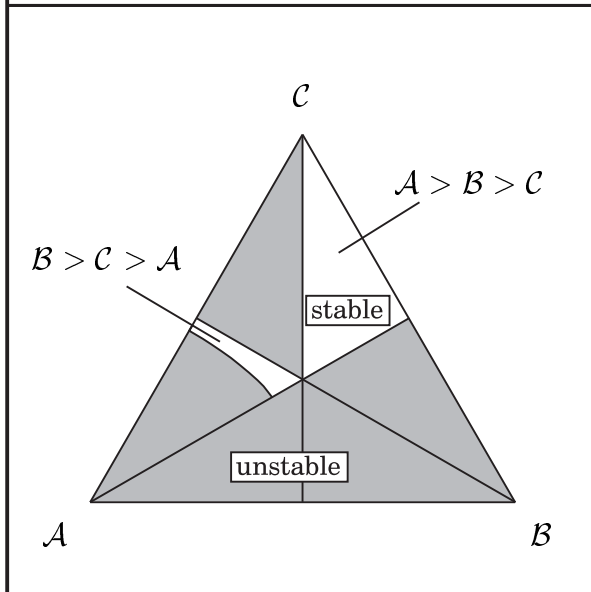
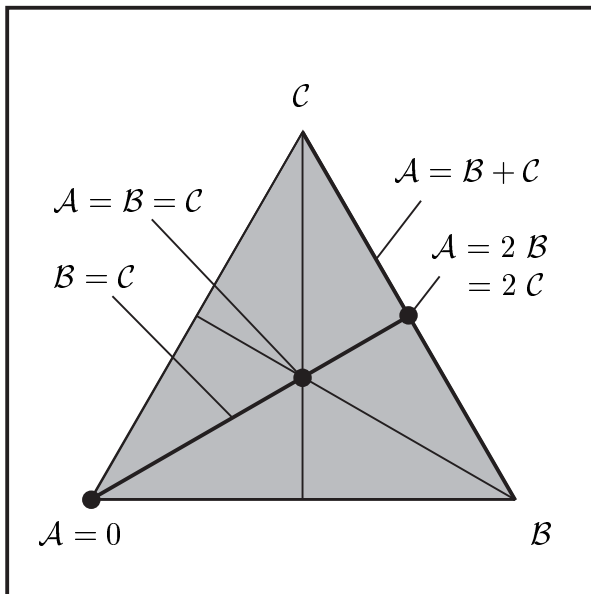
This criterion is equivalent to the statement that \mathcal{A} must be either the maximum or the minimum moment of inertia. We can combine criterion (6-117) with criterion (6-116) and conclude that stability requires either configurations $\mathcal{A} > \mathcal{B} > \mathcal{C}$ or configurations $\mathcal{B} > \mathcal{C} > \mathcal{A}$. The third stability criterion $p > 2\sqrt{q}$ becomes:

$$\left[1 + 3\frac{\mathcal{A} - \mathcal{C}}{\mathcal{B}} + \frac{\mathcal{A} - \mathcal{C}}{\mathcal{B}} \cdot \frac{\mathcal{A} - \mathcal{B}}{\mathcal{C}} \right] > 2 \sqrt{4 \frac{\mathcal{A} - \mathcal{C}}{\mathcal{B}} \cdot \frac{\mathcal{A} - \mathcal{B}}{\mathcal{C}}} \quad (6-118)$$

While this criterion has no effect on the configurations $\mathcal{A} > \mathcal{B} > \mathcal{C}$, it cancels out a large part of the possible configurations $\mathcal{B} > \mathcal{C} > \mathcal{A}$. However, also the configurations $\mathcal{B} > \mathcal{C} > \mathcal{A}$ which are not violated by criterion (6-118), are not advisable for the construction of a gravity gradient satellite. The gravity gradient satellite rotates with the rotation velocity ω_{orbit} on the 1-axis of its body frame of reference, and \mathcal{A} is the moment of inertia for the rotation on this axis. When the satellite is not a perfectly rigid body and energy dissipation is present, the rotation is just stable if \mathcal{A} is the maximum moment of inertia. We can conclude that the correct construction of a gravity gradient satellite requires the condition $\mathcal{A} > \mathcal{B} > \mathcal{C}$. The principal axis of the maximum moment of inertia \mathcal{A} must be aligned with the orbit normal, the axis of the intermediate moment of inertia \mathcal{B} must be aligned with the velocity vector and the axis of the minimum moment of inertia \mathcal{C} must point to the gravitational center. Natural satellites (moons) in the solar system with a "captured" rotation rate have usually an asymmetric (not exactly spherical) ellipsoid of inertia. They have dissipated the entire energy of the relative rotation during the astronomically long time of their past life. After having orbited very often around the planet, their orientation has finally assumed the stable attitude of a gravity gradient satellite.

Stability diagram. It is a convenient way to make the conditions of stability for the gravity gradient satellite more transparent by showing a graphical representation. According to their conditional equation (6-103), the moments of inertia \mathcal{A} , \mathcal{B} and \mathcal{C} may not assume arbitrary values, they are subjected to certain conditions. The moments of inertia must all be positive, then the sum of two elements must always be greater than the third element (in other words $\mathcal{A} + \mathcal{B} > \mathcal{C}$, $\mathcal{B} + \mathcal{C} > \mathcal{A}$, $\mathcal{A} + \mathcal{C} > \mathcal{B}$). If one element is approximately zero (the prolate stick rotor), then the other two elements must be nearly equal (for example when \mathcal{A} equals zero then \mathcal{B} equals \mathcal{C}).

A triangle with three equal sides has the appropriate shape for the stability diagram. It has the corner points \mathcal{A} , \mathcal{B} and \mathcal{C} . The moments of inertia accept their minimum value at (or near) a corner. For example, the corner location \mathcal{A} represents a stick-shaped rotor with $\mathcal{A} = 0$ and $\mathcal{B} = \mathcal{C}$. The moments of inertia



accept their maximum value on the side lines of the triangle, where a side line in opposition to a corner belongs to the same moment of inertia. For example, locations on the line between \mathcal{B} and \mathcal{C} represent disk-shaped rotors with $\mathcal{A} = \mathcal{B} + \mathcal{C}$. Axisymmetric rotors are located on the bisecting lines of the corner angles. For example, the bisecting line of the angle at the corner \mathcal{A} represents rotors with $\mathcal{A} \neq \mathcal{B} = \mathcal{C}$. The middle point of the triangle is the location for rotors with a spherical ellipsoid of inertia, where $\mathcal{A} = \mathcal{B} = \mathcal{C}$.

The stability triangle is capable to contain inside its border lines the possible configurations of the moments of inertia \mathcal{A} , \mathcal{B} and \mathcal{C} for all rotors (the border lines themselves are actually not legal locations). When we draw the regions for stable motion of the gravity gradient satellite in the stability triangle, criterion (6-116) eliminates the lower right half of the triangle (all regions below the bisecting line of the angle at corner \mathcal{A}). Criterion (6-117) leaves over two smaller triangular regions, an upper one where $\mathcal{A} > \mathcal{B} > \mathcal{C}$ and a lower one where $\mathcal{B} > \mathcal{C} > \mathcal{A}$. Finally, evaluation of criterion (6-118) eliminates most of the lower region.

6.4.3 The Rotating Space Station

The axisymmetric space station. When we look at the dynamical Euler equations (6-106) once again, we can see that the gravity gradient satellite is not the only equilibrium solution. Obviously, a satellite with an exactly spherical ellipsoid of inertia ($\mathcal{A} = \mathcal{B} = \mathcal{C}$) can rotate stable about any space-fixed axis. More interesting in practice is the case of a large space station with axisymmetric ellipsoid of inertia ($\mathcal{A} \neq \mathcal{B} = \mathcal{C}$). An equilibrium motion is possible when the station rotates with a constant rate ω_1 on the body-fixed 1-axis, provided that the body-fixed 1-axis remains always perpendicular with the orbital plane. In this case we have $\vec{\omega} = (\omega_1, 0, 0)$, $\vec{g}_{orbit}/g_{orbit} = (0, \sin((\omega_1 - \omega_{orbit})t), \cos((\omega_1 - \omega_{orbit})t))$. The unit vector that aims at the gravitational center rotates in the body frame of reference with the rate $\omega_1 - \omega_{orbit}$. The spin rate of the space station ω_1 may be greater than ω_{orbit} , for example to generate artificial gravity for the crew.

Equations of motion for small perturbations. We want to calculate the motion in the vicinity of the equilibrium solution. When we simplify the system (6-110), we have to observe that now Ω and its time derivative $\dot{\Omega}$ are not small any more. However, we assume that the two other Cardan angles Φ and Θ and their time derivatives $\dot{\Phi}$ and $\dot{\Theta}$ are still very small. Again we replace the sine function of a small angle by the angle itself; replace the cosine function of a small angle by the value 1; and neglect all terms which involve products of small values. The first equation of the system confirms that $\dot{\Omega}$ is constant (we learn nothing else from this equation). The result of this simplification procedure for the second and the third equation of system (6-110) is the following second-order differential equation system:

$$\begin{aligned}
 & \mathcal{B} \cdot (\ddot{\Phi} \sin \Omega + \dot{\Phi} \dot{\Omega} \cos \Omega - \ddot{\Theta} \cos \Omega + \dot{\Theta} \dot{\Omega} \sin \Omega + \\
 & \quad \omega_{orbit} (-\dot{\Phi} \cos \Omega + \dot{\Omega} \Phi \sin \Omega - \dot{\Theta} \sin \Omega - \dot{\Omega} \Theta \cos \Omega)) + \\
 & (\mathcal{A} - \mathcal{B}) \cdot (\dot{\Omega} + \omega_{orbit}) \cdot (+\dot{\Phi} \cos \Omega + \dot{\Theta} \sin \Omega + \omega_{orbit} (\Phi \sin \Omega - \Theta \cos \Omega)) \\
 & \quad = +3 \omega_{orbit}^2 (\mathcal{A} - \mathcal{B}) \Theta \cos \Omega \\
 & \mathcal{B} \cdot (\ddot{\Phi} \cos \Omega - \dot{\Phi} \dot{\Omega} \sin \Omega + \ddot{\Theta} \sin \Omega + \dot{\Theta} \dot{\Omega} \cos \Omega + \\
 & \quad \omega_{orbit} (+\dot{\Phi} \sin \Omega + \dot{\Omega} \Phi \cos \Omega - \dot{\Theta} \cos \Omega + \dot{\Omega} \Theta \sin \Omega)) + \\
 & (\mathcal{A} - \mathcal{B}) \cdot (\dot{\Omega} + \omega_{orbit}) \cdot (-\dot{\Phi} \sin \Omega + \dot{\Theta} \cos \Omega + \omega_{orbit} (\Phi \cos \Omega + \Theta \sin \Omega)) \\
 & \quad = -3 \omega_{orbit}^2 (\mathcal{A} - \mathcal{B}) \Theta \sin \Omega
 \end{aligned} \tag{6-119}$$

The nonlinear equation system is still complicated. Fortunately, we can transform it:

$$\begin{aligned}
 & \ddot{\Phi} + \left(\frac{\mathcal{A}\omega_1}{\mathcal{B}} - 2\omega_{orbit} \right) \cdot \dot{\Theta} + \left(\frac{\mathcal{A}\omega_1\omega_{orbit}}{\mathcal{B}} - \omega_{orbit}^2 \right) \cdot \Phi = 0 \\
 & \ddot{\Theta} - \left(\frac{\mathcal{A}\omega_1}{\mathcal{B}} - 2\omega_{orbit} \right) \cdot \dot{\Phi} + \left(\frac{\mathcal{A}\omega_1\omega_{orbit}}{\mathcal{B}} - \left(4 - 3\frac{\mathcal{A}}{\mathcal{B}} \right) \omega_{orbit}^2 \right) \cdot \Theta = 0
 \end{aligned} \tag{6-120}$$

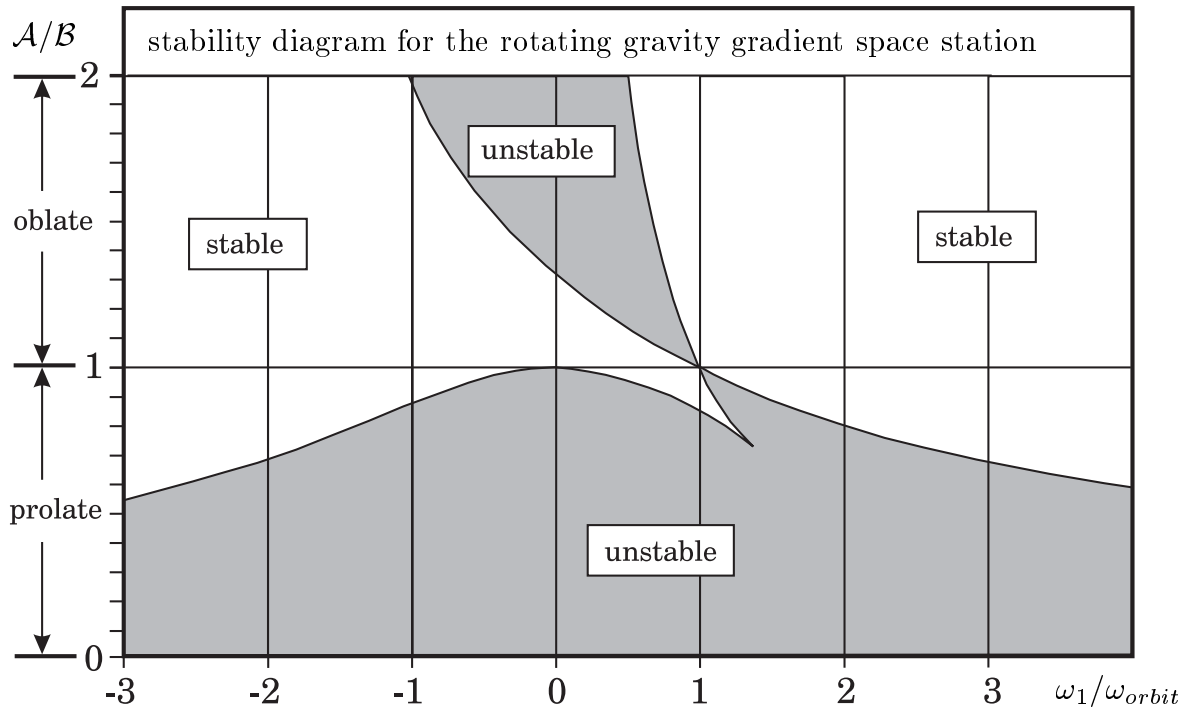
To find the system above we have to add the equations of system (6-119) two times (using the first time the factors $\sin \Omega$ and $\cos \Omega$; the second time $-\cos \Omega$ and $\sin \Omega$). Furthermore, we replace the expression $(\dot{\Omega} + \omega_{orbit})$ by the constant value ω_1 .

Stability of the rotating gravity gradient satellite. The linear system (6-120) is much better conditioned for a stability analysis than the nonlinear system (6-119).

We insert the trial solutions $\Phi = C_1 \cdot e^{\omega_{orbit}\lambda t}$ and $\Theta = C_2 \cdot e^{\omega_{orbit}\lambda t}$, eliminate the coefficients C_1 and C_2 from the equation system and get the characteristic equation as:

$$\lambda^4 + \left[\left(\frac{\mathcal{A}}{\mathcal{B}} \frac{\omega_1}{\omega_{orbit}} \right)^2 - 2 \left(\frac{\mathcal{A}}{\mathcal{B}} \frac{\omega_1}{\omega_{orbit}} \right) + 3 \frac{\mathcal{A}}{\mathcal{B}} - 1 \right] \cdot \lambda^2 + \left[\left(\frac{\mathcal{A}}{\mathcal{B}} \frac{\omega_1}{\omega_{orbit}} \right)^2 + \left(3 \frac{\mathcal{A}}{\mathcal{B}} - 5 \right) \cdot \left(\frac{\mathcal{A}}{\mathcal{B}} \frac{\omega_1}{\omega_{orbit}} \right) - 3 \frac{\mathcal{A}}{\mathcal{B}} + 4 \right] = 0 \quad (6-121)$$

Again, the characteristic equation (6-121) for the coupled motion Θ and Φ is a quadratic equation of the type $\lambda^4 + p\lambda^2 + q = 0$. For motion stability it is required that $q > 0$ and $p > 2\sqrt{q}$. When we evaluate the characteristic equation (6-121) numerically, we can use the ratios ω_1/ω_{orbit} and \mathcal{A}/\mathcal{B} as parameters for the diagram:



It must be emphasized that the diagram analyzes exclusively the effect of the gravity gradient vector on the stability of a rotating axisymmetric satellite (considering the satellite as a perfectly rigid body and ignoring the effect of energy dissipation). The satellite moves on a circular orbit and rotates on the axis which is perpendicular to the orbital plane. An oblate satellite ($\mathcal{A} > \mathcal{B}$) is stable when it rotates faster than the orbit ($\omega_1 > \omega_{orbit}$). However, also a prolate satellite ($\mathcal{A} < \mathcal{B}$) can be stable, provided that it rotates fast enough. A non-rotating satellite ($\omega_1 = 0$) is just stable when it is slightly oblate ($1 < \mathcal{A}/\mathcal{B} < 1.33$). In this case a disk-shaped satellite ($\mathcal{A}/\mathcal{B} = 2$) is unstable, as well as a stick-shaped satellite ($\mathcal{A}/\mathcal{B} = 0$). A satellite that accepts always the same orientation with respect to the gravitational center ($\omega_1 = \omega_{orbit}$) is stable when it is either oblate ($\mathcal{A}/\mathcal{B} > 1$) or slightly prolate ($0.86 < \mathcal{A}/\mathcal{B} < 1$).

7. Launcher Dynamics

The ancient German V2, a missile of the second world war, was actually the first rocket that became operational. Its shape has still an influence on the general image of conventional space launchers: the V2 weapon was a cigar-shaped tube with a conic tip and fins at the rear end. Today aerodynamic fins are not used anymore, but conventional space launchers have still much in common with the first operational rocket: space launchers are slender bodies with a more or less tubular shape. The propellant is stored in cylindrical tanks with hemispherical ends (also spherical tanks are used). The payload is carried on top of the vehicle; and the propulsion system that drives the vehicle is mounted at the rear end. Conventional space launchers are multi-stage rockets, composed of stages with the shape of a cylinder. The upper stage has usually a smaller diameter than the rest of the vehicle. The payload is protected against aerodynamic loads by a jettisonable shroud (the payload shroud has usually a larger diameter than the upper stage). Often strap-on booster motors are used to assist the launch phase; however, these boosters are nothing else than cylindrical rocket stages.

Rocket engines are used to propel the vehicle, but they are also used to control the trajectory and to stabilize the attitude during the ascent flight phase. The thrust direction can be deviated from the nominal centerline in a way that the engines generate control torques about the center of mass of the vehicle. By this the guidance system is able to control the attitude and simultaneously the trajectory of the launcher. Thus, when we want to analyze the problem of how to control a space launcher we are actually faced with three quite independent partial problems: first, we have to find an optimized nominal reference trajectory; second, we have to find a navigation strategy that makes the launcher follow the predetermined reference trajectory; and third we have to find a way to stabilize the attitude of the flexible body of the vehicle.

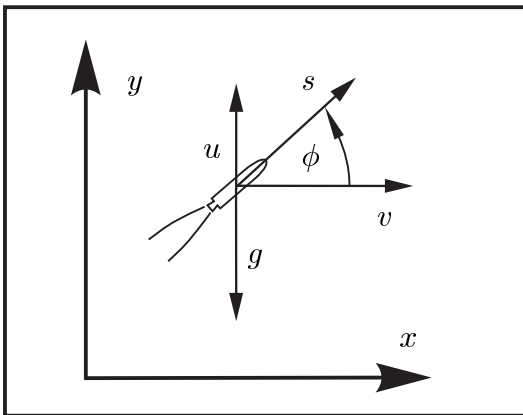
The dynamics and control of conventional space launchers is the content of the seventh chapter of this book. The first section of this chapter is concerned with the problem of ascent trajectory optimization, taking surface-launched and airborne-launched rockets as examples (airborne launchers use an aircraft as a launch platform). The optimal ascent trajectory of a space launcher determines important design aspects, such as the payload capacity and the stress environment for the structure. Once we have established a nominal reference trajectory, the next problem is how to implement an optimal navigation strategy. Now we simulate the reference trajectory and consider the disturbances which can occur during the real flight. The objective of the navigation strategy is to guide the launcher as close as possible to the desired terminal orbit (it is not important anymore to maximize the payload capacity). Finally, when we have determined and tested an appropriate steering strategy, we consider the launcher as a flexible material body with finite dimensions and calculate its behaviour as a function of the control torques which occur during the flight. The trajectory simulations of this chapter are results of a special computer program [“Skynav’s User Manual”, Ingenieurbüro Dr.Schlingloff, Munich-Regensburg, 1993].

7.1. Thrust Vector Optimization

The optimization of a launcher demands an optimal performance of the propulsion system: the thrust vector (determined by the thrust direction, the mass flow rate and the mixture ratio) has to be controlled appropriately during the ascent flight phase. We will use the Hamilton Lagrange theory to find optimal time functions for the variables that determine the behaviour of the thrust vector. The application of the Hamilton Lagrange theory is explained in more detail in the chapter five of this book.

7.1.1 Control of the Thrust Angle

Uniform gravitational field without drag. Initially, when the space launcher departs from the launch pad, the thrust vector points vertically upwards; and finally, when the launcher enters a circular orbit, the thrust vector points in a more or less horizontal direction. The tipping movement of the thrust direction during the ascent flight phase is approximately a linear function of time (also called “tilt-function”). We will show the application of the Hamilton-Lagrange method for the optimization of the thrust direction during the ascent flight of a launcher using a simplified scenario: we take the equations of motion of a rocket that moves in a uniform gravitational field without atmosphere. The simplified scenario allows us to obtain analytical solutions (the “real problem” of space launcher trajectory optimization is more complicated and requires the application of a computer). We use the coordinate x to define the horizontal distance from the launch pad and the coordinate y to define the flight altitude. When we assume that the constant gravitational acceleration g points downwards, the equations of motion for the non-atmospheric flight of a space launcher follow as:



$$\begin{aligned}
 \dot{u} &= s \sin \phi - g \\
 \dot{v} &= s \cos \phi \\
 \dot{y} &= u \\
 \dot{x} &= v
 \end{aligned} \tag{7-1}$$

The velocity vector of the vehicle is given by the vertical component u and the horizontal component v ; the angle ϕ determines the direction of the thrust vector. Our objective is to calculate an optimal control strategy for the thrust angle ϕ ; considering the thrust acceleration s as a predetermined function of time.

Optimization of the thrust direction. To find an optimal control strategy for the thrust angle ϕ we follow the mathematical scheme of the Hamilton-Lagrange theory. Therefore we introduce Lagrange multipliers λ and establish the Hamiltonian \mathcal{H} as:

$$\mathcal{H}(t) = (s \sin \phi - g) \lambda_u + s \cos \phi \lambda_v + u \lambda_y + v \lambda_x \quad (7-2)$$

The value of the Hamiltonian changes in time when the thrust acceleration s is an explicit time function (the thrust acceleration of a launcher is usually not constant). The behaviour of the Lagrange multipliers λ is defined by the Lagrange equations. These differential equations take a simple form for the motion of a rocket in a uniform gravitational field without drag. We can readily integrate the equations analytically:

$$\begin{aligned} \dot{\lambda}_u &= -\frac{\partial \mathcal{H}}{\partial u} = -\lambda_y \\ \dot{\lambda}_v &= -\frac{\partial \mathcal{H}}{\partial v} = -\lambda_x \\ \dot{\lambda}_y &= -\frac{\partial \mathcal{H}}{\partial y} = 0 \\ \dot{\lambda}_x &= -\frac{\partial \mathcal{H}}{\partial x} = 0 \end{aligned} \quad \rightarrow \quad \begin{aligned} \lambda_u &= -\lambda_{y_0} \cdot t + \lambda_{u_0} \\ \lambda_v &= -\lambda_{x_0} \cdot t + \lambda_{v_0} \end{aligned} \quad (7-3)$$

The initial values of the Lagrange equations (terms λ_{u_0} , λ_{v_0} , λ_{x_0} and λ_{y_0}) are also the integration constants. Then the optimal thrust angle control follows as:

$$\frac{\partial \mathcal{H}}{\partial \phi} = \lambda_u s \cos \phi - \lambda_v s \sin \phi = 0 \quad \rightarrow \quad \tan \phi = \frac{-\lambda_{y_0} \cdot t + \lambda_{u_0}}{-\lambda_{x_0} \cdot t + \lambda_{v_0}} \quad (7-4)$$

Our next problem is to find the correct values for the constants λ_{u_0} , λ_{v_0} , λ_{x_0} and λ_{y_0} . The theory demands that the initial values of the Lagrange equations are adjusted appropriately: the integration of the trajectory must satisfy certain final conditions.

Thrust angle control for intercept maneuvers. Let us first consider the simple case that a missile intercepts a target object. When the intercept velocity is not restricted (when the velocity components u and v are free at the final time t_{end}), then the multipliers λ_u and λ_v vanish at the end of the trajectory. We can denote:

$$\begin{aligned} \lambda_{u,end} &= -\lambda_{y_0} \cdot t_{end} + \lambda_{u_0} = 0 \\ \lambda_{v,end} &= -\lambda_{x_0} \cdot t_{end} + \lambda_{v_0} = 0 \end{aligned} \quad (7-5)$$

When we insert these relationships (7-5) into the thrust angle control equation (7-4) we find that the thrust angle ϕ is constant for this type of trajectory. Thus:

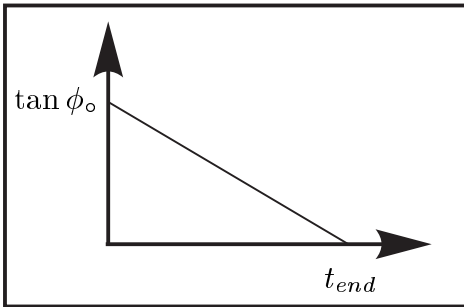
$$\tan \phi = \frac{\lambda_{y_0}}{\lambda_{x_0}} = \text{constant} \quad (7-6)$$

It is easy to integrate the equations of motion (7-1) when the thrust angle ϕ is constant in time: the optimal intercept trajectory is a curve that joins the starting place and the location where the missile hits the target object.

Thrust angle control for the ascent to orbit. Let us now apply the equations of motion (7-1) to the ascent of a space launcher to a circular orbit: a predefined horizontal velocity at a certain altitude is now the final condition for the trajectory. In contrast to an intercept maneuver the final values of the multipliers λ_u and λ_v are not automatically zero now (the velocity components u and v are predetermined at the final time). However, the range x is not restricted, and therefore the concerning multiplier λ_x vanishes at the end of the trajectory. We can conclude:

$$\lambda_{x,end} = \lambda_{x_0} = 0 \quad \rightarrow \quad \tan \phi = -\frac{\lambda_{y_0}}{\lambda_{v_0}} \cdot t + \frac{\lambda_{u_0}}{\lambda_{v_0}} \quad (7-7)$$

The thrust angle control $\tan \phi$ is a linear function of time, defined by the initial value $\tan \phi_0 = \lambda_{u_0}/\lambda_{v_0}$ and the constant “tilt rate” $d \tan \phi / dt = -\lambda_{y_0}/\lambda_{v_0}$. The final altitude y is maximized when we adjust these values in a way that the final values of the velocity vector u and v are satisfied. However, a maximization of the final altitude y requires automatically that the final vertical velocity u equals zero; and therefore we may consider the final value of the velocity u as unrestricted. Also the Lagrange multiplier λ_u vanishes at the final instant of the ascent trajectory:



$$\lambda_{u,end} = -\lambda_{y_0} \cdot t_{end} + \lambda_{u_0} = 0 \quad (7-8)$$

It follows that the launcher thrusts horizontally at the final time of the ascent trajectory:

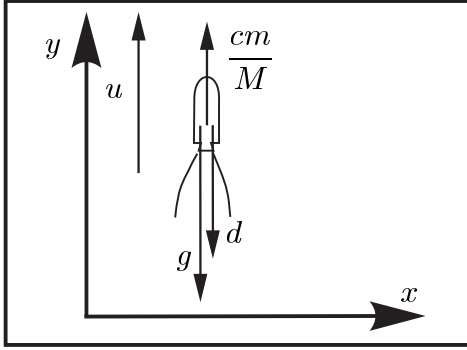
$$\tan \phi = \tan \phi_0 \cdot \left(1 - \frac{t}{t_{end}}\right) \quad (7-9)$$

The trajectory integration requires an adjustment of the initial thrust angle and the flight time: we must choose ϕ_0 and t_{end} in a way that the launcher reaches the circular horizontal velocity v exactly at the moment when the vertical velocity u vanishes.

7.1.2 Control of the Thrust Magnitude

Vertical flight through the atmosphere. Let us now concentrate on whether it is sensible to control also the thrust magnitude during the ascent flight of a launcher. Liquid rocket engines are sometimes throttleable (or can be switched off and on again); solid motors can have a predefined time function for the thrust magnitude. The structure of the launcher is exposed to aerodynamic stresses during the atmospheric flight phase; and probably it makes sense to reduce the thrust in order to limit these loads. Obviously, a thrust reduction reduces the drag losses but increases the gravity losses. To analyze the influence of a reduction of the lift-off thrust on the payload of a space launcher we will use again a simplified model that allows some analytical results. Let us consider again a uniform field of gravity where a rocket ascends vertically through the atmosphere; now the vehicle is exposed to an aerodynamic drag force.

We are allowed to disregard the constant range variable x and the horizontal velocity. However, it is necessary that we introduce the actual rocket mass M as a new state variable, because the thrust acceleration $s = c \cdot m/M$ is a function of the mass flow rate m (term c is the constant exhaust velocity). Thus we can formulate the equations of motion by the following system of first-order differential equations:



$$\begin{aligned}\dot{u} &= \frac{cm}{M} - d - g \\ \dot{y} &= u \\ \dot{M} &= -m\end{aligned}\quad (7-10)$$

The flight altitude y , the vertical velocity u and the rocket mass M are state variables; the mass flow rate m is the only control variable. The launcher is pushed upwards by the thrust force, but gravity and drag pull it down. We assume in a simplification that the gravitational acceleration g is constant, however, the drag deceleration d depends on aerodynamic pressure and air density and is therefore a quadratic function of the velocity u and an exponentially declining function of the altitude y . We can write:

$$d = \frac{C_D}{2} \frac{u^2}{M} \cdot \rho_0 e^{-y/h_{scale}} \quad (7-11)$$

Term C_D is the drag coefficient; term ρ_0 is the air density at sea-level; and h_{scale} is the atmospheric scale height (we consider these terms as approximately constant).

Optimization of the mass flow rate. Now we construct the Hamiltonion \mathcal{H} and rearrange it considering all terms which are functions of the control variable m . Thus:

$$\mathcal{H} = \left(\frac{cm}{M} - d - g\right)\lambda_u + u\lambda_y - m\lambda_M = -(d+g)\lambda_u + u\lambda_y + m \left[\frac{c}{M}\lambda_u - \lambda_M\right] = 0 \quad (7-12)$$

Hamiltonian and Lagrange multipliers are continuous functions of time, also when the control variable takes a discontinuous course. We denote the partial derivatives $\partial d/\partial u = 2d/u$, $\partial d/\partial y = -d/h_{scale}$, $\partial d/\partial M = -d/M$ and the Lagrange equations:

$$\begin{aligned}\dot{\lambda}_u &= -\frac{\partial \mathcal{H}}{\partial u} = \frac{2d}{u} \cdot \lambda_u - \lambda_y \\ \dot{\lambda}_y &= -\frac{\partial \mathcal{H}}{\partial y} = -\frac{d}{h_{scale}} \cdot \lambda_u \\ \dot{\lambda}_M &= -\frac{\partial \mathcal{H}}{\partial M} = \left(\frac{cm}{M^2} - \frac{d}{M}\right) \cdot \lambda_u\end{aligned}\quad (7-13)$$

The Hamiltonion is constant in time since it is not an explicit function of the time t ; its value is constantly zero when we consider the case of an unrestricted final time t_{end} . Then the equation (7-12) constitutes a first integral of the Lagrange equations (7-13).

Discontinuous control of the mass flow rate. It can be the case that the optimal trajectory consists exclusively of flight periods with wide open throttle and flight periods with switched-off engine. Therefore we introduce the switch function \mathcal{K} as:

$$\mathcal{K} = \frac{\lambda_M}{\lambda_u} - \frac{c}{M} \quad (7-14)$$

The Lagrange multipliers λ remain continuous time functions also at moments of discontinuous control; and the Hamiltonian \mathcal{H} is just constantly zero when the value of the continuous switch function \mathcal{K} vanishes at the switching instances. As the next step we form the time derivative of \mathcal{K} and eliminate the Lagrange multipliers. The result of the procedure is a first-order differential equation for the switch function:

$$\dot{\mathcal{K}} = \left(\mathcal{K} + \frac{c}{M}\right) \cdot \frac{g - d + m \cdot \mathcal{K}}{u} - \frac{d}{M} \quad (7-15)$$

The equation above is the control law that we need to calculate the optimal switch function for the mass flow rate m . The function \mathcal{K} is positive during the initial flight phase; however, the drag grows while the launcher climbs through the atmosphere. The motor must be switched-off when the switch function runs through zero and its time derivative is negative ($\mathcal{K} = 0, \dot{\mathcal{K}} < 0$). This demands the following condition:

$$d > \frac{g}{1 + u/c} \quad (7-16)$$

The drag deceleration must exceed a certain value, otherwise it cannot be optimal to switch-off the engine in order to maximize the payload capacity of the vehicle. We should just switch off the engine when the condition (7-16) is fulfilled. As a consequence the drag diminishes in the following flight phase and we can switch on the engine when \mathcal{K} becomes positive. However, it makes just sense to reduce the thrust for a certain time interval when a fast sounding rocket is under consideration: condition (7-16) will probably never be fulfilled during the ascent flight of a heavy space launcher, and a reduction of the thrust would lead to a smaller payload capacity.

Continuous control of the mass flow rate. It is just of theoretical interest what happens when switch function \mathcal{K} and its time derivative $\dot{\mathcal{K}}$ vanish simultaneously. Then $\partial\mathcal{H}/\partial m = 0$, and we consider the singular case of a so-called intermediate thrust arc with a mass flow rate control $m(t)$ between zero thrust and maximum thrust. Such an intermediate thrust arc can only take place when the condition (7-16) becomes valid for the first time; and in order to keep $\dot{\mathcal{K}}$ at zero it is necessary that we throttle the motor exactly in a way that the condition remains valid as an equation:

$$d = \frac{g}{1 + u/c} \quad (7-17)$$

We can differentiate this equation to find the control function for the mass flow rate:

$$m(t) = M \cdot \frac{u^2(u + c)/h_{scale} + (2c + 3u)(d + g)}{2c^2 + 4uc + u^2} \quad (7-18)$$

However, it will usually be impossible to accomplish the condition (7-17), because the actual drag force of the launcher is probably always smaller than its actual weight.

7.1.3 Control of the Exhaust Velocity

Variable mixture ratio. Liquid rocket engines are usually bipropellant systems (for example, engines which operate with hydrogen as fuel and oxygen as oxidizer). The performance is determined by the relative velocity c of the exhausted gases, and the best performance demands that the engine works with a certain mixture of fuel and oxidizer (usually not a stoichiometric mixture). However, sometimes rocket engines work with a mixture ratio that deviates slightly from the performance optimum, for example when the mixture with the highest exhaust velocity involves the disadvantage of comparatively heavy tanks. Since liquid rocket engines can be operated at least in some limits with a variable mixture ratio, the question arises whether it is sensible to vary the mixture ratio of a space launcher during the mission.

We have to take the energy of the propellant under consideration when we want to analyze the influence of a variation of the mixture ratio on the performance of the space launcher. The propellant energy useful for propulsion is equal to the kinetic energy of all the exhausted masses dM , we can write the differential form down as:

$$dE_{jet} = -\frac{1}{2}dM c^2 \quad (7-19)$$

The negative sign in the equation above comes from the definition of the mass flow rate m as $m = -dM/dt$; the term dM refers to the mass of the launcher and is negative by definition. The rate of expenditure of this energy is called “power of the jet”:

$$\dot{E}_{jet} = \frac{1}{2} m c^2 \quad (7-20)$$

Thus, while the thrust acceleration $s = c m/M$ is a linear function of both, mass flow rate m and exhaust velocity c , the power of the jet is a linear function of the mass flow rate but a square-law function of the exhaust velocity.

Let us use again the term v to characterize the horizontal velocity of the rocket powered vehicle. When the vehicle accelerates exclusively under the influence of the thrust force, its motion is governed by the following differential equation system:

$$\dot{v} = \frac{c \cdot m}{M} \quad (7-21)$$

$$\dot{M} = -m \quad (7-22)$$

$$\dot{E}_{jet} = \frac{c^2 \cdot m}{2} \quad (7-23)$$

We can readily insert the equation (7-22) into the equation (7-21) and integrate, provided that c is constant. Then the equation (7-23) is valid but can be ignored; and the well-known result of the integration (the so-called “Ciolkovskij equation”) is valid for any time function of the mass flow rate m . However, the Ciolkovskij equation is obviously not valid anymore when c is a function of time. In this case it is necessary to find optimal control functions for the exhaust velocity c and the mass flow rate m .

To calculate the conditions for an optimal control of mass flow rate m and exhaust velocity c we make again use of the Hamilton-Lagrange theory. This time the optimization objective is the “maximization of the final velocity”. The velocity v is determined by the equation of motion (7-21); the conditions (7-22) and (7-23) however, must be considered as “restrictions in differential form”. The term \dot{M} is actually just another name for the mass flow rate, but the equation $\dot{M} = -m$ states that the consumeable propellant mass is limited; and the equation $\dot{E}_{jet} = c^2 m/2$ states that the energy available for propulsion is limited. Then the Hamiltonian takes the form of:

$$\mathcal{H} = \frac{c m}{M} \cdot \lambda_v - m \cdot \lambda_M + \frac{c^2 m}{2} \cdot \lambda_E \quad (7-24)$$

We form the partial derivatives of the Hamiltonian with respect to the state variables:

$$\dot{\lambda}_v = -\frac{\partial \mathcal{H}}{\partial v} = 0 \quad \rightarrow \quad \lambda_v = constant \quad (7-25)$$

$$\dot{\lambda}_M = -\frac{\partial \mathcal{H}}{\partial M} = \frac{c m}{M^2} \cdot \lambda_v \quad (7-26)$$

$$\dot{\lambda}_E = -\frac{\partial \mathcal{H}}{\partial E_{jet}} = 0 \quad \rightarrow \quad \lambda_E = constant \quad (7-27)$$

The condition for an optimal control of the control variable m can be denoted as:

$$\frac{\partial \mathcal{H}}{\partial m} = \frac{c}{M} \cdot \lambda_v - \lambda_M + \frac{c^2}{2} \cdot \lambda_E = 0 \quad (7-28)$$

This condition is equivalent to the condition $\mathcal{H} = 0$ along the optimal trajectory: the Hamiltonian is singular regarding the mass flow rate m . The equation (7-28) can be used as a replacement for the differential form (7-26) in two occasions: either when the exhaust velocity c is constant or when c is not constant but optimally controlled. Thus, the condition for an optimal control of the variable c is either $\mathcal{H} = 0$ or:

$$\frac{\partial \mathcal{H}}{\partial c} = \frac{m}{M} \cdot \lambda_v + c m \cdot \lambda_E = 0 \quad (7-29)$$

This demands that either the mass flow rate m vanishes (trivial in this case), or that the three Lagrange multipliers vanish simultaneously (also trivial), or, since λ_v and λ_E are constant, that the expression $(c \cdot M)$ is constant along the optimal trajectory. Also in this nontrivial case the mass flow rate m is an arbitrary function of time.

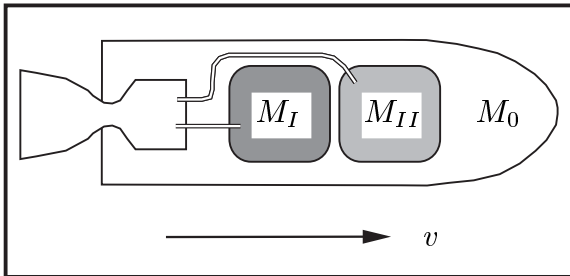
Let us finally integrate the equation of motion for the nontrivial case. The exhaust velocity c grows inversely proportional to the mass M , and the expression $c \cdot M = c_0 M_0$ is constant (c_0 is the initial exhaust velocity and M_0 the initial mass). We can write:

$$v_{end} - v_0 = \int_0^{end} \frac{cm}{M} dt = c_0 M_0 \int_0^{end} \frac{m}{M^2} dt = c_0 M_0 \left(\frac{1}{M_{end}} - \frac{1}{M_0} \right) = c_{end} - c_0 \quad (7-30)$$

The equation above is a replacement for the Ciolkovskij equation. It is valid when the exhaust velocity is a variable and the energy available for propulsion is limited.

Mixed mode. We have seen that the Ciolkovskij equation is not valid anymore when the vehicle follows an optimized thrusting profile and the exhaust velocity is controllable. However, we have still not answered the question whether it is sensible to equip a space launcher with engines constructed for variable exhaust velocity. The operation with a high exhaust velocity requires the excessive consumption of jet power (for example the use of high energetic propellant); the operation with low exhaust velocity saves propulsion energy (utilizing low energetic propellant). When we simply offered the operation with high exhaust velocity without giving a penalty for the consumption of energy, we would get the trivial result that operation with high energetic propellant is preferable for a high final velocity. The use of high energetic propellant is usually preferable, indeed, but the operation of a space launcher with low energetic propellant has the advantage that the tanks for storing the propellant are considerably smaller than the tanks for high energetic propellant.

To analyze the problem we will use the following approach: we consider the stage of a space launcher that can burn simultaneously two different propellant combinations in the same engine, low energetic propellant with the exhaust velocity c_I and high energetic propellant with the exhaust velocity c_{II} . At the moment of its ignition the stage carries in its tanks a certain predetermined amount M_I of the low energetic propellant and a certain amount M_{II} of the high energetic propellant. The exhaust velocity c is constant in single-mode operation (then $c = c_I$ when the engine burns exclusively low energetic propellant, or $c = c_{II}$ when the engine burns high energetic propellant). However, the exhaust velocity c is a function of the actual mixture ratio when the engine operates in mixed-mode. The energy flow in the thrust chamber determines the effective exhaust velocity c as a function of the mass flow rates m_I and m_{II} :



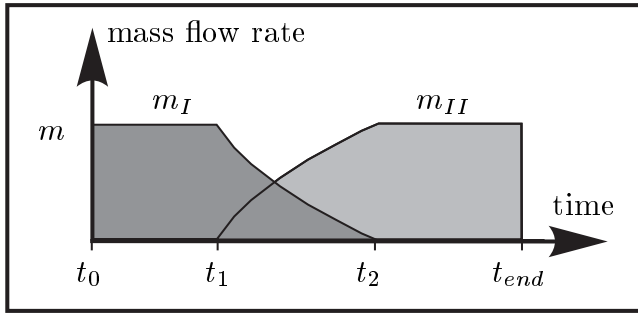
$$c = \sqrt{\frac{c_I^2 \cdot m_I + c_{II}^2 \cdot m_{II}}{m_I + m_{II}}} \quad (7-31)$$

$$m = m_I + m_{II} \quad (7-32)$$

$$M = M_I + M_{II} + M_0 \quad (7-33)$$

Note that the exhaust velocity c is a nonlinear function of the mass flow rate m_I and the mass flow rate m_{II} . For example, the condition (7-31) is valid when the low energetic propellant M_I is the combination kerosene/LOX and the high energetic propellant M_{II} is the combination LH2/LOX (this tripropellant system would require the construction of a special engine that can burn kerosene and hydrogen simultaneously with oxygen). A rocket stage that uses exclusively the bipropellant combination LH2/LOX achieves the highest exhaust velocity in “fuel-rich” operation, however, the fuel requires larger tanks per unit of mass than the oxidizer. Then we can use the symbol M_I to characterize a certain amount of the oxidizer combined with a small amount of the fuel and the symbol M_{II} to determine another amount of the oxidizer combined with the amount of fuel that gives the highest exhaust velocity.

Mass flow rate profile. The problem is now to find the optimal way to operate the engine; optimization objective is again the maximization of the final velocity v_{end} . The first section of the chapter two of this book shows that it is efficient to expend during the initial flight phase low energetic propellant and during the final flight phase high energetic propellant. Thus, the optimal way might be that the launcher expends in single-mode operation first all the low energetic propellant and switches then instantaneously to the operation with high energetic propellant (“bang-bang control”).



However, this is not the optimal way to control the mixture ratio. We can use the Hamilton-Lagrange method to demonstrate that the optimal control of the mixture rate involves a flight phase ($t_1 \rightarrow t_2$) with mixed-mode operation between the initial flight phase ($t_0 \rightarrow t_1$) and the final flight phase ($t_2 \rightarrow t_{end}$).

Application of the optimization theory. Let us use again a simple model to examine the case of mixed mode operation because it allows us to find analytical results. We consider just the thrust force and ignore all other forces which act on the vehicle: then the time derivative of the velocity is simply the thrust acceleration $s = c m/M$:

$$\dot{v} = \frac{\sqrt{(c_I^2 \cdot m_I + c_{II}^2 \cdot m_{II}) \cdot (m_I + m_{II})}}{M_0 + M_I + M_{II}} \quad (7-34)$$

$$\dot{M}_I = -m_I \quad (7-35)$$

$$\dot{M}_{II} = -m_{II} \quad (7-36)$$

The three equations above constitute the equations of motion. The exhaust velocity c is constant when the engine works in single mode operation (when either m_I or m_{II} equals zero); and then we can easily integrate the equation (7-34) analytically to find the famous Ciolkovskij equation. However, it is also possible to integrate this equation analytically when the engine works in optimal mixed mode operation. To find conditions for an optimal control of the mixed mode operation we follow the principles of the Hamilton-Lagrange theory and construct the Hamiltonian \mathcal{H} as:

$$\mathcal{H} = \frac{\sqrt{(c_I^2 \cdot m_I + c_{II}^2 \cdot m_{II}) \cdot (m_I + m_{II})}}{M_0 + M_I + M_{II}} \cdot \lambda_v - m_I \cdot \lambda_{M_I} - m_{II} \cdot \lambda_{M_{II}} \quad (7-37)$$

Optimization objective is the maximization of the velocity v_{end} exactly at the instant when the entire low energetic propellant M_I and the entire high energetic propellant M_{II} have been consumed. Therefore it is necessary that we take the equations (7-35) and (7-36) as restrictions in the form of differential equations into the considerations, otherwise we would get the trivial result that, in order to reach a high final velocity, it is more efficient to spend high energetic propellant than low energetic propellant.

The theory demands that the Hamiltonian \mathcal{H} accepts a constant value when it is not an explicit function of the time, and that \mathcal{H} vanishes at the final instant when a trajectory with unrestricted final time is under consideration. We can conclude that in our case the Hamiltonian \mathcal{H} (equation 7-37) is constantly zero when the mass flow rates m_I and m_{II} are controlled between the values zero and constant maximum (but \mathcal{H} is a time function when the maximum of the mass flow rate is not constant). As the next step we write down the Lagrange equations. Here they take the form of:

$$\dot{\lambda}_v = -\frac{\partial \mathcal{H}}{\partial v} = 0 \quad \rightarrow \quad \lambda_v = \text{constant} \quad (7-38)$$

$$\dot{\lambda}_{MI} = -\frac{\partial \mathcal{H}}{\partial M_I} = \frac{\sqrt{(c_I^2 \cdot m_I + c_{II}^2 \cdot m_{II}) \cdot (m_I + m_{II})}}{(M_0 + M_I + M_{II})^2} \cdot \lambda_v \quad (7-39)$$

$$\dot{\lambda}_{MII} = -\frac{\partial \mathcal{H}}{\partial M_{II}} = \dot{\lambda}_{MI} \quad \left(= \frac{cm}{M^2} \lambda_v \right) \quad (7-40)$$

We can observe that the multiplier λ_v is constant and that the time derivatives of the multipliers λ_{MI} and λ_{MII} are equal. It is obviously easy to integrate these equations when the engine works in single mode operation. However, let us pay attention to the optimal control of mixed mode operation, when the following conditions are valid:

$$\frac{\partial \mathcal{H}}{\partial m_I} = \frac{c_I^2(m_I + m_{II}) + (c_I^2 m_I + c_{II}^2 m_{II})}{2\sqrt{(c_I^2 \cdot m_I + c_{II}^2 \cdot m_{II}) \cdot (m_I + m_{II})} (M_0 + M_I + M_{II})} \cdot \lambda_v - \lambda_{MI} = 0 \quad (7-41)$$

$$\frac{\partial \mathcal{H}}{\partial m_{II}} = \frac{c_{II}^2(m_I + m_{II}) + (c_I^2 m_I + c_{II}^2 m_{II})}{2\sqrt{(c_I^2 \cdot m_I + c_{II}^2 \cdot m_{II}) \cdot (m_I + m_{II})} (M_0 + M_I + M_{II})} \cdot \lambda_v - \lambda_{MII} = 0 \quad (7-42)$$

$$\left(\text{or : } \lambda_{MI} = \frac{c_I^2 + c^2}{2 c M} \cdot \lambda_v ; \quad \lambda_{MII} = \frac{c_{II}^2 + c^2}{2 c M} \cdot \lambda_v \right)$$

The Lagrange equations (7-38), (7-39) and (7-40) are valid along the whole optimal trajectory (for single mode operation as well as for mixed mode operation); however, the conditions (7-41) and (7-42) are valid for mixed mode operation only. The multiplier λ_v is constant; and while the time derivatives of the multipliers λ_{MI} and λ_{MII} are equal, their difference $\lambda_{MI} - \lambda_{MII}$ must also be a time invariant constant value. We get the following equation when we subtract equation (7-42) from equation (7-41):

$$\lambda_{MI} - \lambda_{MII} = \frac{c_I^2 - c_{II}^2}{2 c M} \cdot \lambda_v \quad \rightarrow \quad c \cdot M = \text{constant} \quad (7-43)$$

Thus, an optimal control of the mass flow rates m_I and m_{II} during the time interval of mixed mode operation demands that the product of the actual mass M and the exhaust velocity c is constant in time. The mass diminishes while the engine expels propellant; and the exhaust velocity grows inversely proportional to the launcher mass. The time interval of mixed mode operation starts with the low exhaust velocity c_I at the end of the phase of burning low energetic propellant exclusively, it ends with the exhaust velocity c_{II} when all the low energetic propellant has been consumed and the launcher continues its flight with burning high energetic propellant exclusively.

Comparison of mixed mode and single mode operation. However, the proof is still missing that the mixed mode operation is better than single mode operation, switching instantaneously between burning low energetic propellant and high energetic propellant: the bang-bang control would also lead to a trajectory that is optimal in the sense of the Hamilton Lagrange theory. The only way to decide the best way of operation is to integrate both trajectories analytically and compare the results. The interval of mixed mode operation starts at the instant t_1 when the launcher has the actual mass M_1 and operates with the exhaust velocity c_I ; it ends at the instant t_2 when the launcher has the actual mass M_2 and operates with the exhaust velocity c_{II} . The product $c \cdot M = c_I \cdot M_1 = c_{II} \cdot M_2$ is constant during that time interval. We can integrate the thrust acceleration $s = cm/M$ to find the velocity increase $\Delta v = v_2 - v_1$:

$$\Delta v = \int_1^2 \frac{cm}{M} dt = c_I M_1 \int_1^2 \frac{m}{M^2} dt = c_I M_1 \left(\frac{1}{M_2} - \frac{1}{M_1} \right) = c_{II} - c_I \quad (7-44)$$

The time interval between t_1 and t_2 is quite short, particularly when the exhaust velocity c_{II} is not much higher than the exhaust velocity c_I . The next problem is to calculate the mass of low energetic propellant M_I and the mass of high energetic propellant M_{II} which the launcher expends during this time interval (now M_I and M_{II} are the propellant masses expended during the time interval $t_2 - t_1$). Therefore:

$$M_I + M_{II} = M_1 - M_2 \quad (7-45)$$

We get another equation to calculate these propellant masses when we integrate the mass flow rates m_I and m_{II} over the time interval $t_2 - t_1$. Therefore we must insert equation (7-31) into the condition $c \cdot M = c_I \cdot M_1 (= c_{II} \cdot M_2)$ and square the result:

$$\int_1^2 (c_I^2 m_I + c_{II}^2 m_{II}) dt = (c_I M_1)^2 \int_1^2 \frac{m}{M^2} dt$$

$$\text{or:} \quad c_I^2 M_I + c_{II}^2 M_{II} = c_I M_1 \cdot (c_{II} - c_I) \quad (7-46)$$

Equation (7-45) together with equation (7-46) allow us to find the propellant masses:

$$M_I = \frac{c_{II} - c_I}{c_{II} + c_I} \cdot M_1 ; \quad M_{II} = \frac{c_{II} - c_I}{c_{II} + c_I} \cdot M_2 \quad (7-47)$$

Once we know the propellant masses M_I and M_{II} it is easy to compare the mass flow rate profiles. Let us assume that $c_I = 3000$ m/s and $c_{II} = 4000$ m/s. When the engine operates in mixed mode, the velocity increases by the value $\Delta v = c_{II} - c_I = 1000$ m/s (equation 7-44). However, when the engine burns first the low energetic propellant M_I and then the high energetic propellant M_{II} , the velocity increases by the value:

$$\begin{aligned} \Delta v &= -c_I \ln \frac{M_1 - M_I}{M_1} + c_{II} \ln \frac{M_2 + M_{II}}{M_2} = -c_I \ln \frac{2c_I}{c_I + c_{II}} + c_{II} \ln \frac{2c_{II}}{c_I + c_{II}} \\ &= 996.578 \text{ m/s} \end{aligned} \quad (7-48)$$

The numerical result proves that the mixed mode operation is indeed the optimum; but the calculation shows also that the advantage is so small that we can ignore it.

7.2 Launcher Trajectory Optimization

We have analyzed the optimal thrust vector control for the ascent trajectory of a space launcher already in the preceding section of this chapter; however, the analysis was based on the unrealistic assumption that the launcher moves in a uniform gravitational field without drag (in this simplified scenario the thrust angle $\tan \phi$ is a linear declining function of time). The problem gets much more complicated when we consider the three-dimensional motion of a space launcher in the central gravitational field with atmosphere, and the solution to this problem requires the application of a computer. Space launcher trajectories are influenced by the atmospheric forces lift and drag. Since the lift force is harmful to the structure, space launchers must be guided through the atmosphere in a way that the lift force is kept below a certain small value (therefore it is necessary to impose a lift limitation for the optimization of a launcher trajectory). The drag force reduces the payload capacity of the launcher; however, the influence of the drag force on the payload capacity is usually small. During the optimization process there is no reason to reduce the thrust in order to maximize the payload capacity (but it is sensible to reduce the thrust in order to keep aerodynamic loads and thrust acceleration in the allowed limits). Our objective is to find an optimal thrusting programme that brings the launcher to the destination orbit and that obeys the restrictions imposed by flight safety and structure firmness.

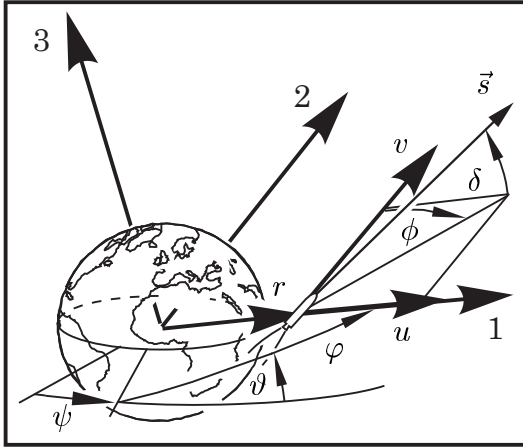
7.2.1 Ascent of Conventional Space Launchers

Trajectory profile. The launcher stands on its launch pad when a command triggers the ignition of the liquid engines. In case the engines show problems, the launcher is not detached from the launch pad (in this case the valves which control the fuel flow are turned off again). However, when the liquid engines operate properly, the solid booster motors are ignited and the vehicle lifts off (we assume that the launcher is equipped with solid boosters). Initially, the launcher ascends vertically for some seconds: the flight control system is concentrated on the lift-off phase and pays attention that the launcher does not tilt over or touches the launch tower; but when the altitude is high enough the vehicle starts slowly to incline. About one minute later, at an altitude between 10 and 20 km, the aerodynamic pressure reaches its maximum (famous Q_{max}). Now the launcher is slaved to ascend with a very small angle of attack, in order to limit the lateral acceleration which could break the fragile structure. A reduction of the motor thrust during the flight phase of high aerodynamic loads is advisable to limit the stresses for the structure of the vehicle. The “gravity tilt” trajectory guides the launcher to higher layers of the atmosphere where the strict limitation of the angle of attack can be released. Now the launcher has lost propellant mass and its thrust acceleration becomes a problematic load for the structure and the payload. It is necessary to limit this load, either by switching-off or throttling the liquid engines or by an appropriately predetermined thrusting programme for the solid booster motors. The boosters and the first stage are dropped when they are empty, and the launcher continues its flight with the propulsion of its second stage.

The separated empty stage continues to climb on its trajectory to space, but later it falls and splashes down in the ocean. The sea area where the reentry decay takes place has to be closed to traffic during the launch period. In some limits it is possible to influence the impact location by letting the launcher follow a certain ground track (a curve called “dog-leg”). The essential acceleration phase to orbital velocity starts when the launcher has left the atmosphere and moves in the vacuum of space (the launcher cannot fly very fast inside the atmosphere, because the thermal load is a problem when the flight altitude is below 100 km). The shroud that protects the payload can be dropped when in all later flight phases the heat load is smaller than a certain value (when the launcher has climbed above an altitude of 100 km, for example). The launch vehicle has attained orbital velocity when its second stage is empty. Even if the risk is in fact small that the reentering second stage means hazard to the surface of the earth when it falls down anywhere, it is not wanted that the empty stage stays on orbit. The stage should either be equipped with a deorbiting system or it should be separated when it is still on an orbit with a well-known impact point. The launcher is usually equipped with one more upper stage when the payload is advanced to geostationary orbit GEO. This upper stage brings the satellite first to the geostationary transfer orbit GTO, then the circularization is either performed by satellite integrated propulsion, or by re-ignition of the upper stage or by a separate “kick” stage. The transfer orbit is inclined with respect to the equatorial plane, however, its apogee and its perigee are located exactly above the equator. In case the capability of the launcher is greater than the payload mass, the launcher annihilates its over-capacity by flying on a non-optimal trajectory (this may include, for example, a higher flight altitude for a better radar visibility or a dog-leg curve for an improved flight safety). Additionally, the launcher needs some over-performance as contingency in order to be able to adjust trajectory disturbances.

Simplified equations. The use of simplified differential equations of motion is usually not advisable for the numerical computation of an accurate launcher trajectory: a computer program has to be used anyway; and then it is simply easier to implement all the complete equations than to justify that a certain simplification is allowed. Therefore it is necessary that we know the behaviour of the launcher accurately; sometimes however, particularly when we perform preliminary design studies, the accurate dynamical behaviour of a launcher is still unknown. Then we are forced to simplify the equations: we may, for example, neglect the lift force or assume constant aerodynamic coefficients, or take a spherical model of the earth with an exponentially declining air density. However, it is not permissible that we assume a non-rotating earth or that we use the model of a plane trajectory, or that we disregard load limits. The use of simplified control functions is less problematic than the use of simplified equations of motion: it is sure that in reality the launcher can follow the computed trajectory, even when the computed trajectory is not 100% optimized. In reality it is not possible that the launcher follows a totally optimized trajectory, because the trajectory is subjected to disturbances which are not known before the flight (for example the weather on the launch date, uncertainties in burn behaviour of solid propellant, the performance reserve which is necessary for stabilization and so on).

Equations of motion. The derivation of the equations of the translational motion for the trajectory of a flying object is explained in the fourth chapter of this book. Six first order differential equations determine the translational motion of the vehicle (which is considered as a mass particle without size); state variables are the velocity components u (vertical), v (horizontal) and the radius r , together with the Euler angles φ , ϑ and ψ , or, alternatively, the Cardan angles Φ , Θ and Ω . We can write:



$$\begin{aligned} \dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} + d_1 + l_1 + s_1 \\ \dot{v} &= -\frac{uv}{r} + d_2 + l_2 + s_2 \\ \dot{r} &= u \\ \dot{\varphi} &= \frac{v}{r} - \frac{d_3 + l_3 + s_3}{v} \cdot \sin \varphi \cot \vartheta \\ \dot{\vartheta} &= \frac{d_3 + l_3 + s_3}{v} \cdot \cos \varphi \\ \dot{\psi} &= \frac{d_3 + l_3 + s_3}{v} \cdot \frac{\sin \varphi}{\sin \vartheta} \end{aligned} \quad (7-49)$$

In these equations γ is the gravitational constant of the earth, $3.9865 \cdot 10^{14} \text{ m}^3/\text{s}^2$. Our objective is to find control functions for the three components s_1 , s_2 and s_3 of the thrust acceleration vector. We assume that a conventional space launcher follows a trajectory with a very small angle of attack, so that we can neglect the components of the lift vector l_1 , l_2 and l_3 . Then the components of the drag deceleration vector d_1 , d_2 and d_3 are just functions of the actual flight velocity and flight altitude.

The drag deceleration. The principal influence of the drag on the steering of a launcher trajectory causes a steep vertical ascent during the initial flight phase: the drag drives the launcher out of the atmosphere. The drag deceleration \vec{d} is a function of the relative velocity \vec{w} of the vehicle with respect to the ambient atmosphere; and, even when a calm atmosphere without wind is under consideration, the atmosphere is in motion because it is carried along with the rotation of the earth ($\omega_{earth} = 2\pi/\text{day}$):

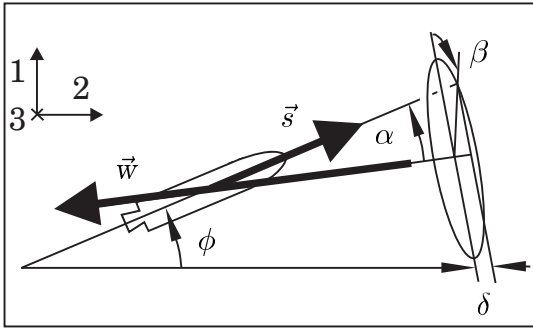
$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} -u \\ r \omega_{earth} \cos \vartheta - v \\ -r \omega_{earth} \sin \vartheta \cos \varphi \end{pmatrix} \quad (7-50)$$

$$\vec{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \frac{1}{2} \varrho \frac{C_D A}{M} w \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \quad (7-51)$$

$$(w = \sqrt{w_1^2 + w_2^2 + w_3^2}; \quad \varrho(r) = \varrho_o e^{-(r - R_{earth})/h_{scale}})$$

Term C_D is the drag coefficient, term A is a reference area and M is the actual mass of the launcher. The air density $\varrho(r)$ is an exponentially declining function of the altitude (we assume a sea-level air density $\varrho_o=1.225 \text{ kg/m}^3$, an equatorial earth radius $R_{earth}=6378 \text{ km}$ and an atmospheric scale height of $h_{scale} = 8.333 \text{ km}$).

The thrust acceleration. We have two options to determine the direction of the thrust acceleration vector \vec{s} : either we use the thrust angles ϕ (in the actual flight plane, measured against the horizontal line) and δ (out of the actual flight plane, measured against the actual flight plane), or, alternatively, we use the angles α (the angle of attack) and β (the banking angle). When the launcher moves in space we better take δ and ϕ because these angles refer to the coordinate system and not to the direction of the incident wind. However, when the launcher climbs through the atmosphere, we better use the angles α and β , because during the atmospheric flight phase it is necessary that we impose a limitation of the lift: when the lift limitation is effective it restricts the angle of attack α but it does not influence the banking angle β . Knowing the actual velocity of the incident wind \vec{w} we can transform a notation of the thrust acceleration vector \vec{s} with the angles α and β into a notation with the angles δ and ϕ (and vice versa). Term s is the length of the thrust acceleration vector:



$$\begin{aligned}
 s_1 &= s \sin \phi \cos \delta \\
 &= s \left(\cos \alpha \frac{-w_1}{w} + \sin \alpha \cos \beta \frac{w_2}{w} \right) \\
 s_2 &= s \cos \phi \cos \delta \\
 &= s \left(\cos \alpha \frac{-w_2}{w} + \sin \alpha \cos \beta \frac{-w_1}{w} \right) \\
 s_3 &= s \sin \delta \\
 &= s \sin \alpha \sin \beta
 \end{aligned} \tag{7-52}$$

We ignore the lateral component w_3 of the incident wind in these relationships since it is much smaller than the other components w_1 (vertical) and w_2 (horizontal). The simplification $w_3 \approx 0$ can be interpreted as uncertainty in the physical meaning of the angles α and β : like ϕ and δ these thrust angles are “100% correct”, however, they represent just approximately the angle of attack and the banking angle.

The lift limitation. The lift force appears when the launcher flies with an angle of attack, and the lift is a dangerous lateral load for the structure of the vehicle (thrust and drag are axial loads). The launcher flies with a small angle of attack, therefore:

$$\vec{l} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} \approx \frac{l}{w} \begin{pmatrix} w_2 \cos \beta \\ -w_1 \cos \beta \\ w \sin \beta \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{7-53}$$

The direction of the lift force is always rectangular to the direction of the incident wind, its magnitude l is approximately proportional to the sinus of the angle of attack (factor $\sin \alpha$). The thrust acceleration \vec{s} (equation 7-52) which is always aligned with the centerline of the vehicle generates a component in direction of the incident wind (proportional to $\cos \alpha$) and a component perpendicular to the direction of the incident wind (proportional to $\sin \alpha$). We make nearly no mistake in neglecting the lift because the lift is represented by the component of the thrust force that is rectangular with the incident wind: in our model we neglect the lift force, however, therefore we incline the thrust vector a little more with respect to the direction of the incident wind as it would be inclined in reality. The actual value α is even smaller than in our model.

Hamilton function and Lagrange equations. In trajectory optimization for conventional space launchers it is allowed to neglect the lift force, but it is not allowed to ignore the lift limitation. If we ignored the lift limitation, the optimization method would lead us to a wrong trajectory which the launcher would never be able to follow in reality (the launcher would not be able to stand the load of the lateral acceleration). The aerodynamic pressure $Q = 1/2 \rho w^2$, multiplied with the angle of attack α (or $\sin \alpha$), may not exceed a certain (predetermined) small value; and we are allowed to neglect the lift acceleration because the launcher will follow a trajectory with a strict lift limitation. The assumption that the lift force is negligibly small is also valid for winged space shuttles; however, this assumption is not valid anymore when we consider future aerospace planes or airborne launchers which require some lift for flying.

We consider the equations of the motion (7-49) and construct the Hamiltonian \mathcal{H} as:

$$\begin{aligned} \mathcal{H} = & \left(\frac{v^2}{r} - \frac{\gamma}{r^2} + d_1 + s_1 \right) \lambda_u + \left(-\frac{uv}{r} + d_2 + s_2 \right) \lambda_v + u \lambda_r + \frac{v}{r} \lambda_\varphi \\ & + \frac{s_3}{v} \cdot \left(\lambda_\varphi \sin \varphi \cot \vartheta + \lambda_\vartheta \cos \varphi + \lambda_\psi \sin \varphi / \sin \vartheta \right) \end{aligned} \quad (7-54)$$

The thrust acceleration s is an explicit function of the time t , and therefore the Hamiltonian \mathcal{H} is a value that changes in time. You can observe that, in order to find optimal functions, we neglect the lift acceleration l and the third component of the drag deceleration d_3 ($w_3 \approx 0$). The Lagrange equations for λ_u , λ_v and λ_r follow as:

$$\begin{aligned} \dot{\lambda}_u &= -\frac{\partial \mathcal{H}}{\partial u} = \frac{v}{r} \lambda_v - \lambda_r - \frac{\partial d_1}{\partial u} \lambda_u - \frac{\partial d_2}{\partial u} \lambda_v - \frac{\partial s_1}{\partial u} \lambda_u - \frac{\partial s_2}{\partial u} \lambda_v \\ \dot{\lambda}_v &= -\frac{\partial \mathcal{H}}{\partial v} = -2\frac{v}{r} \lambda_u + \frac{u}{r} \lambda_v - \frac{1}{r} \lambda_\varphi - \frac{\partial d_1}{\partial v} \lambda_u - \frac{\partial d_2}{\partial v} \lambda_v - \frac{\partial s_1}{\partial v} \lambda_u - \frac{\partial s_2}{\partial v} \lambda_v + \frac{s_3}{v^2} (\dots) \\ \dot{\lambda}_r &= -\frac{\partial \mathcal{H}}{\partial r} = \left(\frac{v^2}{r^2} - 2\frac{\gamma}{r^3} \right) \lambda_u - \frac{uv}{r^2} \lambda_v + \frac{v}{r^2} \lambda_\varphi - \frac{\partial d_1}{\partial r} \lambda_u - \frac{\partial d_2}{\partial r} \lambda_v \end{aligned} \quad (7-55)$$

The components d_1 and d_2 of the drag deceleration vector (equations 7-50 and 7-51) are functions of the incident wind and the flight altitude. Approximately is valid: $w_1 = -u$, $w_2 = v_o - v$, $w_3 = 0$ and $w = \sqrt{u^2 + (v_o - v)^2}$. The components of the velocity vector of the launcher u (vertical) and v (horizontal) refer to inertial space; and the term v_o represents the inertial velocity of the launch pad (the launcher has the same inertial velocity as the ambient atmosphere when it stands on the launch pad). The factor v_o is approximately constant during the atmospheric flight phase, because the drag deceleration causes a steep ascent through the atmosphere. We can assume:

$$\begin{aligned} \frac{\partial d_1}{\partial u} &= -\left(\frac{w_1}{w^2} + \frac{1}{w_1} \right) d_1 & \frac{\partial d_2}{\partial u} &= -\frac{w_1}{w^2} d_2 \\ \frac{\partial d_1}{\partial v} &= -\frac{w_2}{w^2} d_1 & \frac{\partial d_2}{\partial v} &= -\left(\frac{w_2}{w^2} + \frac{1}{w_2} \right) d_2 \\ \frac{\partial d_1}{\partial r} &= -\frac{d_1}{h_{scale}} & \frac{\partial d_2}{\partial r} &= -\frac{d_2}{h_{scale}} \end{aligned} \quad (7-56)$$

The drag and its derivatives vanish when the vehicle flies out of the atmosphere.

The Lagrange equations (7-55) are still not complete because we have still not considered the partial derivatives of the thrust acceleration vector with respect to the direction of the incident wind. A dependency of the thrust direction on the direction of the incident wind is not present when the launcher moves in the vacuum of space. In space the lift limitation is not effective and we can define the unconstrained thrust direction using the thrust angles δ and ϕ . However, the lift limitation restricts the thrust direction during the atmospheric flight phase; and in the atmosphere it is advisable that we use the angles α and β to determine the thrust direction. The lift limitation restricts just the angle of attack α but not the banking angle β . We insert the thrust vector according to its definition (7-52) and differentiate:

$$s_1 \lambda_u + s_2 \lambda_v = -s \cos \alpha \left(\frac{w_1}{w} \lambda_u + \frac{w_2}{w} \lambda_v \right) + s \sin \alpha \cos \beta \left(\frac{w_2}{w} \lambda_u - \frac{w_1}{w} \lambda_v \right)$$

$$\frac{\partial(s_1 \lambda_u + s_2 \lambda_v)}{\partial u} = -s \cos \alpha \left(-\frac{w_2^2}{w^3} \lambda_u + \frac{w_1 w_2}{w^3} \lambda_v \right) + s \sin \alpha \cos \beta \left(\frac{w_1 w_2}{w^3} \lambda_u + \frac{w_2^2}{w^3} \lambda_v \right) \quad (7-57)$$

$$\frac{\partial(s_1 \lambda_u + s_2 \lambda_v)}{\partial v} = -s \cos \alpha \left(\frac{w_1 w_2}{w^3} \lambda_u - \frac{w_1^2}{w^3} \lambda_v \right) + s \sin \alpha \cos \beta \left(-\frac{w_1^2}{w^3} \lambda_u - \frac{w_1 w_2}{w^3} \lambda_v \right) \quad (7-58)$$

The partial derivatives (7-57) and (7-58) vanish when the vehicle leaves the atmosphere, but they are not zero when the angle of attack α is restricted during the atmospheric flight phase. Then it is necessary that we consider these terms in the Lagrange equations, otherwise the launcher would follow a non-optimal trajectory.

Let us consider again the Hamiltonian \mathcal{H} (equation 7-54). The third component of the thrust acceleration vector s_3 is multiplied by an expression that is a function of the Euler angles φ , ϑ and ψ . We differentiate the Hamiltonian with respect to the Euler angles to find the Lagrange equations for the multipliers λ_φ , λ_ϑ and λ_ψ . Thus:

$$\begin{aligned} \dot{\lambda}_\varphi &= -\frac{\partial H}{\partial \varphi} = \frac{s_3}{v} (\lambda_\vartheta \sin \varphi + \lambda_\varphi \cos \varphi \cot \vartheta - \lambda_\psi \frac{\cos \varphi}{\sin \vartheta}) \\ \dot{\lambda}_\vartheta &= -\frac{\partial H}{\partial \vartheta} = \frac{s_3}{v} \frac{\sin \varphi}{\sin^2 \vartheta} (\lambda_\psi \cos \vartheta - \lambda_\varphi) \\ \dot{\lambda}_\psi &= -\frac{\partial H}{\partial \psi} = 0 \end{aligned} \quad (7-59)$$

Fortunately we can integrate these differential equations analytically and write down:

$$\begin{aligned} \dot{\lambda}_\varphi &= -\partial H / \partial \varphi \quad \rightarrow \quad \lambda_\varphi = (C_1 \sin \psi - C_2 \cos \psi) \sin \vartheta + C_3 \cos \vartheta \\ \dot{\lambda}_\vartheta &= -\partial H / \partial \vartheta \quad \rightarrow \quad \lambda_\vartheta = C_1 \cos \psi + C_2 \sin \psi \\ \dot{\lambda}_\psi &= -\partial H / \partial \psi \quad \rightarrow \quad \lambda_\psi = C_3 \end{aligned} \quad (7-60)$$

The terms C_1 , C_2 and C_3 in the equation system above are the integration constants.

Control of the banking angle. The ground track of the ascent trajectory is mainly a function of the banking angle β . The time function of the angle β is responsible for the final values of the inclination angle ϑ , the node angle ψ and the path angle φ . In order to guide the launcher into the desired final orbit we have to adjust the initial values for the Lagrange multipliers λ_φ , λ_ϑ and λ_ψ appropriately (or the integration constants C_1 , C_2 and C_3). The Hamiltonian \mathcal{H} (7-54) contains an expression in parenthesis that is a function of the Euler angles and the concerning multipliers. When we insert the integrals (7-60) into this expression we find the relationship:

$$\begin{aligned} & \lambda_\varphi \sin \varphi \cot \vartheta + \lambda_\vartheta \cos \varphi + \lambda_\psi \sin \varphi / \sin \vartheta \\ & = C_1 \{ \cos \varphi \cos \psi - \sin \varphi \sin \psi \cos \vartheta \} + C_2 \{ \cos \varphi \sin \psi + \sin \varphi \cos \psi \cos \vartheta \} \\ & \quad + C_3 \{ \sin \varphi \sin \vartheta \} \approx C_1 \quad (7-61) \\ & (= C_1 \cos \Phi \cos \Theta + C_2 \sin \Phi \cos \Theta + C_3 \sin \Theta) \end{aligned}$$

The relationship above contains three expressions in curly braces with trigonometric functions of the Euler angles φ , ϑ and ψ . If we want we can replace these expressions by notations which use trigonometric functions of the Cardan angles Φ and Θ (the geometry of these angles is explained in detail in the chapter four of this book).

The final inclination ϑ of the ascent trajectory is usually predefined; but the final node angle ψ and the final path angle φ are usually free. In this case we can replace the whole expression (7-61) in an approximation by the single constant C_1 (the angles ψ and φ are small, $C_3 = 0$ and $C_2 = C_1 \tan \psi_{end}$). We have to adjust only the value of C_1 ; and the ground track of the trajectory is nearly a straight line. However, when we want that the ground track of the trajectory is a “dog leg” curve, we must use the more complicated form and adjust also the constants C_2 and C_3 .

Let us assume the simple form of the expression (7-61) and rewrite the Hamiltonian using the definition of the thrust acceleration vector (the equations 7-52). Thus:

$$\begin{aligned} \mathcal{H} = & \left(\frac{v^2}{r} - \frac{\gamma}{r^2} + d_1 \right) \lambda_u + \left(-\frac{uv}{r} + d_2 \right) \lambda_v + u \lambda_r + \frac{v}{r} \lambda_\varphi \quad (7-62) \\ & + s \left[-\cos \alpha \left(\frac{w_1}{w} \lambda_u + \frac{w_2}{w} \lambda_v \right) + \sin \alpha \cos \beta \left(\frac{w_2}{w} \lambda_u - \frac{w_1}{w} \lambda_v \right) + \sin \alpha \sin \beta \cdot \frac{C_1}{v} \right] \end{aligned}$$

Term s is the value of the thrust acceleration (the thrust acceleration is usually a discontinuous function of time when an engine is switched-off or a stage is separated). The optimal control for the banking angle β requires a maximization of the Hamiltonian \mathcal{H} at every instant; and, since β is not restricted and a continuous time function, the maximum principle is equivalent to the condition $\partial \mathcal{H} / \partial \beta = 0$. We can conclude:

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial \beta} = & s \left[-\sin \beta \sin \alpha \left(\frac{w_2}{w} \lambda_u - \frac{w_1}{w} \lambda_v \right) + \cos \beta \sin \alpha \frac{C_1}{v} \right] = 0 \\ \rightarrow \quad \tan \beta = & \frac{C_1 w}{v(w_2 \lambda_u - w_1 \lambda_v)} \quad (7-63) \end{aligned}$$

The lift limitation has no influence on the optimal control of the banking angle β .

Control of the angle of attack. The maximum principle states that Hamiltonian has to be maximized with respect to all possible control functions; and this principle is also valid to find the optimal control law for the angle of attack α . Therefore:

$$\frac{\partial \mathcal{H}}{\partial \alpha} = s \left[-\sin \alpha \left(-\frac{w_1}{w} \lambda_u - \frac{w_2}{w} \lambda_v \right) - \cos \alpha \cos \beta \left(\frac{w_2}{w} \lambda_u - \frac{w_1}{w} \lambda_v \right) + \cos \alpha \sin \beta \frac{C_1}{v} \right] = 0$$

$$\rightarrow \tan \alpha = \frac{w_1 \lambda_v - w_2 \lambda_u}{\cos \beta (w_1 \lambda_u + w_2 \lambda_v)} \quad (7-64)$$

In case the lift limitation is not violated we have to control the angle of attack according to the condition above. However, it can be the case that the launcher is exposed to high aerodynamic loads and that the angle of attack proposed by the equation above violates the lift limitation. The lift limitation requests that the angle α must be smaller than a certain value α_{max} (usually, the product of the aerodynamic pressure ($Q = 1/2 \rho w^2$) and the sine of the angle of attack is kept constant during the aerodynamic flight phase). When the lift limitation is effective we are forced to ignore condition (7-64) and control the angle of attack at the borders of the limitation.

We have seen that during the aerodynamic flight phase with an effective lift limitation the angle of attack α is predetermined and the condition (7-63) is valid to determine the banking angle β . The condition (7-64), however, is just valid during flight phases when it does not violate the lift limitation. During the flight phases when the lift limitation is not effective we can replace the angles β and α by notations using the thrust angles ϕ and δ . Then we can consider the relationships (7-52) and transform the control laws (7-63) and (7-64) into the following conditions for $\tan \delta$ and $\tan \phi$:

$$\tan \phi = \frac{\lambda_u}{\lambda_v} ; \quad \tan \delta = \frac{C_1}{v (\lambda_u \sin \phi + \lambda_v \cos \phi)} \quad (7-65)$$

A replacement of the angles α and β is sensible for the flight in the vacuum of space where the incident wind is not present anymore, because the thrust angles δ and ϕ refer directly to the attitude of the moving coordinate system (ϕ with respect to the horizontal line and δ with respect to the actual flight plane). It is easy to verify that we would find the same conditions (7-65) when we used the thrust angles ϕ and δ from the beginning and formed the partial derivatives $\partial \mathcal{H} / \partial \phi = 0$ and $\partial \mathcal{H} / \partial \delta = 0$ of a notation of the Hamiltonian with the thrust angles ϕ and δ .

Initial conditions. We have to integrate the equations of motion (7-49) simultaneously with the Lagrange equations (7-55), therefore it is necessary that we adjust the initial values of the three Lagrange multipliers λ_u , λ_v , λ_r and the constant C_1 : the trajectory must comply with the desired final state variables u , v , r and ϑ . Actually it is necessary that we find only three correct initial conditions for these control variables, because the values of these multipliers in relation to each other are decisive (and not their absolute values). We can choose the initial value of one of these multiplies arbitrarily, for example $\lambda_u = 1$ at the initial time of the ascent trajectory. However, we are also forced to adjust the payload mass of the launcher appropriately, otherwise the numerical trajectory integration cannot satisfy the four final conditions of the target orbit (the trajectory maximizes the payload mass of the launcher).

Control of the thrust magnitude. Up to now we have considered the length of the thrust acceleration vector \vec{s} as a predetermined function of time. The thrust acceleration of a space launcher accepts a discontinuous behaviour when an engine is switched off or when a stage is separated. Sometimes it is necessary to switch-off the thrust completely ($s = 0$), for example when the space launcher needs a coasting flight period in order to reach the altitude of the final orbit or to reach the position above the equator for the injection into the geostationary transfer orbit. Usually, a coasting period takes place when the launcher has left the atmosphere and moves in the vacuum of space. The instant when the ignition of an engine terminates the coasting period is usually an optimization parameter (the beginning of the coast arc is also an optimization parameter when it is caused by the shut down of a re-ignitable engine).

Let us assume that the thrust acceleration can be controlled either by a throttling of liquid engines during the flight or by the determination of an appropriate burning profile for the solid booster motors before the flight. The launcher loses weight while it expends propellant mass; and the thrust acceleration s is often a critical load shortly before the separation of a stage. Then we have to limit the thrust acceleration:

$$s \leq s_{max}$$

When a throttling of engines is not possible and the stage is equipped with several engines, the thrust acceleration load is simply reduced by switching-off an engine. We have already seen that it is not sensible to reduce the thrust during the flight phase of high aerodynamic stresses in order to improve the payload capacity of the launcher (because the gravitational losses are usually higher than the drag losses). Let us consider now a reduction of the thrust acceleration $s(t)$ with the intention to limit the maximum of the aerodynamic pressure. The actual aerodynamic pressure Q , determined by the following expression, should be smaller than the value Q_{max} . Thus:

$$Q = \frac{1}{2} \rho w^2 \approx \frac{1}{2} \rho_0 e^{-(r - R_{earth})/h_{scale}} (u^2 + (v_0 - v)^2) \leq Q_{max} \quad (7 - 66)$$

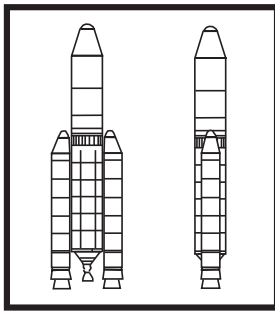
Initially, at the launch pad, the launcher experiences no aerodynamic pressure Q (because $w = 0$); and Q vanishes also in space (where $\rho = 0$); but Q causes serious loads at flight altitudes between 10 and 20 km. Often it is necessary that we observe a maximum permissible load Q_{max} carefully and reduce the thrust acceleration when the maximum value is reached. A constant value Q demands the following condition:

$$\dot{Q} = \frac{dQ(u, v, r)}{dt} = \frac{\partial Q}{\partial u} \cdot \dot{u} + \frac{\partial Q}{\partial v} \cdot \dot{v} + \frac{\partial Q}{\partial r} \cdot \dot{r} = 0 \quad (7 - 67)$$

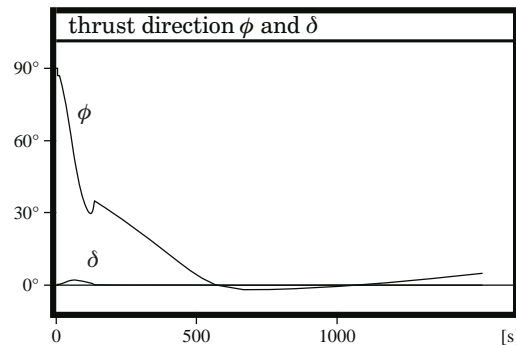
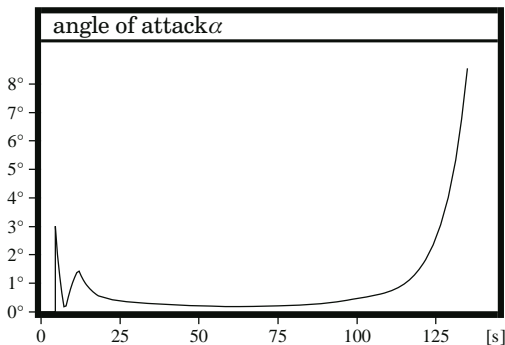
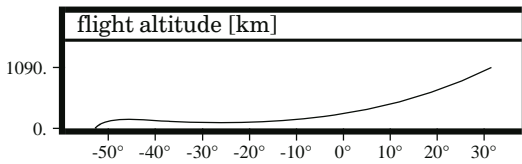
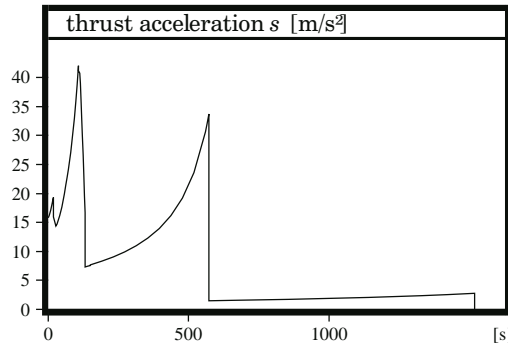
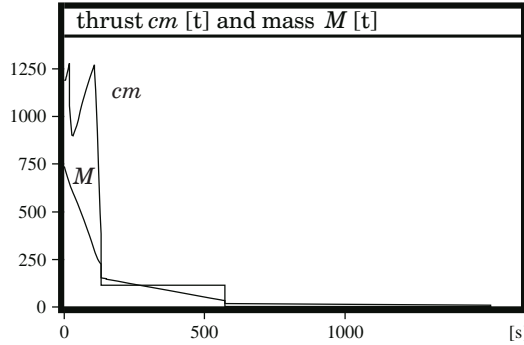
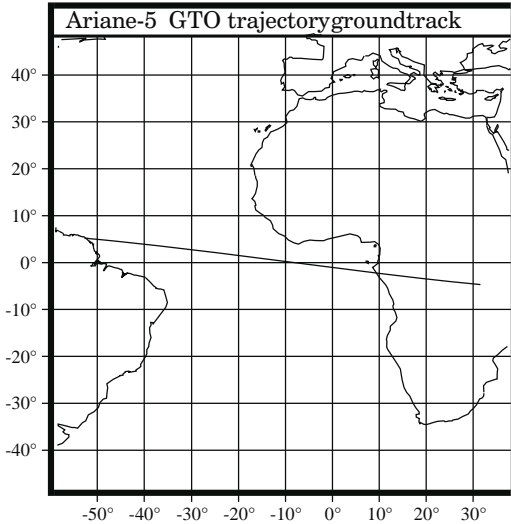
When we consider that the launcher thrusts during that time interval nearly exactly into the direction opposite to the incident wind, we can insert the equations of the motion (7-49) into the condition above and find a control function for the thrust acceleration s that keeps the aerodynamic pressure Q approximately constant:

$$s \approx \frac{u}{w} g + d + \frac{u w}{2h_{scale}} \quad (7 - 68)$$

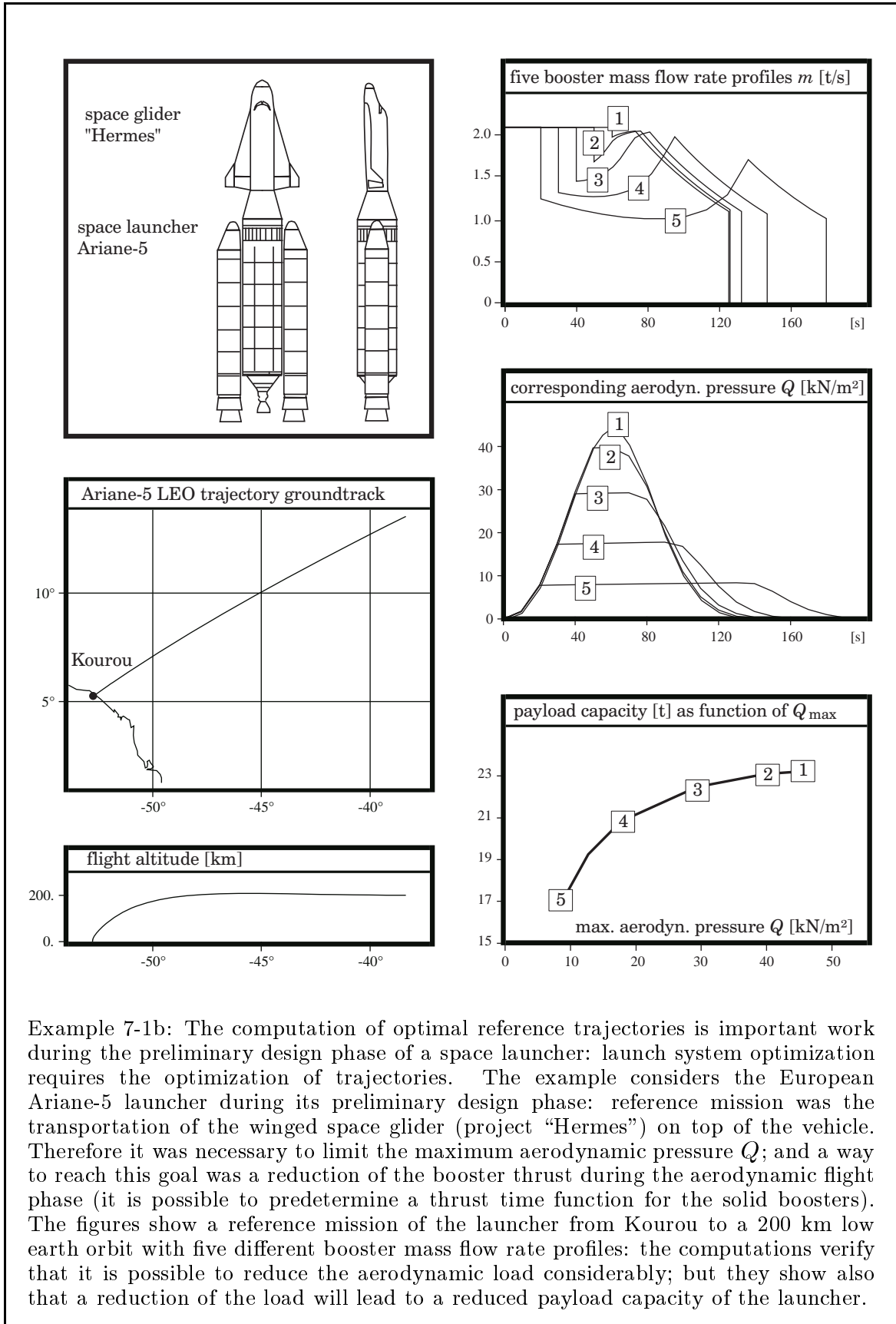
Term $g = 9.8066 \text{ m/s}^2$ is the gravitational acceleration and term d the actual drag deceleration; $h_{scale} = 8.333 \text{ km}$ is the value of the atmospheric scale height.



	core stage	boosters	upper stage
propellant	LOX/LH2	Solid	MMH/N2O4
ignition mass [t]	168	274	11.5
burnout mass [t]	15	36	3.1
thrust [kN]	801(sl)	5000(sl)	27(v)
c (vacuum) [m/s]	4310	2735	3100
c (sea level) [m/s]	3362	2407	
diameter [m]	5.4	3.1	5.4
height [m]	30.0	30.0	4.5



Example 7-1a: The figures show computational data of the ascent of an Ariane-5 launcher from Kourou (52.760° West; 5.230° North) to a geostationary transfer orbit (apogee at 35775 km, perigee at 300 km, $\vartheta = 7^\circ$). The boosters are dropped after 131 seconds at an altitude of 46 km; the payload shroud of 2.32 tons is discharged after 150 seconds at an altitude of 90 km. The launcher has a payload capacity of just 6.7 tons because it has to fly a funny curve to bring its core stage to a controlled re-entry in front of the coast of Peru. The lift limitation keeps the angle of attack extremely small ($\alpha < 0.3^\circ$) during the aerodynamic flight phase ($Q_{max} = 30 \text{ kN/m}^2$).



Example 7-1b: The computation of optimal reference trajectories is important work during the preliminary design phase of a space launcher: launch system optimization requires the optimization of trajectories. The example considers the European Ariane-5 launcher during its preliminary design phase: reference mission was the transportation of the winged space glider (project “Hermes”) on top of the vehicle. Therefore it was necessary to limit the maximum aerodynamic pressure Q ; and a way to reach this goal was a reduction of the booster thrust during the aerodynamic flight phase (it is possible to predetermine a thrust time function for the solid boosters). The figures show a reference mission of the launcher from Kourou to a 200 km low earth orbit with five different booster mass flow rate profiles: the computations verify that it is possible to reduce the aerodynamic load considerably; but they show also that a reduction of the load will lead to a reduced payload capacity of the launcher.

Velocity requirement for the ascent to orbit. It is well-known that the circular velocity on an orbit around the earth is a function of the altitude of this orbit. For example, the circular velocity amounts 7785 m/s at an altitude of 200 km above the surface; and the circular velocity amounts 3072 m/s at an altitude of 35872 km (geostationary position). A space launcher must have a higher Δv capability than what is necessary to establish the altitude and the velocity of the destination orbit. The table below lists the Δv distribution for a typical space launcher:

circular velocity at zero altitude:	\approx	7905 m/s
launch in eastern direction:	\approx	-464 m/s
Hohmann transfer to a 200 km orbit:	\approx	+121 m/s
gravitational losses:	\approx	+1500 m/s
drag losses:	\approx	+100 m/s
aerostatic engine losses:	\approx	+250 m/s
thrusting direction losses:	\approx	+100 m/s
<hr/>		
total Δv to LEO:	\approx	9512 m/s
transfer from LEO to GTO:	\approx	+2454 m/s
<hr/>		
total Δv to GTO:	\approx	11966 m/s
transfer from GTO to GEO:	\approx	+1477 m/s
<hr/>		
total Δv to GEO:	\approx	13443 m/s

The entire Δv capability which a space launcher needs to ascend to a certain orbit depends on the construction of the launcher and is a function of many parameters. The accurate value can only be determined by a computation of the ascent trajectory. However, we can assume in an approximation that the velocity requirement Δv is composed of the following parts: the launcher has to establish the circular velocity of 7905 m/s at the surface of the earth, but the velocity requirement for this maneuver is actually smaller because the launcher starts already with a velocity caused by the rotation of the earth (about 464 m/s). Additionally, the launcher has to establish the altitude of the parking orbit (the Δv requirement for a Hohmann transfer from sea-level to 200 km altitude amounts 121 m/s). The trajectory is subjected to gravitational losses, drag losses and engine losses (the engines work for some time inside the atmosphere where the exhaust velocity is smaller than in space); and the thrust is not always aligned with the optimal direction. Thus, a space launcher for low earth orbit transportation needs a Δv capability of about 9.5 km/s. Since a single rocket stage is usually not capable to reach such a high burnout velocity even when it carries no payload, LEO launchers are usually two stage vehicles or single stage vehicles supported by strap-on booster motors (sometimes, when the performance of the stages is bad, three or even four stages are used). A GTO launcher needs one more upper stage for having a Δv capacity of about 12 km/s. There are no drag losses in space. We can calculate the Hohmann transfer once again to find the velocity requirement for the transfer from LEO to GTO and from GTO to GEO. The gravitational losses are much smaller now, provided that the engine of the upper stage is not too weak.

7.2.2 Ascent Trajectories of Airborne Launchers

Benefits of using an aircraft as a launch platform. A launcher that ascends from the launch pad to its destination in space has to establish the velocity and the altitude of the target orbit; however, the mission involves losses (gravity losses, drag losses and a reduced efficiency of the engines due to the ambient aerostatic pressure). The launcher has already an initial velocity and an initial altitude when it is launched horizontally from a flying jet aircraft; and additionally the trajectory losses decrease greatly. The airborne launched vehicle has in fact the double payload capacity as a similar surface launched vehicle. Another benefit of the method of airborne launching is that the launch service provider is independent of a launch site in an equatorial region of the earth. However, the method of airborne launching involves two disadvantages: it requires the application of an expensive big jet aeroplane and wings for the initial pull-up maneuver of the rocket after the separation from the aircraft.

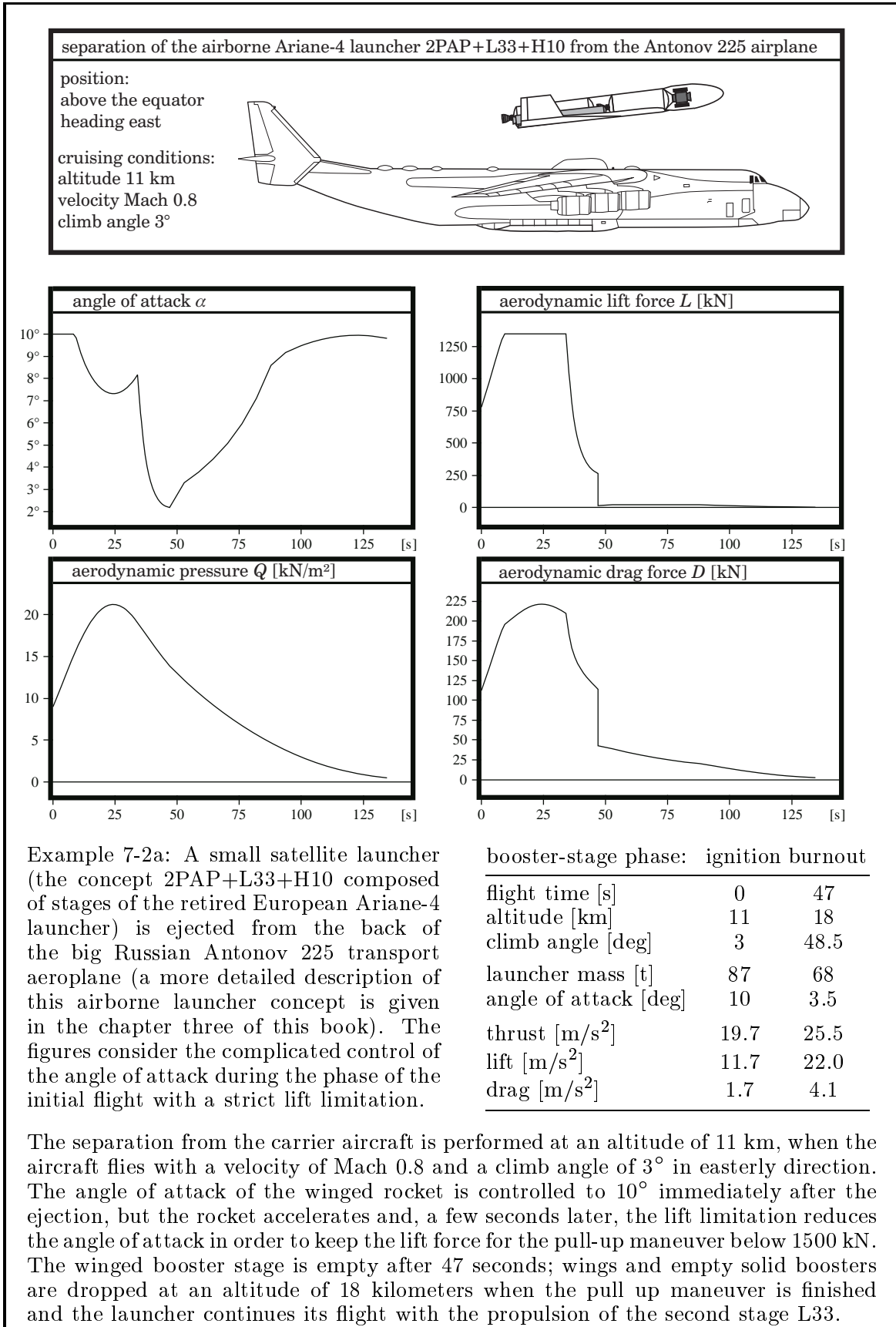
Wings for the initial pull-up maneuver. The wings cause a lift force and an additional (induced) drag force when the vehicle is flying with an angle of attack α (the lift is approximately a linear function of α and the induced drag is approximately a quadratic function of α); both forces are proportional to the aerodynamic pressure $Q = \frac{1}{2}\rho w^2$. We can write for the lift acceleration l and the drag deceleration d :

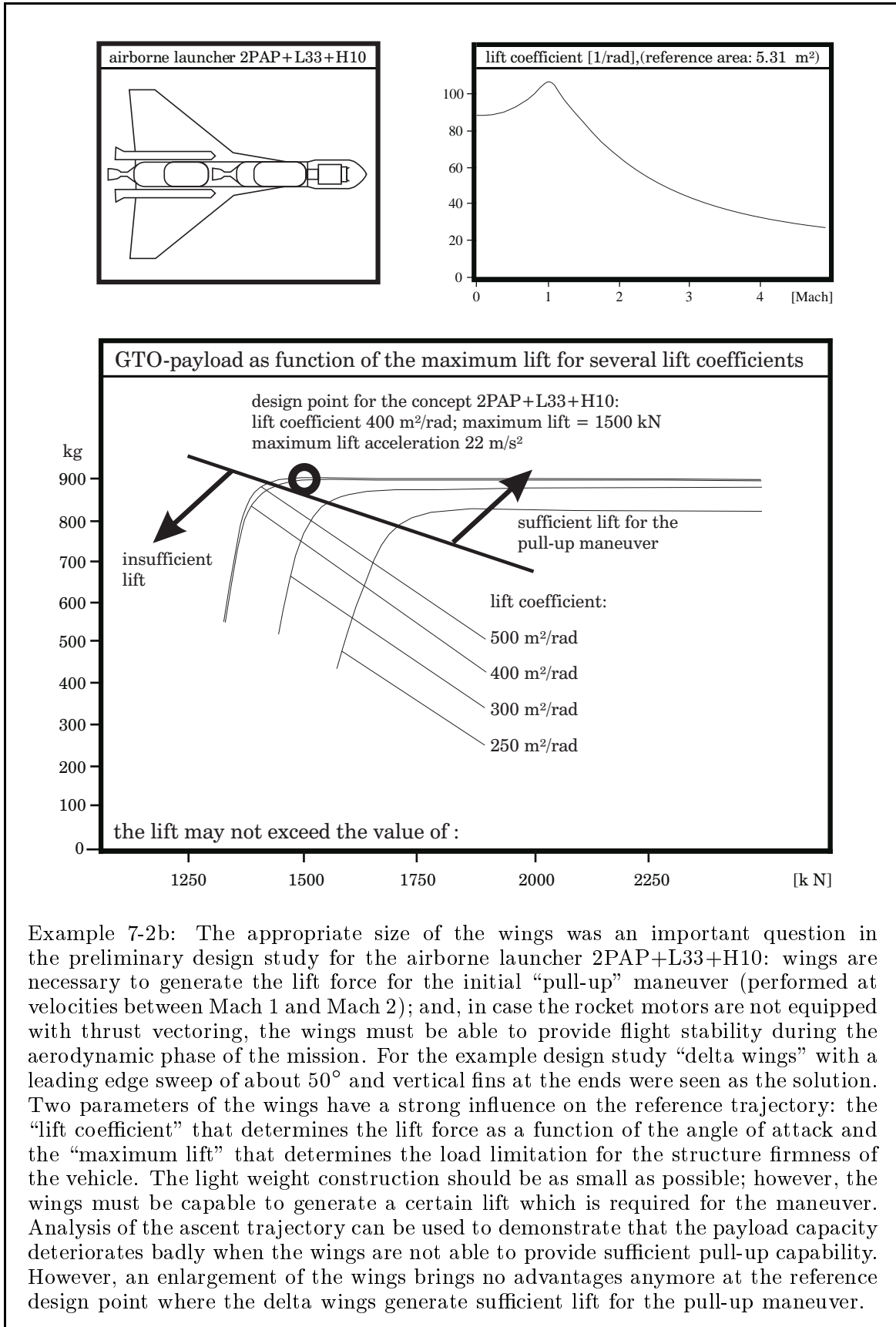
$$l = \frac{1}{2} \rho w^2 \frac{(C_{L\alpha} \cdot \alpha) A}{M} \quad (7-69)$$

$$d = \frac{1}{2} \rho w^2 \frac{(C_{D0} + C_{D\alpha} \cdot \alpha^2) A}{M} \quad (7-70)$$

The aerodynamic coefficients $C_{L\alpha}$, C_{D0} and $C_{D\alpha}$ are functions of the Mach number. Often these coefficients are inaccurately known during the preliminary design phase and later provided from wind channel experiments in form of diagrams or tables. For the calculation of the trajectory the vector form of the lift acceleration (7-53) and the drag deceleration (7-51) has to be implemented in the equations of motion (7-49).

Control of the initial pull-up maneuver. Let us assume that the carrier aircraft is capable to establish the appropriate initial conditions for the ascent trajectory of the airborne launcher. When a geostationary position is the final satellite destination, the ejection is performed exactly above the equator flying in easterly direction. In case the destination is an inclined or polar low earth orbit, however, the ejection is usually performed at a location distant from the equator; but at this position the aircraft must turn until it stays on a course with the correct heading angle. Immediately after the separation from the aircraft the airborne launcher has to start with the pull-up maneuver, therefore it flies with the maximum angle of attack and with a vanishing banking angle ($\alpha = \alpha_{max}$, $\beta = 0$). The lateral (pitch) acceleration is about 22 m/s²; and the trajectory control is determined by the maximum aerodynamic load that is allowed during this flight phase. The pull-up maneuver is finished when the launcher climbs with an angle of about 50° through the upper layers of the atmosphere. Now the lift force is not necessary anymore; the airborne launcher can drop the wings and continue its flight like a conventional surface launched vehicle.





Example 7-2b: The appropriate size of the wings was an important question in the preliminary design study for the airborne launcher 2PAP+L33+H10: wings are necessary to generate the lift force for the initial “pull-up” maneuver (performed at velocities between Mach 1 and Mach 2); and, in case the rocket motors are not equipped with thrust vectoring, the wings must be able to provide flight stability during the aerodynamic phase of the mission. For the example design study “delta wings” with a leading edge sweep of about 50° and vertical fins at the ends were seen as the solution. Two parameters of the wings have a strong influence on the reference trajectory: the “lift coefficient” that determines the lift force as a function of the angle of attack and the “maximum lift” that determines the load limitation for the structure firmness of the vehicle. The light weight construction should be as small as possible; however, the wings must be capable to generate a certain lift which is required for the maneuver. Analysis of the ascent trajectory can be used to demonstrate that the payload capacity deteriorates badly when the wings are not able to provide sufficient pull-up capability. However, an enlargement of the wings brings no advantages anymore at the reference design point where the delta wings generate sufficient lift for the pull-up maneuver.

7.3. Space Launcher Navigation

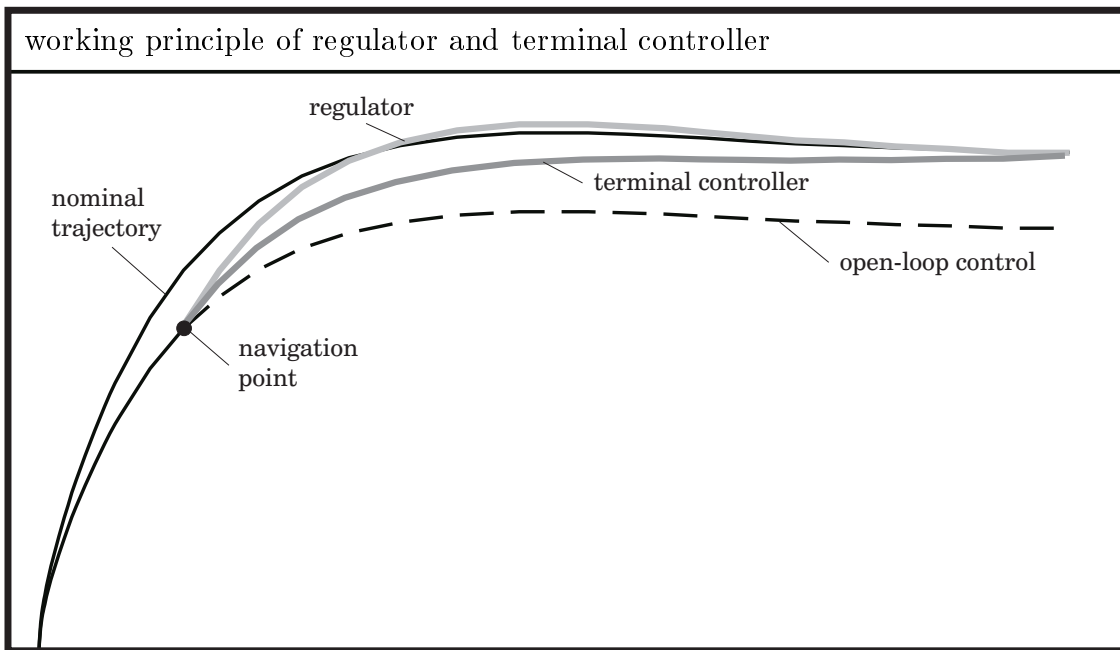
The ability to calculate optimal ascent trajectories is particularly important during the preliminary design phase of a new space launcher, because in this phase the trajectory optimization serves as a basis for the optimization of the launcher system. When the construction of the launch system has been defined, the trajectory optimization serves as a basis for the implementation of a navigation strategy. For the real flight mission it is not sufficient to control the thrust vector in accordance with the data of the precalculated optimal trajectory: the real trajectory of the launcher would soon deviate substantially from the precalculated way, because the mathematical model is never 100% accurate and the real mission is influenced by effects which are not known before the mission (the weather on the launch date, for example). The duty of the navigation strategy is to keep the launcher close to the precalculated reference trajectory.

7.3.1 Guidance Systems

Stabilization. Let us consider a space launcher that is actually on its way to space. The launcher should follow its optimized reference trajectory; therefore it is necessary that the vehicle has been equipped with an attitude control system that stabilizes at every instant the predetermined reference attitude. The control of the equilibrium state of the attitude is called “stabilization” of the launcher. An inertial sensor measures the actual attitude; the attitude control system compares the actual attitude with the reference attitude and finds control functions for the appropriate thrust vectoring (it is very important that the attitude stabilization system compels compliance with the correct attitude at every instant of the flight, because the launcher is usually destroyed within a few seconds when the attitude is lost).

Navigation. The sensor of the navigation system measures time, actual position and velocity and transfers these data to the navigation computer (the sensor for measuring these data is usually an inertial platform). The computer of the navigation system compares the measured data with the stored data of the precalculated trajectory. In case that deviations from the nominal trajectory are detected, the navigation computer has to determine a new reference attitude for the attitude control system. The method of detecting deviations from the actual position and finding the appropriate way back to the nominal trajectory is called “navigation” of the launcher: the on-board computer regulates the change of operating point conditions in order to recover the reference trajectory (“piloting and guidance”). In this connection let us define two more terms which are often used in the field of control engineering: we say that the trajectory is controlled “open-loop” when the navigation system ignores disturbances and controls the thrust angles in accordance to the reference trajectory; we say that the trajectory is controlled “closed-loop” when the navigation system measures deviations and tries to re-establish the precalculated reference trajectory.

Regulators and terminal controllers. Thus, in order to handle the disturbances which occur during the real flight mission, it is necessary that we implement a closed-loop guidance strategy into the navigation computer system of the space launcher. Such a strategy is called “regulator” when it keeps the launcher close to the reference trajectory during the entire flight time; the strategy is called “terminal controller” when it brings the launcher close to the reference trajectory only at the final instant of the flight time. Regulators are only preferable when it is necessary to avoid substantial deviations from the reference trajectory all the time; otherwise there is no reason to guide the launcher back to the reference trajectory. It is important that the navigation strategy brings the launcher as near as possible to the final destination orbit; and it should be avoided that the navigation strategy consumes a lot of propulsion. A disturbance that occurs at a certain instant of the mission has a negative effect on the accuracy of the payload delivery; however, the negative effect can be annihilated during the entire following flight time. An “asymptotic optimal regulator”, in fact nothing else than a terminal controller, regulates only the final state of the trajectory and is conditioned for saving propulsion in doing this task.



Trajectory re-optimization. A simple but naive idea for the implementation of a terminal controller is to let the navigation computer update the reference trajectory at certain moments by finding always totally new optimized trajectories for the remaining flight time. Even though this idea is convincing at the first look, the practical application involves serious problems: first, it is maybe difficult for the on-board computer to calculate during the mission optimal trajectories (the navigation algorithm should be “fast and robust”); second, the optimization objective is not a “maximization of the final mass” anymore but a “minimization of the final deviation”. It is obviously impossible to change the payload mass during the mission; however, it is essential that the payload reaches safely and accurately its destination orbit.

Iterative path adaptive guidance. The navigation method that was used in the Saturn launch vehicles of the American moon landing programme can be seen as a trajectory re-optimization strategy, even though the simple steering law that was used is actually not the optimum. We have seen that the optimal control law for the thrust angle ϕ is approximately a linear function of time when the launcher flies in the vacuum of space over a flat (non-spherical) earth. Therefore equation (7-7) states:

$$\phi \approx \tan \phi = C_o - C_1 \cdot t \quad (7-71)$$

Here term C_1 is the constant “tilt-rate” and C_o is the thrust angle ϕ at the initial instant when $t = 0$. The launcher cannot use this control law for its flight through the atmosphere, because the lift limitation forces the vehicle to follow a “gravity tilt trajectory” with an extremely small angle of attack. Consequently, the launcher is controlled “open-loop” during its atmospheric flight phase. When the launcher leaves the atmosphere the thrust angle ϕ is already small, and the law (7-71) constitutes a good approximation for the steering of the optimal thrust direction in the actual flight plane. However, now it is necessary to update the constants C_o and C_1 (because perturbations deformed the precalculated trajectory); and the update can be done “on-board” by an integration of the following flight phase (therefore the Saturn launchers used simplified equations of motions to approximate the flight over the oblate earth with a realistic gravitational field). The constants C_o and C_1 have to be adjusted in a way that the “on-board” trajectory integration complies with the desired final values of altitude and velocity. The method is called “iterative path adaptive guidance”, because the thrust angle ϕ is continuously updated during the non-atmospheric flight.

However, the control law (7-71) cannot be used anymore during the final flight phase. Near burnout, shortly before the vehicle reaches orbital velocity, the navigation method would give a large thrust angle even for a small altitude error. The altitude of the destination orbit is reached early before burnout, and the altitude constraint can be dropped during the navigation of the final flight period. Approximately one minute before the cutoff the launcher replaces the equation (7-71) by an even simpler law:

$$\phi = C_o \quad (7-72)$$

The turning rate of the vehicle is small near the final orbit; and the navigation strategy ignores the rate term C_1 and concentrates on the adjustment of the final velocity.

The updating of the thrust angle ϕ is continued until about 10 seconds before the final cutoff, but during the last seconds the thrust angle is kept constant. Finally, a simple velocity (or energy) presetting triggers the cutoff of the engines. Even though this path adaptive guidance strategy is extremely simple, the Saturn-5 launch vehicles could achieve a typical injection accuracy of better than half a kilometer in altitude and better than 1 m/s in velocity for missions to low earth orbit (and the guidance system could cope with a premature shut-down of an engine of the second stage that occurred during the flight of Apollo 13). However, the disadvantage of the iterative path adaptive guidance strategy is that this method makes the space launcher follow an ascent trajectory which is quite far away from the real performance optimum.

Linear proportional terminal control. A more robust and even simpler strategy than the re-optimization algorithm is called the “linear terminal controller”. The linear proportional terminal controller regulates the final state of the trajectory and pays attention to the limits in which the thrust vector is controllable. The method works as follows: we have to find time functions for certain navigation coefficients before the actual flight and implement these functions in the on-board navigation computer. When the navigation computer senses deviations in the actual values of position and velocity, it simply multiplies these deviations with the navigation coefficients to get the control data for the correction of the nominal attitude. Thus, when r is the altitude, u the vertical velocity, v the horizontal velocity and ϑ the inclination angle, then the terms Δr , Δu , Δv and $\Delta\vartheta$ express the actual deviation from the nominal trajectory at a certain navigation point (time t). The nominal attitude of the vehicle is determined by the thrust direction, for example using the thrust angles ϕ and δ (or α and β). The thrust angle δ (out of the actual flight plane) is mainly responsible for the control of the inclination; the thrust angle ϕ (in the actual flight plane) is mainly responsible for the control of the altitude and the velocity (we can at least in an approximation consider the control of the inclination as independent of the control of the altitude and the velocity). Then the guidance system simply calculates the navigation thrust angles $\Delta\delta$ and $\Delta\phi$ and adds these angles to the nominal attitude (defined by the nominal thrust direction δ and ϕ). Thus:

$$\Delta\delta = N_{\vartheta} \cdot \Delta\vartheta \quad (7 - 73)$$

$$\Delta\phi = N_r \cdot \Delta r + N_u \cdot \Delta u + N_v \cdot \Delta v \quad (7 - 74)$$

Terms N_{ϑ} , N_r , N_u and N_v are the navigation coefficients. These coefficients are functions of the construction of the launcher; and they are functions of the individual navigation strategy that we want to implement in the guidance system of the launcher. However, these navigation coefficients are not functions of actual perturbations; and we can calculate them before the mission as addition to the nominal trajectory.

Once we have found the navigation coefficients, the stability and the accuracy of this navigation strategy can be tested with a computer that simulates the perturbed flight of the space launcher. Later, when experience with the navigation strategy is present, we can easily improve the linear terminal controller by adding navigation coefficients which consider also terms of nonlinear deviations from the reference trajectory.

Navigation accuracy. The recovery of the reference trajectory (or the regulation of the final state of the trajectory) by a closed-loop navigation strategy has an influence on the optimal performance of the launcher. However, the deviation of the nominal thrust direction in order to compensate perturbations cannot lead to a reduced payload capacity, because it is practically impossible to change the payload mass during the flight. Instead of this, the reduced performance can be transformed into a reduced velocity at the final instant of the reference trajectory; and the reduced final velocity can be made up by additional (contingency) propulsion. Alternatively, in case the launcher uses a solid upper stage where additional propulsion is not available, there is also the option to reduce the altitude of the destination orbit during the flight.

7.3.2 The Lambda-Matrix Method

Variation of state variables and control variables. The problem is now how we can calculate the time functions for the navigation coefficients for such a linear proportional terminal controller (equations 7-73 and 7-74). The following section explains a method that has become known in the literature as the “lambda matrix method” (referring to the Lagrange multipliers λ of the trajectory optimization theory):

We can consider a perturbation at a certain instant of the flight (navigation point t) as a deviation of the vector of the state variables $\Delta\vec{x}$ and a deviation of the vector of the control variables $\Delta\vec{q}$. The actual perturbation will later have an influence on the final state variables \vec{x}_{end} . There is approximately a linear relationship between the perturbation and the final deviation $\Delta\vec{x}_{end}$ for small perturbations. We can write:

$$\Delta\vec{x}_{end} = \left[\frac{\partial x_{end}}{\partial q} \right] \Delta\vec{q} + \left[\frac{\partial x_{end}}{\partial x} \right] \Delta\vec{x} \quad (7-75)$$

Our intention is to find a control vector $\Delta\vec{q}$ that lets the final deviation $\Delta\vec{x}_{end}$ vanish. Therefore we have to invert the matrix $[\partial x_{end}/\partial q]$. The procedure is just possible when the matrix is not singular and contains the same number of rows and columns. For example, we need three actual control variables \vec{q} when we want to adjust three final state variables $\Delta\vec{x}_{end} = 0$. When it is possible to invert the matrix we can write:

$$\Delta\vec{q} = - \left[\frac{\partial x_{end}}{\partial q} \right]^{-1} \cdot \left[\frac{\partial x_{end}}{\partial x} \right] \cdot \Delta\vec{x} \quad (7-76)$$

The exponent -1 in the equation above indicates the inverted matrix that determines the behaviour of the final deviation of the trajectory in the linear vicinity as a function of variations of the actual control variables. If we used the thrust angles ϕ and δ as the actual control variables we were just able to adjust two final state variables. The “lambda-matrix method” is preferable because it uses the Lagrange multipliers as control variables (these multipliers λ determine the thrust direction directly, the elements of vector λ correspond to the thrust angles and their time derivatives).

Trajectory control with the Hamilton-Lagrange theory. Let us briefly recall the optimization theory that we have used in the preceding section of this chapter (therefore we list all the important equations in their abstract mathematical form):

$$\begin{aligned} \dot{x}_i &= f_i(\vec{x}, \vec{q}, t) \quad ; \quad i = 1, n \\ \mathcal{H} &= \sum_{i=1}^n \lambda_i f_i \\ \dot{\lambda}_i &= - \frac{\partial \mathcal{H}}{\partial x_i} \quad ; \quad 0 = \frac{\partial \mathcal{H}}{\partial q} \end{aligned} \quad (7-77)$$

We constructed the Hamiltonian \mathcal{H} (7-54) to the equations of motion (7-49) and derived the adjointed system of Lagrange equations (7-55 and 7-59). The control of the thrust direction was then defined by the equations (7-63) and (7-64) (or 7-65).

The “lambda-matrix-method” regulates the thrust direction indirectly by an update of the Lagrange multipliers. Let us first consider the control of the inclination and assume that the control of the inclination is independent of the control of the altitude and velocity. The actual value of the multiplier λ_ϑ influences the the thrust angle δ ; and δ is the thrust angle that guides the launcher into the desired orbital inclination. We find the navigation coefficient N_ϑ for the control of the inclination ϑ_{end} simply as:

$$\begin{aligned} \Delta\delta &= \frac{\partial\delta}{\partial\lambda_\vartheta} \cdot \Delta\lambda_\vartheta ; & \Delta\lambda_\vartheta &= -\frac{\partial\vartheta_{end}/\partial\vartheta}{\partial\vartheta_{end}/\partial\lambda_\vartheta} \cdot \Delta\vartheta \\ \rightarrow N_\vartheta &= -\frac{\partial\delta}{\partial\lambda_\vartheta} \cdot \frac{\partial\vartheta_{end}/\partial\vartheta}{\partial\vartheta_{end}/\partial\lambda_\vartheta} \quad \left(= -\frac{\partial\delta}{\partial\vartheta_{end}} \cdot \frac{\partial\vartheta_{end}}{\partial\vartheta} \right) \end{aligned} \quad (7-78)$$

The control of the altitude and velocity is more complicated since the thrust angle ϕ is determined by the ratio of λ_u and λ_v (equation 7-65); the multiplier λ_r , however, has no influence on the actual thrust direction (the derivative $\partial\phi/\partial\lambda_r$ equals zero). We can use a variation of λ_v to determine a variation of the thrust direction $\Delta\phi$:

$$\Delta\phi = \frac{\partial\phi}{\partial\lambda_v} \cdot \Delta\lambda_v \quad (7-79)$$

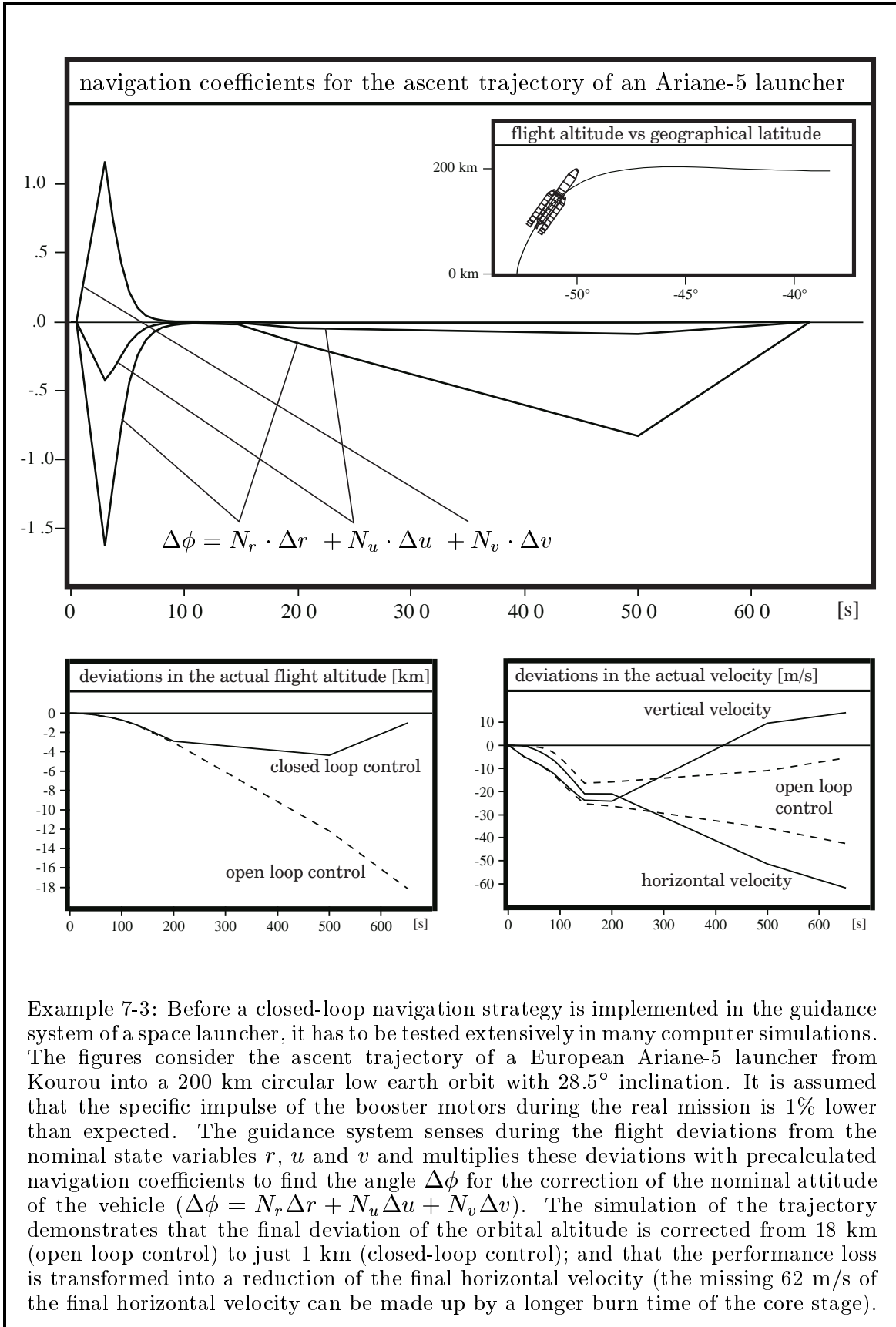
Now we have to calculate the deviations of the final state variables u_{end} and r_{end} (or alternatively u_{end} and v_{end}) as functions of a variation of the actual state variables u , v and r and as functions of the actual control variables λ_v and λ_r . Then we form:

$$\begin{pmatrix} \Delta\lambda_v \\ \Delta\lambda_r \end{pmatrix} = - \begin{pmatrix} \frac{\partial u_{end}}{\partial\lambda_v} & \frac{\partial u_{end}}{\partial\lambda_r} \\ \frac{\partial r_{end}}{\partial\lambda_v} & \frac{\partial r_{end}}{\partial\lambda_r} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial u_{end}}{\partial u} & \frac{\partial u_{end}}{\partial v} & \frac{\partial u_{end}}{\partial r} \\ \frac{\partial r_{end}}{\partial u} & \frac{\partial r_{end}}{\partial v} & \frac{\partial r_{end}}{\partial r} \end{pmatrix} \cdot \begin{pmatrix} \Delta u \\ \Delta v \\ \Delta r \end{pmatrix} \quad (7-80)$$

Finally, we find after some transformations the navigation coefficients N_u , N_v and N_r :

$$\begin{aligned} N_u &= \frac{\partial\phi}{\partial\lambda_v} \cdot \left(\frac{\partial u_{end}}{\partial u} \cdot \frac{\partial r_{end}}{\partial\lambda_r} - \frac{\partial r_{end}}{\partial u} \cdot \frac{\partial u_{end}}{\partial\lambda_r} \right) / \left(\frac{\partial u_{end}}{\partial\lambda_v} \cdot \frac{\partial r_{end}}{\partial\lambda_r} - \frac{\partial u_{end}}{\partial\lambda_r} \cdot \frac{\partial r_{end}}{\partial\lambda_v} \right) \\ N_v &= \frac{\partial\phi}{\partial\lambda_v} \cdot \left(\frac{\partial u_{end}}{\partial v} \cdot \frac{\partial r_{end}}{\partial\lambda_r} - \frac{\partial r_{end}}{\partial v} \cdot \frac{\partial u_{end}}{\partial\lambda_r} \right) / \left(\frac{\partial u_{end}}{\partial\lambda_v} \cdot \frac{\partial r_{end}}{\partial\lambda_r} - \frac{\partial u_{end}}{\partial\lambda_r} \cdot \frac{\partial r_{end}}{\partial\lambda_v} \right) \\ N_r &= \frac{\partial\phi}{\partial\lambda_v} \cdot \left(\frac{\partial u_{end}}{\partial r} \cdot \frac{\partial r_{end}}{\partial\lambda_r} - \frac{\partial r_{end}}{\partial r} \cdot \frac{\partial u_{end}}{\partial\lambda_r} \right) / \left(\frac{\partial u_{end}}{\partial\lambda_v} \cdot \frac{\partial r_{end}}{\partial\lambda_r} - \frac{\partial u_{end}}{\partial\lambda_r} \cdot \frac{\partial r_{end}}{\partial\lambda_v} \right) \end{aligned} \quad (7-81)$$

The coefficients above used in the equation (7-74) adjust the final values u_{end} and r_{end} ; they transfer the performance loss to a deteriorated final horizontal velocity v_{end} (assuming that the launcher makes up the missing velocity by an extra burn time). When we want to adjust the final horizontal velocity instead of the final altitude we have simply to replace the term r_{end} by the term v_{end} in the systems (7-80) and (7-81). Using the lambda-matrix method we can also adjust other trajectory parameters at instances different from the final instant (for example, we can adjust the apogee altitude exactly above the equator for a mission to the geostationary transfer orbit). However, a control strategy has no effect on the injection accuracy anymore when the launcher is already near the final orbit; and finally the method must be suspended (or replaced by another method) because the equation system (7-81) gets singular.

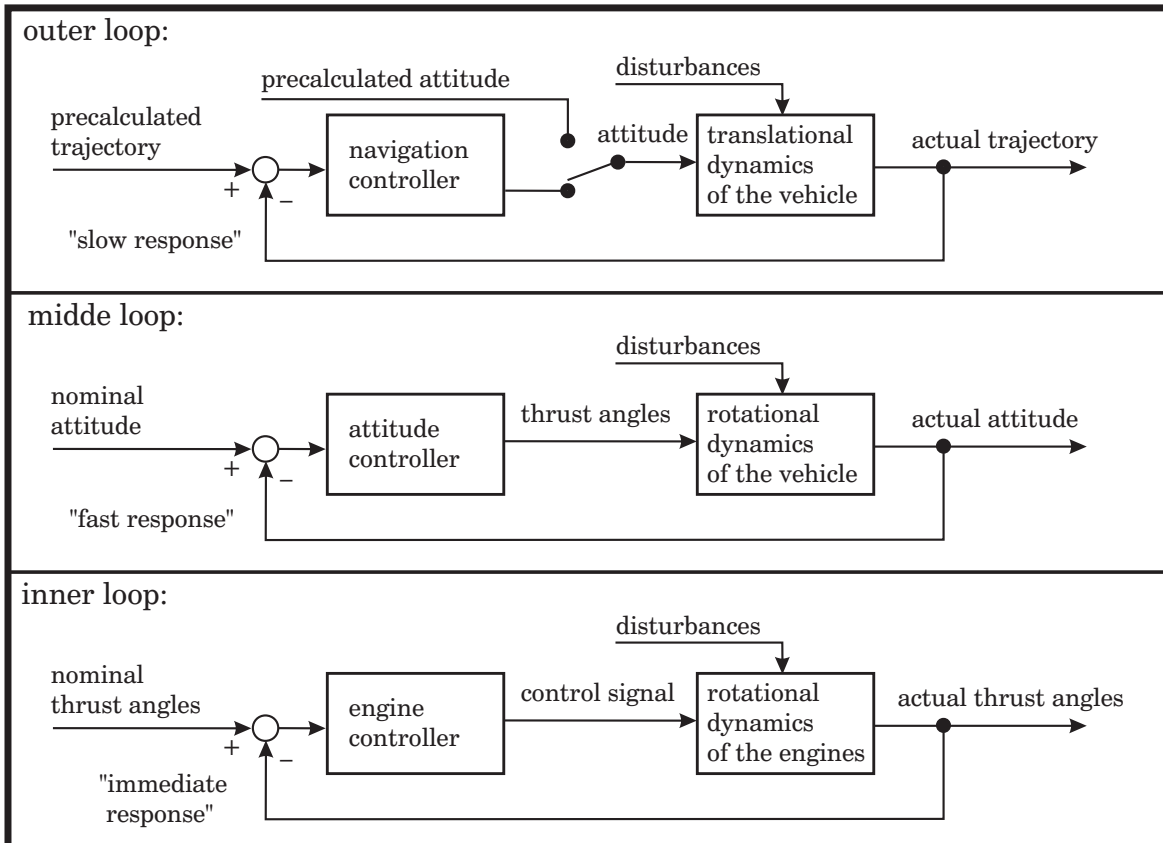


7.4. Space Launcher Attitude Stabilization

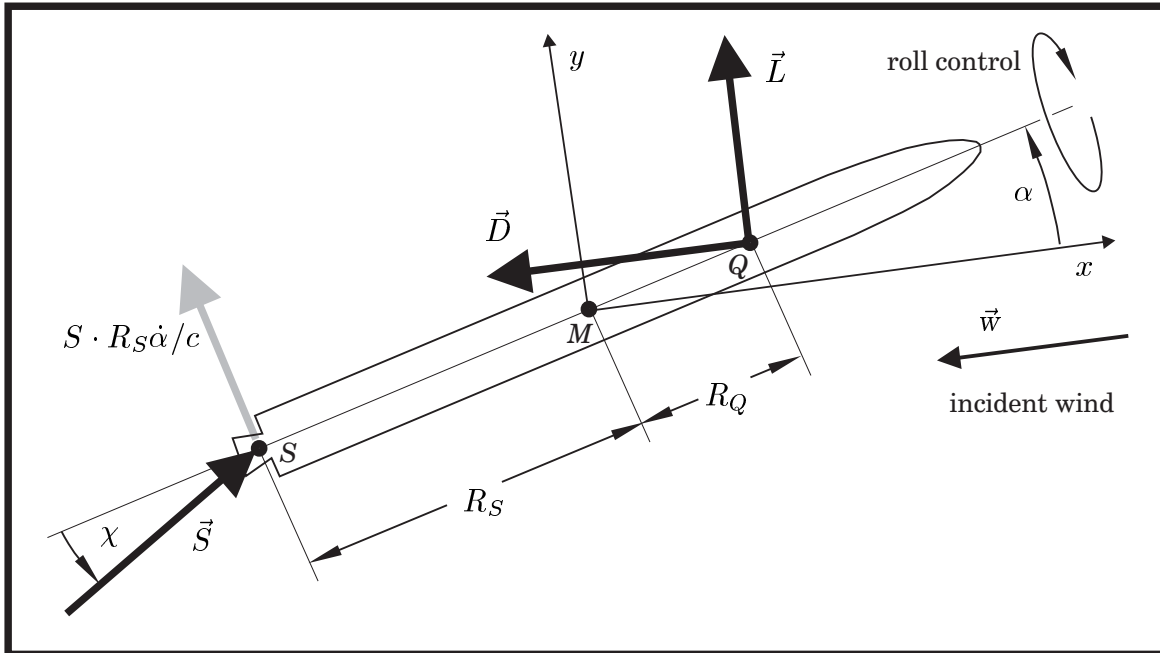
The ascending launcher is a material body with finite dimensions and thus it can simultaneously translate and rotate. Both, translational motion and rotational motion, are controlled by an appropriate steering of the thrust vector (“thrust vectoring”). The content of the last section of this chapter is the rotational motion of conventional space launchers and the control of this motion.

7.4.1 Control of the Angular Motion

Cascade control. The piloting system of a space launcher is actually a cascade controller that consists of three merged control loops. Its outer loop realizes the navigation strategy and has a comparatively slow response (we have seen that the guidance strategy can switch between “open-loop control” for the atmospheric flight phase and “closed-loop control” for the accurate navigation of the final flight phase). The inner control loop steers the actuators of the engines. However, the moment of inertia of the engines is usually very small in comparison with the rest of the vehicle; and, provided that the engines react much faster than the vehicle, we can assume an immediate response of the thrust direction and ignore the rotational dynamics of the engines in the calculation of the rotational dynamics of the vehicle. Then the attitude dynamics of the vehicle is determined by the middle loop of the piloting system.



Rotational dynamics of the vehicle. In the analysis of the attitude control loop of a space launcher we assume that the nominal attitude (determined by the angle of attack and the banking angle, for example) changes just slowly in time. Also the mass of the vehicle, the location of the center of mass and the moment of inertia about the center of mass are constant during the short time interval that is now under consideration. Then we place a coordinate system on the center of mass of the launcher and align the x -axis with the nominal direction of the incident wind (the opposite drag direction) and the y -axis with the nominal direction of the lift. The coordinate system keeps its direction with respect to inertial space during the short time interval under consideration; and it is an inertial system since we disregard the gravitational attraction and the centrifugal force along the ballistic trajectory (the coordinate system can be used to describe the rotational motion with respect to inertial space and the translational motion with respect to a ballistic trajectory).



The point of application Q of the aerodynamic forces drag \vec{D} and lift \vec{L} is usually located in front of the center of mass M (displacement R_Q); and consequently the non-powered motion is unstable when the launcher climbs through the atmosphere. The aerodynamic torque is approximately a linear function of the angle of attack α . For aerodynamic stability we could either shift the center of the aerodynamic pressure Q back using fins at the rear end, or we could shift the point M ahead using ballast at the front. However, every additional weight deteriorates the performance of a space launcher considerably, and therefore it is much better to establish stability of the attitude via the method of thrust vectoring: the point of application S of the thrust force \vec{S} is located at the rear end of the vehicle (distance R_S); and when we deviate the thrust force \vec{S} from the centerline of the vehicle it generates a torque about the center of mass M . For flight stability it is necessary that the control torque caused by the thrust misalignment (angle χ) is greater than the aerodynamic moment.

Jet damping. The rear end of the launcher accepts a velocity component $R_S \cdot \dot{\alpha}$ when the launcher rotates on the z -axis with the rotational velocity $\dot{\alpha}$; and, as a consequence, a small momentum is exchanged between the launcher and the propellant that leaves the nozzle. The thrust force S is simply the product of mass flow rate m and exhaust velocity c , thus $S = m \cdot c$. We can see that the force caused by the rotation of the exhaust plume $S_{jet\ damping} = m \cdot R_S \dot{\alpha} = S \cdot R_S \dot{\alpha} / c$ is much smaller than the thrust S . This small force is called “jet damping” because it acts against the direction of the rotational motion; but it is so small that we can neglect it.

Roll control. The rotation of the vehicle on its symmetry axis is indifferent (neither stable nor unstable), and obviously the roll angle will drift away from its nominal value when a small disturbing torque acts on this axis. The launcher needs the ability to control this angle. However, the required control torque is quite small; and in case of several engines the roll control can be performed via an opposite thrust misalignment of two parallel working engines. In case of a single gas generator cycle engine the roll angle is usually controlled via an appropriate pivoting of the off-center gas-generator exhaust plume (the plume is normally parallel with the thrust direction).

Attitude control. The control of the attitude of the launcher with respect to the incident wind is more critical, because the aerodynamic torque tries to tilt the vehicle. The launcher must establish and maintain a certain angle of attack and therefore deviate the thrust from the centerline in order to generate a counteracting torque. The angular motion of the launcher on the y -axis and on the z -axis is unstable (without active control a small attitude disturbance would grow rapidly); and control of the rotation on these two axes is required. Thrust vectoring is usually performed by a gimbal deflection of the engines or their nozzles. A typical value for the thrust vectoring capability is a deflection of the misalignment angle χ between $\pm 10^\circ$ (in horizontal and vertical direction) and a misalignment velocity of $\dot{\chi} \approx 10^\circ/\text{second}$.

Equations of motion. We use Newton’s law to establish the equations of motion: the time derivative of the angular momentum vector is equal to the vectors of all external torques; and the time derivative of the linear momentum vector is equal to the vector of all external forces. Then we get the equations for the rotation on the z -axis and for the translation in x and y -direction (multiplied by the launcher mass M) as:

$$\mathfrak{S} \cdot \ddot{\alpha} = L \cdot R_Q \cos \alpha + D \cdot R_Q \sin \alpha - S \cdot R_S \sin \chi \quad (7-82)$$

$$M \cdot \ddot{x} = -D + S \cdot \cos(\alpha + \chi) \quad (7-83)$$

$$M \cdot \ddot{y} = L + S \cdot \sin(\alpha + \chi) \quad (7-84)$$

You can see that we still consider the launcher as a rigid body; the term \mathfrak{S} is the moment of inertia for the rotation on the z -axis. We can derive a similar (or identical) equation system for the rotation on the y -axis and the translation in z -direction; however, for small aerodynamic incident angles this other equation system is decoupled from the equation system (7-82) and (7-84). The launcher oscillates simultaneously on the z -axis and on the y -axis; but let us now analyze the stability and control of the rotation on the z -axis as a representative for the angular motion of a space launcher.

The attitude controller. The aerodynamic torque that acts on the center of mass of the vehicle is a linear function of the angle of attack α (provided that the angle of attack remains small, for example below 10°); and we can abbreviate the aerodynamic torque as $L \cdot R_Q \cos \alpha + D \cdot R_Q \sin \alpha = \mathcal{M}_\alpha \cdot \alpha$. The restoring torque is a nearly linear function of the thrust misalignment angle χ (for small angles $\sin \chi \approx \chi$). We can include a perturbation $\mathcal{M}_{disturbance}$ and rewrite the equation (7-82) as:

$$\mathfrak{S} \cdot \ddot{\alpha} = \mathcal{M}_\alpha \cdot \alpha - R_S S \cdot \chi + \mathcal{M}_{disturbance} \tag{7-85}$$

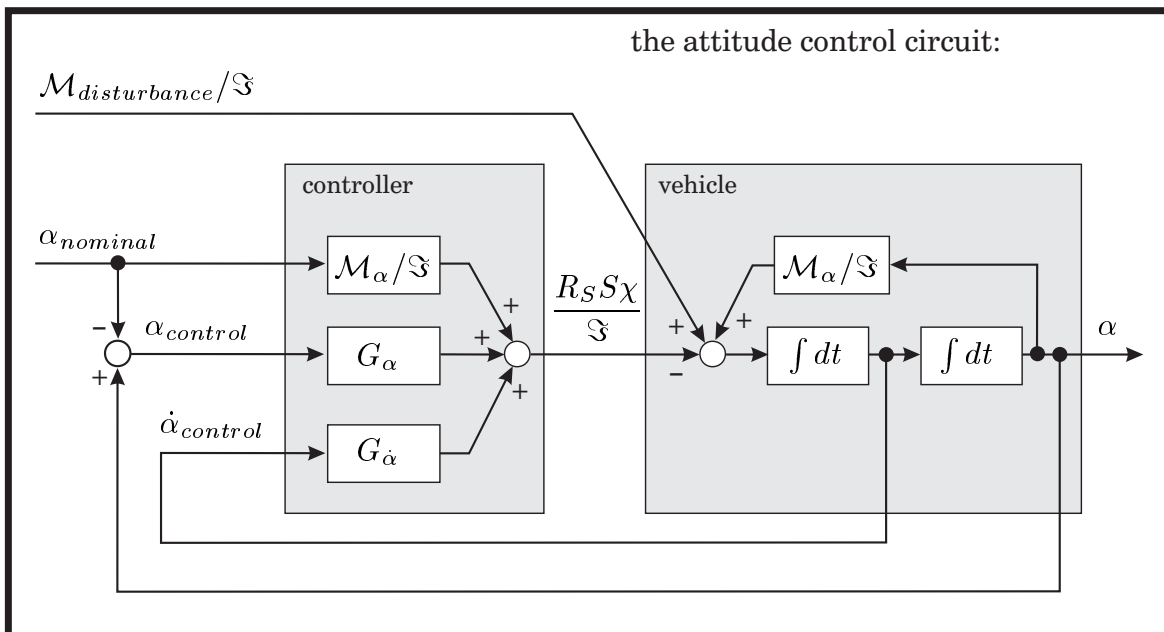
The term \mathcal{M}_α in the equation above is the unit aerodynamic moment and, in case \mathcal{M}_α is positive and the control moment $S R_S \cdot \chi$ is zero, the equation determines an exponentially growing function of the angle of attack α for the excitement with an infinitesimally small perturbation torque $\mathcal{M}_{disturbance}$. The nominal angle of attack is always small but usually not exactly zero (the navigation controller requires that the launcher flies with a small angle of attack); and we can write the angle α as:

$$\alpha = \alpha_{nominal} + \alpha_{control} \tag{7-86}$$

During the short time interval that is under consideration the term $\alpha_{control}$ is rapidly changing; the term $\alpha_{nominal}$, however, is constant and its time derivatives equal zero. Actually, the nominal attitude of the vehicle changes just slowly during the entire flight time, and the attitude control system is rather a disturbance controller than a follow-up controller. Thus, the misalignment of the thrust force generates an open-loop control moment $\mathcal{M}_\alpha \cdot \alpha_{nominal}$ that establishes the nominal attitude and a closed-loop control moment that stabilizes the nominal attitude. We can write:

$$R_S S \cdot \chi = \mathcal{M}_\alpha \cdot \alpha_{nominal} + \mathfrak{S}(G_\alpha \cdot \alpha_{control} + G_{\dot{\alpha}} \cdot \dot{\alpha}_{control}) \tag{7-87}$$

The terms G_α and $G_{\dot{\alpha}}$ are the so-called “gain factors” for α and its time derivative $\dot{\alpha}$. The following connection diagram illustrates the working principle of the controller:



The controller that we analyze in this section for the rotational stability of the launcher is a linear controller with proportional and differential behaviour. We can insert the control torque (7-87) into the equation of motion (7-85) to find the condition:

$$\ddot{\alpha}_{control} + G_{\dot{\alpha}} \cdot \dot{\alpha}_{control} + (G_{\alpha} - \mathcal{M}_{\alpha}/\mathfrak{S}) \cdot \alpha_{control} = \mathcal{M}_{disturbance}/\mathfrak{S} \quad (7-88)$$

This second order linear differential equation describes a harmonic oscillation for the control angle $\alpha_{control}$ that is excited by a perturbation, the function $\mathcal{M}_{disturbance}/\mathfrak{S}$.

The steady state performance. The dynamical system given by equation (7-88) is stable as long as the proportional gain factor G_{α} is greater than the term $\mathcal{M}_{\alpha}/\mathfrak{S}$. We find the steady state performance of the controller when we consider a constant disturbance and calculate the state of the system after an infinitely long time interval:

$$\alpha_{control} = \frac{\mathcal{M}_{disturbance}}{\mathfrak{S} \cdot G_{\alpha} - \mathcal{M}_{\alpha}} \quad (7-89)$$

You can see that a constant disturbance causes a constant attitude control angle, and that the steady state performance improves when we use a larger gain factor (the deviation $\alpha_{control}$ from the nominal attitude gets smaller for larger values of G_{α}).

The transient performance. The word transient refers to the time interval when the system is changing from one steady state to another one; for example when the system is excited by a step input function $\mathcal{M}_{disturbance}/\mathfrak{S}$. To find the solution to the equation (7-88) for the excitement with a step input function we insert the trial solution $\alpha_{control} = C \cdot e^{\lambda t}$ (with the amplitude C , the multiplier λ and the time t) and calculate the characteristic equation of the homogeneous system. The result is:

$$\lambda^2 + G_{\dot{\alpha}} \cdot \lambda + G_{\alpha} - \mathcal{M}_{\alpha}/\mathfrak{S} = 0 \quad (7-90)$$

or:

$$\lambda_{1/2} = -\frac{G_{\dot{\alpha}}}{2} \pm \sqrt{\frac{G_{\dot{\alpha}}^2}{4} - \left(G_{\alpha} - \frac{\mathcal{M}_{\alpha}}{\mathfrak{S}}\right)} \quad (7-91)$$

The differential equation (7-88) can have three different solutions, depending on the sign of the argument of the square root function of its characteristic equation (7-91).

When the square root argument is negative the multipliers λ are complex numbers (with negative real parts and conjugate imaginary parts). Then the solution is:

$$\alpha_{control} = \frac{\mathcal{M}_{disturbance}}{\mathfrak{S} \cdot G_{\alpha} - \mathcal{M}_{\alpha}} \cdot \left[1 - e^{-(G_{\dot{\alpha}}/2)t} \cdot \left(\cos \nu t + \frac{G_{\dot{\alpha}}}{2\nu} \sin \nu t \right) \right] \quad (7-92)$$

with the frequency:

$$\nu = \sqrt{-\frac{G_{\dot{\alpha}}^2}{4} + \left(G_{\alpha} - \frac{\mathcal{M}_{\alpha}}{\mathfrak{S}}\right)} \quad (7-93)$$

The solution describes a damped oscillation with the damping factor $(G_{\dot{\alpha}}/2)$ and the frequency ν . At the initial time ($t = 0$) the excursion (and its velocity) equals zero; at the final time ($t \rightarrow \infty$) the excursion accepts the new steady state (equation 7-89).

When the square root argument of the equation (7-91) is positive both eigenvalues are real numbers with negative values. The solution to the equation (7-88) is then:

$$\alpha_{control} = \frac{\mathcal{M}_{disturbance}}{\mathfrak{S} \cdot G_{\alpha} - \mathcal{M}_{\alpha}} \cdot \left[1 - e^{-(G_{\dot{\alpha}}/2)t} \cdot \left(\cosh \mu t + \frac{G_{\dot{\alpha}}}{2\mu} \sinh \mu t \right) \right] \quad (7-94)$$

with the factor:
$$\mu = \sqrt{\frac{G_{\dot{\alpha}}^2}{4} - \left(G_{\alpha} - \frac{\mathcal{M}_{\alpha}}{\mathfrak{S}} \right)} \quad (7-95)$$

The solution describes an overdamped oscillation without overshoot. Once again, the excursion equals zero at the initial time and accepts finally the new steady state.

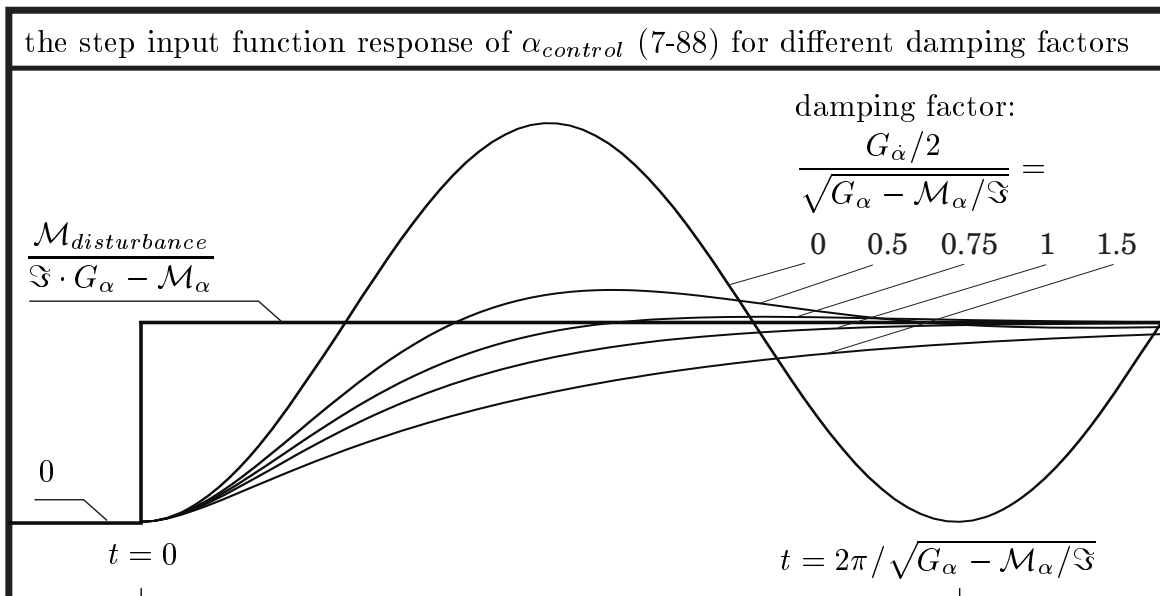
The third solution is more of theoretical interest, because the square root argument is never exactly zero. In this case both multipliers λ are identical negative numbers:

$$\alpha_{control} = \frac{\mathcal{M}_{disturbance}}{\mathfrak{S} \cdot G_{\alpha} - \mathcal{M}_{\alpha}} \cdot \left[1 - \left(1 + \frac{G_{\dot{\alpha}}}{2} t \right) \cdot e^{-(G_{\dot{\alpha}}/2)t} \right] \quad (7-96)$$

in case:
$$G_{\dot{\alpha}} = 2\sqrt{G_{\alpha} - \mathcal{M}_{\alpha}/\mathfrak{S}} \quad (7-97)$$

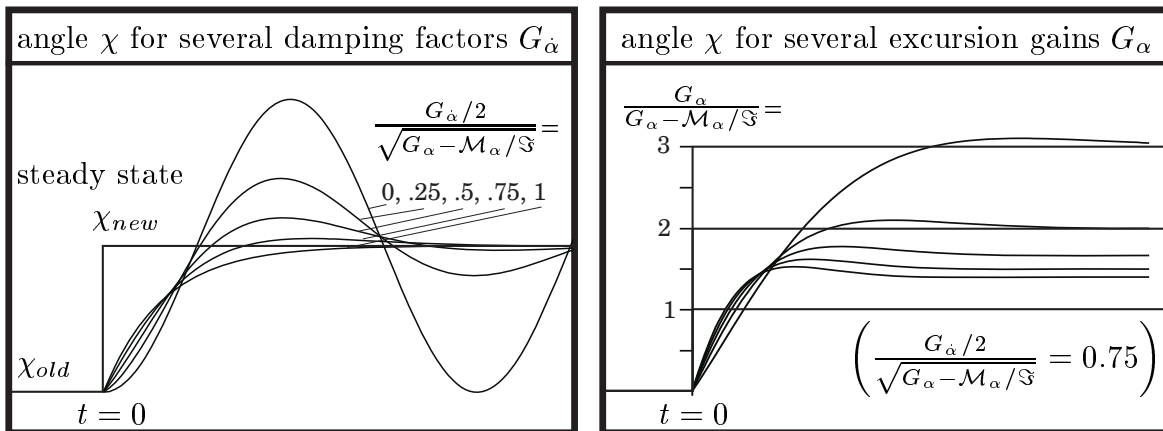
The excursion (7-96) as response to the step input function corresponds to the critically damped case, where any further decrease in damping would result in overshoot; its graphical representation is actually very similar to the overdamped case (7-94).

The figure below illustrates the response of the control system (7-88) to the step input function. The undamped system ($G_{\dot{\alpha}} = 0$) oscillates with the frequency $\nu_o = \sqrt{G_{\alpha} - \mathcal{M}_{\alpha}/\mathfrak{S}}$ about the final steady state condition $\mathcal{M}_{disturbance}/(\mathfrak{S}\nu_o^2)$. The underdamped system ($G_{\dot{\alpha}}/2\nu_o = 0.5$ or 0.75) has a comparatively short response time but it overshoots the final steady state condition. The critically damped system ($G_{\dot{\alpha}} = 2\nu_o$) and the overdamped system ($G_{\dot{\alpha}}/2\nu_o = 1.5$) approach the final steady state condition without overshoot but slower than the underdamped system.



The gain factors. The problem is now the selection of appropriate values for the gain factors of the control system (7-88). The search is rather an empirical procedure (based on computer simulations) than an analytical derivation. However, the right value of rate gain factor $G_{\dot{\alpha}}$ must stand in a certain relation with the natural frequency of the undamped system $\nu_o = \sqrt{G_{\alpha} - \mathcal{M}_{\alpha}/\mathfrak{S}}$: when the damping is too small, the system reacts on a step input disturbance with a large overshoot and a long oscillation period before the new equilibrium is reached; but the transient response time is also long when the damping is high. Experience shows that a good solution requires a slight overshoot, for example the damping ratio $G_{\dot{\alpha}}/(2\nu_o) \approx 0.75$. The left figure below shows the thrust angle χ during the transient period for several damping factors.

Let us now consider the excursion gain factor G_{α} . To ensure static stability of the control system this factor must be greater than the term $\mathcal{M}_{\alpha}/\mathfrak{S}$. The response time of the system (determined by the frequency 7-93) and its steady state performance (determined by the equation 7-89) as well improve when we increase the factor G_{α} . The right figure below shows the transition of the thrust misalignment angle χ to its new steady state for several gain factors; and we can observe that also the thrust misalignment reduces for larger gain factors. This all brings us to the conclusion that the gain factor should be as high as possible. However, the launcher is not a rigid body, it is a flexible body and it can vibrate. We must avoid resonance conditions where the frequency of the transient oscillation of the controller is near the natural frequencies of the launcher or its component parts. A large gain factor is desirable, however, the value is limited by the possible interference with the body of the launcher.



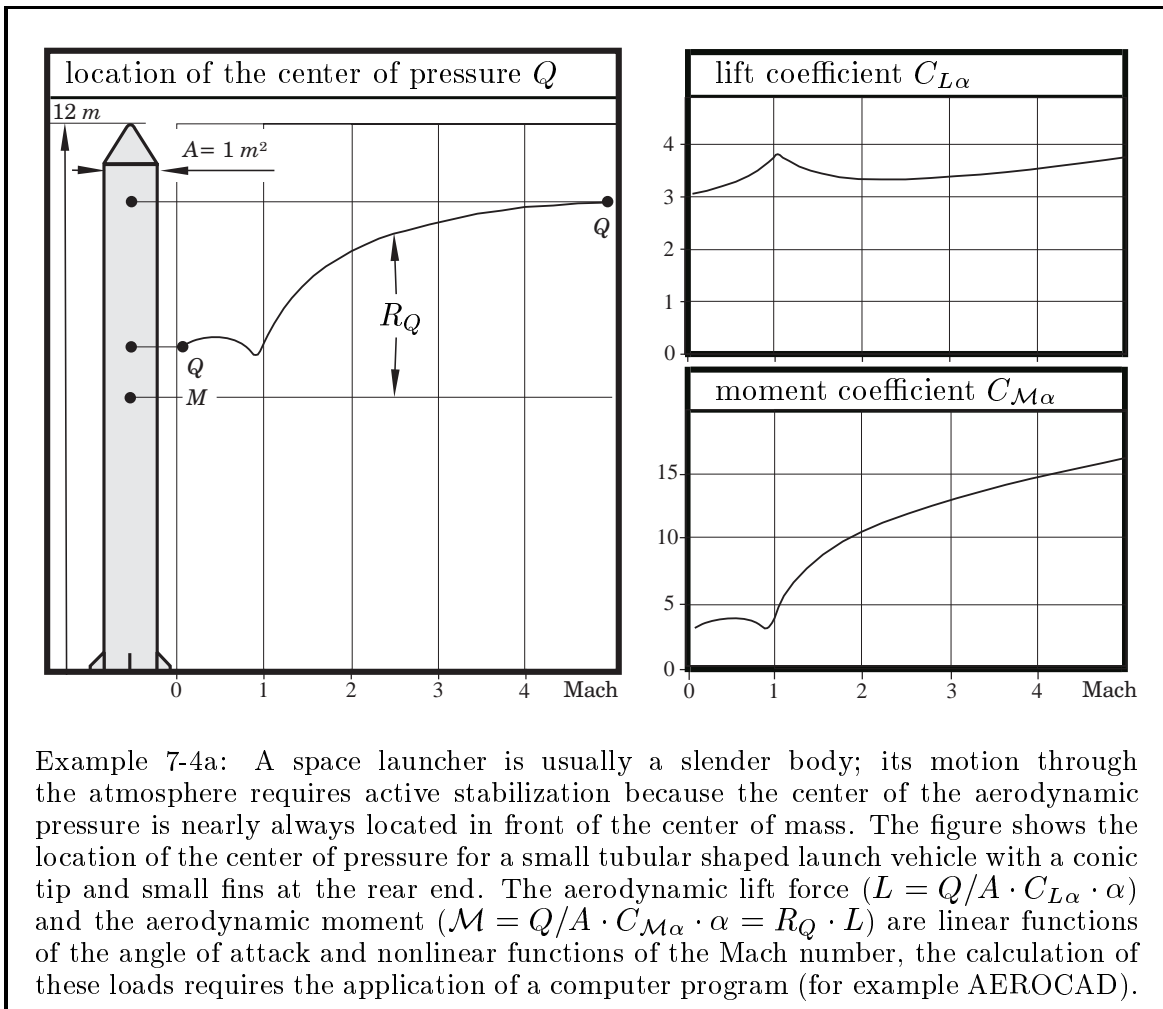
When the attitude of the launcher with a proportional differential controller (7-88) is disturbed by a constant torque (the step input function), the new steady state is actually an error in the attitude at the end of the transition period. However, the outer guidance loop with its slow response compensates steady state attitude errors. When we expect constant perturbations we can also use “error integral compensation” in the fast attitude control loop and include an additional feedback control proportional to the factor $\int \alpha_{control} dt$ in the equation (7-87) to eliminate the final steady state attitude error. The control torque developed by the integral error reacts slowly, but it leaves the transient performance of the control system essentially unchanged.

Acceleration control. Let us now rewrite the equations of the translational motion (7-83 and 7-84) for a small angle of attack α and a small thrust misalignment angle χ (the cosine of a small angle equals 1 and the sine equals the angle, approximately). The aerodynamic lift L is approximately a linear function of the angle of attack α (we can write $L = L_\alpha \cdot \alpha$ where the term L_α represents the “unit lift force”). The aerodynamic drag D , however, is not a function of the angle of attack. Therefore:

$$M \cdot \ddot{x} = -D + S \quad (7-98)$$

$$M \cdot \ddot{y} = L_\alpha \cdot \alpha + S \cdot (\alpha + \chi) \quad (7-99)$$

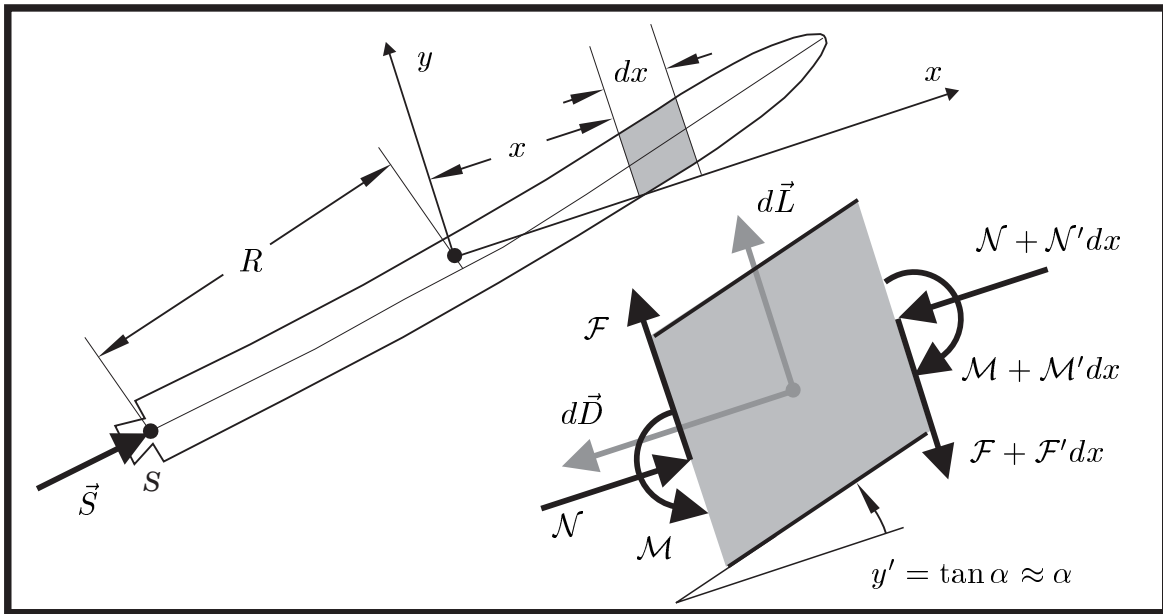
Equation (7-98) tells us that the acceleration \ddot{x} in flight direction is nearly not a function of the angle of attack, but the thrust misalignment angle χ in equation (7-99) is a linear function of α and its first time derivative $\dot{\alpha}$; and, consequently, the lateral acceleration \ddot{y} oscillates with the frequency of the angle α . The attitude control systems of space launchers use sometimes additionally acceleration feedback control in order to limit the structural loads during the atmospheric flight phase. However, accelerometer control means a destabilization of the attitude control loop, and a compromise between stress reduction and attitude stability must be achieved.



Example 7-4a: A space launcher is usually a slender body; its motion through the atmosphere requires active stabilization because the center of the aerodynamic pressure is nearly always located in front of the center of mass. The figure shows the location of the center of pressure for a small tubular shaped launch vehicle with a conic tip and small fins at the rear end. The aerodynamic lift force ($L = Q/A \cdot C_{L\alpha} \cdot \alpha$) and the aerodynamic moment ($\mathcal{M} = Q/A \cdot C_{M\alpha} \cdot \alpha = R_Q \cdot L$) are linear functions of the angle of attack and nonlinear functions of the Mach number, the calculation of these loads requires the application of a computer program (for example AEROCAD).

7.4.2 The Flexible Rocket

Bending vibrations. We have seen that high gain factors are desirable for a fast reaction of the attitude stabilization feedback control system of the space launcher, but the body of the vehicle (or parts of it) may not resonate in response to the attitude control loop. Let us now consider the launcher as a flexible slender body. In order to calculate its natural frequencies we have to slice the body into infinitely thin disks of the length dx and the mass dM and establish the conditions of equilibrium for every disk (the method of dissection). The aerodynamic loads which act on the disk are dL (lift) and dD (drag). The internal loads at the section x are \mathcal{N} (the normal force in x -direction), \mathcal{F} (the shear force in y -direction) and \mathcal{M} (the bending moment in z -direction); and, since dx is infinitesimally small, these loads change proportionally to the length dx when we move the section from the location x to the new location $x + dx$ (we use the prime to indicate a differentiation with respect to the distance d/dx and a dot to indicate a differentiation with respect to time d/dt). The internal loads on the disk at the location $x + dx$ act in the negative direction.



We denote the equilibrium of torques for the center of mass of the infinitely thin disk (neglecting its rotational moment of inertia $d\mathcal{S}$ and considering $dx^2 \approx 0$) and Newton's linear momentum theorem for the translational motion in x and y -direction:

$$0 = \mathcal{N} y' dx - \mathcal{F} dx - \mathcal{M}' dx \tag{7-100}$$

$$dM \cdot \ddot{x} = -dD - \mathcal{N}' dx \tag{7-101}$$

$$dM \cdot \ddot{y} = +dL - \mathcal{F}' dx \tag{7-102}$$

Note that we do not have to consider an aerodynamic torque in the equation (7-100): the torque $\mathcal{M}_\alpha \cdot \alpha$ with respect to the center of mass M of the vehicle is the result of the integration of all the torques $x \cdot dL$ over the length of the entire launch vehicle.

The partial differential equation for the bending vibrations. The internal bending moment that is necessary to deform a beam is proportional to the curvature of the beam; and this curvature is nearly (but not exactly) a linear function of the second order derivative y'' of the elastic deflection curve y . The proportional coefficient is called EI (it consists actually of two parts, the modulus of elasticity in tension or compression E and the bending moment of inertia I about the neutral axis). The deformation is a function of the bending moments and shearing forces, however, the shearing strains can be neglected when the beam is long enough (for example, when its ratio of length to diameter is greater than the factor 4). Then we can write:

$$\mathcal{M} = -EI y'' \quad (7-103)$$

Let us return to the equations (7-100), (7-101) and (7-102). We find a system of three partial differential equations when we divide these relationships by the length dx :

$$\mathcal{F} = \mathcal{N} y' - \mathcal{M}' = \mathcal{N} y' + (EI y'')' \quad (7-104)$$

$$M' \cdot \ddot{x} = -D' - \mathcal{N}' \quad (7-105)$$

$$M' \cdot \ddot{y} = +L' - \mathcal{F}' \quad (7-106)$$

The acceleration in flight direction \ddot{x} is not a function of the location x or the time t (we consider bending vibrations during a certain short time interval of the mission). Equation (7-98) tells us that $\ddot{x} = (S - D)/M$; and we can integrate equation (7-105) at once to find the normal force \mathcal{N} inside the vehicle as a function of the location x :

$$\mathcal{N}(x) = S - M(x) \ddot{x} - D(x) \quad (7-107)$$

We can readily verify that the thrust force S is the integration constant. For example, the internal normal force \mathcal{N} equals S at the location $x = -R$ where $M(x) = 0$ and $D(x) = 0$; it vanishes at the tip of the vehicle where $M(x) = M$ and $D(x) = D$. To find the equation for the bending vibration \ddot{y} we differentiate the equation (7-104) with respect to the coordinate x and insert the result into the equation (7-106). Thus:

$$M' \cdot \ddot{y} = L' - (\mathcal{N} y')' - (EI y'')'' \quad (7-108)$$

The equation above is called a “partial differential equation” because it contains coordinate derivatives (the primes) and time derivatives as well (the dots). Bernoullies approach separates the time derivatives from the coordinate derivatives. We can try:

$$y(t, x) = C_{\circ} \cdot \sin \omega(t - t_{\circ}) \cdot f(x) + g(x) \quad (7-109)$$

The bending vibration is an oscillation with the frequency ω and the amplitude C_{\circ} ; but the functions $f(x)$ and $g(x)$ are both just functions of x and not functions of t .

$$\begin{aligned} C_{\circ} \cdot \sin \omega(t - t_{\circ}) \cdot \left[-M' \omega^2 f(x) + (\mathcal{N} f(x)')' + (EI f(x)'')'' \right] \\ = \left\{ L' - (\mathcal{N} g(x)')' - (EI g(x)'')'' \right\} \end{aligned} \quad (7-110)$$

We find the condition above when we differentiate Bernoullies approach (7-109) twice with respect to coordinate and time and insert the results into the equation (7-108).

The trial solution (7-109) separates the time dependent behaviour of the curve y from the static behaviour: the expression in curly braces in the condition (7-110) equals zero when the oscillation amplitude vanishes ($C_o = 0$). Therefore we can conclude:

$$L' = (\mathcal{N} g(x)')' + (EI g(x)'')'' \quad (7 - 111)$$

The equation (7-111) above is the differential equation for the elastic curve $g(x)$ of a non-vibrating flexible rocket that is exposed to the external length specific load L' . We get the same law for y when we remove the term $M'\ddot{y}$ from the equation (7-108). This equation allows us to calculate the equilibrium deformation of the flexible rocket; but the equation for the bending oscillation is more interesting at the moment. Therefore we have to find a function $f(x)$ that complies with equation (7-110) when the amplitude is not zero ($C_o = 1$) and the equilibrium condition (7-111) is satisfied:

$$M'\omega^2 f(x) = (\mathcal{N} f(x)')' + (EI f(x)'')'' \quad (7 - 112)$$

Finding a solution to this equation in the general case requires the application of a numerical iteration scheme on a computer. However, an analytical solution is possible when we assume constant mean values for the length specific mass M' , the flexural strength EI and the internal normal force \mathcal{N} . With these constant mean values the equation (7-112) simplifies to a fourth order linear differential equation. We can write:

$$EI f(x)'''' + \mathcal{N} f(x)'' - M'\omega^2 f(x) = 0 \quad (7 - 113)$$

The solution to the equation (7-113) is determined by the characteristic equation:

$$EI \lambda^4 + \mathcal{N} \lambda^2 - M'\omega^2 = 0 \quad \rightarrow \quad \lambda^2 = \frac{-\mathcal{N} \pm \sqrt{\mathcal{N}^2 + 4 EI M'\omega^2}}{2 EI} \quad (7 - 114)$$

$$\lambda_{1/2} = \pm \sqrt{\frac{-\mathcal{N} + \sqrt{\mathcal{N}^2 + 4 EI M'\omega^2}}{2 EI}} = \pm \mu$$

$$\lambda_{3/4} = \pm i \sqrt{\frac{+\mathcal{N} + \sqrt{\mathcal{N}^2 + 4 EI M'\omega^2}}{2 EI}} = \pm i \nu$$

The characteristic equation has exactly four solutions (“roots”) for the multiplier λ ; two of them are real (abbreviated as μ) and the two others are conjugate imaginary (abbreviated as ν). The general solution to the differential equation (7-113) is then:

$$f(x) = C_1 \cosh \mu x + C_2 \sinh \mu x + C_3 \cos \nu x + C_4 \sin \nu x \quad (7 - 115)$$

The equation above contains five unknown quantities: the four integration constants C_1 , C_2 , C_3 and C_4 and the frequency ω of the bending vibration. To find integration constants we have adapt the solution (7-115) to its special boundary conditions (the bearing conditions of the vibrating beam). The amplitude of this vibration, however, is actually not a function of the bearing conditions (constants C_1 to C_4); and we find the natural frequencies ω of the bending vibrations of the flexible launcher when we check under which conditions the vibration amplitude is undetermined.

The natural frequencies of the bending vibration. In order to calculate the integration constants C_1 to C_4 in equation (7-115) we have to consider the special boundary conditions (or bearing conditions) for the flying launcher. The second order derivative of the elastic curve $f(x)''$ is approximately proportional to the bending torque \mathcal{M} (compare equation 7-103); and the third order derivative $f(x)'''$ is approximately proportional to the shearing force (compare equation 7-104). The bending moment and its first derivative vanish at the tip of the launcher and at the rear end, because no forces act on the tip of the launcher and the thrust force is parallel with the curve $f(x)$ at the rear end. We can conclude that the functions $f(x)''$ and $f(x)'''$ equal zero at the two locations $x = \pm R$ (the center of mass M is located exactly in the middle of the launcher when the length-specific mass M' is constant). Differentiation of the equation (7-115) brings us to the following equation system:

$$\begin{aligned} f(-R)'' = 0 : & \quad +C_1 \mu^2 \cosh \mu R - C_2 \mu^2 \sinh \mu R - C_3 \nu^2 \cos \nu R + C_4 \nu^2 \sin \nu R = 0 \\ f(+R)'' = 0 : & \quad +C_1 \mu^2 \cosh \mu R + C_2 \mu^2 \sinh \mu R - C_3 \nu^2 \cos \nu R - C_4 \nu^2 \sin \nu R = 0 \\ f(-R)''' = 0 : & \quad -C_1 \mu^3 \sinh \mu R + C_2 \mu^3 \cosh \mu R - C_3 \nu^3 \sin \nu R - C_4 \nu^3 \cos \nu R = 0 \\ f(+R)''' = 0 : & \quad +C_1 \mu^3 \sinh \mu R + C_2 \mu^3 \cosh \mu R + C_3 \nu^3 \sin \nu R - C_4 \nu^3 \cos \nu R = 0 \end{aligned} \quad (7-116)$$

Transformed:

$$\begin{aligned} C_3 = C_1 \cdot \frac{\mu^2 \cosh \mu R}{\nu^2 \cos \nu R}; \quad C_4 = C_2 \cdot \frac{\mu^2 \sinh \mu R}{\nu^2 \sin \nu R} \\ C_1 \cdot (\mu \sinh \mu R \cos \nu R + \nu \cosh \mu R \sin \nu R) = 0 \quad (7-117) \\ C_2 \cdot (\mu \cosh \mu R \sin \nu R - \nu \sinh \mu R \cos \nu R) = 0 \quad (7-118) \end{aligned}$$

The condition (7-117) has only then the solution $C_1 \neq 0$ when the expression in parenthesis vanishes, and the condition (7-118) demands then that constant $C_2 = 0$; or, alternatively, the expression in parenthesis in condition (7-118) vanishes, and that means $C_1 = 0$ and $C_2 \neq 0$ (otherwise the equation system has just the trivial solution where all integration constants equal zero). Nontrivial solutions exist only if:

$$\tanh \mu R = -\frac{\nu}{\mu} \tan \nu R \quad \text{or:} \quad \tanh \mu R = +\frac{\mu}{\nu} \tan \nu R \quad (7-119)$$

The two equations above determine the natural frequencies ω of the flexible launcher, together with the multipliers μ and ν according to the definition in equation (7-114). An infinite set of natural frequencies ω fulfills these transcendental equations, but we have to find them numerically on a computer. However, we can approximate the solution when we neglect the influence of the internal normal force \mathcal{N} . Then we have:

$$\mu = \nu = \sqrt[4]{M'\omega^2/EI} \quad (7-120)$$

The hyperbolic tangent function is nearly 1 for values larger than 2; and this means that the first conditions (7-119) has solutions approximately for $\nu R = \frac{3}{4}\pi, \frac{7}{4}\pi, \frac{11}{4}\pi, \dots$ and the second condition, however, has solutions approximately for $\nu R = \frac{5}{4}\pi, \frac{9}{4}\pi, \dots$. We can write the solutions uniformly as $\nu R = \frac{1}{2}(k + \frac{1}{2})\pi$ when we introduce the quantity k as the oscillation mode of the bending vibration of the flexible launcher.

The natural frequency ω of the mode k is then determined by the following equation:

$$\omega_k = \sqrt{\frac{EI}{M'}} \cdot \left(\frac{(k + \frac{1}{2}) \pi}{2 R} \right)^2 \tag{7-121}$$

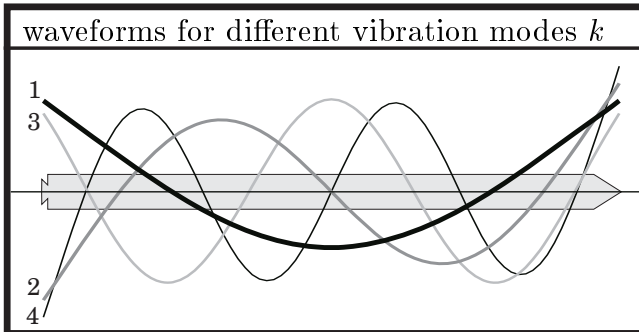
It must be emphasized that this relationship is only valid for a launcher with the overall length $2R$ that has a constant length-specific mass distribution M' and a constant flexural stiffness EI and is not exposed to a load \mathcal{N} in the direction of its axis.

The waveforms of the bending vibrations. Still missing in our discussion are the waveforms (also called “eigenforms”) which belong to the individual vibration modes. The modes with odd k number ($k = 1, 3, 5, \dots$) satisfy the first condition (7-119) and switch-off the constants C_2 and C_4 . The corresponding curves are consequently:

$$f(x, k) = C_1 \cdot \left(\cosh \mu x + \frac{\mu^2 \cosh \mu R}{\nu^2 \cos \nu R} \cos \nu x \right) \tag{7-122}$$

The modes with even k number ($k = 2, 4, 6, \dots$) satisfy the second condition (7-119) and switch-off the constants C_1 and C_3 . The vibration curves are then:

$$f(x, k) = C_2 \cdot \left(\sinh \mu x + \frac{\mu^2 \sinh \mu R}{\nu^2 \sin \nu R} \sin \nu x \right) \tag{7-123}$$



The figure on the left illustrates the different waveforms for the bending vibrations of the space launcher. Waveforms with an odd vibration mode number k resemble the cosine function centered in the middle of the launcher; waveforms with an even vibration mode number k resemble the sine function.

length: $2 R_S = 12 \text{ m}$

length specific mass: $M' = 2000 \text{ kg/m}$

radius: $r = 0.564 \text{ m}$

area: $A = 1 \text{ m}^2$

wall thickness: $t = 1 \text{ mm}$

wall material steel: $E = 200 \text{ kN/mm}^2$

Example 7-4b: A small space launcher with tubular shape (length $2 \cdot R = 12 \text{ m}$ and cross section area $A = 1 \text{ m}^2$) utilizes solid propellant (density of 1800 kg/m^3). The walls of the vehicle are made of steel. The geometrical moment of inertia I is then approximately $I = \pi r^3 t = 5.642 \cdot 10^8 \text{ mm}^4$; the flexural stiffness $EI = 11254 \text{ kN/m}^2$ and the length-specific mass $M' \approx 2000 \text{ kg/m}$. The natural frequency of the first vibration mode ($k = 1$) of the flexible vehicle is then $\omega = 36.6 \text{ rad/s}$ or 5.83 Hz .

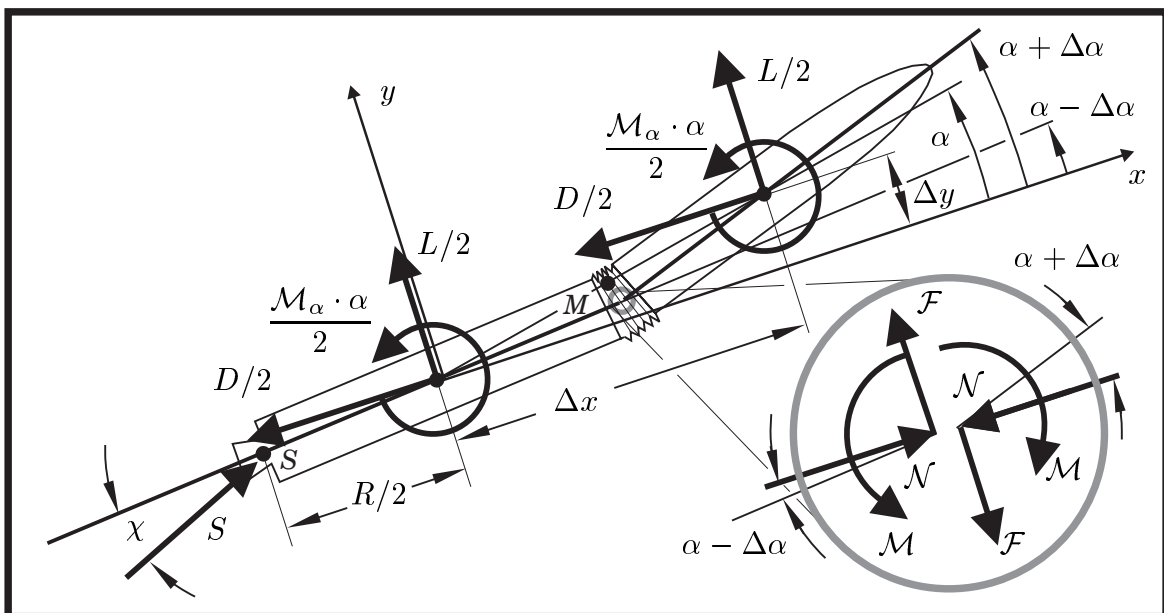
Superposition of bending vibrations. The actual oscillation of a flexible launcher can be considered as a superposition of the waveforms of all bending vibration modes, however, the bending oscillation has for every waveform k its individual amplitude. The “acoustic pattern” of the launcher depends on the stimulation of the oscillation (its excitation) and on the way the oscillation energy is absorbed from the structure of the vehicle (the damping). The energy inside the vibrating beam oscillates between two energy forms, the maximum potential energy and the maximum kinetic energy:

$$E_{potential} = \frac{1}{2} \int EI (f(x)'')^2 dx \tag{7 - 124}$$

$$E_{kinetic} = \frac{1}{2} \omega_k^2 \int M' f(x)^2 dx \tag{7 - 125}$$

We find these energies via integration, the expression $(f(x)'')^2$ is proportional to the square of the bending moment (7-103) and the expression $(\omega_k f(x))^2$ proportional to the square of the velocity of the particle dM at zero crossing (compare 7-109). The amplitudes of higher modes (harmonic frequencies) are usually much smaller than the amplitude of the first mode, because energy and bending moment associated with a certain amplitude increase considerably with the frequency of the vibration. Even though the harmonic frequencies ω_k of bending vibrations can produce “loud noise” (and structural loads), the displacements caused by these higher vibration modes are actually very small. The first vibration mode $k = 1$, however, is a phenomenon that can cause a serious interaction with the attitude control system of the space launcher.

Attitude control. The equation of motion for the bending vibrations of a flexible space launcher (7-108) is a partial differential equation that allows just in some special occasions analytical solutions. To analyze the interaction of the attitude control system with the body of launcher we will use a better conditioned model: we consider the launcher as composed of two rigid parts connected in the middle by a flexible adapter.



The flexible connection between the two rigid parts of the space launcher allows the vehicle to oscillate with the frequency ω_1 of the first bending vibration mode. We will still continue to consider the slender body of the vehicle as homogeneous beam with the mass M , the overall length $2R$ and the constant length-specific mass M' . The moment of inertia of the vehicle for rotations about its center of mass M is then:

$$\mathfrak{S} = M' \int_{-R}^{+R} x^2 dx = \frac{1}{3}MR^2$$

For example, our model would fit exactly for a two-stage launcher with two identical stages and a flexible adapter between the two stages. Each one of the two rigid parts has the length R , the mass $M/2$ and the moment of inertia $MR^2/24$ (for rotation about its own center of mass). The flexible connection carries the internal loads \mathcal{F} (the shearing force), \mathcal{N} (the normal force) and \mathcal{M} (the bending moment) from one part to the other one. Let us consider the amplitude of the bending vibration as so small that it does not influence the magnitude of the aerodynamic loads L (lift), D (drag) and $\mathcal{M}_\alpha \cdot \alpha$ (torque). Our intention is to analyze the bending vibration, therefore we can distribute the aerodynamic loads equally among the two parts. Newton's law for the motion in x - and y -direction applied to every part yields then:

first part (rear):

second part (front):

$$\frac{M}{2}\ddot{x} = -\frac{D}{2} + \mathcal{N} + S; \quad \frac{M}{2}(\ddot{x} + \Delta\ddot{x}) = -\frac{D}{2} - \mathcal{N} \quad (7-126)$$

$$\frac{M}{2}\ddot{y} = \frac{L}{2} + \mathcal{F} + S \cdot (\chi + \alpha - \Delta\alpha); \quad \frac{M}{2}(\ddot{y} + \Delta\ddot{y}) = \frac{L}{2} - \mathcal{F} \quad (7-127)$$

The mass multiplied by the acceleration is equal to all forces which act on the mass. The angles χ , α and $\Delta\alpha$ are small, and, approximately, the sine of an angle is the angle itself and the cosine is 1. A kinematical relationship determines the distance of the center of mass of the second part from the center of mass of the first part. Thus:

$$\Delta x = R \quad (7-128)$$

$$\Delta y = R \cdot \alpha \quad (7-129)$$

It follows that $\Delta\ddot{x} = 0$ and that $\Delta\ddot{y} = R\ddot{\alpha}$; and we can use the equations (7-126) to calculate the internal normal force $\mathcal{N} = -S/2$ and the equations (7-127) to find a relationship for the internal shearing force \mathcal{F} . The theorem of the angular momentum applied to the rotational motion of the first part (rotation angle $\alpha - \Delta\alpha$) and the second part (angle $\alpha + \Delta\alpha$) brings us the following differential equation system:

$$\frac{Mr^2}{24}(\ddot{\alpha} - \Delta\ddot{\alpha}) = \frac{\mathcal{M}_\alpha \cdot \alpha}{2} + \frac{R}{2}(\mathcal{F} - \mathcal{N} \cdot (\alpha - \Delta\alpha)) + \mathcal{M} - \frac{R}{2}S \cdot \chi \quad (7-130)$$

$$\frac{MR^2}{24}(\ddot{\alpha} + \Delta\ddot{\alpha}) = \frac{\mathcal{M}_\alpha \cdot \alpha}{2} + \frac{R}{2}(\mathcal{F} - \mathcal{N} \cdot (\alpha + \Delta\alpha)) - \mathcal{M} \quad (7-131)$$

Reference point is in both cases the center of mass of the concerning part of the body.

The next steps in our analysis are the addition and the subtraction of the equations (7-130) and (7-131): the intention is to transform these equations in a way that they describe the motion of the bending vibration $\Delta\alpha$ separately from the rotational motion α of the vehicle. We use again the moment of inertia $\mathfrak{S} = \frac{1}{3}MR^2$ and write down:

$$\frac{1}{4}\mathfrak{S}\Delta\ddot{\alpha} = \frac{R \cdot S}{2}\Delta\alpha - 2\mathcal{M} + \frac{R}{2}S \cdot \chi \tag{7-132}$$

$$\mathfrak{S} \ddot{\alpha} = \mathcal{M}_\alpha \cdot \alpha + \frac{R \cdot S}{2}\Delta\alpha - R S \cdot \chi \tag{7-133}$$

The equations above describe a coupled linear oscillation for the motions α and $\Delta\alpha$. Let us first concentrate our attention on the bending vibration $\Delta\alpha$: we know that the internal moment \mathcal{M} is a linear function of $\Delta\alpha$ and we are interested in eliminating this term from the equation system. \mathcal{M} must satisfy the following condition when the beamlike vehicle oscillates with the frequency ω_1 of the first bending vibration mode:

$$\omega_1^2 = \frac{8\mathcal{M}}{\mathfrak{S}\Delta\alpha} - \frac{2R S}{\mathfrak{S}}$$

The equation (7-132) can be formulated as a constrained (or forced) linear oscillation:

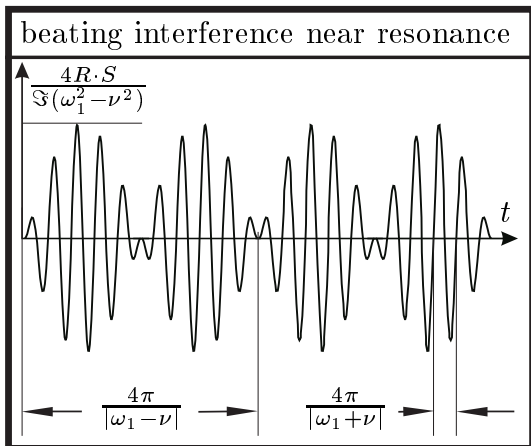
$$\Delta\ddot{\alpha} + \omega_1^2 \cdot \Delta\alpha = \frac{2R \cdot S}{\mathfrak{S}} \cdot \chi_{max} \cos \nu t \tag{7-134}$$

The bending vibration of the launcher is excited by the oscillating thrust direction. Let us assume that the thrust misalignment angle $\chi(t)$ oscillates with the natural frequency ν (equation 7-93) of the closed-loop attitude control system of the launcher (the term χ_{max} characterizes the amplitude of the oscillation of the thrust angle χ). In this case it is possible to write the general solution to equation (7-134) in the form:

$$\Delta\alpha = C_1 \sin \omega_1 t + C_2 \cos \omega_1 t + \frac{2R \cdot S}{\mathfrak{S}(\omega_1^2 - \nu^2)} \cdot \cos \nu t \tag{7-135}$$

We can find the integration constants C_1 and C_2 when we assume initial conditions for the vibration, for example $\Delta\alpha = 0$ and $\Delta\dot{\alpha} = 0$ at the initial instant $t = 0$. Thus:

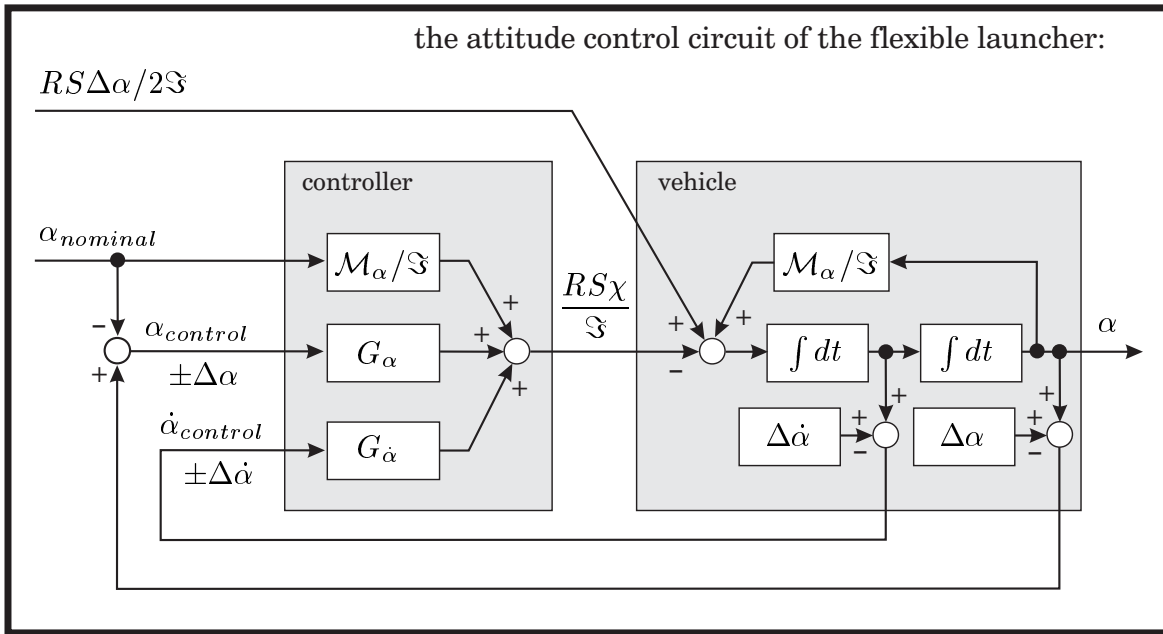
$$\Delta\alpha = \frac{Rr \cdot S}{\mathfrak{S}(\omega_1^2 - \nu^2)} \cdot (\cos \nu t - \cos \omega_1 t) = \frac{4R \cdot S}{\mathfrak{S}(\omega_1^2 - \nu^2)} \cdot \sin \frac{(\omega_1 - \nu)t}{2} \cdot \sin \frac{(\omega_1 + \nu)t}{2} \tag{7-136}$$



The result is the superposition of two oscillations with the same amplitude but with different frequencies (“beating interference”); the amplitude gets extremely large for the situation that $\nu \approx \omega_1$ (called “resonance”). Resonance must be avoided in any case during the entire flight time, otherwise the launcher is destroyed within a few seconds. The figure on the left shows the beating vibration of the angle $\Delta\alpha$ (equation 7-136) near the resonance situation as a slow amplitude oscillation (frequency: $|\omega_1 - \nu|/2$).

When we compare the equation (7-133) for the rotational motion of the flexible launcher with the equation (7-85) for the rotational motion of the rigid launcher we can observe that now the term $RS\Delta\alpha/2$ corresponds to a perturbation $\mathcal{M}_{disturbance}$. The twisting angle $\Delta\alpha$ of the rear end of the vehicle causes an additional thrust torque about the center of mass M . However, this influence of the vibration on the stability of the vehicle is usually insignificant (provided that the vibration amplitude is small).

Structural damping of the bending vibrations (when a small part of the internal torque \mathcal{M} is proportional to the rate $\Delta\dot{\alpha}$) has the effect that the oscillation amplitude decays slowly when the stimulation stops, however, the structural bending is usually not strong enough to prevent excessive amplitudes in the case of resonance. The resonance situation can just be avoided when the natural frequency of the attitude control loop of the launcher ν is during the entire flight time substantially below the frequency ω_1 of the first bending vibration mode (and the harmonic content). However, even when ν is smaller than ω_1 , an incorrect measurement of the angle α in the attitude control loop can cause a stimulation of the bending vibration $\Delta\alpha$. To analyze this interference let us look again at the attitude control circuit:



It is the point in question whether the sensors for measuring the attitude α and the attitude rate $\dot{\alpha}$ are installed in the front part of the vehicle (near to the payload) or in the rear end (near to the engine). The control unit of the vehicle that is used during the entire mission is usually installed inside the upper stage (otherwise it would get lost when an empty stage is dropped); and we would naturally prefer to install the sensors next to the control unit. However, the sensors measure actually the incorrect attitude $\alpha + \Delta\alpha$ and the incorrect rate $\dot{\alpha} + \Delta\dot{\alpha}$ when they are installed inside the front part of the vehicle; but they measure also the incorrect values $\alpha - \Delta\alpha$ and $\dot{\alpha} - \Delta\dot{\alpha}$ when they are installed in the rear part near to the engine. The vibration angle $\Delta\alpha$ and its rate $\Delta\dot{\alpha}$ act as disturbances on the measurement unit of the space launcher.

The perturbations $\Delta\alpha$ and $\Delta\dot{\alpha}$ are transmitted from the sensors to the controller of the vehicle and transformed with the gains G_α and $G_{\dot{\alpha}}$ into a faulty control torque. Instead of using the equation (7-87) for the closed-loop control we have to write now:

$$RS \cdot \chi = \mathcal{M}_\alpha \cdot \alpha_{nominal} + \mathfrak{S}(G_\alpha \cdot (\alpha_{control} \pm \Delta\alpha) + G_{\dot{\alpha}} \cdot (\dot{\alpha}_{control} \pm \Delta\dot{\alpha})) \quad (7-137)$$

The positive sign is valid when the sensors are mounted in the front part of the vehicle, the negative sign is valid when the sensors are mounted in the rear part. The effect of the additional (wrong) sensor signal on the bending vibrations is a modification of the natural frequency and the damping properties (compare equation 7-134). Therefore:

$$\begin{aligned} \Delta\ddot{\alpha} + \omega_1^2 \cdot \Delta\alpha &= 2 (\mathcal{M}_\alpha / \mathfrak{S}) \cdot \alpha_{nominal} \\ &+ 2 G_\alpha \cdot (\alpha_{control} \pm \Delta\alpha) + 2 G_{\dot{\alpha}} \cdot (\dot{\alpha}_{control} \pm \Delta\dot{\alpha}) \quad (7-138) \end{aligned}$$

This second order differential equation consists of a homogeneous part which describes the natural resonance vibration and an inhomogeneous part which describes the forced vibration. We have to establish the characteristic equation of the homogeneous part of the oscillation to find the properties of the free oscillation $\Delta\alpha$. Thus:

$$\lambda^2 \mp 2 G_{\dot{\alpha}} \lambda + (\omega_1^2 \mp 2 G_\alpha) = 0 \quad (7-139)$$

The positive sign in the equation corresponds now to the case that the sensors are mounted in the rear end of the vehicle. In this case the multiplier λ takes the value of:

$$\lambda = -G_{\dot{\alpha}} \pm \sqrt{G_{\dot{\alpha}}^2 - (\omega_1^2 + 2 G_\alpha)} \quad (7-140)$$

We can see that now the natural frequency of the bending vibration is higher than before and that the rate gain factor damps down the bending vibration. The negative sign in the equation (7-139), however, corresponds to the case that the sensors are mounted in the front end of the vehicle; and in this case the multiplier λ becomes:

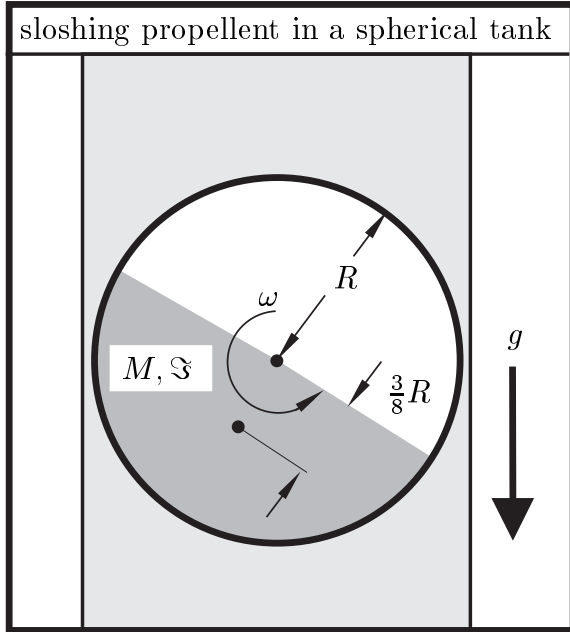
$$\lambda = +G_{\dot{\alpha}} \pm \sqrt{G_{\dot{\alpha}}^2 - (\omega_1^2 - 2G_\alpha)} \quad (7-141)$$

The false signal amplifies now the bending vibration and lowers its natural frequency.

We can conclude that it is not sufficient that a single sensor package for measuring the attitude angle α and its rate $\dot{\alpha}$ is installed in the upper stage of a space launcher. Any incorrect information about the total behaviour of the vehicle can cause a self-excitation and instability of the attitude control system. The control unit of a space launcher is usually a complex multi-sensor system equipped with an inertial platform supported by several (redundant) rate gyros and several accelerometers mounted inside different parts of the vehicle. The actual signal for the control of the attitude of the vehicle is then generated via a filtering network on a digital computer. An appropriate shaping network can damp down the bending vibrations without affecting the stability of the attitude control function. The control gains which determine the frequency of the rigid body must be high enough to guarantee a fast response of the vehicle in case of attitude errors, however, the rigid-body control frequency must also be sufficiently below the frequency of the first bending mode.

7.4.3 Sloshing Propellents

The motion of liquids in tanks. Many parts of the launch vehicle oscillate during the mission, but an appropriate fixing of the masses is usually easy and the natural frequencies are then much higher than the frequency of the attitude control system (the attitude control loop oscillates with a natural frequency of 0.1 to 1 Hz, typically). However, an interaction can take place with the oscillation of liquid propellant in partially empty tanks (sloshing). It is obviously not possible to fix these masses, and the natural frequency of the free oscillation of the propellant is often near the



frequency of the attitude control system. Oscillating propellant masses are usually modeled as pendulums in the simulations: for example, consider a spherical tank that is actually half empty (radius R). When we considered the propellant mass M as a rigid hemispherical body, the moment of inertia with respect to the center of the tank would be $\mathfrak{S} = \frac{2}{5}MR^2$. The distance between its center of mass and the center of the tank is then $\frac{3}{8}R$, and the oscillation frequency of the "rigid body" pendulum follows as $\omega = \sqrt{15g/16R}$. Term g is the gravitational acceleration (for comparison: a cord pendulum of a particle mass and a cord of the length R oscillates with the frequency $\omega = \sqrt{g/R}$).

Metal sheets inside the tanks provide damping of the oscillation ("tank baffling"). The damping attenuates the amplitude of the sloshing motion considerably; and the damping has also the effect that the propellant does not swing like a cord pendulum: the oscillation frequency deviates often substantially from the estimate $\omega = \sqrt{g/R}$, particularly when the liquid is stored in a cylindrical tank with hemispherical ends (the behaviour of sloshing liquids is a complicated problem in computational fluid dynamics: the liquid oscillates often by a factor two faster than a cord pendulum).

Modeling the sloshing motion. The forces caused by sloshing propellents inside partially empty tanks can be considered as disturbances for the attitude control system of the launcher: we remove the sloshing mass M_P from the equation system and replace it by a force \mathcal{F} that acts on the vehicle (the point of application is the distance R_P from the center of mass M of the vehicle). The equations of motion are then:

$$\mathfrak{S} \cdot \ddot{\alpha} = \mathcal{M}_\alpha \cdot \alpha - S R_S \cdot \chi + R_P \cdot \mathcal{F}_y \tag{7-142}$$

$$M \cdot \ddot{x} = -D + S + \mathcal{F}_x \tag{7-143}$$

$$M \cdot \ddot{y} = L_\alpha \cdot \alpha + S \cdot (\alpha + \chi) + \mathcal{F}_y \tag{7-144}$$

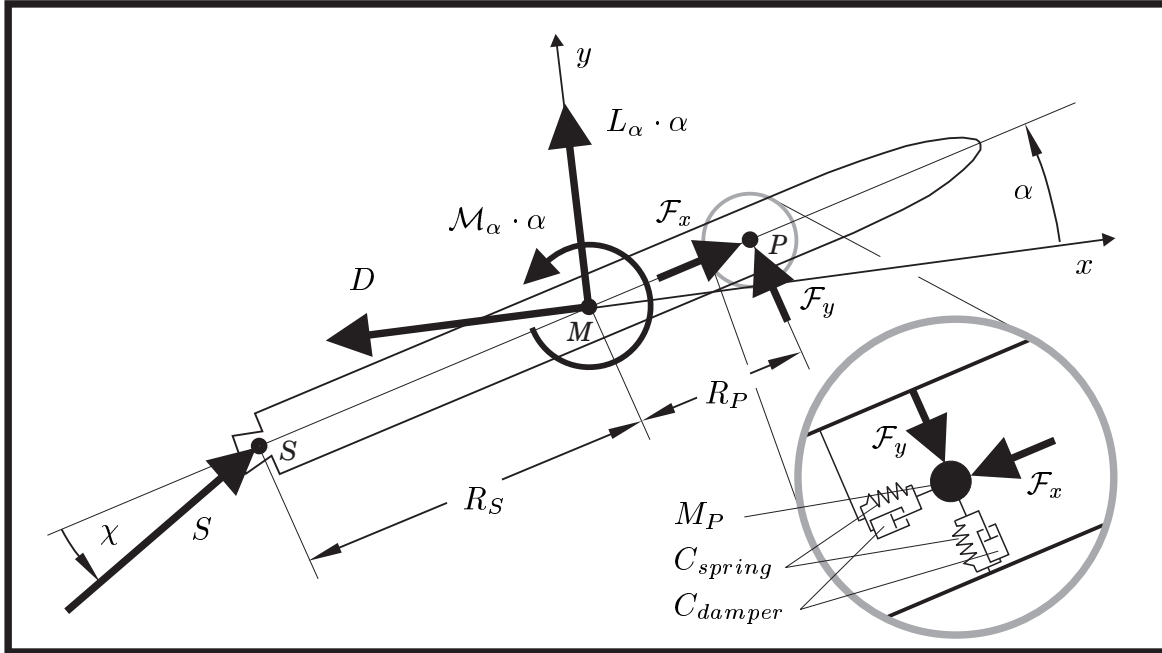
The sloshing propellant mass M_P causes the forces \mathcal{F}_x and \mathcal{F}_y and the torque $R_P \cdot \mathcal{F}_y$.

The easiest way to incorporate the forces \mathcal{F}_x and \mathcal{F}_y in the dynamical model of the vehicle is the use of damped spring mechanisms. The effect on the mass M_P is then:

$$\mathcal{F}_x = M_P \cdot (\ddot{x} + \Delta\ddot{x}) = C_{spring} \cdot \Delta x + C_{damper} \cdot \Delta\dot{x} \quad (7-145)$$

$$\mathcal{F}_y = M_P \cdot (\ddot{y} + \Delta\ddot{y}) = C_{spring} \cdot (\alpha R_P + \Delta y) + C_{damper} \cdot (\dot{\alpha} R_P + \Delta\dot{y}) \quad (7-146)$$

The equations (7-142) to (7-146) provide a system of five second order differential equations to calculate numerically the five unknown quantities α , x , y , Δx and Δy . The result is a coupled linear oscillation of the propellant mass and the space launcher.



The launcher as a multi-body system. We have seen that the frequency of the closed-loop attitude control motion of the flexible space launcher must be lower than the frequency of the first bending mode (for example four times lower) to prevent dangerous resonance conditions throughout the flight. However, the aerodynamic loads L and D , the center of pressure Q , the mass of the launcher M , its mass distribution M' , the location of the center of mass M and the moment of inertia \mathfrak{S} are all quantities which change their values substantially during the mission. It is consequently necessary to adapt the gain factors G_α and $G_{\dot{\alpha}}$ which determine the frequency of the control oscillation continuously to every new situation. The accuracy of the orbital injection and the safety margins for attitude stability are tested in extensive computer simulations during the qualification tests of a new guidance and control system for a space launcher. The launcher is modeled as a mechanical “multibody system” (considering its elasticity, vibrations in the engines, sloshing propellents and so on). These computer simulations must prove that the control system of the launcher is under all conditions stable and capable to handle all external (and internal) disturbances that can be encountered during all phases of the mission. The search for the appropriate control gains is rather an empirical procedure than a theoretical derivation, where a compromise has to be found between flight safety and injection accuracy.

8. Planetary Missions

The American “Apollo”-project is with no doubt the highlight of past astronautics. Never before man had traveled so far away from home, and managed all the difficulties which are associated with manned space flight. The first part of this chapter treats lunar missions, or, more in general, trajectories which are formed by the attraction of two celestial bodies. The subject is known as the “restricted-three-body-problem” of celestial mechanics (restricted because the third body, the spacecraft, is of negligible mass). The gravitational field of earth and moon together allows a variety of funny shaped orbits: spacecraft can move on periodic trajectories and flyby repeatedly earth and moon; “halo orbits” can be established around the libration points of the earth-moon system; and spacecraft can leave the gravitational field of the earth after having executed a lunar swingby maneuver.

“Deep space” is the destination for some scientific missions. When the spacecraft wants to leave the vicinity of the earth and travel to a distant planet, it has to enter an earth escape orbit. Astonishingly, a journey to a distant planet requires not much more propulsion than a mission to our moon. The gravitation of planets can be used as a propulsion system for planetary space missions. This “transport system” makes it possible that all planets of our solar system can be visited at the price of a Mars or Venus flyby. In fact, a space launcher has the same payload capacity for the transport of a satellite into geostationary position as it has for the transport of a probe to the far distant planet Pluto. Planets are visited on the way, their attraction works as a “propulsion system”, which does not consume propellant and will not fail to work. Nowadays, nearly all planetary deep space missions utilize gravity assist maneuvers. Planetary flights and the theory of gravity propelled missions are the content of the second and the third section of this chapter.

Planet Mars is the most attractive candidate for a future manned exploration of the solar system. A manned Mars mission is possible with existing technology and will not necessarily cost much more than the American Apollo project. The propulsion requirement for a Mars return mission is analyzed in the fourth section of this chapter.

The foundation of two-body celestial mechanics was laid more than two centuries ago, by Euler, Lagrange (1772) and Jacobi (1836). The science came early to a high developed state, since accurate observation of planets was useful for nautical purposes. Later it became known that comets change their orbital elements when they flyby Jupiter (Tisserand, 1889). The fascinating options which planetary swingbys offer to astronautics were first seen by M.Minovitch (1961). Fundamental work about halo orbits was done by R.W.Farquhar (1966) and refined by D.L.Richardson (1980). The books of V.Szebehely [“Theory of Orbits”, Academic Press, New York, 1967], J.M.A.Danby [“Celestial Mechanics”, Willmann-Bell, Richmond, Virginia, 1988], and Krafft A. Ehricke [Space Flight, D.Van Nostrand Company, Inc., Princeton, New Jersey, 1962] are recommended as fundamental literature in this field.

8.1. Lunar Trajectories

Two-body celestial mechanics became important to astronautics when lunar trajectories had to be studied for the American Apollo project. The project practiced successfully manned rendezvous maneuvers in lunar orbit (where a landing vehicle docked on a return spacecraft). The project is still the highlight of all past astronautics.

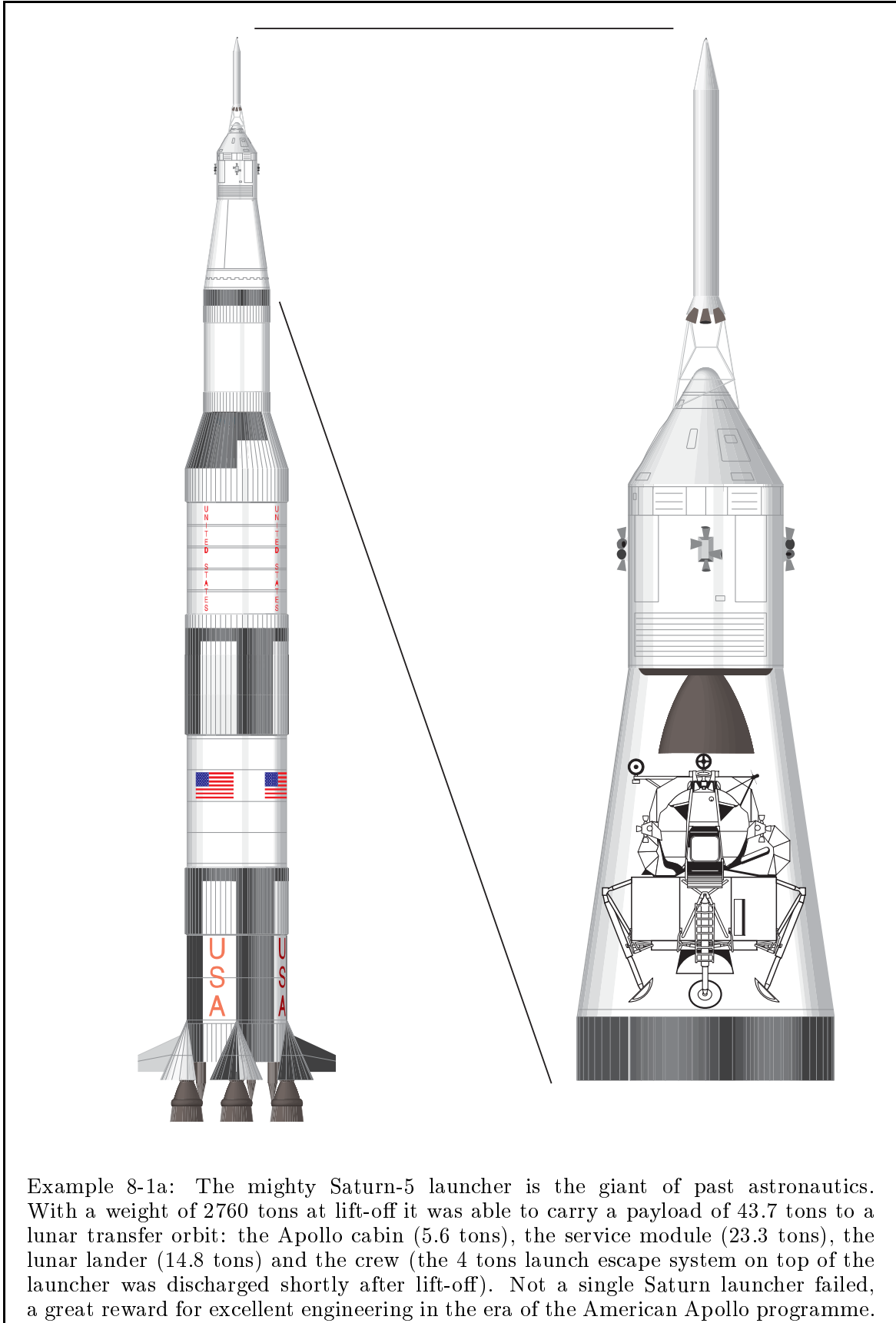
8.1.1 Velocity Requirement for a Lunar Mission

Hardware. The American Apollo project was planned and executed in the time interval between 1961 and 1972. In those days space flight was new, and the project had to face many difficult technical problems with unknown solutions. New equipment with extraordinary performance had to be developed in a short time interval. The mission required a huge space launcher (Saturn-5); a spacecraft with reentry capsule (Apollo) and a moon landing vehicle (Lunar Module). The equipment was developed and tested during a preparatory phase of the project. Access to space and return to the surface of the earth were tested with the Mercury capsule (one astronaut). Orbital maneuvering, rendezvous, docking and extra-vehicular action were trained with the Gemini capsule (two astronauts). A smaller derivative of the moon rocket, the Saturn-1b launcher, was used to test the Apollo capsule (three astronauts). A Lunar Rover for more efficient ground activities was developed in a later phase of the project, when it became sure that the Saturn-5 launcher had a sufficient payload reserve. Finally, a left over Saturn-5 launcher was used to launch the large space station Skylab, a laboratory that was implemented in the third stage of the launcher.

Progress by precaution and redundancy. The success of the Apollo project was based on a careful preceding in small steps from “easy” to “more difficult”. The safety of the astronauts was all the time in the center of the attention:

- the capsule was carried on top of the launcher (recoverable by a rescue rocket);
- the launcher had redundant engines (in its first and in its second stage);
- the launcher had fins for attitude stabilization in case of engine malfunction;
- the capsule used three parachutes for splash-down watering, one was redundant;
- Apollo 11 and 12 used “free-return” trajectories on their way to the moon;
- one astronaut prevailed in the orbiting spacecraft during the landing mission.

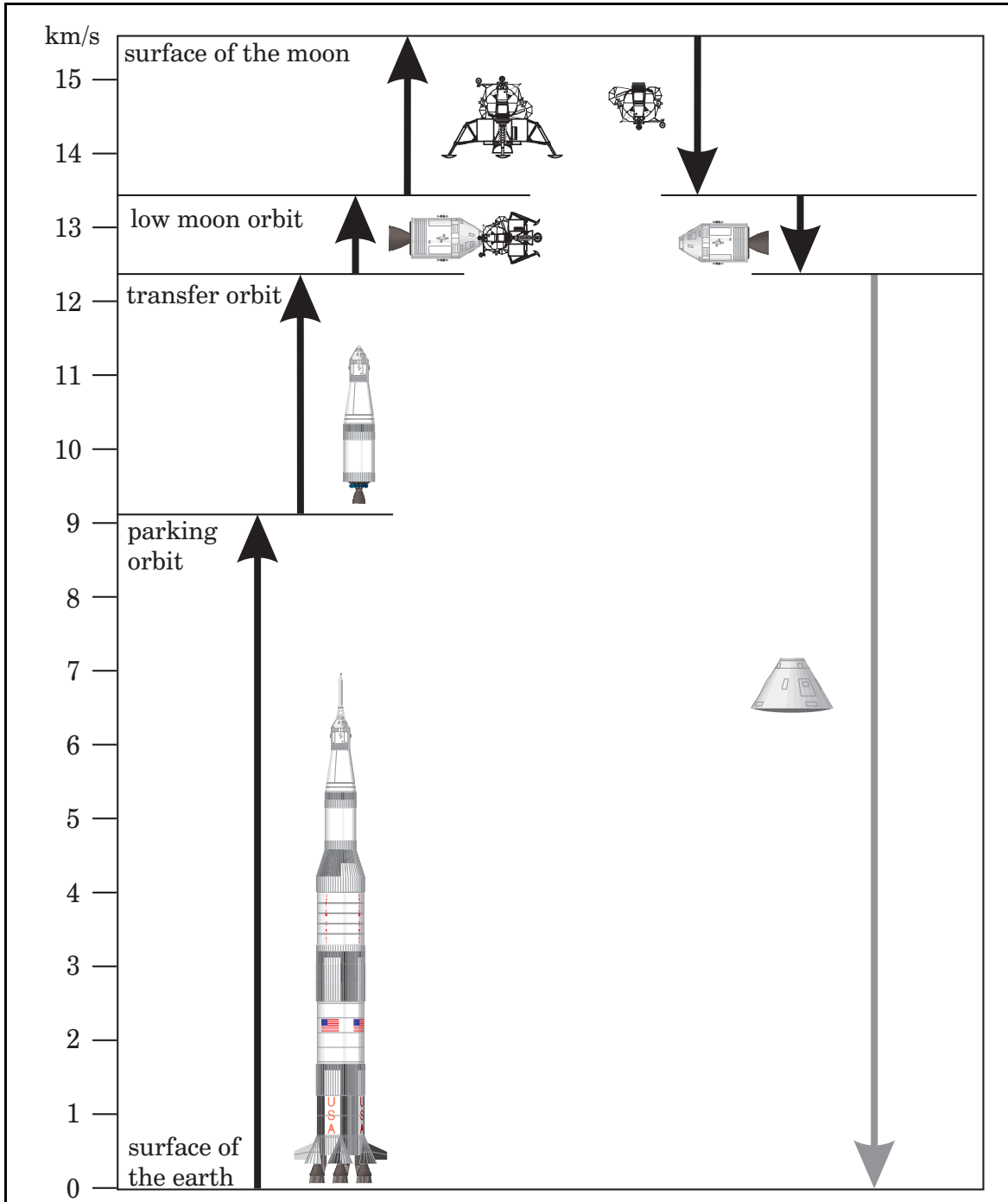
The project became an extraordinary success as a reward for precautions and careful engineering: nine spacecraft with a crew of three astronauts circumnavigated the moon; and twelve astronauts had the opportunity to visit the surface of the moon. All astronauts returned home safely. The greatest success was the rescue of the crew of Apollo 13 (their spacecraft experienced a serious damage when a fuel cell exploded on the way to the moon). Due to engine redundancy, not a single Saturn launcher failed. However, all precautions could not prevent the death of three astronauts in a fire accident that inflamed the capsule during an apparently harmless ground test.



Execution of the lunar missions. Lunar Modules of the Apollo project landed six times on the moon (Apollo 11 to Apollo 17, with the exception of Apollo 13). A typical mission required a time interval of nine days, where the astronauts were able to examine the surface of moon for two days. As the first step of the mission, the Saturn-5 launcher entered a low earth orbit (“parking orbit”). Some hours later the third stage of the launcher performed the injection into the lunar transfer orbit (therefore it was necessary to measure the data of the “parking orbit” accurately). After burnout of the third stage, the Apollo spacecraft separated from the stage, turned around, docked on the Lunar Module and withdrew it from the payload compartment of the empty stage. The transfer time to the moon took 63 hours; the three astronauts lived in the Apollo capsule during the whole time period. Several midcourse maneuvers were necessary to adjust the transfer trajectory. Finally, when the flight configuration arrived at the other side of the moon, a braking maneuver of the Apollo spacecraft established an orbit surrounding the moon. Two astronauts moved into the Lunar Module; one astronaut stayed in the orbiter. Then the Lunar Module separated from the Apollo spacecraft, reduced its velocity and attained an elliptic orbit. The actual descent maneuver of the Lunar Module began about 15 km above the surface by an essential burn maneuver of the first stage, the touch down was about 8 minutes later. The prevail time on the surface was not the same for the six missions, it ranged from 22 hours (Apollo 11) to 75 hours (Apollo 17). The two astronauts used the upper stage of the Lunar Module to return to orbit. They performed a rendezvous maneuver with the orbiter and moved back into the Apollo cabin. After the separation of the Lunar Module, the Apollo spacecraft started its engine for earth return. The return flight took 86 hours. The service module was dropped from the capsule shortly before the reentry phase. The mission was completed with the parachute watering of the capsule and the rescue of the astronauts.

initial mass on the launch pad (launcher+payload)	≈ 2760 t
mass after the launch to the parking orbit ($\Delta v \approx 9100$ m/s)	≈ 125 t
mass on the way to the moon after the injection ($\Delta v \approx 3300$ m/s)	≈ 43.6 t
mass on the lunar orbit after the braking maneuver ($\Delta v \approx 1000$ m/s)	≈ 30.0 t
initial mass of the lander after the separation from the orbiter	≈ 15.0 t
mass on the surface after the landing ($\Delta v \approx 2200$ m/s)	≈ 6.6 t
mass of the second stage when it lifts-off from the surface:	≈ 4.7 t
mass of the stage after ascent to orbit ($\Delta v \approx 2200$ m/s)	≈ 2.1 t
mass of the configuration after docking and separation of the lander	≈ 15.1 t
mass after the departure from the lunar orbit ($\Delta v \approx 1000$ m/s)	≈ 10.5 t
mass of the reentry capsule:	≈ 5.9 t

The table above shows that just a very small part of the mass of the launcher at lift-off reached the surface of the moon ($\approx 0.24\%$); another very small part returned from the lunar parking orbit back to the earth ($\approx 0.22\%$). We can observe that the payload of a big launcher for a hard impact on the moon is approximately 1.6%; however, the payload for a soft landing amounts just about 0.5%. The payload for a flight to the surface of the moon with return ticket to the earth is incredibly small: the only component of the Apollo project which reached the surface of the moon and returned back to the earth was actually nothing else than two astronauts.

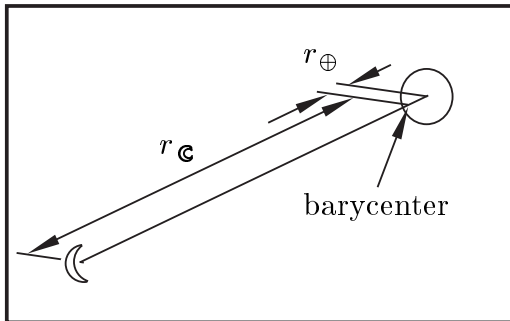


Example 8-1b: A manned moon landing mission needs an immense amount of rocket propulsion. The figure above shows that the principal share is the launch of the mission ($\Delta v \approx 12.4$ km/s), to bring the spacecraft on its way to the moon. Operations on lunar orbit require a Δv capacity of about 1 km/s for each main maneuver (arrival and departure). The descent from orbit to the surface requires a Δv capability of about 2.2 km/s; and the ascent to orbit needs the same amount. Fortunately, the return to the surface of the earth is “cheap”: rocket propulsion is not required because an atmospheric reentry maneuver decelerates the capsule.

8.1.2 Lunar Circumnavigation

The earth-moon gravitational system. We may consider the trajectory of an earth satellite as a conic orbit when it is lower than the geostationary position. The orbit is subjected to perturbations (there is influence of rest atmosphere, asymmetry of the gravitational field, solar pressure and so on). More distant than geostationary position, the gravitational influence of other celestial bodies becomes significant. Gravitation of the sun forces the earth-moon “two body system” to move on a nearly circular orbit around the sun (orbital period $T = 1$ year). When a spacecraft moves on an orbit around the earth, the gravity of the sun is neutralized by a nearly equivalent centrifugal force in opposite direction. Thus, the gravitational field in near-earth space is mainly formed by earth and moon. The sun acts as dominant perturbation.

The center of mass of several celestial bodies is called “barycenter”. Barycenter of earth and moon is located on the straight line between earth and moon, but still inside the earth. Earth and moon move around their barycenter; and, seen from this point, earth and moon are always in opposition to each other (the gravitational influence of the sun can be neglected, and no other external forces act on the system). We find the location of the barycenter when we consider the distance of the two bodies and their mass ratio. When r_{\oplus} is the distance between the center of the earth and the barycenter, and $r_{\mathbb{C}}$ the corresponding distance of the moon, then we have:



$$\gamma_{\mathbb{C}} \cdot r_{\mathbb{C}} = \gamma_{\oplus} \cdot r_{\oplus} \quad (8-1)$$

$$\gamma_{\oplus} = 3.986 \cdot 10^{14} \text{ m}^3/\text{s}^2$$

$$\gamma_{\mathbb{C}} = 4.903 \cdot 10^{12} \text{ m}^3/\text{s}^2$$

$$r_{\mathbb{C}} \approx 380000 \text{ km}$$

$$r_{\oplus} \approx 4670 \text{ km}$$

Terms γ_{\oplus} and $\gamma_{\mathbb{C}}$ are gravitational constants, they are proportional to the mass of the concerning celestial body. Regarding its mass, the moon is 81.301 times smaller than the earth (equatorial earth radius: 6378 km, moon radius: 1741 km).

The mutual attraction between two spherical celestial bodies is inversely proportional to the square of distance (measured from the center of mass). The acceleration of the moon $\ddot{\vec{r}}_{\mathbb{C}}$ is exactly the acceleration caused by the gravitational attraction:

$$\ddot{\vec{r}}_{\mathbb{C}} = - \frac{\gamma_{\oplus}}{(\vec{r}_{\mathbb{C}} + \vec{r}_{\oplus})^2} \cdot \frac{\vec{r}_{\mathbb{C}}}{|\vec{r}_{\mathbb{C}}|} = - \frac{\gamma_{\oplus}^3}{(\gamma_{\mathbb{C}} + \gamma_{\oplus})^2} \cdot \frac{\vec{r}_{\mathbb{C}}}{r_{\mathbb{C}}^3} \quad (8-2)$$

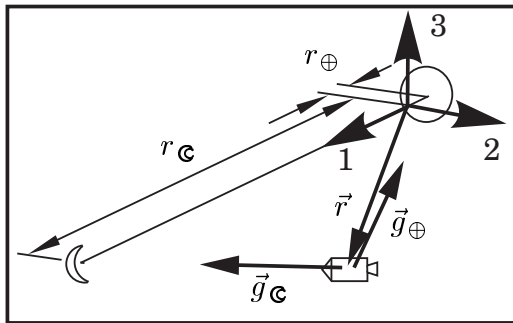
The solution to this equation determines a conic orbit. The moon moves on an elliptic orbit; the barycenter is one of the focal points. The eccentricity of the orbit of moon is low ($\varepsilon = 0.065$); and the distance earth-moon ($r_{\mathbb{C}} + r_{\oplus}$) is approximately constant.

Let us regard in an approximation a constant distance earth-moon. Then the angular velocity ω_{orbit} of the earth-moon two-body system is also approximately constant. Centrifugal and gravitational acceleration must be equal, and we can find ω_{orbit} as:

$$\begin{aligned} \omega_{orbit}^2 \cdot r_{\mathbf{C}} &= \frac{\gamma_{\oplus}}{(r_{\mathbf{C}} + r_{\oplus})^2} \\ \omega_{orbit} &= \sqrt{\frac{\gamma_{\oplus}}{r_{\mathbf{C}}(r_{\mathbf{C}} + r_{\oplus})^2}} = \frac{2\pi}{27.322[\text{days}]} \end{aligned} \tag{8-3}$$

The time interval of 27.322 days is called sidereal period of the moon (sidereal month).

The “restricted three body problem”. The problem is now to calculate the motion of a spacecraft that coasts in the gravitational field of the earth-moon system. The problem is called “restricted three body problem” of celestial mechanics. The problem is “restricted” because the mass of the third body (the spacecraft) is negligibly small in comparison with the masses of the two other bodies (“primaries”). We will use a moving coordinate system to formulate the motion of the spacecraft. The origin is placed on the barycenter; and the 1-direction aims always at the moon. Earth and moon move always in the 1-2-plane. Thus, the vector of the angular velocity $\vec{\omega}_{orbit}$ aims in 3-direction. The vector that locates the spacecraft $\vec{r} = (x, y, z)$ determines the three-dimensional motion. The trajectory of the spacecraft is just influenced by the gravitational acceleration of the moon $\vec{g}_{\mathbf{C}}$ and of the earth \vec{g}_{\oplus} :



$$\begin{aligned} \vec{g}_{\mathbf{C}} &= \frac{-\gamma_{\mathbf{C}}}{|\vec{r} - \vec{r}_{\mathbf{C}}|^3} \begin{pmatrix} x - r_{\mathbf{C}} \\ y \\ z \end{pmatrix} \\ \vec{g}_{\oplus} &= \frac{-\gamma_{\oplus}}{|\vec{r} - \vec{r}_{\oplus}|^3} \begin{pmatrix} x + r_{\oplus} \\ y \\ z \end{pmatrix} \end{aligned} \tag{8-4}$$

When we differentiate the location \vec{r} to find the velocity $\dot{\vec{r}}$ and the acceleration $\ddot{\vec{r}}$, we have to obey the rules for vector differentiation in a rotating coordinate system:

$$\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}; \quad \dot{\vec{r}} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \omega_{orbit} \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \dot{x} - y \omega_{orbit} \\ \dot{y} + x \omega_{orbit} \\ \dot{z} \end{pmatrix} \tag{8-5}$$

$$\ddot{\vec{r}} = \begin{pmatrix} \ddot{x} - \dot{y} \omega_{orbit} \\ \ddot{y} + \dot{x} \omega_{orbit} \\ \ddot{z} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \omega_{orbit} \end{pmatrix} \times \begin{pmatrix} \dot{x} - y \omega_{orbit} \\ \dot{y} + x \omega_{orbit} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} \ddot{x} - 2\dot{y} \omega_{orbit} - x \omega_{orbit}^2 \\ \ddot{y} + 2\dot{x} \omega_{orbit} - y \omega_{orbit}^2 \\ \ddot{z} \end{pmatrix}$$

The notation for the acceleration $\ddot{\vec{r}}$ would be slightly more complicated when we considered the angular velocity ω_{orbit} of the system as a function of time ($\dot{\omega}_{orbit} \neq 0$).

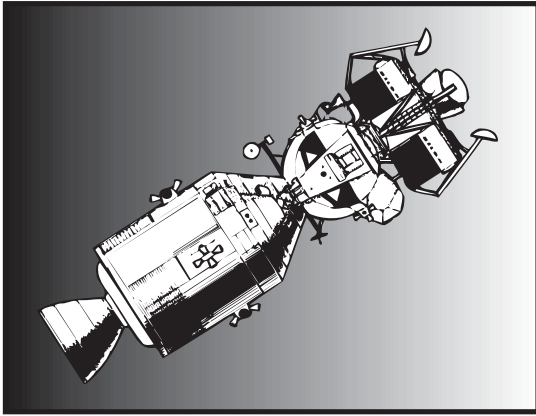
Now the equations of motion ($\ddot{\vec{r}} = \vec{g}_{\mathbf{C}} + \vec{g}_{\oplus}$) can be written in a component notation:

$$\begin{aligned}\ddot{x} - 2\dot{y} \omega_{orbit} - x \omega_{orbit}^2 &= -\frac{\gamma_{\mathbf{C}} \cdot (x - r_{\mathbf{C}})}{\sqrt{(x - r_{\mathbf{C}})^2 + y^2 + z^2}^3} - \frac{\gamma_{\oplus} \cdot (x + r_{\oplus})}{\sqrt{(x + r_{\oplus})^2 + y^2 + z^2}^3} \\ \ddot{y} + 2\dot{x} \omega_{orbit} - y \omega_{orbit}^2 &= -\frac{\gamma_{\mathbf{C}} \cdot y}{\sqrt{(x - r_{\mathbf{C}})^2 + y^2 + z^2}^3} - \frac{\gamma_{\oplus} \cdot y}{\sqrt{(x + r_{\oplus})^2 + y^2 + z^2}^3} \\ \ddot{z} &= -\frac{\gamma_{\mathbf{C}} \cdot z}{\sqrt{(x - r_{\mathbf{C}})^2 + y^2 + z^2}^3} - \frac{\gamma_{\oplus} \cdot z}{\sqrt{(x + r_{\oplus})^2 + y^2 + z^2}^3}\end{aligned}$$

$$\left(|\vec{r} - \vec{r}_{\mathbf{C}}| = \sqrt{(x - r_{\mathbf{C}})^2 + y^2 + z^2} ; \quad |\vec{r} - \vec{r}_{\oplus}| = \sqrt{(x + r_{\oplus})^2 + y^2 + z^2} \right) \quad (8-6)$$

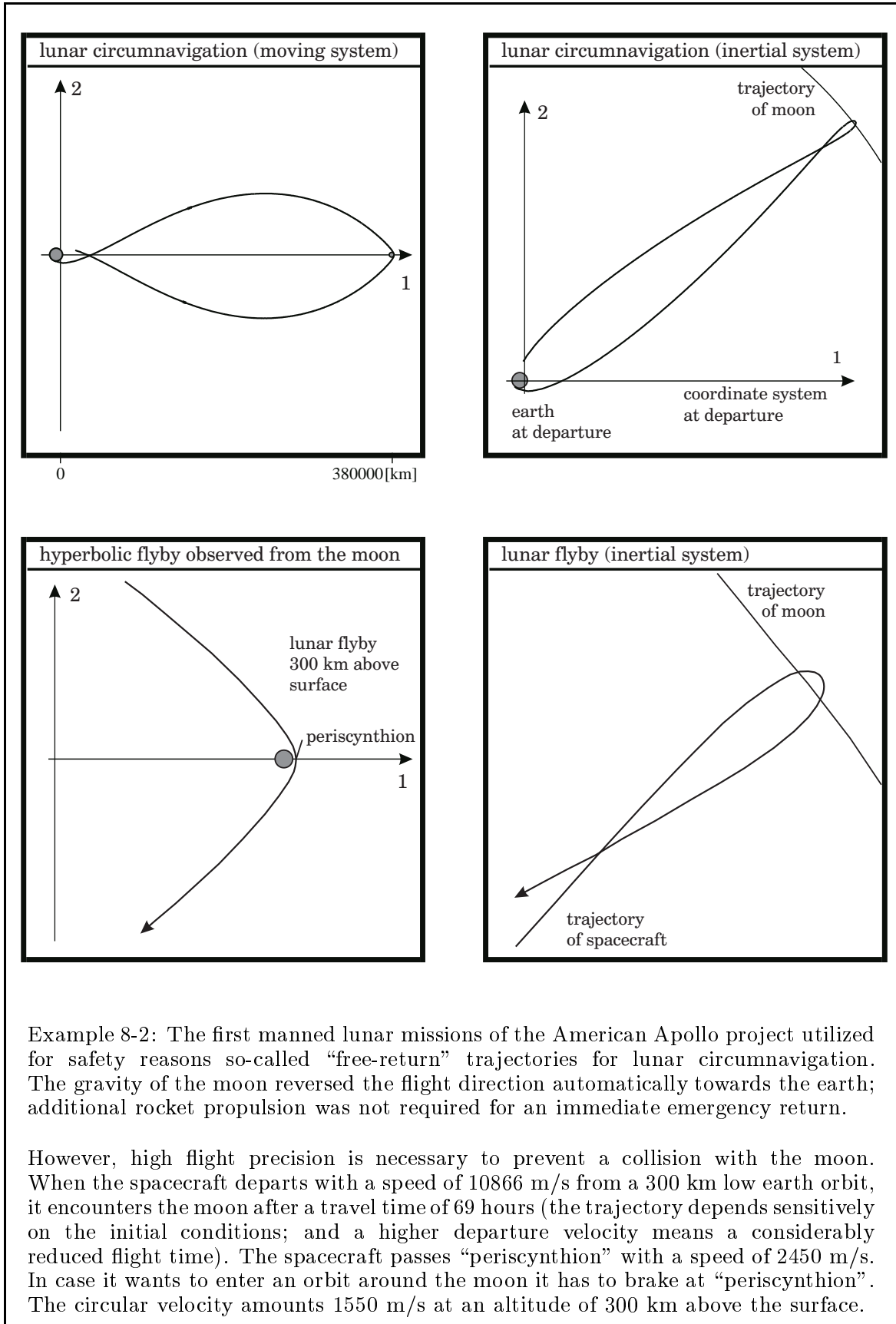
The gravity of the moon is small on the surface of the earth ($g_{\mathbf{C}} \approx 3.3 \cdot 10^{-5} \text{ m/s}^2$), about 300000 times lower than the gravity of the earth (x between $-4670 \pm 6378 \text{ km}$). The centrifugal effects are here of the same magnitude ($1.2 \cdot 10^{-5} \text{ m/s}^2$ on the side of the earth that is close to the moon, and $7.8 \cdot 10^{-5} \text{ m/s}^2$ on the other side of the earth). The lunar attraction together with centrifugal forces is responsible for the tides, but in the calculation of low earth orbits it can be neglected.

Lunar rendezvous. It was a safety reason when so-called “free return trajectories” were used for the first manned lunar missions of the Apollo project: a free-return transfer trajectory does not require propulsion for the circumnavigation of the moon.



In case a mission had to be terminated before landing on the moon, the free return trajectory would bring the spacecraft automatically back to the earth. Since the use of free-return trajectories means a restriction concerning possible landing areas on the moon, the safety method was dropped for later flights. (therefore Apollo 13 required some thrust for circumnavigation, provided by the engine of the Lunar Landing Module).

Consider a spacecraft that moves on a parking orbit around the earth, for example at 300 km altitude. To visit the moon the spacecraft has to increase its speed nearly to earth escape velocity (10926 m/s). When the spacecraft departs with a velocity of 10846 m/s it arrives at the moon 100 hours later. The travel time reduces to 82 hours when the spacecraft uses a little more departure speed (10852 m/s); and the flight takes just 69 hours with a departure speed of 10866 m/s. The transfer trajectory depends sensitively on the initial conditions: slightly more departure speed reduces the travel time considerably. Very accurate navigation is required, otherwise the spacecraft will collide with the moon (or will at least not return to earth anymore). The nearest distance to the moon, called “periscynthion”, is reached when the spacecraft arrives at the opposite side of the moon ($y = 0$). Here, at an altitude of 300 km above the surface, the relative speed amounts 2450 m/s. The spacecraft has to brake 900 m/s when it intends to establish a circular moon orbit (velocity: 1550 m/s).

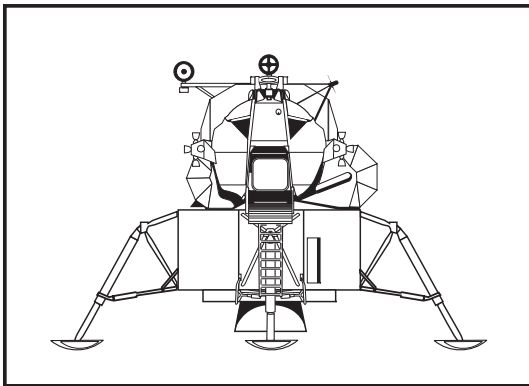


Example 8-2: The first manned lunar missions of the American Apollo project utilized for safety reasons so-called “free-return” trajectories for lunar circumnavigation. The gravity of the moon reversed the flight direction automatically towards the earth; additional rocket propulsion was not required for an immediate emergency return.

However, high flight precision is necessary to prevent a collision with the moon. When the spacecraft departs with a speed of 10866 m/s from a 300 km low earth orbit, it encounters the moon after a travel time of 69 hours (the trajectory depends sensitively on the initial conditions; and a higher departure velocity means a considerably reduced flight time). The spacecraft passes “periscynthion” with a speed of 2450 m/s. In case it wants to enter an orbit around the moon it has to brake at “periscynthion”. The circular velocity amounts 1550 m/s at an altitude of 300 km above the surface.

Landing on the moon. A mission profile was selected for the American Apollo project which included the rendezvous between a moon orbiter and a moon lander. In comparison with the alternative, the “direct landing” of the entire spacecraft, the “lunar orbit rendezvous” has the advantage that the heavy return vehicle stays safely on an orbit around the moon. The disadvantage is that on orbit rendezvous and docking are difficult and risky maneuvers (but experience proved the feasibility).

The Apollo spacecraft stayed on orbit while the landing vehicle performed its mission. The moon lander reduced its velocity and descended to the surface (the landing maneuver consumed the propellant of the first stage). After the visit of the surface the astronauts returned to orbit with the second stage of the landing vehicle.

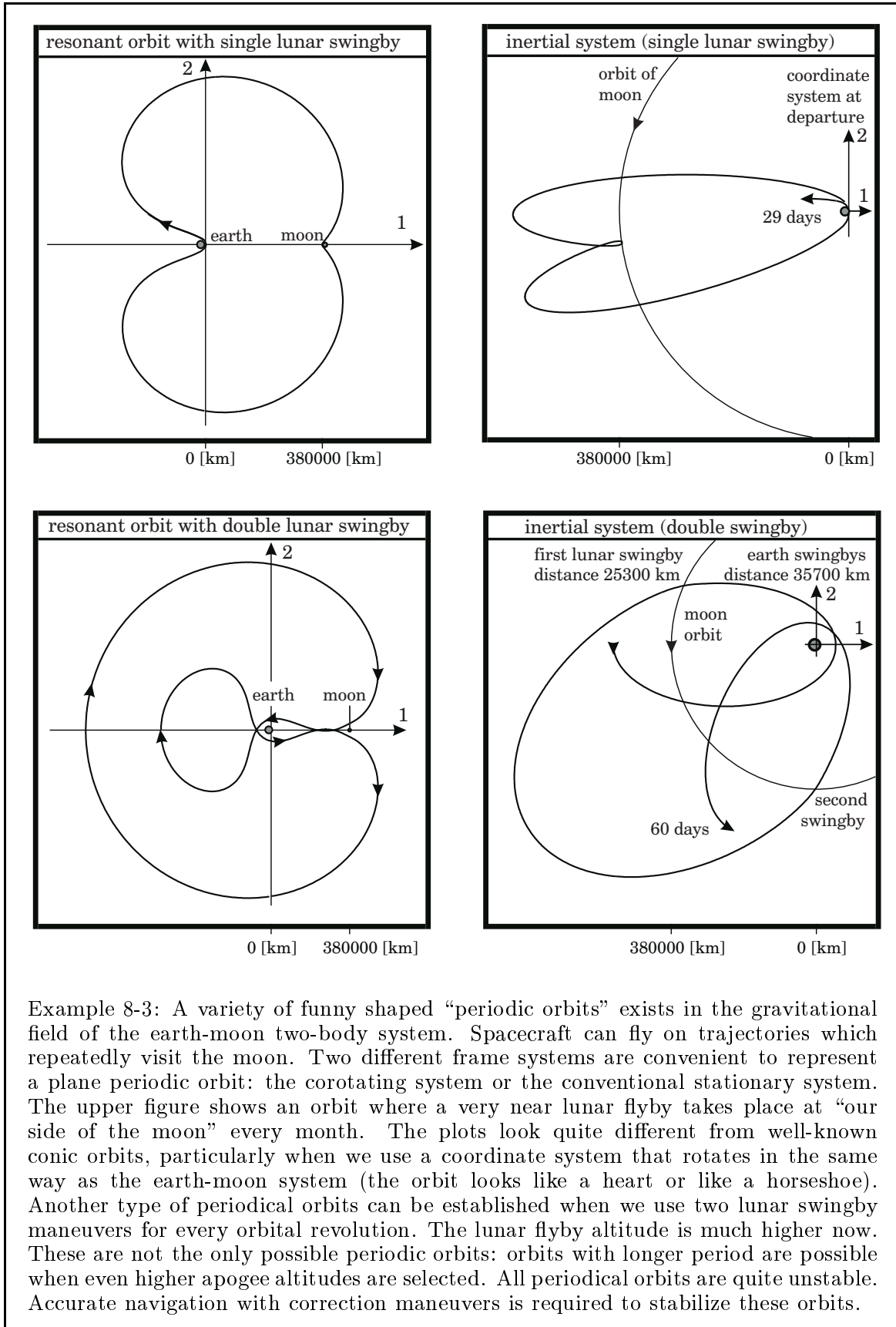


Both maneuvers (descent and ascent) required a Δv capability of about 2 km/s (the Lunar Landing Module had about 10% reserve). The lander docked on the orbiter and the two astronauts moved back to the Apollo cabin. Then the spacecraft discharged the empty second stage of the lander and departed for the earth. The empty stage of the lander stayed on orbit and collided with the surface of the moon after a braking maneuver.

Lunar orbits are quite unstable, due to perturbations that come from the irregular gravitational field of the moon. As a consequence, the moon does not have natural satellites (in the entire solar system no moon has a natural satellite, “submoons” do not exist). A spacecraft that intends to stay for an extended time interval on an orbit around the moon needs rocket propulsion for orbit stabilization.

Periodic orbits in the earth-moon system. Let us have again a look at the equation system (8-6). We can see that trajectories exclusively in the 1-2 plane are possible ($z = 0$, $\dot{z} = 0$ and $\ddot{z} = 0$). It is interesting to identify periodic orbits in the earth-moon system (“long-duration” orbits which oscillate between earth and moon can be important for a scientific probe or for the planning of a manned space-station). There are several opportunities to synchronize the plane trajectory of a spacecraft with the monthly rotation of the earth-moon system. Every orbital revolution the spacecraft has to execute a single or a double swingby maneuver at the moon.

Periodic orbits in the earth-moon system are unstable, particularly when the orbit approaches the moon closely. The word “unstable” has many interpretations: here it means that it is necessary to stabilize a periodic orbit by the appropriate application of propulsive maneuvers. The trajectory is continuously subjected to perturbations: the moon moves on an eccentric orbit and has an irregular gravitational field. Any small trajectory deviation will be amplified considerably after a near moon flyby. It is required to correct the trajectory permanently, otherwise the spacecraft collides with the moon after a relatively short period of time (after some revolutions).



Example 8-3: A variety of funny shaped “periodic orbits” exists in the gravitational field of the earth-moon two-body system. Spacecraft can fly on trajectories which repeatedly visit the moon. Two different frame systems are convenient to represent a plane periodic orbit: the corotating system or the conventional stationary system. The upper figure shows an orbit where a very near lunar flyby takes place at “our side of the moon” every month. The plots look quite different from well-known conic orbits, particularly when we use a coordinate system that rotates in the same way as the earth-moon system (the orbit looks like a heart or like a horseshoe). Another type of periodical orbits can be established when we use two lunar swingby maneuvers for every orbital revolution. The lunar flyby altitude is much higher now. These are not the only possible periodic orbits: orbits with longer period are possible when even higher apogee altitudes are selected. All periodical orbits are quite unstable. Accurate navigation with correction maneuvers is required to stabilize these orbits.

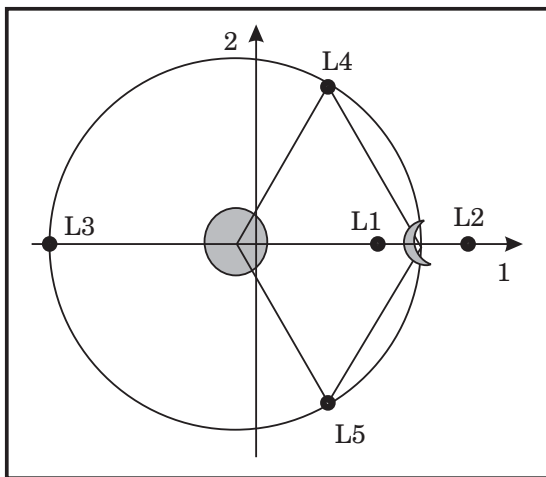
8.1.3 Liberation Points of the Earth-Moon System

The three collinear liberation points L1, L2 and L3. The 1-axis is the straight line which connects the earth with the moon (compare equation 8-6). A spacecraft that stays on this line experiences three forces: gravity from the earth's attraction, gravity from the moon's attraction, and also a centrifugal force caused by the rotation of the coordinate system. There are three locations on the 1-axis where the three forces neutralize each other, called "liberation points" or "Lagrangian points". A spacecraft placed at one of these points does not notice any acceleration with respect to the moving coordinate system. The equation of condition for the collinear liberation points L1, L2 and L3 follows from the system (8-6): all coordinates and their time derivatives stand at zero, with exception of the coordinate x . Therefore:

$$-x \omega_{orbit}^2 = -\frac{\gamma_{\oplus} \cdot (x - r_{\oplus})}{|x - r_{\oplus}|^3} - \frac{\gamma_{\ominus} \cdot (x + r_{\oplus})}{|x + r_{\oplus}|^3} \quad (8-7)$$

The evaluation of the equation (8-7) requires a numerical iteration scheme. We find as solution for L1 $x = +322000$ km, for L2 $x = +444650$ km, and for L3 $x = -386700$ km. Point L1 is located between earth and moon in "cislunar space", comparatively close to the moon. Point L2 is located on the other side of the moon. Point L3 is in opposition to the position of the moon, nearly at the same distance from the earth.

The triangular liberation points L4 and L5. Let us consider again the equation system (8-6). There are two more locations in the 1-2 plane where acceleration and velocity vanish ($\ddot{x} = \ddot{y} = 0$, $\dot{x} = \dot{y} = 0$). These solutions are called "triangular liberation points L4 and L5". The locations of L4 and L5 form with the locations of earth and moon equilateral triangles. Thus, we have $|\vec{r} - \vec{r}_{\oplus}| = |\vec{r} - \vec{r}_{\ominus}| = r_{\oplus} + r_{\ominus}$. We can compute the triangular points analytically, in contrast to the collinear points:



$$x = \frac{(r_{\oplus} - r_{\ominus})}{2} \quad (8-8)$$

$$y = \pm \sqrt{3} \frac{(r_{\oplus} + r_{\ominus})}{2}$$

From the astronomical point of view, compared to the collinear positions L1 and L2, the triangular positions L4 and L5 are less interesting (but, for example, maybe it is sensible to station a radio telemetry satellite there).

Stability of the liberation points. Soon after discovery of the liberation points (in the 18th century) the question came up whether these positions are stable. It was assumed that in case of stability celestial material would allocate at these points.

In astronautics the liberation points are interesting as parking places for spacecraft. The point in question is whether a spacecraft can be parked there without spending propellant for “station keeping”. The collinear points L1, L2 and L3 are unstable, unfortunately (a parking spacecraft will drift away from the position after a short period of time). Under certain conditions the triangular points L4 and L5 are stable. Stability is not present for the earth-moon system, due to perturbations from the sun. The conditions for stability of L4 and L5 are fulfilled for the Sun-Jupiter system, and indeed minor celestial bodies were found which oscillate around these locations (the famous group of “Trojan asteroids”).

The demonstration of instability for the points L1, L2 or L3 is a common task in the field of celestial mechanics. Considered is motion in the “linear vicinity”, where the nonlinear right sides of the system (8-6) can be replaced by linear terms. Near the points L1, L2 or L3, the location vector $\vec{r} = (x, y, z)$ is replaced in an approximation by the expression $\vec{r} \approx ([x] + \Delta x, \Delta y, \Delta z)$. The square brackets shall indicate that now $[x]$ is not a variable any more, but the distance of the concerning liberation point from the barycenter. Employing linearization, the function $f(x, y, z)$ becomes $f \approx [f] + [\partial f / \partial x] \cdot \Delta x + [\partial f / \partial y] \cdot \Delta y + [\partial f / \partial z] \cdot \Delta z$. All expressions in square brackets are constants, their values have to be calculated for the linearization point. We get for liberation points L1, L2 and L3 the following linear equation system:

$$\begin{aligned} \Delta \ddot{x} - 2\omega_{orbit} \Delta \dot{y} - \omega_{orbit}^2 \Delta x &= (2C_{\mathbf{C}} + 2C_{\oplus}) \Delta x \\ \Delta \ddot{y} + 2\omega_{orbit} \Delta \dot{x} - \omega_{orbit}^2 \Delta y &= -(C_{\mathbf{C}} + C_{\oplus}) \Delta y \\ \Delta \ddot{z} &= -(C_{\mathbf{C}} + C_{\oplus}) \Delta z \end{aligned} \tag{8-9}$$

with the constants: $C_{\mathbf{C}} = [\gamma_{\mathbf{C}} / |r_{\mathbf{C}} - x|^3]$; $C_{\oplus} = [\gamma_{\oplus} / |r_{\oplus} + x|^3]$ (8-10)

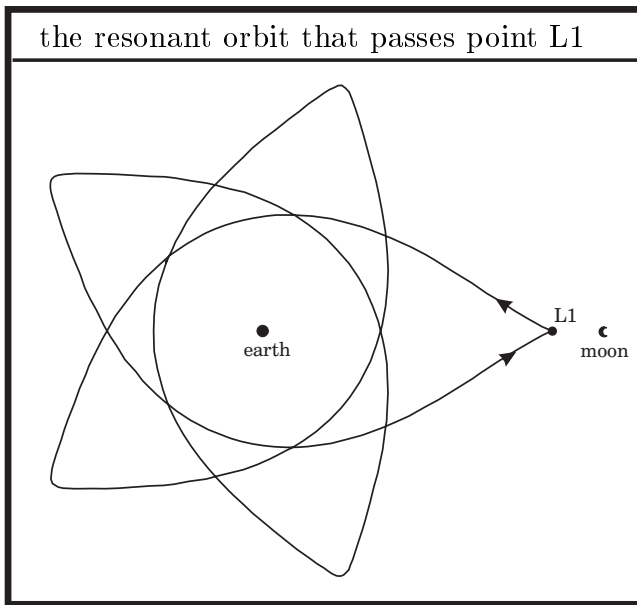
In contrast to the more accurate nonlinear representation (8-6), the linear equation system (8-9) has the only advantage that now an analytical solution can be obtained. Since expression $C_{\mathbf{C}} + C_{\oplus}$ is positive; the motion for Δz is stable and not coupled with the motion of Δx and Δy any more. For linear stability it is sufficient to consider the motion in the 1-2 plane. The procedure works as follows: insert trial solutions $\Delta x = C_1 \cdot e^{\lambda t}$ and $\Delta y = C_2 \cdot e^{\lambda t}$, eliminate from the equation system the amplitude coefficients C_1 and C_2 , and get the characteristic equation of the system (8-9) as:

$$\lambda^4 + (2\omega_{orbit}^2 - C_{\mathbf{C}} - C_{\oplus}) \cdot \lambda^2 + (\omega_{orbit}^2 + 2C_{\mathbf{C}} + 2C_{\oplus})(\omega_{orbit}^2 - C_{\mathbf{C}} - C_{\oplus}) = 0 \tag{8-11}$$

The “roots” λ of the characteristic equation are complex numbers, usually. They determine the solution to the equation system. For stability it is necessary that not a single root λ has a positive real part. Using mathematics of complex numbers we can find out that $(\omega_{orbit}^2 - C_{\mathbf{C}} - C_{\oplus}) > 0$ is a condition for stability. This condition is violated for all collinear liberation points, indicating that these locations are unstable. Even though stable periodic oscillations exist also in the linear vicinity, some roots have a positive real part. Small perturbations will also excite these unstable solutions, and spacecraft parked in L1, L2 or L3 will not stay there for a longer period of time.

We have seen that the conventional way to analyze the stability of a mechanical system is to establish the equations of motion, then approximate these equations in the vicinity of a reference point by a linear system and find the solution to the linearized system of the equation of motion. The roots of the characteristic equation are complex numbers: the imaginary part determines a periodic oscillation and the real part determines the corresponding amplitude function. The motion is just stable when all roots of the characteristic equation have negative real parts. The linearization method involves the disadvantage that it ignores possibly dominant nonlinear effects.

The simple numerical integration of the more accurate nonlinear system of the equations of motion avoids these disadvantages. When we place the spacecraft at the liberation point L1 and simulate the trajectory on a computer using the equation



system (8-6), we can observe how the spacecraft drifts away from the unstable position. Initially, the spacecraft stays for about half a month in the vicinity of the liberation point. Then it moves slowly towards the earth, later it performs five revolutions on an elliptic orbit around the earth. Astonishingly, after 3 months (82 days) it returns back to the initial position, the liberation point L1. The unstable position L1 is a part of a periodic orbit in the earth-moon system. Presented in a rotating coordinate system, the path of the trajectory resembles a flower with five petals.

We want to complete the stability analysis with the calculation of the linear system of the equations of motion in the vicinity of the triangular liberation points L4 and L5. The variables are now $x = (r_{\text{☾}} - r_{\text{⊕}})/2 + \Delta x$; $y = \pm\sqrt{3}(r_{\text{☾}} + r_{\text{⊕}})/2 + \Delta y$; $z = \Delta z$. It is easy to introduce the new variables Δx , Δy , Δz into the left-hand side of equation system (8-6), because in the co-rotating coordinate system the expressions for velocity and acceleration are linear terms. The replacement of the nonlinear right-hand sides of the equation system (8-6) by linear terms is a little more complicated, because the coordinate y is not zero anymore. Again, we form the derivatives of these nonlinear functions with respect to the three coordinates x , y and z ; and replace a function $f(x, y, z)$ by its linear approximation $f \approx [f] + [\partial f/\partial x] \cdot \Delta x + [\partial f/\partial y] \cdot \Delta y + [\partial f/\partial z] \cdot \Delta z$. The expressions in square brackets are constant values taken at the liberation point. We have to consider that the distances of the triangular liberation points L4 and L5 from the primaries are $\sqrt{(x - r_{\text{☾}})^2 + y^2 + z^2} = \sqrt{(x + r_{\text{⊕}})^2 + y^2 + z^2} = r_{\text{⊕}} + r_{\text{☾}}$. The gravitational constants of moon and earth can be eliminated from the equation system using the relationships $\gamma_{\text{☾}} = r_{\text{⊕}} \omega_{\text{orbit}}^2 (r_{\text{⊕}} + r_{\text{☾}})^2$ and $\gamma_{\text{⊕}} = r_{\text{☾}} \omega_{\text{orbit}}^2 (r_{\text{⊕}} + r_{\text{☾}})^2$.

Then, motion in the vicinity of L4 and L5 is determined by the linear equation system:

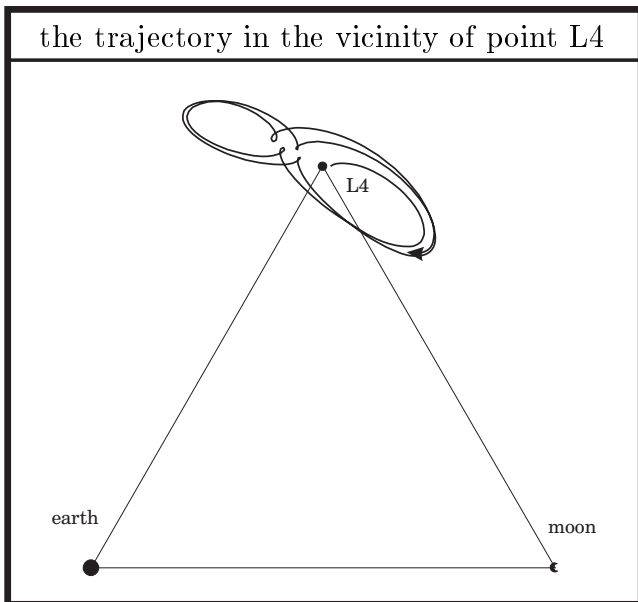
$$\begin{aligned} \Delta\ddot{x} - 2\omega_{orbit}\Delta\dot{y} - \omega_{orbit}^2\Delta x &= \left(\frac{3}{4} - 1\right) \omega_{orbit}^2 \Delta x \pm \frac{3\sqrt{3}}{4} \frac{(r_{\mathbf{c}} - r_{\oplus})}{(r_{\mathbf{c}} + r_{\oplus})} \omega_{orbit}^2 \Delta y \\ \Delta\ddot{y} + 2\omega_{orbit}\Delta\dot{x} - \omega_{orbit}^2\Delta y &= \left(\frac{9}{4} - 1\right) \omega_{orbit}^2 \Delta y \pm \frac{3\sqrt{3}}{4} \frac{(r_{\mathbf{c}} - r_{\oplus})}{(r_{\mathbf{c}} + r_{\oplus})} \omega_{orbit}^2 \Delta x \\ \Delta\ddot{z} &= -\omega_{orbit}^2 \Delta z \end{aligned} \tag{8-12}$$

The system (8-12) is derived from a linearization of the nonlinear system (8-6); the sign + refers to the liberation point L4 and the sign - to the liberation point L5. The motion Δz is entirely decoupled from the motions Δx and Δy ; coordinate Δz oscillates stable with the period ω_{orbit} (the angular velocity of the earth-moon system).

We insert the solutions $\Delta x = C_1 e^{\omega_{orbit} \lambda t}$, $\Delta y = C_2 e^{\omega_{orbit} \lambda t}$ in the first and second equation of system (8-12); eliminate the amplitude coefficients C_1 and C_2 from the equation system and get (after some transformations) the characteristic equation as:

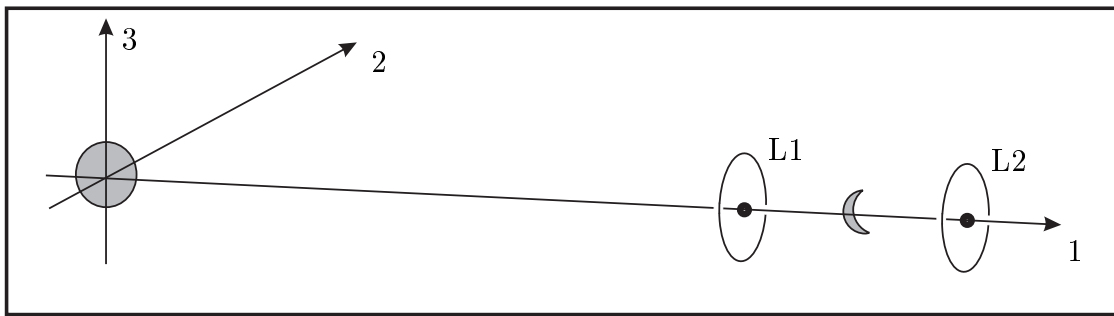
$$\lambda^4 + \lambda^2 + \frac{27}{4} \frac{r_{\mathbf{c}} \cdot r_{\oplus}}{(r_{\mathbf{c}} + r_{\oplus})^2} = 0 \quad ; \text{or:} \quad \lambda^2 = -\frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{27}{4} \frac{r_{\mathbf{c}} \cdot r_{\oplus}}{(r_{\mathbf{c}} + r_{\oplus})^2}} \tag{8-13}$$

For the earth-moon two-body system the ratio $(r_{\mathbf{c}}/r_{\oplus})$ is 81.301; and both solutions to the square of the multiplier λ are purely negative. The root $\lambda^2 = -0.911$ or $\lambda = \pm 0.955\sqrt{-1}$ corresponds to a stable oscillation with the period of nearly one month (exactly $0.995 \omega_{orbit}$); this “short-period mode” can be interpreted as a slight eccentricity of the underlying circular orbit of the liberation point around the earth. The other root $\lambda^2 = -0.089$ or $\lambda = \pm 0.298\sqrt{-1}$ corresponds to a stable oscillation with the period of about three months; this “long-period mode” is a movement of the position toward and away from the moon. The actual trajectory is a superposition of the two modes, the motion depends on the initial deviation from the liberation point.



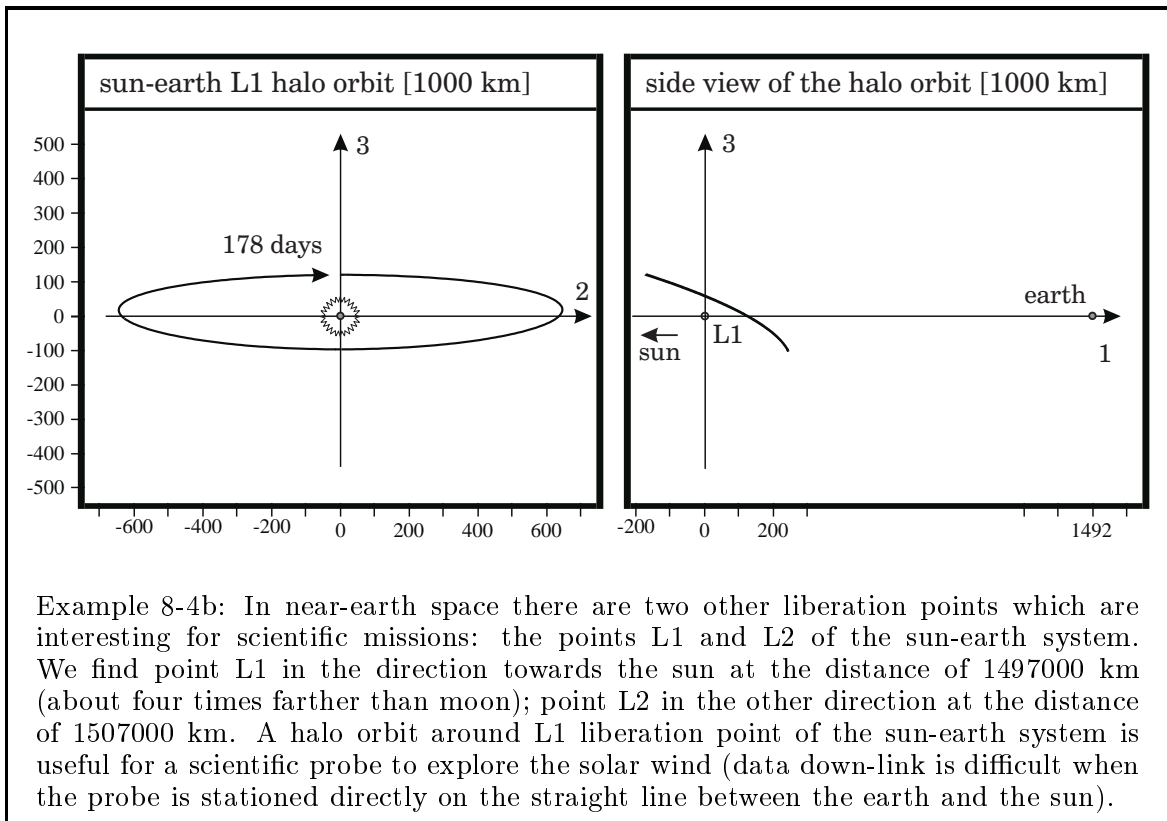
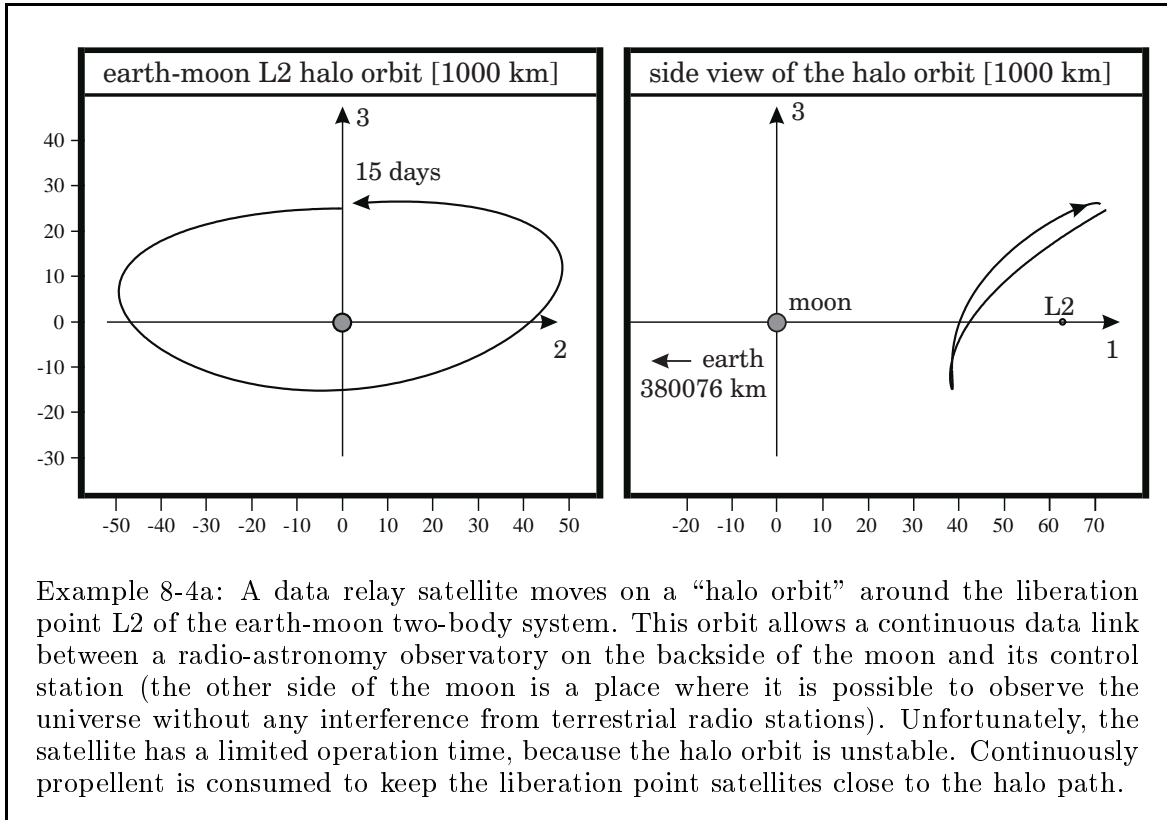
The motion is apparently stable when we integrate the nonlinear system (8-6) numerically to simulate the trajectory of a spacecraft that is placed into the vicinity of point L4. Presented in a co-rotating system, the trajectory resembles clouds that surround the point L4. However, the stability of the triangular liberation points of the earth-moon system is not correctly represented when we ignore the gravitational influence of the sun. When we include the influence of the sun, the spacecraft departs from the L4 in a comparatively short time interval.

Halo orbits. A spacecraft can establish an orbit around a liberation point even when the location is unstable. Three-dimensional periodic orbits around the liberation points L1 and L2 are called “halo orbits”, because they are not centered but aside of the celestial body, like a halo around a head. Halo orbits are not perfectly periodic orbits and unstable like their liberation point: station keeping for a “liberation point satellite” needs deterministic and stochastic trajectory correction maneuvers. For example, a halo orbit around the L2 liberation point of the earth-moon system could be used by a data relay satellites for a radio telescope station on the other side of the moon (for observations without interference from terrestrial radio stations); and a halo orbit around the L1 liberation point of the sun-earth two-body system is useful for sun observatory satellites (to explore the solar radiation at a distance where it is still undisturbed by the magnetic field of the earth).



Equation (8-9) shows that in the linear vicinity of the liberation point the oscillations Δx and Δy are coupled and have the same frequency; the oscillation Δz is decoupled and has another (but similar) frequency. More distant from the liberation points nonlinear effects become stronger. Let us now consider that the rotation of the coordinate system stops suddenly ($\omega_{orbit} = 0$), when an imaginary hand stops the motion of the earth and the moon and keeps them in their positions. Naturally, this is not at all an allowed simplification, in reality the two-body system would collapse; but the model assumption makes the properties of halo orbits more transparent. The liberation point L1 (between earth and moon) moves to the position where the gravity attraction of earth is exactly equal to the attraction of the moon. You can easily verify that this position is unstable: a parking spacecraft that moves slowly to the earth or to the moon experiences in any case an increasing gravity attraction. You can also see that it is possible to lay a circular (or an elliptic) “halo orbit” around the point L1; where the x coordinate is constant (the x -axis is the orbit normal). The composed gravity vector (attraction from the earth and from the moon) is all the time in the orbital plane, rectangular to the x -axis for all locations on the orbit.

More accurate analytical description of halo orbits is a quite complicated procedure. A linear model is insufficient since the problem is dominated by nonlinear effects; but numerical computation of the equation system (8-6) is comparatively easy. The problem is to find an orbit stabilization strategy which consumes propellant sparingly. Halo orbits have been established by American and European probes around the L1 liberation point of the sun-earth system (the sun explorers ISEE-3 and SOHO). Station keeping required two maneuvers every revolution ($\Delta v \approx 10$ m/s every year).



8.1.4 Escape from the Earth-Moon System

Jacobi's integral. There exists one analytical integral for the equation system (8-6):

$$J = \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{2} - \frac{\omega_{orbit}^2(x^2 + y^2)}{2} \quad (8-14)$$

$$- \frac{\gamma_{\mathbf{e}}}{\sqrt{(x - r_{\mathbf{e}})^2 + y^2 + z^2}} - \frac{\gamma_{\oplus}}{\sqrt{(x + r_{\oplus})^2 + y^2 + z^2}}$$

Term J is constant for a spacecraft that coast in the two-body gravitational field. Note that the Jacobi integral (8-14) is not a representative of the mechanical energy: the third part and the fourth part of the right hand side of the equation are indeed the potential energies of the two gravitational fields; but $\dot{x}^2 + \dot{y}^2 + \dot{z}^2 - \omega_{orbit}^2(x^2 + y^2)$ is not the square of the inertial velocity $d\vec{r}/dt$ (compare equation 8-5).

However, the integral J represents the potential with respect to the moving system: expression $(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)/2$ is the "specific kinetic energy" of the relative motion, and expression $-\omega_{orbit}^2(x^2 + y^2)/2$ is the "potential of the centrifugal acceleration".

We can use the integral J to write down the equations of motion (8-6) in the form:

$$\begin{aligned} \ddot{x} - 2\dot{y} \omega_{orbit} &= -\partial J / \partial x \\ \ddot{y} + 2\dot{x} \omega_{orbit} &= -\partial J / \partial y \\ \ddot{z} &= -\partial J / \partial z \end{aligned} \quad (8-15)$$

Specific mechanical energy. The next point in question is the relation between the Jacobi integral J and the specific (mass related) energy e . We may expect that the notation for e resembles the notation for J . The energy of a spacecraft that coasts in the two-body field consist of kinetic energy and potential energy, where the specific kinetic energy is the square of the inertial velocity, divided by factor 2. Thus:

$$e = \frac{(\dot{x} - \omega_{orbit}y)^2 + (\dot{y} + \omega_{orbit}x)^2 + (\dot{z})^2}{2} \quad (8-16)$$

$$- \frac{\gamma_{\mathbf{e}}}{\sqrt{(x - r_{\mathbf{e}})^2 + y^2 + z^2}} - \frac{\gamma_{\oplus}}{\sqrt{(x + r_{\oplus})^2 + y^2 + z^2}}$$

We insert equation (8-16) into equation (8-14) to find the following relationship:

$$\begin{aligned} J &= e - \omega_{orbit} \cdot \{x\dot{y} - y\dot{x} + \omega_{orbit}(x^2 + y^2)\} \\ &= e - \omega_{orbit} \cdot h \cos \vartheta \end{aligned} \quad (8-17)$$

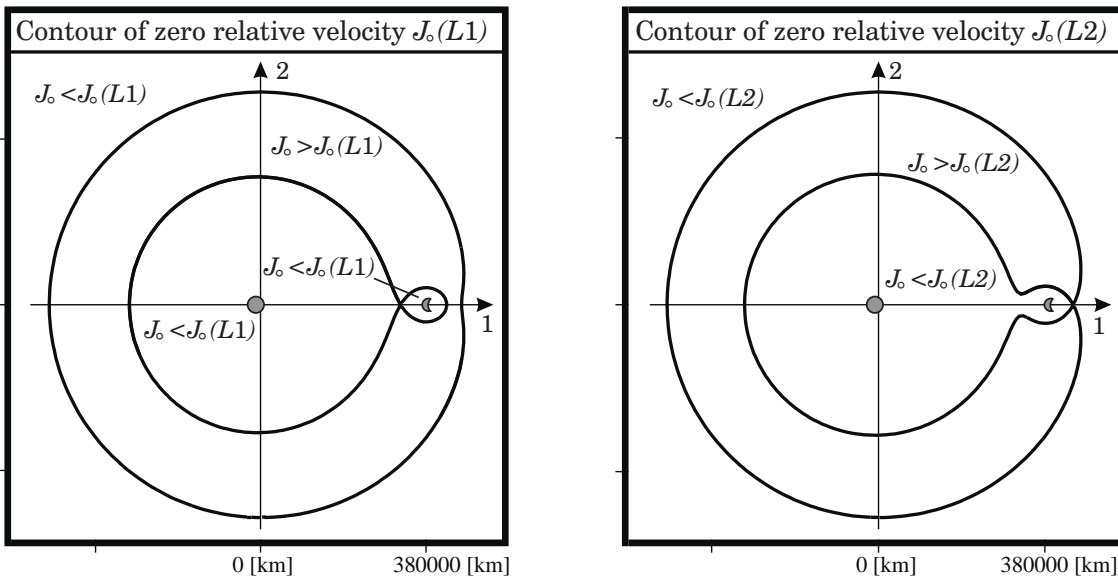
The equation above uses an expression in curly braces; this expression can be interpreted as third component of the angular momentum vector $\vec{h} = \vec{r} \times \vec{v}$. The third component of the angular momentum vector is simply $h \cos \vartheta$, when h is the length of the angular momentum vector and ϑ the inclination angle of the flight plane. Equation (8-17) is the theoretical basis for "Tisserand's criterion" (a rule to identify lost comets which have changed their orbital elements after a near Jupiter passage).

Equation (8-17) determines the relationship between the value of the Jacobi integral J and three other parameters: energy e , angular momentum h and inclination angle ϑ . The three parameters e , h and ϑ are constant values when a spacecraft coasts in the gravitational field of a single celestial body (the “restricted two-body problem”). However, these values are not conserved when a spacecraft coasts in the gravitational field of two celestial bodies (the “restricted three-body problem”). For trajectories in the two-body system the Jacobi integral J is the only analytical conservation law. Here it is a necessary condition that a spacecraft coasts with constant Jacobi integral (satisfying the complete set of equations of motion). Like the energy integral of the one-body system, the Jacobi integral provides no information on the flight time.

Contours of zero relative velocity. We can set the relative velocity to zero ($\dot{x} = 0, \dot{y} = 0, \dot{z} = 0$) and evaluate the equation (8-14) for every location $\vec{r} = (x, y, z)$ in the two-body system. We get for the liberation points of the earth-moon system:

$$J_o(L1) = -1.672 \text{ km}^2/\text{s}^2 ; \quad J_o(L2) = -1.663 \text{ km}^2/\text{s}^2 ; \quad J_o(L3) = -1.579 \text{ km}^2/\text{s}^2$$

$$J_o(L4) = J_o(L5) = -1.567 \text{ km}^2/\text{s}^2$$



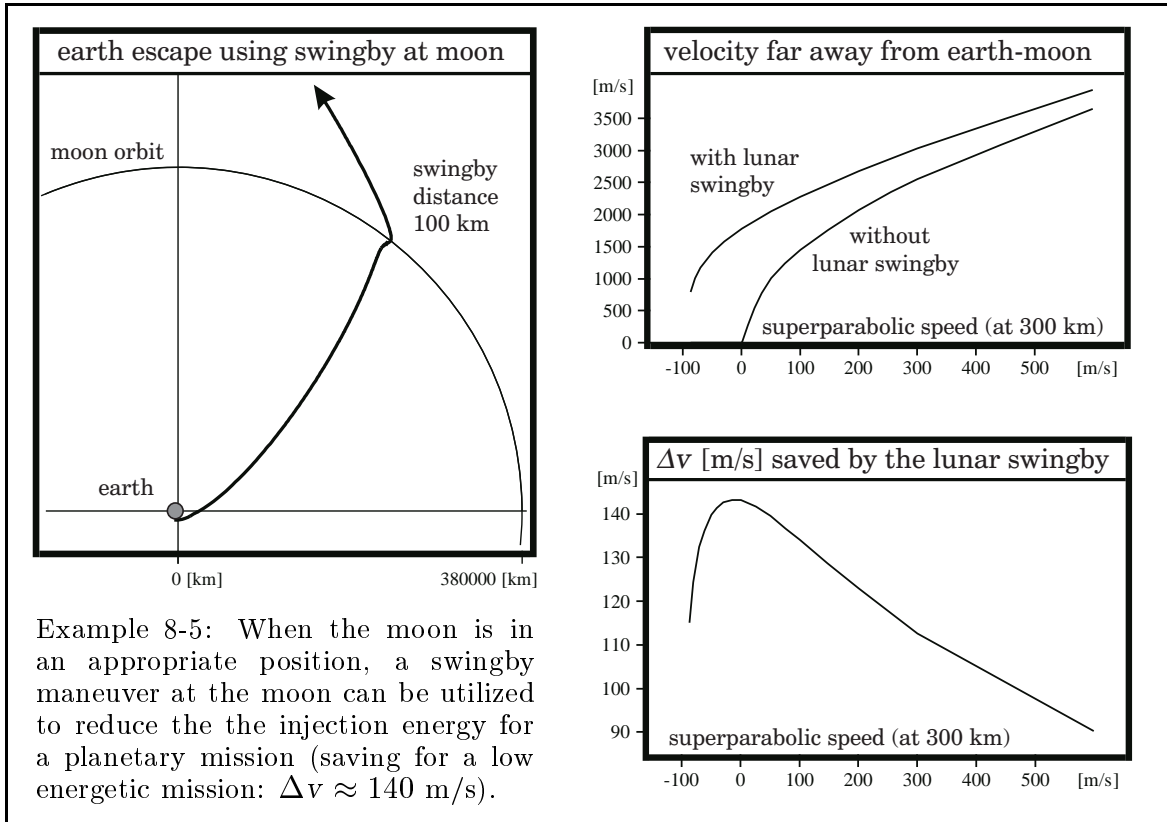
Term J_o is the value of the Jacobi integral at a certain position without any velocity relative to the moving coordinate system. Let us now regard the 1-2-plane of the earth-moon system. When we mark all positions for a certain (predetermined) value of J_o , we get the concerning “line of zero relative velocity”. Note that the integral J involves just a quadratic form of the the relative velocity ($\dot{x}^2 + \dot{y}^2 + \dot{z}^2$). When a spacecraft arrives at a certain line J_o and has still a relative velocity there, then the value J of its flight path must be higher than the value J_o of the intercepted line of zero relative velocity; and when the spacecraft has arrived at this line and rests there (with zero relative velocity), the spacecraft cannot move from this line to another line with a higher values of J_o . It must fall back to lines where the values of J_o are lower. Thus, these contours J_o represent border lines for trajectories with predetermined value of J , which cannot be crossed without propulsion.

Accessible regions. The contours of zero relative velocity for the value $J_o(L1)$ form three curves: there is a drop-shaped curve around the earth; there is another but much smaller drop-shaped curve around the moon (both are connected at L1); and finally there is a bigger curve that encircles the entire earth-moon system. The three curves are nearly circles. Inside the two smaller curves and outside the bigger curve the value J_o of zero relative velocity is smaller than the value of $J_o(L1)$. The trajectory of a spacecraft that coasts from the earth to the moon needs at least to have the value of $J_o(L1)$. A spacecraft that coasts with this value will pass the liberation point L1 on its way from the earth to the moon. When it arrives at L1, it will rest there with “zero relative velocity” (later the spacecraft will drift away because the parking position L1 is unstable). With the value $J_o(L1)$ it is impossible to reach the liberation point L2. While outside again a region with the same value $J_o(L1)$ exists, it will not be possible to cross from inside to outside.

The value $J_o(L2)$ is the minimum value of the Jacobi integral which allows trajectories from inside to outside. If we draw the contours of the value $J_o(L2)$, a small bottleneck opens at point L1. A spacecraft that wants to escape from the earth-moon system with the minimum value of the Jacobi integral will have to pass L1 closely and L2 exactly. The value $J_o(L2)$ is just slightly higher than the value $J_o(L1)$; it indicates that a moon passage cost nearly as much propulsion as an earth escape maneuver.

Escape utilizing a lunar gravity assist maneuver. The Jacobi integral provides just limited help to find an appropriate minimum energy escape trajectory. First, the Jacobi integral is not an expression for the energy of the trajectory. Second, there is no evidence that it is possible to find for given value of J an appropriate trajectory. The contours of zero relative velocity J_o are not at all trajectories. In any case it is necessary to take the system of the equations of motion (8-6) under consideration. In practical astronautics the question is important whether the moon can be helpful to save propellant for planetary missions. Computational analysis shows that a close fly-by at the moon can save propellant equivalent to a Δv consumption of up to 140 m/s. This looks promising, but usually the moon will not be in an appropriate position. Theoretically, it is even possible to assist an earth escape maneuver by several lunar swingbys: the spacecraft can return to the earth on an interplanetary trajectory; but practically the benefit is small and the price is high (increased mission time). We may conclude that a lunar swingby is impractical as an initial step for a planetary mission. Just in some special situations a lunar swingby maneuver can be helpful (for example, when a low energy escape is required to flyby a near-earth asteroid).

Sphere of lunar influence. The potential of gravitation is inversely proportional to the distance, and therefore the gravitational field of any celestial body is unlimited. Strictly speaking, a “sphere of gravitational influence” does not exist. We can define it, assuming that outside this sphere the gravity is smaller than a certain limit. Inside this “sphere of influence” the gravity of the celestial body is dominant, outside the sphere it can be neglected. Since any limit can be taken, the radius of the sphere is more or less arbitrarily defined. The “sphere of influence” serves just as a model, but for trajectory integration this model is usually not accurate enough.



Let us now assume that the gravity of the moon is negligibly small in comparison with the gravity of the earth (the hypothetical case that $\gamma_{\text{C}} \ll \gamma_{\oplus}$). The barycenter of the earth-moon system moves to the center of the earth; and the “sphere of lunar influence” becomes an infinitely small ball around the moon. A spacecraft that moves in this gravitational field coasts on a conic orbit, with the earth in one of the focal points. When the spacecraft intercepts the “sphere of lunar influence”, however, it experiences an impulse and changes instantaneously the mechanical energy e , the angular momentum h and the inclination ϑ of its orbit. Condition (8-17) determines the relationship between these values, where term J is constant. Equation (8-14) is the equation of condition for the value J . Obviously, the location vector $\vec{r} = (x, y, z)$ has the same value for every location on the border of the infinitely small “sphere of lunar influence”. Since J is conserved, the length of the velocity vector $\vec{v} = (\dot{x}, \dot{y}, \dot{z})$ (relative to the moon) does not change during the impulsive lunar flyby maneuver. However, the length of the absolute (geocentric) velocity is not the same any more: after the maneuver the spacecraft moves with the absolute velocity of the moon, increased by the relative velocity before the maneuver. The direction of the relative velocity after the maneuver depends on which side (and how near) moon is passed.

The mass of moon has no influence on the maneuver because we considered the moon as a particle without physical size, but the mass of the moon is not at all concentrated on a small ball with negligible radius. The moon has a radius of 1741 km; and a closer flyby distance is impossible. The mass affects indirectly the efficiency of the maneuver, because the trajectory deflection angle diminishes with the flyby distance.

8.2. Planetary Flights

After having left the gravitational field of the earth-moon system on an escape orbit, the spacecraft moves on a heliocentric trajectory. When the motor is not switched-on, the interplanetary trajectory of the spacecraft is a conic orbit with the sun in one of the focal points. Solar gravity is the dominant force; and the gravitational influence of planets may be neglected. Approaching the target planet, the spacecraft enters the planet's "sphere of influence", where the gravity of the planetary system is dominant and the gravity of the sun may be neglected. Thus, in an approximation, we may consider planetary transfer trajectories as composed of conic sections.

8.2.1 Earth Departure

Hyperbolic earth escape velocity. We regard a spacecraft that moves on a circular low earth orbit. The rocket motor is ignited and the spacecraft accelerates. When it is intended that the vehicle escapes from the gravitational field of the earth, the spacecraft must at least accelerate until it has reached "parabolic escape speed". The burnout velocity must be higher than parabolic escape speed to enable the spacecraft to travel to other planets. This velocity is called "hyperbolic escape velocity"; the spacecraft is by the quantity "superparabolic speed" faster than parabolic speed:

$$v_{circular} = \sqrt{\gamma_{\oplus}/r_p}$$

$$v_{parabolic} = \sqrt{2} \cdot v_{circular} \quad (8 - 18)$$

$$v_{hyperbolic} = v_{parabolic} + v_{superparabolic} \quad (8 - 19)$$

In the equations above term γ_{\oplus} is the the gravitational constant of earth; and term r_p is the distance of the spacecraft from the gravitational center at the moment of the burn maneuver. After burnout the spacecraft departs from the gravitational field; the mechanical energy ($e = v^2/2 - \gamma_{\oplus}/r$) of the escape trajectory is a constant value. We can use the constant energy to calculate the velocity at infinite distance v_{∞} :

$$v_{\infty} = \sqrt{2 v_{parabolic} v_{superparabolic} + v_{superparabolic}^2} \quad (8 - 20)$$

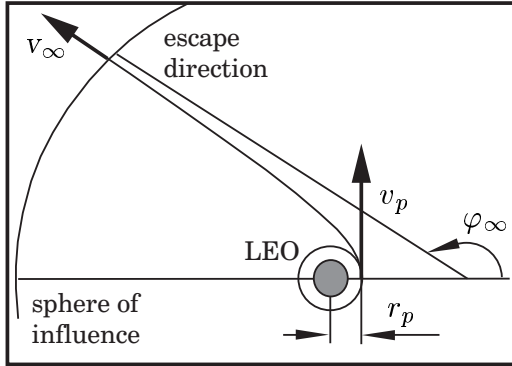
The velocity v_{∞} (far away from the earth) is higher than the velocity $v_{superparabolic}$. Once the vehicle has attained escape speed, the gravity helps: it is much more efficient to accelerate a spacecraft near the gravitational center than far away from it.

Geometry of the escape orbit. Knowing pericenter radius r_p and escape velocity v_{∞} , we can calculate eccentricity ε and pericenter velocity v_p of the escape trajectory:

$$\begin{aligned} \varepsilon - 1 &= r_p v_{\infty}^2 / \gamma_{\oplus} \\ \varepsilon + 1 &= r_p v_p^2 / \gamma_{\oplus} \end{aligned} \quad (8 - 21)$$

Since the escape trajectory is a hyperbolic orbit, we must have $\varepsilon > 1$ and $v_p > v_{\infty}$.

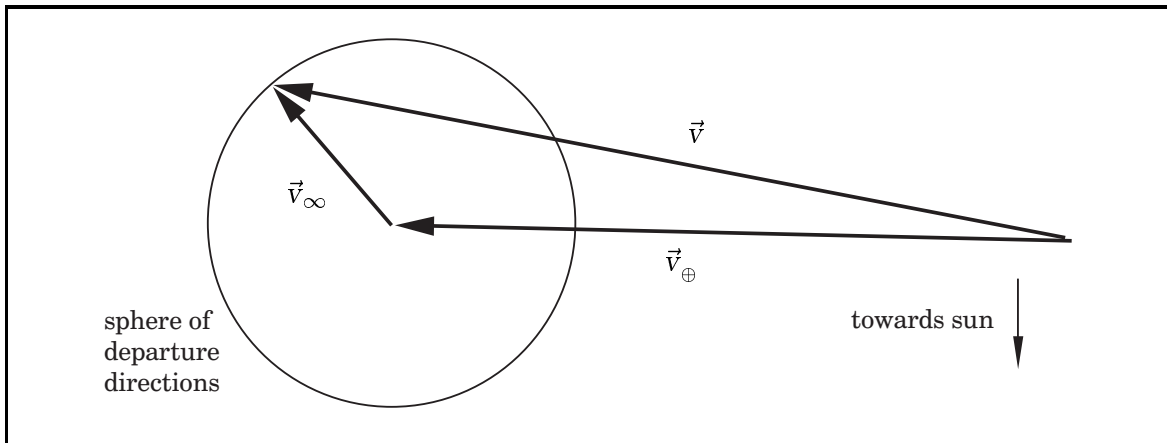
The distance from the gravitational center $r(\varphi)$ is known to be inversely proportional to the expression $1 + \varepsilon \cos \varphi$, where ε is the eccentricity and φ the path angle of the conic orbit (Kepler’s “true anomaly”). The spacecraft moves to an infinite distance ($r \rightarrow \infty$) when $\varphi = \arccos(-1/\varepsilon)$. The asymptotic escape direction φ_∞ follows as:



$$\begin{aligned} \varphi_\infty &= \arccos\left(\frac{-\gamma_\oplus}{r_p v_\infty^2 + \gamma_\oplus}\right) \\ &= \arccos\left(\frac{-\gamma_\oplus}{r_p v_p^2 - \gamma_\oplus}\right) \end{aligned} \tag{8 - 22}$$

We want that the spacecraft moves on a heliocentric trajectory towards the target planet after earth departure. Therefore it is necessary that we adjust three values appropriately: the velocity, the direction and the inclination of the escape orbit.

Earth departure seen from the heliocentric system. A “heliocentric observer” sees the velocity of the earth \vec{v}_\oplus (≈ 39.787 km/s); and he sees the heliocentric velocity of the spacecraft \vec{v} while it departs from the earth. When the spacecraft has left the sphere of earth influence, its velocity relative to the earth is \vec{v}_∞ (aligned with the asymptotic escape direction). We simply have to add the velocity vectors \vec{v}_\oplus and \vec{v}_∞ to obtain the heliocentric velocity \vec{v} of the spacecraft after earth departure.



When the spacecraft wants to leave for one of the outer planets (for example Mars) it has to increase its heliocentric velocity \vec{v} . At earth departure it is necessary that the asymptotic escape direction is more or less parallel with the heliocentric velocity of the earth (the velocity vectors \vec{v}_∞ and \vec{v}_\oplus must point in similar directions). However, when the destination is Venus or Mercury, the spacecraft must reduce its heliocentric velocity at earth departure. To visit one of the inner planets, an escape maneuver in a direction opposite to the heliocentric velocity of the earth is required.

8.2.2 Planetary Flyby

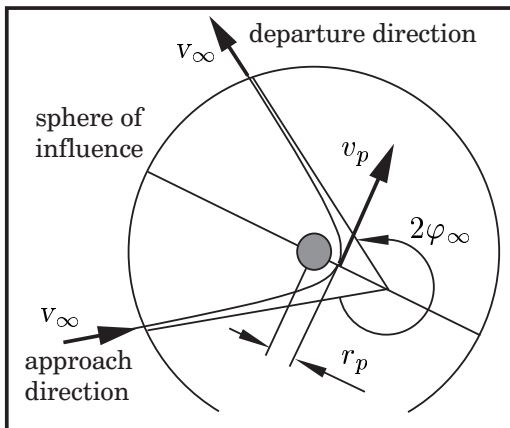
Flyby velocity. Now we consider that the spacecraft approaches the target planet on a heliocentric trajectory. Then it enters the planet's "sphere of influence" with the relative velocity v_∞ . Since inside this sphere of influence the gravitational attraction of the planet is effective, the velocity of the spacecraft relative to the planet $v(r)$ is higher than its approach velocity v_∞ . The velocity $v(r)$ is a function of the distance r , and it depends on the mass of the planet (on its gravitational constant γ_{planet}):

$$v(r) = \sqrt{v_\infty^2 + \frac{2\gamma_{planet}}{r}} \quad (8 - 23)$$

At arrival and at departure we have $r \rightarrow \infty$; thus the relative velocity of the spacecraft is exactly v_∞ on the border of the sphere of influence. A spacecraft that shall enter an orbit around the planet has to reduce its velocity at pericenter passage ($r = r_p$). To stay in the gravity field of the planet it must slow down to a velocity smaller than parabolic speed (parabolic speed is a function of r). The spacecraft enters a circular orbit when, after braking, its velocity is reduced to 70.71% of the parabolic speed:

$$v_{circular} = \sqrt{\gamma_{planet}/r_p} = v_{parabolic}/\sqrt{2} \quad (8 - 24)$$

Deflection of the heliocentric orbit. In case the spacecraft is not decelerated, it will flyby the planet and leave again the sphere of gravitational influence. Then it departs with the relative velocity v_∞ in another direction. The angle between arrival direction and departure direction is two times the path angle φ_∞ . We can denote:



$$\cos \varphi_\infty = \frac{-\gamma_{planet}}{r_p v_\infty^2 + \gamma_{planet}} \quad (8 - 25)$$

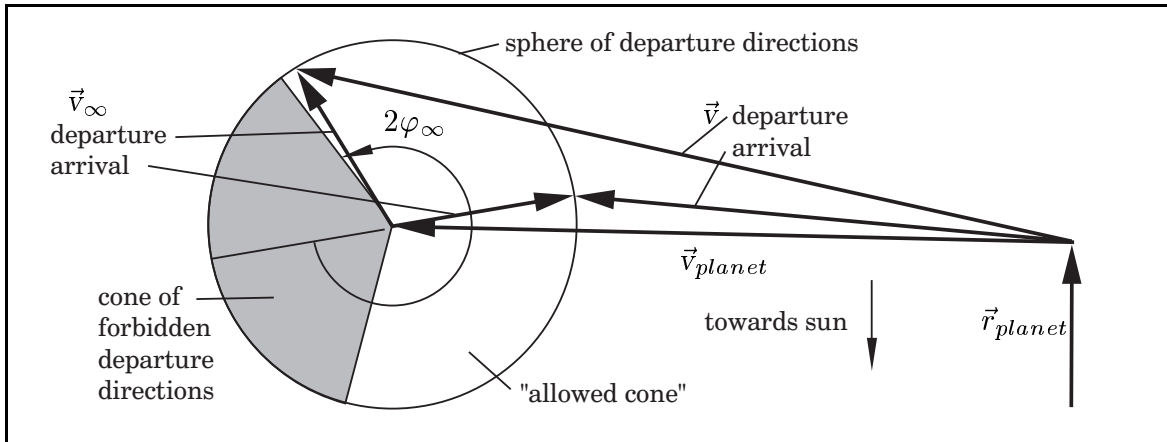
$$r_p = \frac{\gamma_{planet}}{v_\infty^2} \left(\frac{-1}{\cos \varphi_\infty} - 1 \right)$$

$$v_p = \sqrt{v_\infty^2 + 2\gamma_{planet}/r_p}$$

The flyby maneuver "deflects" the heliocentric trajectory of the spacecraft; where the efficiency of the maneuver is characterized by the deflection angle $2\varphi_\infty$. An infinite flyby distance causes no deflection ($r_p \rightarrow \infty$ and $2\varphi_\infty = 180^\circ$); an infinitely small flyby distance reverses the flight direction (impossible $r_p \rightarrow 0$ and $2\varphi_\infty = 360^\circ$). Obviously, in practice a minimum flyby distance has to be observed. The nearest flyby distance is determined by the radius of the planet and the size of its atmosphere (sometimes also by the size of a radiation belt which is dangerous for a spacecraft).

Equation (8-25) shows that the deflection angle is not just a function of the flyby radius but also a function of the flyby velocity: the faster the spacecraft passes the planet, the smaller is the deflection of its trajectory. No deflection is made at infinite speed ($v_\infty \rightarrow \infty$), the flight direction is reversed at “zero” speed ($v_\infty = 0$).

Heliocentric view of the flyby maneuver. An observer on the planet sees that the spacecraft arrives and leaves with the same hyperbolic velocity. Between approach direction and departure direction is the trajectory deflection angle $2\varphi_\infty$. The deflection of the heliocentric trajectory is limited since the maneuver has to observe the nearest flyby altitude. Above this altitude the restriction is ineffective. Any altitude is allowed, provided we are sure that we do not violate the restriction. Usually we are also free in choosing on which side the planet is passed. When we imagine a three-dimensional figure that describes the relationships between the velocity vectors, then the “sphere of departure directions” is a ball with the radius v_∞ . The flyby distance restriction cuts the ball into two pieces: all possible departure directions lie inside the “allowed cone”; departure directions outside this cone form the “forbidden cone”. The centerline (symmetry axis) of the allowed cone is parallel with the approach direction; the cone angle is determined by the maximum deflection angle $2\varphi_\infty$. Therefore the cone angle is a function of the relative velocity v_∞ .



The heliocentric velocity \vec{v} of the spacecraft is a result of a vector addition. Before and after the flyby maneuver the velocities form vector triangles: $\vec{v} = \vec{v}_{planet} + \vec{v}_\infty$. Note that the maneuver does not only change the direction of \vec{v} , but also the length.

Remember that the actual flyby velocity v_p is higher than the relative velocity v_∞ (equation 8-25). We can calculate the direction of the actual relative velocity at pericenter passage \vec{v}_p/v_p and the unit vector that locates the pericenter \vec{r}_p/r_p with respect to the planet when we know the relative velocity \vec{v}_∞ at approach and at departure:

$$\frac{\vec{v}_p}{v_p} = \frac{\vec{v}_{\infty, approach} + \vec{v}_{\infty, departure}}{|\vec{v}_{\infty, approach} + \vec{v}_{\infty, departure}|}, \quad \frac{\vec{r}_p}{r_p} = \frac{\vec{v}_{\infty, approach} - \vec{v}_{\infty, departure}}{|\vec{v}_{\infty, approach} - \vec{v}_{\infty, departure}|} \quad (8 - 26)$$

The heliocentric velocity of the spacecraft at pericenter passage is then $\vec{v}_{planet} + \vec{v}_p$; and the location of the pericenter with respect to the sun is $\vec{r}_{planet} + \vec{r}_p$.

Changing of orbital elements. Let us return to the Jacobi integral (8-14) and consider that the sun and the flyby-visited planet are a two-body system of celestial mechanics. The Jacobi integral must be a constant value for a spacecraft that coasts in this gravitational field. It indicates that the relative velocity v_∞ is the same for all locations on the surface of the small “sphere of gravitational influence” of the planet. Alternatively we can express the Jacobi integral (8-14) in the form of equation (8-17), using the specific energy e , angular momentum h , and inclination ϑ of the orbit. The angular velocity $\omega_{orbit} = \sqrt{\gamma_\odot/r^3}$ of a planet orbit is approximately constant (γ_\odot is the gravity constant of the sun; r is the distance of the planet from the sun).

Conveniently, the elements of heliocentric orbits are provided in terms of semi-major axis a and eccentricity ε (rather than using e and h), but the laws of orbital mechanics let us readily transform the equation (8-17). With $e = -\gamma_\odot/(2a)$, $h = \sqrt{a\gamma_\odot(1 - \varepsilon^2)}$ we get:

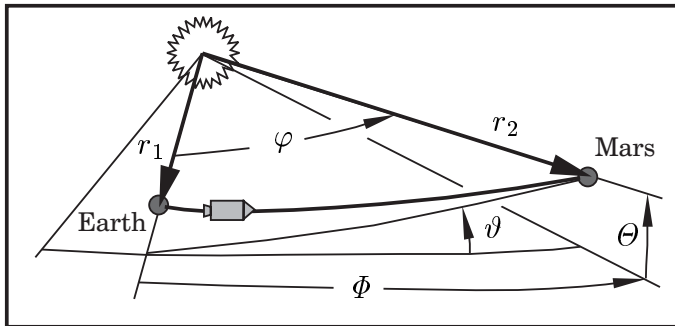
$$-\frac{2J}{\gamma_\odot} = \frac{1}{a} + \frac{2\sqrt{a(1 - \varepsilon^2)} \cos \vartheta}{\sqrt{r^3}} \quad (8 - 27)$$

The right hand side of the equation above involves the orbital elements a , ε and ϑ . The left side of the equation is constant. The relationship between a , ε , ϑ and r (the right side) must accept the same value before and after a planet flyby maneuver.

Equation (8-27) is the criterion due to Tisserand, constituted more than a century ago for the identification of lost comets. Comets can change their orbital elements when they pass a planet closely. When it is not sure whether a comet is a new one or just an old one that has visited Jupiter, Tisserand’s criterion (8-27) may be helpful.

8.2.3 Heliocentric Transfer Trajectories

Inclination of the transfer orbit. Let us have a closer look now at the junction line between departure planet and target planet (for example the trajectory of a spacecraft that coasts from earth to Mars). The orbital plane of the target planet is never exactly the same plane as the orbital plane of the departure planet. To reach planet Mars, the spacecraft has to depart from the earth with an inclination angle ϑ . When the rendezvous location is determined by the Cardan angles Φ and Θ , we have:



$$\begin{aligned} \tan \vartheta &= \tan \Theta / \sin \Phi \\ \sin \varphi &= \sin \Theta / \sin \vartheta \\ \cos \varphi &= \cos \Phi \cos \Theta \end{aligned} \quad (8 - 28)$$

$$(-90^\circ < \vartheta < +90^\circ)$$

$$(-90^\circ < \Theta < +90^\circ)$$

$$(0 < \Phi < 360^\circ)$$

The relationships are simple laws of spherical geometry (fundamental mathematics).

An important special case is that the target planet is encountered at the other side of the sun. We know that the so-called Hohmann orbit is an economical transfer trajectory; and it consists of two tangential impulsive burn maneuvers connected by a 180° coast arc. Thus, the final point of the transfer trajectory should be in “opposition” to the starting point. When the angle Φ is exactly 180° , the equation (8-28) offers just two possible geometries: either angle Θ is exactly 0° (then the inclination angle ϑ is undefined and we may choose a plane transfer orbit without inclination change), or, in the general case, the angle Θ is not 0° (then $\vartheta = \pm 90^\circ$ and the plane of the transfer orbit is perpendicular to the initial flight plane). We can conclude that in three-dimensional space a transfer orbit with a 180° coast arc is not an economical trajectory, because immense rocket propulsion would be required to change the inclination substantially. Either we select a rendezvous location where $\Phi \neq 180^\circ$ and ϑ is small; or we use a midcourse impulse to adjust the inclination.

Let us focus on the geometry of the general case, where $\Phi \neq 180^\circ$ and Θ is small. The geometry of the transfer trajectory is not uniquely determined: a second solution to equation (8-28) shows up when we add (or subtract) 180° to the inclination angle ϑ . In practice these “retro-grade” transfer orbits are of no importance. The inclination is uniquely determined when we disregard the unimportant case $\vartheta > 90^\circ$ (or $\vartheta < -90^\circ$).

Eccentricity and pericenter location. Our next problem is to put the appropriate section of a conic orbit between starting point and rendezvous location. Inclination angle ϑ and path angle φ are determined by equation (8-28). Now we have to find the geometry of a conic orbit that connects the initial point at distance r_1 with the final point at distance r_2 ; the angle between radial r_1 and radial r_2 is the path angle φ . The distance r from the gravitational center (the sun) is a function of the path angle:

$$r_1 = \frac{(h^2/\gamma_\odot)}{1 + \varepsilon \cos(-\varphi_p)} ; \quad r_2 = \frac{(h^2/\gamma_\odot)}{1 + \varepsilon \cos(\varphi - \varphi_p)} \quad (8 - 29)$$

The two relationships above involve three unknown elements of the transfer orbit: the specific angular momentum h , the eccentricity ε , and the path angle between initial point and perihelion φ_p (the solar gravitational constant γ_\odot is well-known). We can eliminate h from the equations and find a relationship between φ_p and ε :

$$r_1(1 + \varepsilon \cos(-\varphi_p)) = r_2(1 + \varepsilon \cos(\varphi - \varphi_p))$$

$$\text{or transformed:} \quad r_1 - r_2 = \varepsilon \cdot (\sin \varphi_p (r_2 \sin \varphi) - \cos \varphi_p (r_1 - r_2 \cos \varphi)) \quad (8 - 30)$$

The equation (8-30) is a condition for the eccentricity ε and the perihelion angle φ_p (the terms r_1 , r_2 and φ are predefined). The eccentricity ε determines the type of transfer orbit: it can be a circle ($\varepsilon = 0$), an ellipse ($0 < \varepsilon < 1$), a parabola ($\varepsilon = 1$) or a hyperbola ($\varepsilon > 1$). Obviously, it is easy to calculate ε when φ_p is given; but it is more complicated to calculate φ_p for a given value ε . When we analyze the equation (8-30) carefully we find out that ε must be positive and not smaller than a certain limit, and that not all values are allowed for φ_p . Furthermore, the equation (8-30) has just one solution for ε when φ_p is given; but two solutions for φ_p when ε is given.

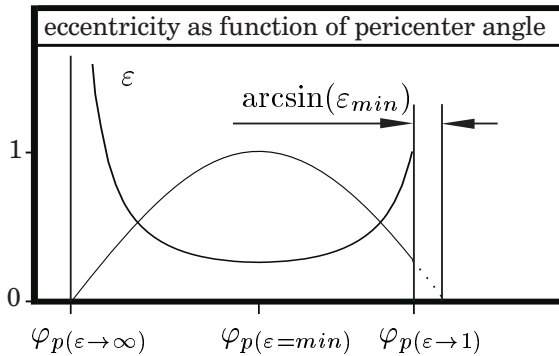
A geometric analysis will make the important relationship (8-30) more transparent. We introduce the distance ρ (according to the theorem of the cosine it is the geometrical distance between initial location and final location on the transfer orbit) and the term $\varphi_{p(\varepsilon \rightarrow \infty)}$ (the perihelion angle of the fastest transfer orbit). Therefore:

$$\rho = \sqrt{r_1^2 + r_2^2 - 2r_1r_2 \cos \varphi} \quad (8-31)$$

$$\sin \varphi_{p(\varepsilon \rightarrow \infty)} = (r_1 - r_2 \cos \varphi) / \rho \quad (8-32)$$

$$\cos \varphi_{p(\varepsilon \rightarrow \infty)} = (r_2 \sin \varphi) / \rho$$

Using these new terms, we can transform the equation (8-30) and rewrite it as:

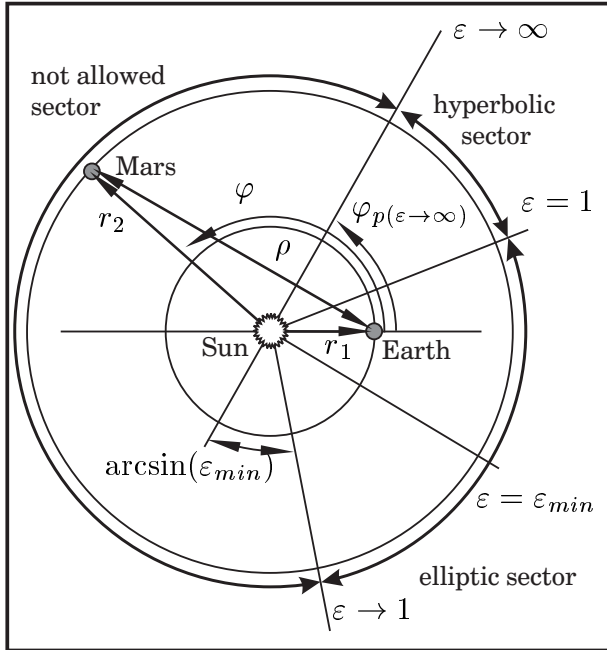


$$\varepsilon \cdot \sin(\varphi_p - \varphi_{p(\varepsilon \rightarrow \infty)}) = \frac{r_2 - r_1}{\rho}$$

$$\varepsilon_{min} = \frac{|r_2 - r_1|}{\rho} \quad (8-33)$$

Let us first consider the case that the transfer angle φ is smaller than 180° . When in equation (8-33) $\varphi_p = \varphi_{p(\varepsilon \rightarrow \infty)}$, the argument of the sinus function vanishes and the eccentricity of the transfer orbit must accept an infinite value ($\varepsilon \rightarrow \infty$). The transfer orbit is nothing else than the straight line (distance ρ) that connects the initial point with the final point, the fastest trajectory between these two points. The transfer orbit is a “degenerated” hyperbola with an infinitely short transfer time; and the radial which belongs to its “perihelion” is rectangular to the straight line ρ . The equation (8-33) shows that the eccentricity ε accepts its minimum value when $|\varphi_p - \varphi_{p(\varepsilon \rightarrow \infty)}| = 90^\circ$. This means that the perihelion radial of the transfer orbit with smallest eccentricity $\varphi_{p(\varepsilon=min)}$ is rectangular to the perihelion radial of the orbit with the fastest transfer time. To find the radial that belongs to ε_{min} , we have to rotate the radial $\varphi_{p(\varepsilon \rightarrow \infty)}$ exactly 90° . We have to rotate this radial clockwise if r_1 is smaller than r_2 ; we have to rotate it counter-clockwise if r_1 is bigger than r_2 . When we continue the rotation in the same direction we come to the radial that belongs to the perihelion of the transfer orbit with longest transfer time (clockwise for $r_1 < r_2$, otherwise counter-clockwise). While we rotate, the expression $\sin(\varphi_p - \varphi_{p(\varepsilon \rightarrow \infty)})$ becomes again smaller and the eccentricity ε grows. Finally the eccentricity arrives at the value $\varepsilon = 1$. Here is the perihelion radial of the orbit with an infinitely long transfer time. It is a parabola, a “degenerated” ellipse where the aphelion is on the way between initial point and final point of the transfer orbit.

When we rotate the perihelion radial from the fastest to the slowest transfer orbit it sweeps out a segment which is smaller than 180° . Other values for φ_p are not allowed: either the equation (8-30) would give a negative solution to the eccentricity ε ; or $\varepsilon > 0$, and the spacecraft would coast to infinity and not return any more.



Thus, all “allowed” locations for the perihelion lie in a sector smaller than 180° . The sector starts at the radial $\varphi_{p(\epsilon \rightarrow \infty)}$ (perihelion location of the fastest transfer orbit), it continues clockwise for $r_1 < r_2$ and counter-clockwise for $r_2 < r_1$. At an angular distance of 90° we find the radial of the transfer orbit with minimum eccentricity. Rotating further we find the parabolic orbit with the longest transfer time. The “allowed” sector is divided into two sub-sector: perihelion locations of hyperbolic orbits (sector angle: $\arcsin(\epsilon_{min})$) and perihelion locations of elliptic orbits (sector angle: $180^\circ - 2 \arcsin(\epsilon_{min})$).

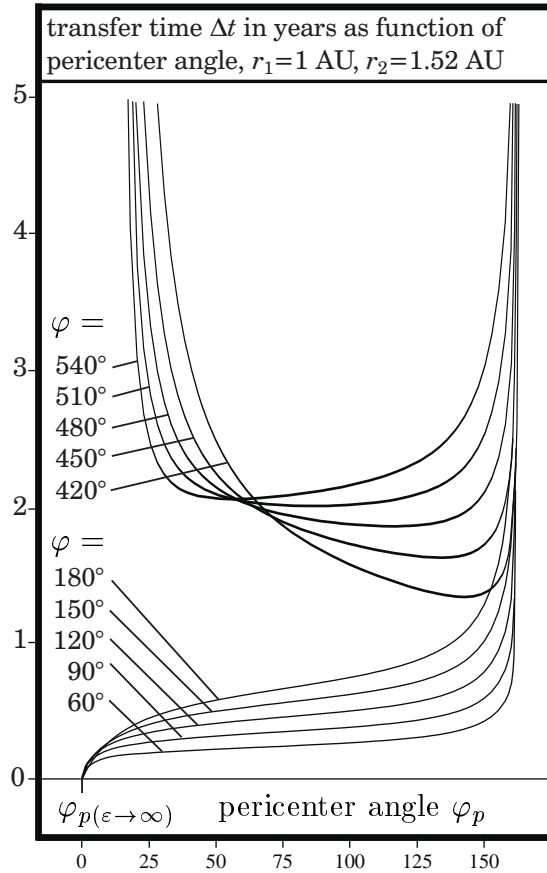
Let us now consider that the transfer angle φ is greater than 180° but still smaller than 360° . The situation is similar, but now the fastest transfer orbit is a degenerated hyperbola which consists of two straight lines (we excluded retrograde orbits from the consideration): the first line connects the initial point with the center of the sun; the second line connects the center of the sun with the final point. The orbit has an infinitely small perihelion radius ($r_p \rightarrow 0, \epsilon \rightarrow -1/\cos(\varphi/2)$) and an infinitely short transfer time; its perihelion radial is located inside the hyperbolic sector at $\varphi_p = \varphi/2$.

When we consider the case that the angle φ is greater than 360° (several revolutions), we have to exclude hyperbolic and parabolic transfer orbits. Only elliptic (or circular) orbits are allowed now, and infinitely short transfer times do not exist any more. All allowed perihelion radial lie inside the elliptic sector. Obviously, the orbit with lowest eccentricity ϵ_{min} has a comparatively short transfer time. We find perihelion radial with higher values of ϵ when we rotate the radial $\varphi_{p(\epsilon=min)}$ either clockwise or counter-clockwise. Anyway, the allowed sector stops when ϵ reaches the value 1. The parabolic transfer orbit has an infinitely long transfer time when at least one “aphelion passage” lies on the way from the initial to the final point.

Angular momentum and energy. The actual problem is to select from the infinite set of transfer orbits between initial and final point exactly the one which satisfies a certain predetermined flight time. When the spacecraft arrives at the rendezvous location it must encounter the target planet. The transfer orbit is entirely determined when we know $\epsilon, \varphi_p,$ and r_1 : we can use equation (8-29) to calculate the specific angular momentum h of the orbit, and then its specific mechanical energy e :

$$h = \sqrt{\gamma_\odot r_1 (1 + \epsilon \cos \varphi_p)} ; \quad e = 0.5 (\epsilon^2 - 1) \cdot (\gamma_\odot / h)^2 \quad (8 - 34)$$

The energy e determines directly semimajor axis a and orbital period T of the orbit.



Transfer time. To calculate the flight time we can use either Kepler's equation or Lambert's theorem, compare chapter four of this book. These transcendental equations allow a direct computation of the flight time between two points on a conic orbit. Our problem is to find the section of a conic orbit that satisfies a predetermined flight time. Therefore we have to apply a numerical iteration process, unfortunately, because we have just a relation between the eccentricity ε and the perihelion angle φ_p (equation 8-30). The problem has just a unique solution when the spacecraft proceeds directly to the target planet, without one or several orbital revolutions before the rendezvous. The diagram on the left shows the flight time Δt for Earth-Mars transfer orbits as a function of the perihelion angle φ_p , with the transfer angle φ as parameter. Longer flight times involve the option of an orbital revolution before the encounter.

Synodic period. The time interval between now and the next constellation where a planet is at the same angle in relation to another planet is called "synodic period". In astronomy it is the time interval that two planets need to move from one opposition (or conjunction) to the next one. In astronautics it is the time between two similar launch windows for the departure for a planet. Planets move on nearly circular orbits; their heliocentric path angle φ is approximately a linear function of time. The planet which is nearer to the sun moves faster, and consequently its orbital period T_1 is shorter than the orbital period T_2 of the more distant planet. After the time interval $T_{synodic}$ the slower orbiting planet arrives at a position with a certain path angle φ ; and the faster orbiting planet arrives at a position with the path angle $\varphi + 2\pi$. Thus:

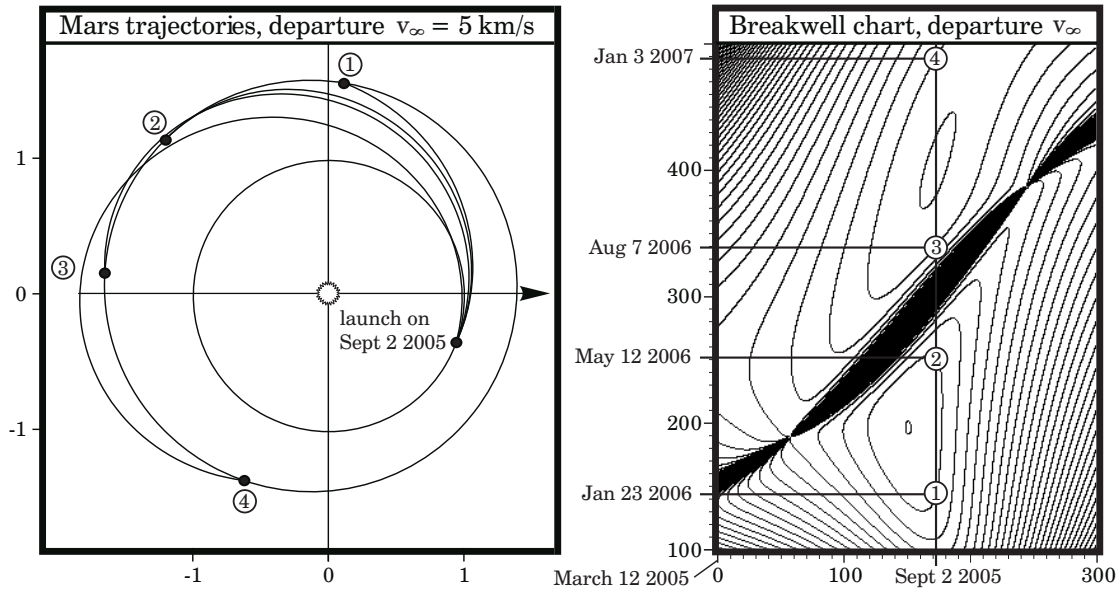
$$2\pi T_{synodic} = (\varphi + 2\pi) T_1 = \varphi T_2 \quad (8-35)$$

We can eliminate φ from the relationship above, and the synodic period follows as:

$$T_{synodic} = \frac{T_1 \cdot T_2}{(T_2 - T_1)} \quad (8-36)$$

For transfers from earth ($T = 1$ year) to Mars ($T = 1.881$ years) we can compute a synodic period of 2.135 years (780 days). For transfers from earth to Venus ($T = 0.615$ years) the synodic period is considerably shorter, $T_{synodic} = 1.599$ years or 584 days. For departure to a distant planet, the synodic period is slightly longer than one year.

The Breakwell diagram. We have seen that the travel time Δt is a function of the path angle φ , which is a function of the Cardan angle Φ that determines the final point of the transfer trajectory with respect to the initial point (compare equations 8-28). To depart from the initial planet the spacecraft must enter a hyperbolic escape orbit: after having left the planet’s “sphere of influence”, the spacecraft moves with the relative velocity v_∞ on the transfer trajectory that will bring it to the target planet. This departure velocity v_∞ is a function of the departure date and the travel time. Often it is intended to find a transfer trajectory that combines a low velocity requirement with a short flight time. Therefore the so-called “Breakwell chart” is helpful:



The Breakwell chart is a map where lines of constant velocity requirement v_∞ are drawn as a function of the departure date (abscisse) and the flight time (ordinate). The result is a figure that looks like an onion cut open into two pieces. The innermost curves determine trajectory opportunities with a comparatively low velocity v_∞ ; the velocity requirement v_∞ grows when trajectories are selected on outer curves. The example shows a diagram for departures from Earth to Mars in the year 2005, considering transfer times between 100 and 500 days (contours of $v_\infty = 4000 \text{ m/s}$, 5000 m/s , 6000 m/s and so on). A spacecraft that wants to travel to Mars with the minimum relative departure velocity has to leave the earth on September 2nd with a relative velocity $v_\infty = 3930 \text{ m/s}$, it will encounter Mars 400 days later, on October 7th 2006. When the spacecraft is able to depart on the same day with a higher relative velocity, for example $v_\infty = 5000 \text{ m/s}$, it can use one of four different transfer trajectories (trajectories 1 to 4). The velocity requirement v_∞ is lower than 5000 m/s for transfer times between trajectory 1 and trajectory 2, and it is also lower than 5000 m/s for transfer times between trajectory 3 and trajectory 4. However, the velocity requirement is considerably higher for transfer times between trajectory 2 and trajectory 3: to encounter the target planet these trajectories use extreme inclination angles ϑ . If we selected a transfer time of 300 days, for example, we would have to adjust the inclination of the trajectory by a midcourse burn maneuver.

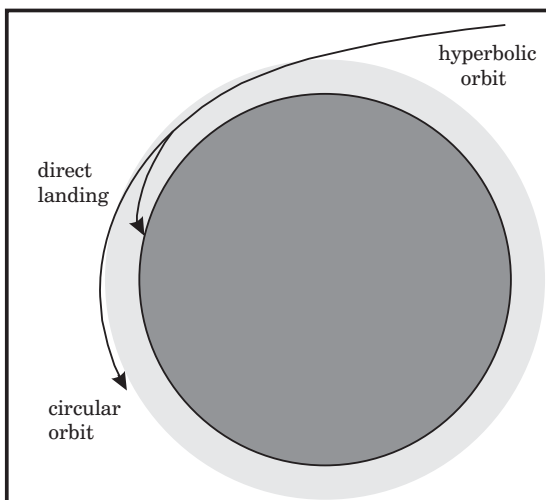
Planetary capture. Finally, at the end of the interplanetary transfer trajectory, the spacecraft arrives at the target planet. Seen from the planet, the spacecraft approaches the planet on a hyperbolic orbit, however, it is still far away from the planet. Then the spacecraft enters the planet's "sphere of influence". Of course, at the border of the sphere of influence the relative speed is different from the relative speed at departure. However, when we draw a Breakwell chart for the relative velocity at arrival, the diagram looks similar to the Breakwell chart for the velocity at departure. Obviously, when the spacecraft departs from a planet with a comparatively low velocity, it arrives at the target planet also with a comparatively low relative velocity.

Inside the "sphere of gravitational influence" the spacecraft moves on a hyperbolic orbit around the planet. The spacecraft speeds up as long as it approaches the pericenter. At the pericenter radius r_p the spacecraft has the horizontal velocity v_p :

$$v_p = \sqrt{v_\infty^2 + \frac{2\gamma_{planet}}{r_p}} \quad (8 - 37)$$

The pericenter velocity v_p depends on the arrival speed v_∞ , on the gravitational constant of the planet γ_{planet} , and on the pericenter radius r_p . The spacecraft has to reduce its speed to establish a periodic orbit around the planet: it must reach a velocity lower than parabolic speed ($v_{parabolic} = \sqrt{2\gamma_{planet}/r_p}$, $v_{circular} = \sqrt{\gamma_{planet}/r_p}$). We can see that the velocity requirement for this maneuver is a nonlinear function of the approach velocity v_∞ . When we try to find the optimal planetary transfer trajectory, we have to consider that the price for the arrival maneuver (in terms of propellant consumption) is quite different from the price of the departure maneuver. This is also the case when the spacecraft uses the same motor for both maneuvers.

Aerocapture. When the target planet has an atmosphere (like Mars or Venus) the spacecraft has the opportunity to execute an aerodynamic braking maneuver. The objective can be to land directly on the planet (with parachutes), or

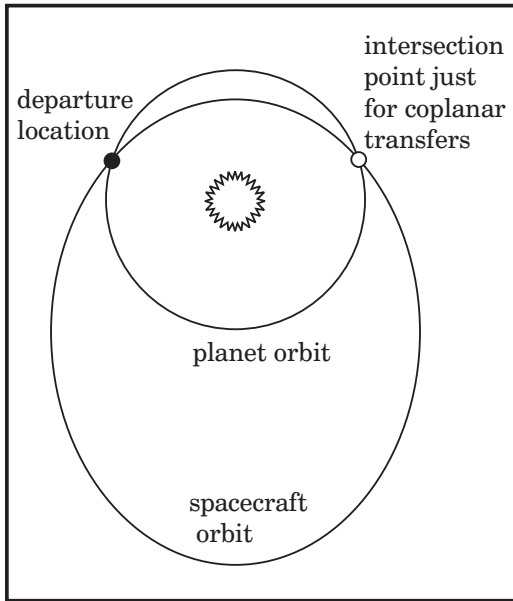


to establish a circular orbit around the planet (with a small impulse after the aerodynamic maneuver). In both cases the essential braking requires heat protection but no propellant. Since the heat shield weighs just a few percent of the mass of the spacecraft, the aerocapture maneuver is much more economical than braking with conventional rocket propulsion (even when a spacecraft is not equipped with a big heat shield, it can utilize the upper layers of the atmosphere for many smaller braking maneuvers, in order to transform a parabolic orbit into a low parking orbit).

8.2.4 Return Orbits

Coplanar return. Destination may be the same planet where the travel starts from. When departure planet and target planet are identical, we talk about “heliocentric return trajectories”. For example, a comet flyby mission can be launched on an “earth return trajectory” (ESA’s project Giotto). Some years later, after having performed the flyby maneuver, the spacecraft will return to the earth automatically.

First, let us consider coplanar return trajectories: these coplanar return trajectories remain all the time in the orbital plane of the planet. The spacecraft departs from the planet (relative velocity \vec{v}_∞) and establishes a heliocentric orbit. This orbit is in the same plane as the heliocentric orbit of the planet; therefore it will intersect the planet’s orbit exactly at two points: at departure location, and at another location (just when the departure velocity \vec{v}_∞ is exactly parallel with the heliocentric velocity of the planet, the two orbits touch each other without intersection.)



Planet and spacecraft run through these two intersection points every orbital revolution. After some years the spacecraft encounters the planet again (planet and spacecraft may have executed several orbital revolutions). Therefore it is necessary that spacecraft and planet run through the intersection point exactly at the same time; a condition which is easily satisfied when the velocity \vec{v}_∞ fits. The spacecraft approaches the planet with the relative velocity \vec{v}_∞ when the arrival takes place at the departure location. At the other location, the spacecraft approaches the planet from another relative direction in space (but also with relative velocity v_∞ , assuming that the planet’s orbit is a circle).

Return to the same point in space. Now let us consider return orbits which are not in the same plane as the orbit of the planet. The spacecraft performs an inclination change maneuver when it departs with a velocity vector \vec{v}_∞ that is inclined with respect to the plane of the planet orbit. It has still the opportunity to encounter the planet on its interplanetary return orbit. However, now there is just one intersection point: the meeting must be exactly at the same location where the departure maneuver took place. After several revolutions (number n), the orbital period T of the return orbit has to satisfy the following condition:

$$T = \frac{n_{planet}}{n} \cdot T_{planet} \tag{8 - 38}$$

Term n_{planet} is the number of revolutions of the planet and T_{planet} its orbital period. The spacecraft encounters the planet exactly with the relative departure velocity \vec{v}_∞ .

When we know the period T we can calculate the semimajor axis a of the orbit:

$$a = \gamma_{\odot} \cdot \sqrt[3]{\left(\frac{T}{2\pi}\right)^2} \quad (8-39)$$

The orbit has the geometry of an ellipse. Remember that the semimajor axis a depends only on the mechanical energy e of the orbit (as well as the period T). We have:

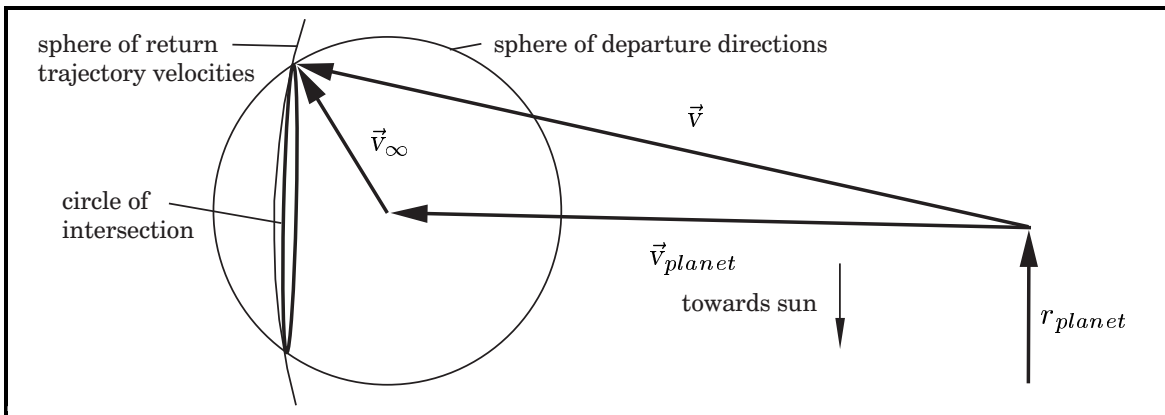
$$e = \frac{|\vec{v}|^2}{2} - \frac{\gamma_{\odot}}{r_{planet}} = -\frac{\gamma_{\odot}}{2a}$$

At the moment the spacecraft departs from the planet, it has the same distance from the center of the sun as the planet (term r_{planet} describes this distance). Since the expression $\sqrt{\gamma_{\odot}/r_{planet}}$ determines the circular velocity at the distance r_{planet} , the vector \vec{v} of the heliocentric velocity at departure must satisfy the following condition:

$$|\vec{v}| = \sqrt{2 - \frac{r_{planet}}{a}} \cdot \sqrt{\frac{\gamma_{\odot}}{r_{planet}}} \quad (8-40)$$

Term \vec{v} is the heliocentric departure velocity as a function of the semimajor axis a of the return orbit. To encounter the planet in a predefined time interval $n \cdot T$, it is just important that the lengths of \vec{v} fits. Surprisingly, the direction of \vec{v} does not matter.

Resonant return. We can follow that return trajectories require a certain heliocentric velocity \vec{v} at the moment of the departure, but it is not necessary that this vector has a predetermined direction. The velocity of the spacecraft \vec{v} must form a triangle with the velocity of the planet \vec{v}_{planet} and the departure velocity \vec{v}_{∞} . Thus, when the length of v_{∞} is given, the tip of v_{∞} ends on a circle which is the intersection line of two spheres: the vector \vec{v}_{planet} points from the center of the “sphere of return trajectory velocities” to the center of the “sphere of departure directions”.



After some time the spacecraft returns to the planet. When it arrives, all vectors of the velocity triangle have the same length and the same direction as they had at departure. A gravity swingby maneuver at the planet can change now the direction of the velocity \vec{v}_{∞} . When the spacecraft changes \vec{v}_{∞} to another direction conserving the length \vec{v} , then the gravity swingby maneuver does not change the orbital period T of the return orbit: the spacecraft enters a so-called “resonant return trajectory”.

8.3. Gravity Propelled Missions

The gravitation of a planet can change the orbital elements of a spacecraft trajectory. When the interplanetary trajectory is carefully directed on a certain planet, a gravity swingby maneuver at this planet can amplify the heliocentric energy of the orbit. In this way it is possible to use the gravitational force of the planets as a “propulsion system” for deep space mission. The “gravity assist transportation system” works in a fantastic manner: it does not consume propellant, it will always be ready, and it will never fail to work. With gravity assist maneuvers it is possible to reach any location in the solar system at the price of a Mars or Venus flyby (including the sun and including trajectories to escape from the solar system). However, launch windows do not show up every time, and the flight time to the destination can be considerably long. In this section we will examine the great possibilities of gravitational maneuvering.

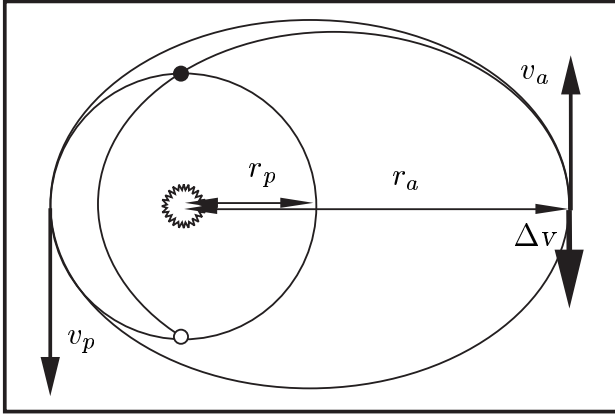
8.3.1 Gravity Assist on Earth-Return Trajectories

Gravity swingby at the earth. Actually, the most important planet for gravitational maneuvering is our own planet. The earth moves with the velocity $v_{\oplus} \approx 29.787$ km/s on a nearly circular orbit around the sun. Consider a spacecraft that is launched on a two-year earth-return trajectory ($T = 2$ years). The semimajor axis a of the return orbit becomes $a = \sqrt[3]{T^2} = 1.5874$ AU (Kepler’s law, equation 8-39). At earth departure we have to increase the heliocentric velocity of the spacecraft, $v/v_{\oplus} = \sqrt{2 - 1/a} = 1.1705$ (equation 8-40). Thus, the minimum (tangential) departure velocity for a two-year earth-return trajectory amounts $v_{\infty} = 5.078$ km/s (the return orbit remains in the ecliptical plane). When the spacecraft encounters the earth exactly two years later, a gravity swingby maneuver cannot increase its heliocentric velocity: the relative velocity at arrival is exactly $v_{\infty} = 5.078$ km/s.

The situation changes when we apply a small “retro-impulse” at aphelion position. The braking deforms the earth-return trajectory a little: now it intersects the orbit of the earth at two locations. We must adjust the departure velocity v_{∞} slightly to encounter the earth at one of these two locations (the flight time on the new earth return trajectory is not exactly two years anymore). The retro-impulse reduces the heliocentric energy of the orbit; and consequently the heliocentric velocity of the spacecraft is lower at arrival than at departure. However, the relative velocity v_{∞} at arrival has grown; and this velocity can be utilized by a gravity swingby maneuver.

We can act in the same way when we take a three-year return trajectory as a basis. The injection velocity is higher now, $v_{\infty} = 6.928$ km/s. However, when the spacecraft departs with just $v_{\infty} = 3.340$ km/s, the orbital period is 1.5 years, and the spacecraft will also encounter the earth after three years. Any departure velocity v_{∞} that is smaller than the solar system’s escape velocity will bring the spacecraft back to earth one day (but coasting on the return trajectory can take a very long time interval). For practical reasons we are advised to select a time interval which is not too long.

Velocity reduction at aphelion position. Let us now regard the effect of a retro impulse at aphelion position a little more in detail. After the interplanetary injection the spacecraft moves at the pericenter of its heliocentric orbit (radius $r_p = 1$ AU). We find the velocity of the spacecraft v_p when we add the relative velocity v_∞ to the heliocentric velocity of the earth v_\oplus . Half an orbital period later the spacecraft reaches the aphelion position (it can also perform several orbital revolutions before it arrives there). At apocenter position (solar distance r_a) its velocity amounts v_a :



$$v_p = v_\infty + v_\oplus$$

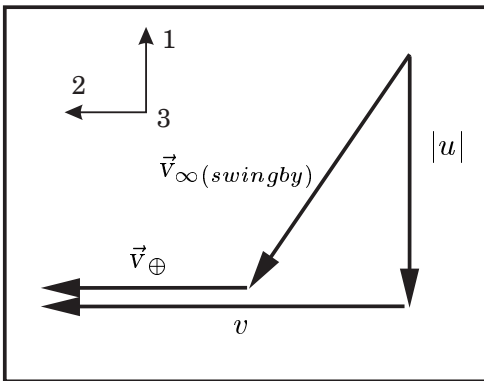
$$v_a = \frac{v_\oplus^2 - 2v_\oplus v_\infty - v_\infty^2}{v_\infty + v_\oplus} \tag{8 - 41}$$

When the spacecraft arrives at aphelion position, a retro-thrust Δv reduces the energy $e = (u^2 + v^2)/2 - \gamma_\odot/r$ and the angular momentum $h = rv$ of the orbit:

$$\Delta e = v_a \Delta v - \Delta v^2/2$$

$$\Delta h = r_a \Delta v \tag{8 - 42}$$

The new energy e and the new angular momentum h are conserved while the spacecraft coasts to the encounter position. The encounter takes place at one of the two points where the return orbit intersects the orbit of the earth, either before \bullet or after \circ perihelion passage. Now, at a distance r_p from the sun, the horizontal component of the heliocentric velocity is v . The vertical component u is negative when the earth is encountered before the perihelion passage \bullet , or u is positive on the other location \circ . We can use Δe and Δh to compute the velocity triangle at the encounter location:



$$u^2 + v^2 = v_p^2 - 2\Delta e = v_p^2 - 2v_a \Delta v + \Delta v^2$$

$$v = v_p - \Delta h/r_p = v_p - \Delta v \cdot v_p/v_a \tag{8 - 43}$$

The relative velocity at encounter follows as:

$$\vec{v}_{\infty(swingby)} = \begin{pmatrix} u \\ v - v_\oplus \\ 0 \end{pmatrix} \tag{8 - 44}$$

We get $\vec{v}_{\infty(swingby)}$ when we subtract the velocity \vec{v}_\oplus from spacecraft velocity $(u, v, 0)$.

Swingby at earth encounter compared with direct injection. Since we applied a small retro-thrust Δv at aphelion position, the relative velocity at arrival $v_{\infty(swingby)}$ is higher than the departure speed v_{∞} . We can now execute a gravitational maneuver at the earth to take advantage of the relative velocity $v_{\infty(swingby)}$:

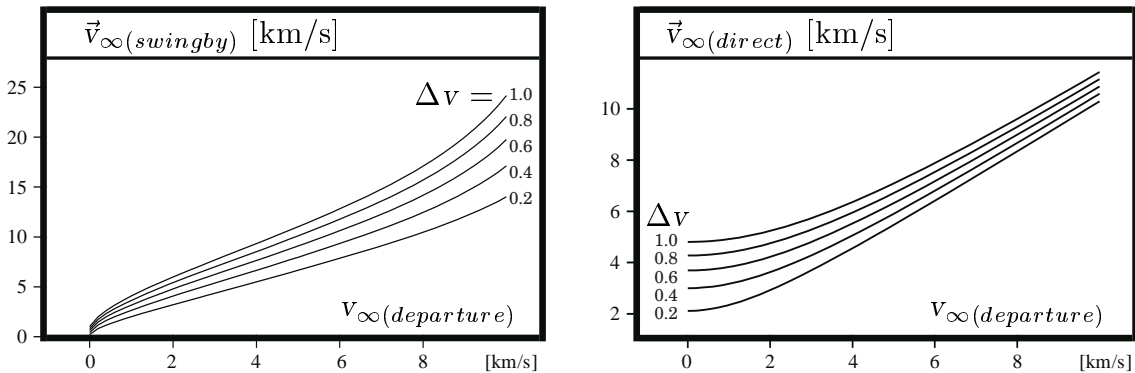
$$v_{\infty(swingby)} = \sqrt{v_{\infty}^2 + \Delta v^2 + 2\Delta v(v_{\oplus} \cdot v_p/v_a - v_a)} \quad (8 - 45)$$

$$v_p = v_{\infty} + v_{\oplus} ; v_a = (2v_{\oplus}^2 - v_p^2)/v_p$$

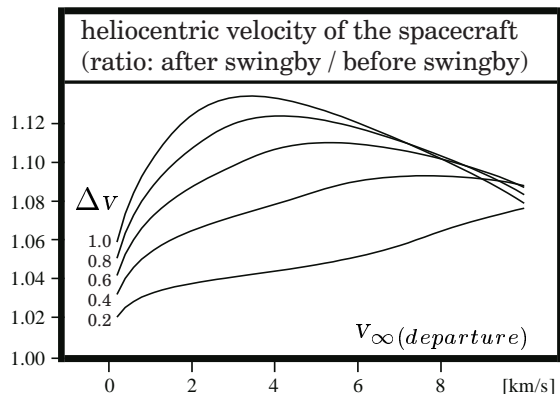
The benefit of the maneuver looks promising, however, the method requires Δv capacity and adds flight time to the planetary mission. The alternative is to spend this Δv capacity directly at departure. When we spend the Δv capacity immediately after the injection burn maneuver, the departure velocity v_{∞} grows to $v_{\infty(direct)}$; and, since the spacecraft is still near the gravitational center of the earth, the improvement of v_{∞} is much higher than the amount of Δv . Furthermore, direct injection has the advantage of a shorter mission duration. The velocity $v_{\infty(direct)}$ is a function of the altitude where the impulse Δv is applied (the lower the better). For example, the parabolic escape velocity $v_{parabolic}$ amounts 11 km/s at 200 km flight altitude.

$$v_{\infty(direct)} = \sqrt{v_{\infty}^2 + \Delta v^2 + 2\Delta v \sqrt{v_{parabolic}^2 + v_{\infty}^2}} \quad (8 - 46)$$

The figures below compare $v_{\infty(swingby)}$ and $v_{\infty(direct)}$ for several values of Δv :



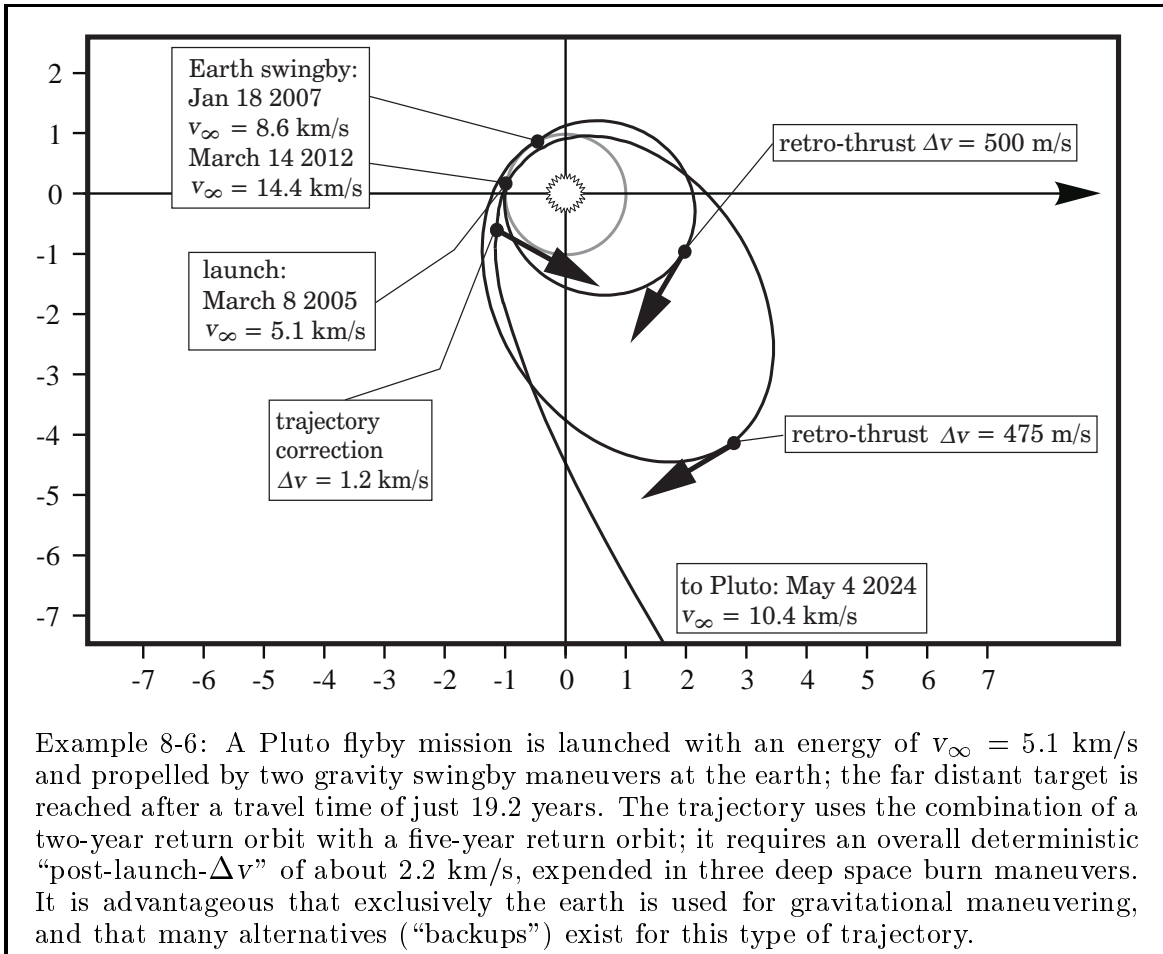
Effect of the flyby distance restriction. The swingby maneuver turns the relative velocity $v_{\infty(swingby)}$ of the spacecraft into a direction which is more or less parallel with the heliocentric velocity of the earth. However, the spacecraft may not pass the earth closer than 300 km above the surface: the flyby distance restriction limits the deflection angle. Particularly when the flyby velocity is high, it will not be possible to turn $v_{\infty(swingby)}$ completely into the parallel direction. The benefit of the maneuver is limited; and the heliocentric velocity increases just about 3.5 km/s, typically.



The following table compares five different earth-return trajectories ($v_{\infty(\text{swingby})}$) with the corresponding maneuvers that apply Δv directly at the injection ($v_{\infty(\text{direct})}$):

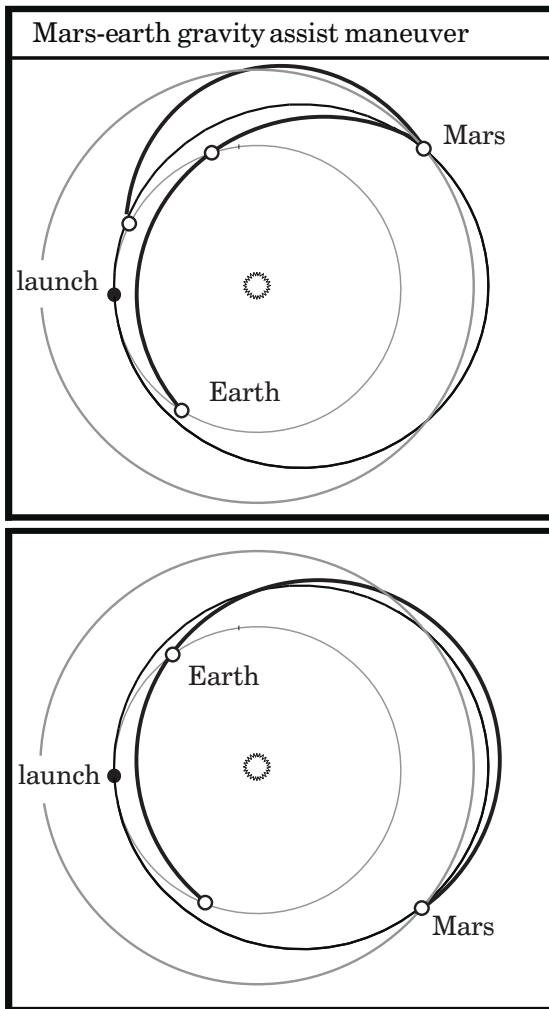
flight time [years]	≈ 5	≈ 4	≈ 3	≈ 3	≈ 2
revolutions	4	3	2	1	1
$v_{\infty(\text{departure})}$ [km/s]	2.3	2.7	3.5	7.0	5.1
Δv [m/s]	238	193	600	500	900
$v_{\infty(\text{direct})}$ [km/s]	3.3	3.4	5.1	7.8	6.9
$v_{\infty(\text{swingby})}$ [km/s]	3.6	3.6	7.0	11.0	11.0

Re-targeting the earth for a gravity swingby maneuver extends the mission time and involves a deep space burn maneuver. A retro-thrust maneuver at aphelion position lowers the heliocentric energy of the return trajectory, but a gravity swingby maneuver at the earth can take advantage of the increased relative flyby velocity. The table above considers realistic values for the midcourse maneuver Δv : it makes no sense to spend more capacity at aphelion position, because the swingby maneuver at the earth has to observe the flyby distance restriction (300 km above the surface). Swingby maneuvers on earth-return trajectories make just sense when the departure velocity $v_{\infty(\text{departure})}$ is higher than 3 km/s; otherwise the benefit is too small, and it is better to spend the capacity of the midcourse maneuver already at the departure.



8.3.2 Mars Swingby on Earth-Return Trajectories

Deflection of the earth-return trajectory. We have seen that it is possible to re-target the earth for a gravity swingby maneuver, however, it includes the disadvantage that the return orbit needs an expensive deep-space burn maneuver. This maneuver is not supposed to put energy into the earth-return trajectory; it must just deform the trajectory: the heliocentric energy of the trajectory is increased by the earth swingby maneuver that follows. We can avoid the deep-space burn maneuver when we let planet Mars do the job. A near Mars flyby can also deform the return orbit. Again, it is not required that the gravity of Mars puts energy into the trajectory: we just need a deflection of the earth-return orbit. The gravity of Mars can do this.

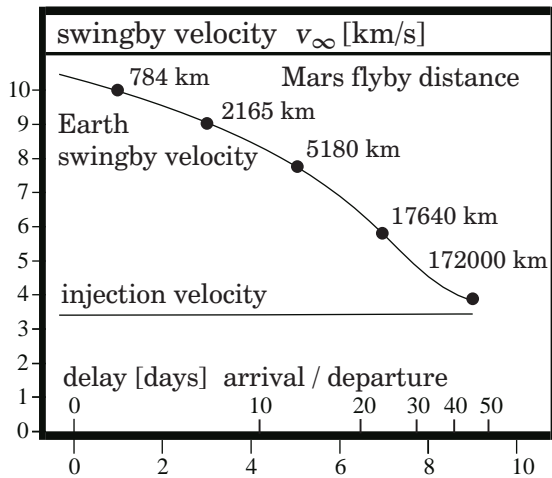
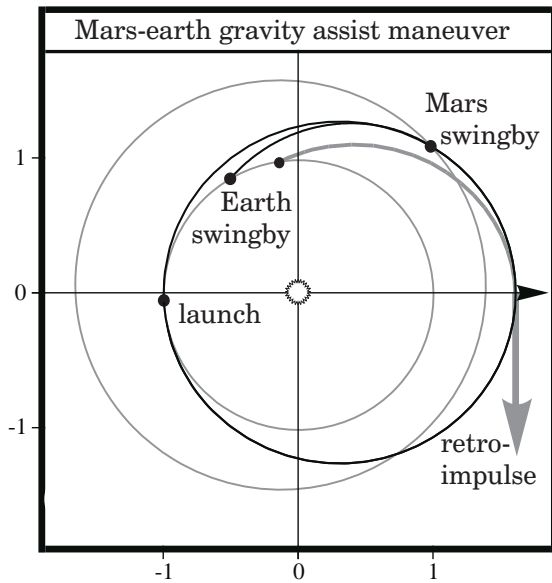


It is obviously desirable to inject with a low launch energy, but we have to insert the spacecraft on the earth-return trajectory which intersects the orbit of Mars ($r_a > r_{Mars} \approx 1.52$ AU). A short mission duration is also desirable; thus, trajectories with 2 revolutions in 3 years ($v_\infty \approx 3.340$ km/s) or with 3 revolutions in 5 years ($v_\infty \approx 4.026$ km/s) are worth to be considered; maybe also trajectories with 4 revolution in 7 years ($v_\infty \approx 4.324$ km/s). Mars can be passed at any location where its orbit intersects the orbit of the spacecraft. It is necessary that the flyby maneuver lowers the perihelion of the earth return trajectory. Sometimes the spacecraft has to pass Mars on the “night side”, sometimes on the “day side”. The earth can be encountered either before or after perihelion passage.

Thus, we have several trajectory options; but it is required that Mars is in an appropriate position; and this is often not the case. It is not always possible to find a launch date for an earth return orbit with Mars in a useful position.

This is the essential disadvantage in comparison with the method of retro-thrusting (which can be applied practically any time); and it makes a difference where Mars is encountered on its orbit around the sun: its orbit is eccentric ($\varepsilon = 0.0934$) and inclined against the ecliptic ($\vartheta = 1.850^\circ$). Not all trajectory options are really good. However, when Mars is passed closely (for example 300 km above its surface), then the spacecraft can return to the earth with a relative velocity of about 10 km/s.

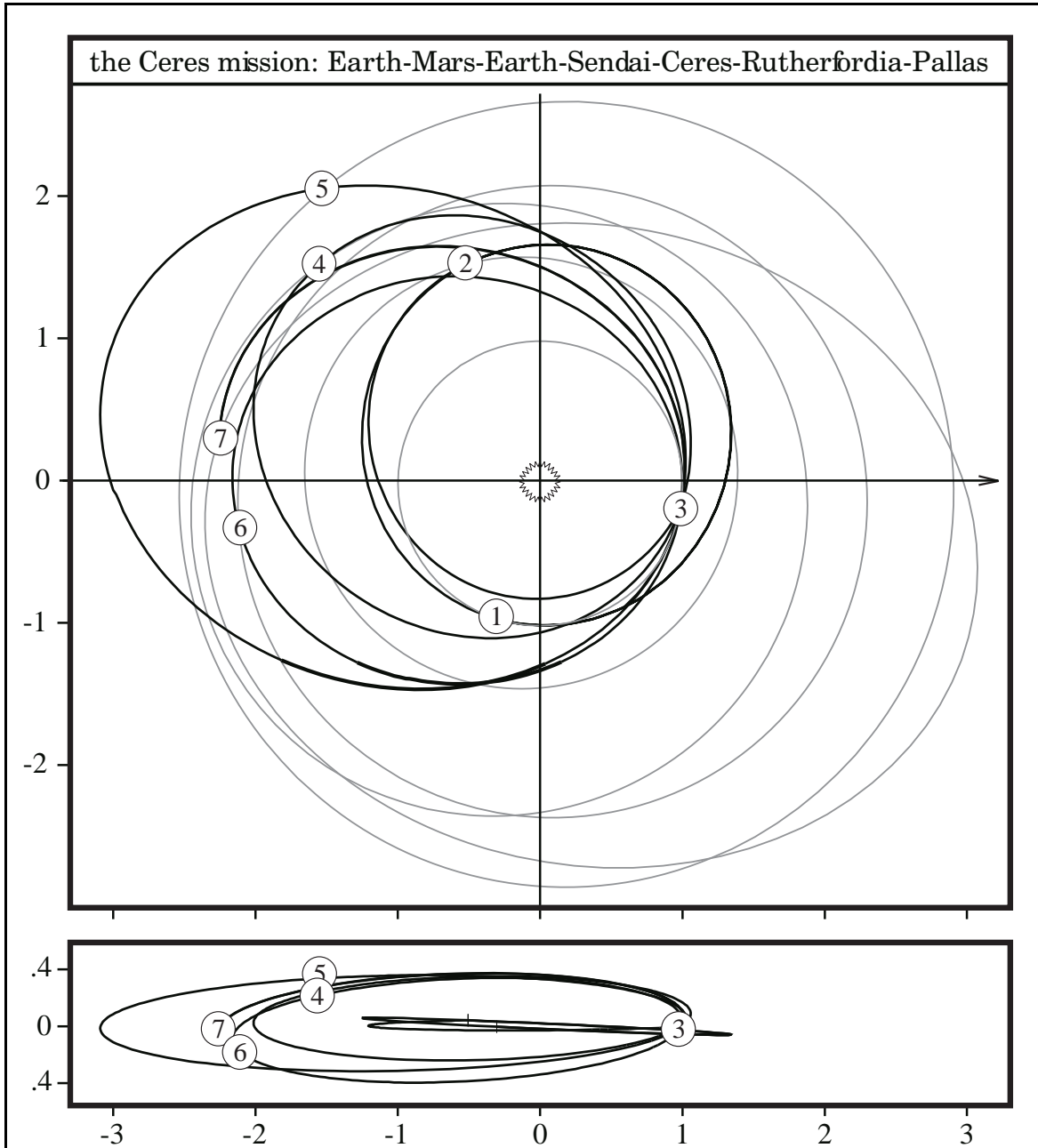
Departure date and flight time. It is important how sensitive the Mars-earth gravity assist maneuver reacts when we change the departure data or the arrival date.



Regarded is a probe launched on March, 24th 2005, on a three year earth-return trajectory (injection $v_\infty = 3.4$ km/s). A Mars flyby (on September 30th 2007) deforms the trajectory so that the earth is encountered on January 21st, 2008. Thus, the flight time is approximately two months shorter than three years. The probe arrives at the earth with a relative velocity of $v_\infty = 10.2$ km/s. The same relative velocity could have been achieved by application of 1.6 km/s retro-thrusting at aphelion position. Thus, the Mars flyby saves propellant equivalent to a Δv capacity of 1.6 km/s.

The earth swingby that follows can use the relative velocity for the continuation of the mission to another target, but the “launch window” for a planetary mission is quite small: changing the schedule has a strong impact on the maneuver. The efficiency deteriorates rapidly when we change the departure date or the arrival date. A spacecraft that departs 10 days later passes Mars at a far distance. Two years later it arrives at the earth, nearly with the same relative velocity that it had for the departure maneuver.

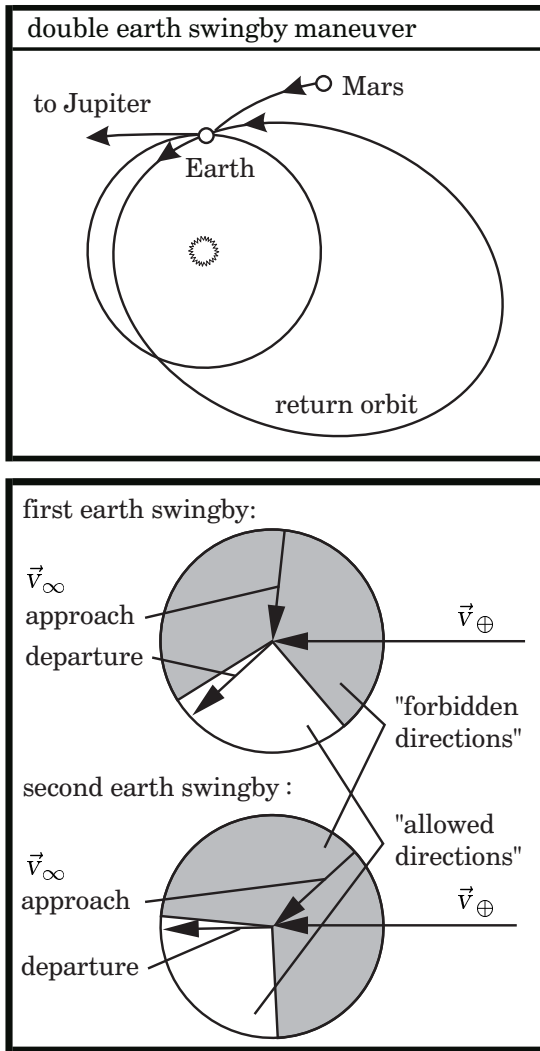
Continuation of the mission. When the spacecraft returns to the earth coming from Mars, it approaches the earth with a relative velocity of about 8 to 10 km/s. A gravity swingby maneuver at the earth turns the relative velocity in a direction more or less parallel with the heliocentric velocity of the earth. In this way the maneuver increases the heliocentric energy of the trajectory. The mission can be continued, for example to a “main-belt” asteroid. Many asteroids are contained in a ring between Mars and Jupiter, called the asteroid main-belt, and we may select one that can be visited using another earth-return trajectory. We may select the departure direction after the the earth-swingby maneuver in a way that the spacecraft visits an asteroid and, additionally, returns back to the earth after exactly two or exactly three years. The swingby maneuver can establish again an earth return orbit. The mission can be continued in this way, repeatedly visiting other asteroids on the resonant trajectory.



Example 8-7: A Mars-Earth gravity assist maneuver can be used to reduce the injection energy for a “multi-asteroid flyby mission” on a resonant earth return orbit. The example shows a tour that includes the asteroids 1-Ceres and 2-Pallas, the two largest bodies in the asteroid main belt, and 3133-Sendai and 1249-Rutherfordia. The return orbit requires an earth-swingby velocity $v_\infty = 9.5$ km/s; but the gravity maneuvers reduce the launch velocity to a comfortable value of just $v_\infty = 4.1$ km/s.

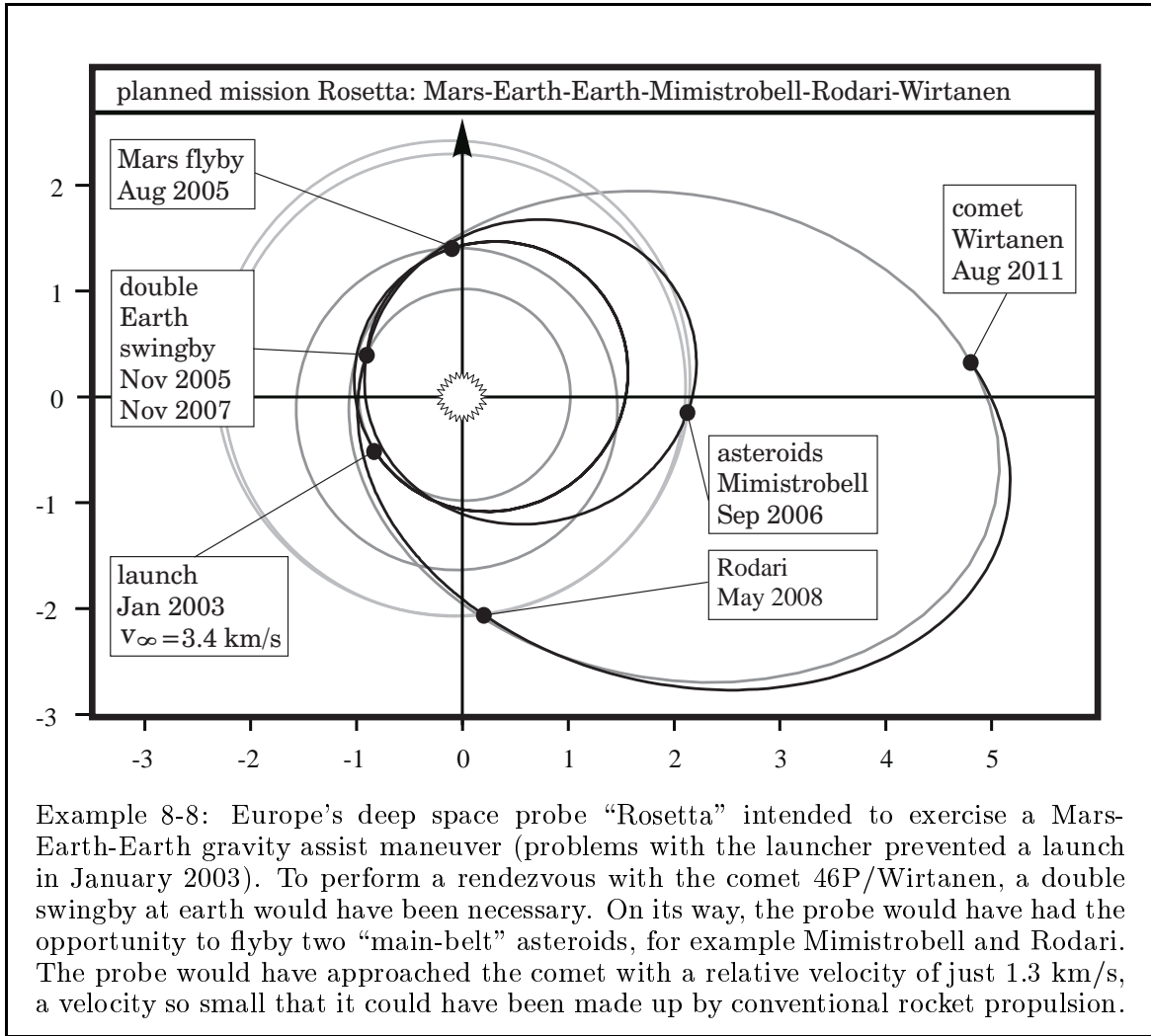
time line		v_∞ [km/s]
1: June 3, 2007	launch	4.1
2: Dec. 13, 2009	Mars	5.0
3: Sep. 12, 2010	Earth	9.5
4: July 25, 2011	Sendai	5.3
3: Sep. 12, 2012	Earth	9.5
5: June 11, 2013	Ceres	10.6
3: Sep. 13, 2015	Earth	9.5
6: Oct. 27, 2016	Rutherfordia	7.2
3: Sep. 13, 2017	Earth	9.5
7: May 21, 2015	Pallas	16.9

Double earth swingby. We have seen that a single gravity swingby maneuver at the earth raises the energy of the Earth-Mars-Earth trajectory sufficiently to visit the asteroid main belt. However, the spacecraft has to execute two earth swingby maneuvers to get sufficient energy for travelling to the outer planets Jupiter or Saturn.



Consider a spacecraft that approaches the earth for a gravity swingby maneuver on a return orbit. Encounter takes place either before or after perihelion passage; the vector of the relative velocity of the spacecraft \vec{v}_{∞} aims either at the sun (when the encounter takes place before perihelion), or it aims away from the sun (when the encounter is after perihelion); but \vec{v}_{∞} is in both cases approximately perpendicular to the vector of the heliocentric velocity of the earth \vec{v}_{\oplus} . It is intended that the swingby maneuver turns the relative velocity \vec{v}_{∞} into a direction parallel with the heliocentric velocity \vec{v}_{\oplus} . However, a deflection angle of 90° is not allowed when the flyby velocity is about 10 km/s: it would violate the "flyby distance restriction" (the flyby radius would be smaller than radius of the earth). Thus, a single earth swingby maneuver cannot take full advantage of the relative velocity \vec{v}_{∞} . The spacecraft enters first a return orbit that brings it back to the earth for a second swingby maneuver after two or after three years; the second swingby maneuver at the earth turns the relative velocity \vec{v}_{∞} parallel with \vec{v}_{\oplus} .

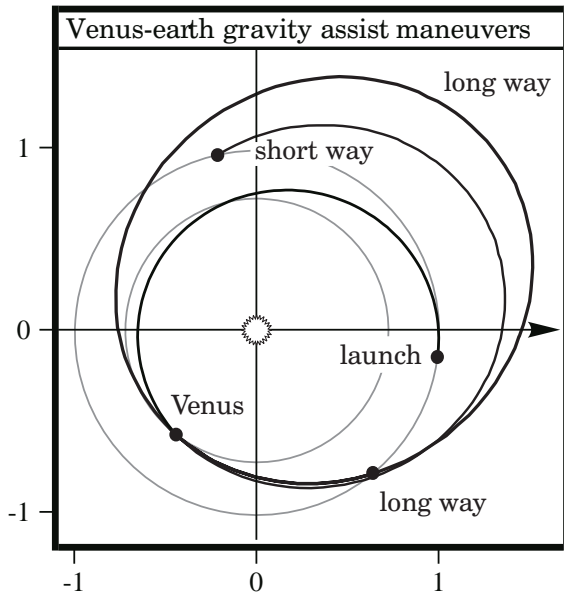
Remember that there are two different kind of earth-return trajectories, both are suitable for double earth swingby maneuvers. There is the type that is not coplanar; where the earth encounter takes place exactly at the same location in space where the departure took place. This return orbit has the advantage that it is comparatively easy to find an asteroid that can be visited on the way through the main belt (the inclination of the return orbit is in some limits a free parameter, any inclination will bring the spacecraft back to the earth). There is also the coplanar type of return orbit that remains in the ecliptic plane, where the earth is encountered at a location different from the departure location. The coplanar return orbit has the advantage that a small thrust maneuver at aphelion position can adjust the arrival date and the arrival velocity. When we design a gravity propelled mission to the outer planets, the adjustment capability makes the coplanar earth return orbit even more valuable.



8.3.3 Gravity Assist from Venus and Earth

Venus swingbys on earth-return trajectories. The principle works similar when we take planet Venus instead of Mars to deform a trajectory that re-targets the earth for a gravity swingby maneuver. We need approximately the same launch energy for the injection of the spacecraft ($v_{\infty} \approx 3.5$ km/s), but Venus is one of the inner planets. The vector of the departure velocity must point in a direction opposite to the direction of the earth’s velocity; and the injection reduces the heliocentric energy of the orbit.

The swingby maneuver at Venus puts back that amount of energy into the trajectory (after the swingby at Venus, the earth-return trajectory has again the orbital period of approximately one year). There are several ways on which the spacecraft can return to the earth; the fastest way re-encounters the earth already ten months after the Venus flyby. A two-year (or a three-year) earth-return trajectory (with two gravity assist maneuvers) utilizes the relative velocity of about 10 km/s at earth encounter.

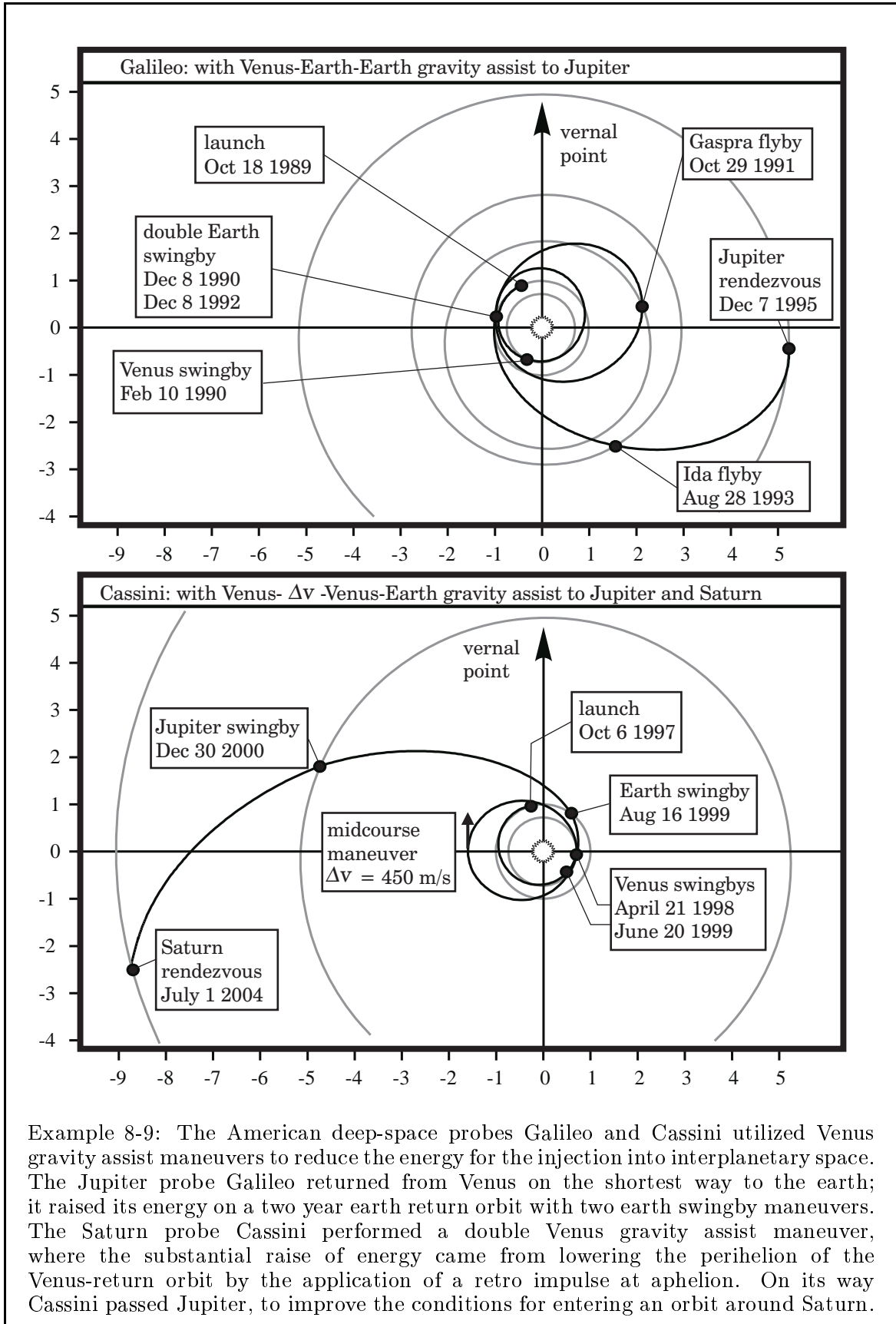


way has the advantage that the earth is encountered with a higher relative velocity: $v_{\infty} = 11.9$ km/s. Since the orbital period of Venus is rather short ($T = 0.806$ years), there are also several options to perform more than one Venus flyby before the probe finally returns to the earth. The perihelion of a coplanar Venus return trajectory can be lowered at aphelion position by a small rocket retro impulse (like it is the case with coplanar earth-return trajectories). This is an efficient way to put more energy into the trajectory, or, in case this is necessary, to adjust the arrival date.

Comparison of Venus-Earth and Mars-Earth maneuvers. We have seen that we can reduce the injection energy for a planetary mission when we execute as a first step a flyby maneuver, alternatively at Mars or at Venus. However, the substantial increase of the heliocentric energy comes from the earth. The method works with the deformation of the earth-return trajectory that is deflected by the flyby at Mars or at Venus. One or two earth gravity assist maneuvers take advantage of the relative velocity. Usually, two earth gravity assist maneuvers are necessary before the mission can be continued to the outer planets (because of the flyby distance restriction).

When we use Mars to deflect the earth-return trajectory, the spacecraft arrives at the earth with a relative speed of 8 to 10 km/s. When the spacecraft comes from Venus, the relative speed is with 10 to 12 km/s a little higher. Trajectories which utilize Venus need a shorter mission time than trajectories which utilize Mars. A duration of less than two years is typical for trajectories with a Venus flyby maneuver, while missions with a Mars flyby maneuver need a flight time of at least about three years (the earth-return trajectory that follows on needs additionally two or three years). Venus is preferable since also the Earth-Venus synodic period is shorter than the Earth-Mars synodic period. However, the solar radiation at Venus (distance 0.720 AU) is about twice as severe as at earth (distance 1 AU); and the thermal environment for a spacecraft is much better on a trajectory that utilizes a Mars swingby maneuver. When the mission duration is not a critical point, the planet Mars is preferable, because the planet Mars itself is certainly a target that is always worth to be visited.

September 15th, 2010, for example, is an appropriate date to leave for Venus. The spacecraft departs with a relative velocity of $v_{\infty} = 3.4$ km/s. Venus is encountered on February 28th, 2011, with a relative velocity of $v_{\infty} = 7.0$ km/s. The spacecraft can be re-directed to the earth on several different ways. Venus is passed at a flyby radius of 18600 km for a return on the shortest way; then the spacecraft arrives on January 4th, 2012, with a relative velocity of $v_{\infty} = 10.3$ km/s. Alternatively, the spacecraft can pass Venus at a radius of 9700 km to return on a longer way, then August 1st, 2012 is the arrival date. The long

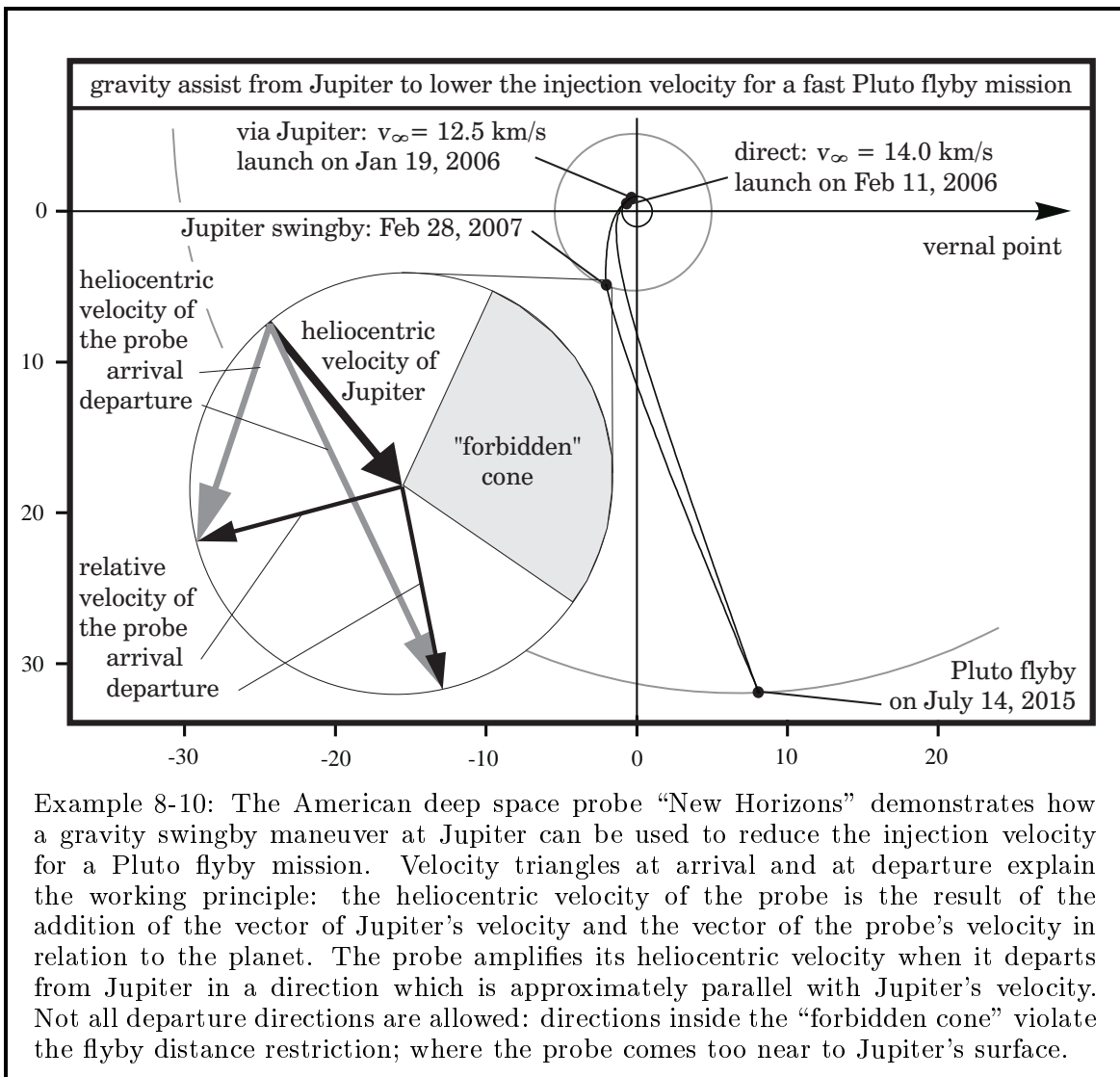


Example 8-9: The American deep-space probes Galileo and Cassini utilized Venus gravity assist maneuvers to reduce the energy for the injection into interplanetary space. The Jupiter probe Galileo returned from Venus on the shortest way to the earth; it raised its energy on a two year earth return orbit with two earth swingby maneuvers. The Saturn probe Cassini performed a double Venus gravity assist maneuver, where the substantial raise of energy came from lowering the perihelion of the Venus-return orbit by the application of a retro impulse at aphelion. On its way Cassini passed Jupiter, to improve the conditions for entering an orbit around Saturn.

8.3.4 Gravity Assist on the Way to the Outer Planets

Jupiter gravity assist on the way to Pluto. A spacecraft has to depart from the earth with a velocity of $v_\infty = (\sqrt{2} - 1) \cdot v_\oplus = 12.34$ km/s when it intends to escape from the solar system on a parabolic orbit. To reach the far remote planet Pluto on a Hohmann-type transfer orbit, the spacecraft needs nearly the same injection velocity ($v_\infty = 11.88 \pm 0.14$ km/s). However, in practice the travel time is “too long” (about 48 ± 17 years, the differences come from the eccentricity of Pluto’s orbit). The required injection velocity grows for faster transfers, for example $v_\infty = 14.8$ km/s for a travel time of 10 years, or $v_\infty = 16.8$ km/s for a travel time of just 7 years.

A gravity assist maneuver at Jupiter can reduce the injection velocity, for example from 14.8 km/s to just 11 km/s, maintaining the flight time of 10 years. The flight mission of the American deep space probe “New Horizons” demonstrates the method.



Example 8-10: The American deep space probe “New Horizons” demonstrates how a gravity swingby maneuver at Jupiter can be used to reduce the injection velocity for a Pluto flyby mission. Velocity triangles at arrival and at departure explain the working principle: the heliocentric velocity of the probe is the result of the addition of the vector of Jupiter’s velocity and the vector of the probe’s velocity in relation to the planet. The probe amplifies its heliocentric velocity when it departs from Jupiter in a direction which is approximately parallel with Jupiter’s velocity. Not all departure directions are allowed: directions inside the “forbidden cone” violate the flyby distance restriction; where the probe comes too near to Jupiter’s surface.

Disadvantages of a Jupiter swingby maneuver. A gravity assist maneuver at Jupiter can greatly reduce the injection energy of a fast mission to the outer planets; however, a near flyby at Jupiter involves also two disadvantages. First, Jupiter has a severe radiation environment; and the spacecraft must be protected against radiation. Second, now the spacecraft cannot be launched “once a year” anymore. For example, the synodic period of Jupiter and Pluto is approximately 13 years; and launching a Jupiter-Pluto mission is just possible in a time interval of three years for every synodic period. Three launch windows exist where the Earth-Jupiter synodic period meets the Jupiter-Pluto synodic period. For example, launching will be possible in November 2015, in December 2016 and in January 2018. The earliest of these launch windows needs the lowest injection velocity but requires the nearest Jupiter flyby distance. The injection velocity grows for the later windows, but the flyby distance gets larger.

mission injection date (v_∞ [km/s])	Jupiter flyby date (radius [km])	Saturn flyby date (radius [km])	overall mission duration [days]
Dec 11 2010 (11.6)	-	Mar 25 2014 (125000)	1200 + 3527
Nov 18 2015 (10.2)	June 10 2017 (172000)	-	570 + 3428
Dec 19 2016 (10.6)	May 23 2018 (491000)	-	520 + 3185
Dec 19 2016 (10.6)	May 23 2018 (183000)	Mar 18 2020 (348000)	520 + 665 + 2992
Jan 20 2018 (11.1)	May 5 2019 (1483000)	-	470 + 3996
Jan 20 2018 (11.1)	May 5 2019 (435000)	Oct 5 2020 (290000)	470 + 519 + 3001

Gravity assistance from Jupiter and Saturn. A question that comes up is whether it is possible to use on-route gravity assistance from other outer planets than from Jupiter. In order to be helpful for a fast Pluto flyby mission, for example, Saturn comes into favorable positions in the time between the years 2010 and 2018. Then it is possible to take Saturn instead of Jupiter, but the replacement does not bring advantages: trajectories which use Saturn require a higher injection velocity than trajectories which use Jupiter. It is an option to include a Saturn swingby in a Jupiter gravity assist trajectory when we see Saturn itself as an interesting target. A near Saturn flyby involves the risky passage through the planet’s ring system. The American space probes Voyager and Cassini crossed the ring plane without damage between the F ring and the G ring; a passage through the bright rings A, B or C however, is probably not possible. Saturn has an equatorial radius of 60268 km.

Saturn	inner radius [km]	outer radius [km]
D-ring	67000	74500
C-ring	74500	92000
B-ring	92000	117500
A-ring	122200	136000
F-ring	140200	140350
G-ring	165800	173800
E-ring	180000	480000

When we plan the fast Jupiter-Pluto flyby mission for the three launch opportunities between 2015 and 2018, the Saturn flyby is an end in itself: a Saturn flyby will neither reduce the time of the mission, nor it will enlarge the flyby distance at planet Jupiter.

Gravity assistance from Jupiter, Saturn and Neptune. The situation is different when we plan a mission to a more distant target, for example to the Kuiper belt object 2003 UB₃₁₃ which has recently been discovered (and nicknamed “Xena”). Considering that Xena is bigger than Pluto and has at least one moon, we may call it the 10th planet of our solar system. However, in comparison with Pluto Xena has presently about the double distance from the sun, and thus the mission time would become too long if we used exclusively the gravitational assistance of Jupiter. Additionally, swingby maneuvers at Saturn and even Neptune are necessary. The table on the preceding page shows that the appropriate constellation of the planets Jupiter and Saturn appears for two launch windows: the first one in December 2016 and the second one in January 2018. Then the mission can be executed as follows:

planet	Earth	Jupiter	Saturn	Neptune	Xena
date	Dec 3 2016	June 4 2018	Apr 13 2020	June 8 2025	June 20 2038
velocity v_∞	11100 m/s	12300 m/s	17600 m/s	22800 m/s	25300 m/s
flyby radius	-	193000 km	125000 km	55000 km	-
date	Jan 12 2018	May 29 2019	Nov 29 2020	Dec 17 2025	Feb 29 2039
velocity v_∞	10800 m/s	13700 m/s	17600 m/s	22400 m/s	24900 m/s
flyby radius	-	529000 km	144000 km	59000 km	-

The second of the two trajectories above is the better option, because the radiation environment for the spacecraft is less crucial when the Jupiter flyby distance is large. The duration of the mission is then roughly 21 years. A mission that proceeds directly from Saturn to Xena (omitting the Neptune swingby maneuver) is not possible: such a mission would pass Saturn closely and intersect one of the bright rings of the planet.

Using additionally Mars or Venus swingby maneuvers. A mission to the outer planets which incorporates a Jupiter swingby maneuver requires an injection velocity v_∞ of about 11 km/s, but 11 km/s is usually “too high” for a reasonable mission. The solution is to improve the trajectory by the addition of an initial maneuver at Mars or at Venus, followed by a double swingby at the Earth, before the spacecraft travels on to Jupiter and finally to the far distant target. Mars or Venus swingby maneuvers on earth return trajectories can increase the earth swingby velocity v_∞ to a value of up to 11 km/s. These maneuvers add in any case flight time to the mission, but they have the capacity to reduce the injection velocity of the mission greatly.

Let us return to our Jupiter-Pluto example mission. The problem is how to combine the Earth-Jupiter-Pluto trajectories with Earth-Mars-Earth-Earth or Earth-Venus-Earth-Earth trajectories. We have to find maneuvers which match the three launch windows for the departure from Earth to Jupiter. The search for opportunities has to be done numerically on a computer, because a large variety of different trajectories exists. All missions depart earlier now; adding Earth-Mars-Earth-Earth or Earth-Venus-Earth-Earth trajectories brings the launch date forward. More opportunities show up when the acceptable duration of the mission becomes longer, but “good” are just those trajectories which combine low injection energy with short flight time.

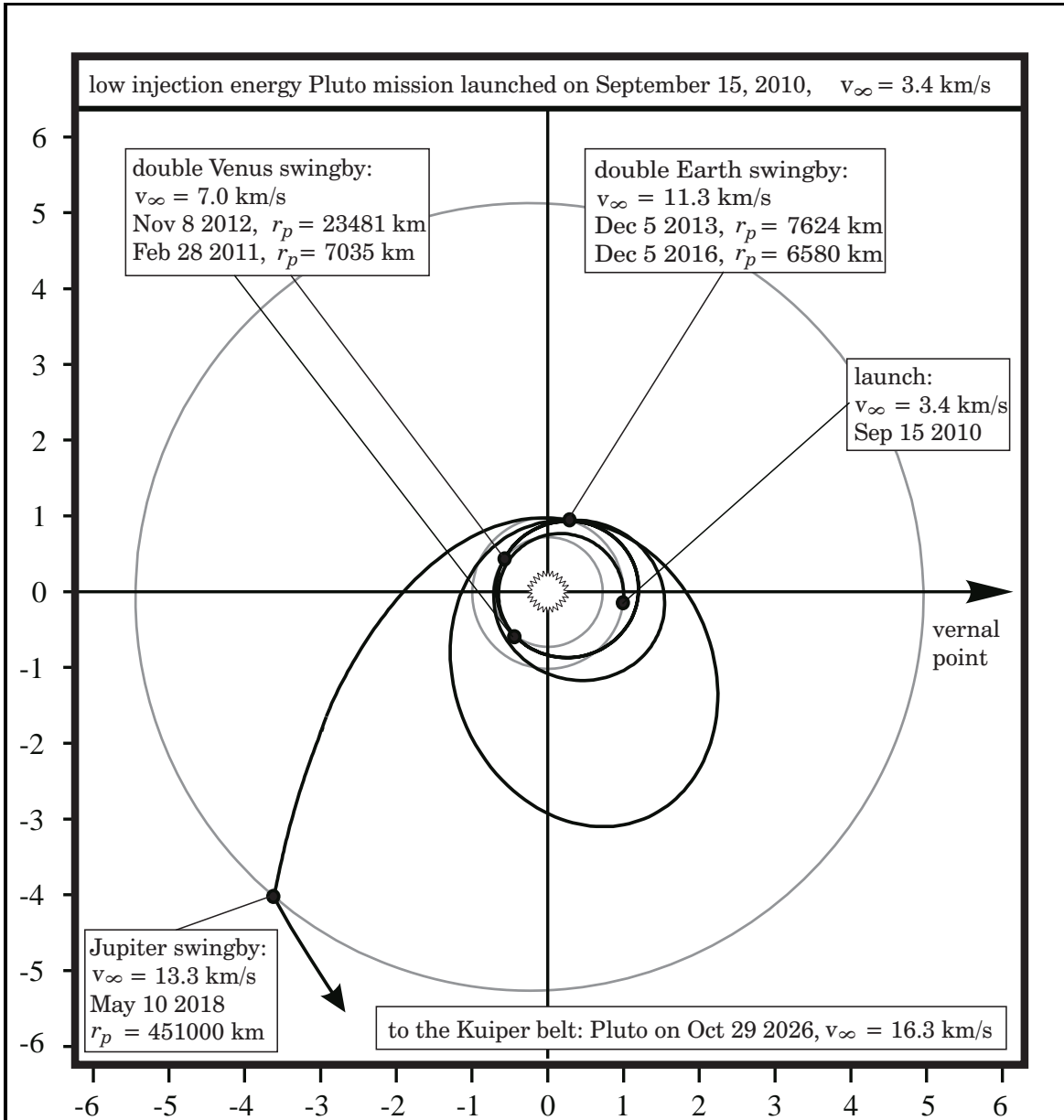
Other criteria that have to be observed are the Jupiter flyby distance (Jupiter should be passed as far as possible) and the amount of deterministic post-launch Δv (the midcourse maneuver of the earth-return orbit should be as small as possible).

mission type	injection	Δv	Jupiter flyby	Pluto flyby
planets	date (v_∞ [m/s])	[m/s]	date (radius [km])	date (time [years])
1 EMEEJP	Jan 26 2005 (4100)	100	May 29 2018(525000)	May 8 2027 (22.3)
2 EMEEJP	Jan 27 2005 (4112)	-	Jul 9 2018 (827000)	Jul 13 2029 (24.5)
3 EMEEJP	Dec 22 2006 (3715)	-	Aug 7 2017 (298000)	Aug 6 2028 (21.5)
4 EMEEJP	Dec 20 2007 (4494)	100	Jul 24 2018 (881000)	Oct 12 2030 (22.7)
5 EMEEJSP	Jan 17 2008 (4283)	200	May 18 2019(482000)	Sep 9 2029 (21.6)
6 EMEEJSP	Feb 28 2009 (3898)	200	Jun 19 2019(625000)	Mar 7 2032 (23.0)
7 EMEEJSP	Mar 2 2009 (3909)	300	May 28 2019(515000)	May 1 2030 (21.1)
8 EVVEEJP	Sep 15 2010 (3391)	-	May 10 2018(451000)	Oct 29 2026 (16.1)
9 EVEEJP	Mar 23 2012 (3488)	-	Jun 4 2018 (588000)	Nov 7 2027 (15.5)
10 EVVEJP	Apr 11 2012 (3584)	320	Jul 9 2017 (196000)	Jun 18 2027 (15.2)
11 EVEEJP	Apr 18 2012 (3291)	200	Aug 10 2017(303000)	Sep 15 2028 (16.5)
12 EVEEJSP	Apr 27 2012 (3542)	-	May 28 2019(537000)	Oct 11 2030 (18.5)

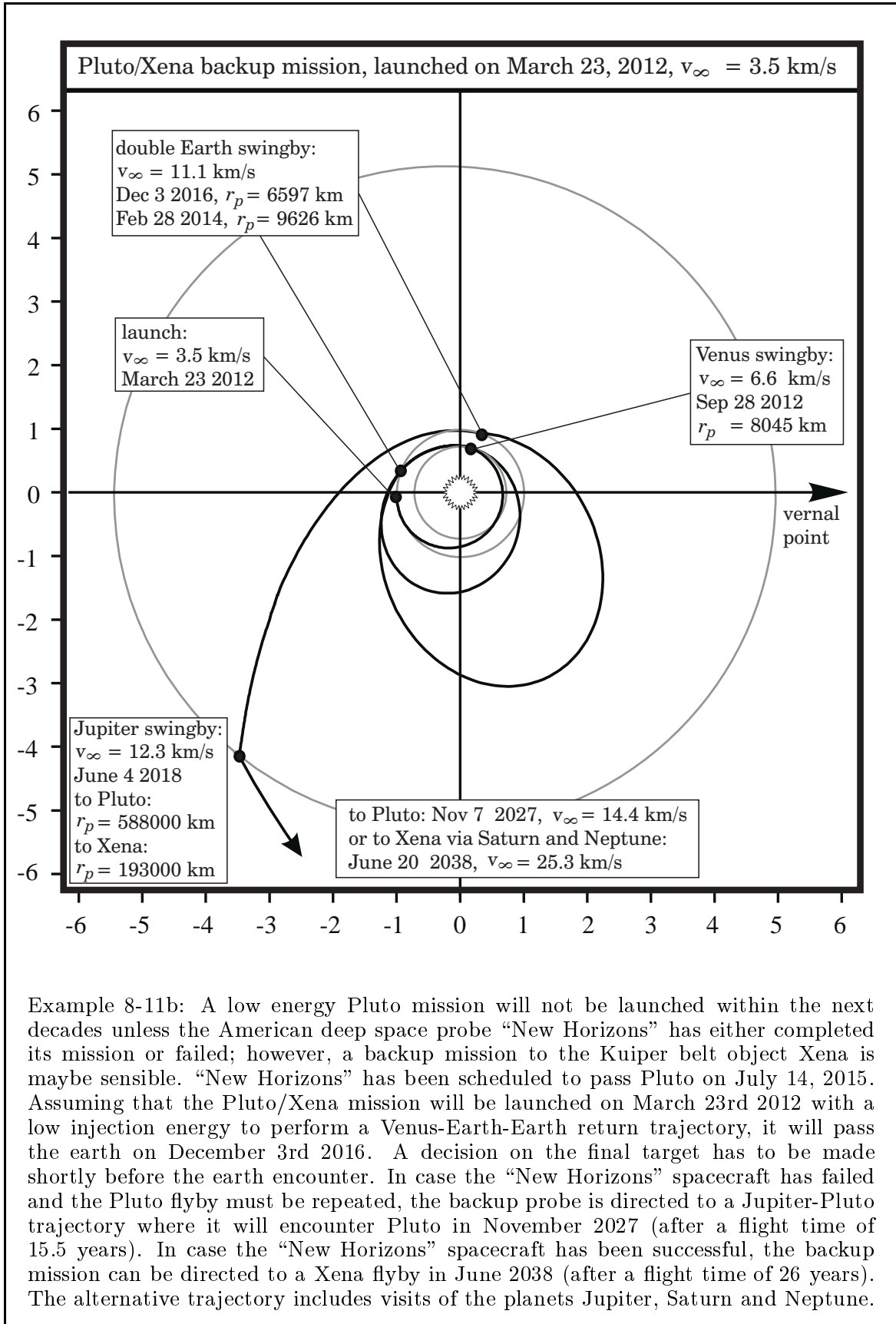
The table above summarizes favorable launch opportunities for the Earth-Jupiter-Pluto constellation which appear in the years between 2015 and 2018. It can be observed that the Jupiter-Pluto synodic period puts the arrival date to the years between 2026 and 2030. All trajectories include a Jupiter flyby, some trajectories visit Saturn additionally. Trajectories which use a Mars maneuver require a mission duration of about 21 to 25 years, while trajectories which employ Venus are much faster (the duration of the mission is just about 16 years). These Venus trajectories are also better concerning the injection velocity v_∞ and the deterministic post-launch Δv .

We can also direct the spacecraft to another target (instead of heading for Pluto), for example to the Kuiper belt object 2003 UB₃₁₃ Xena, provided that the final earth swingby maneuver takes place in December 2016 or in January 2018. This is the case for all trajectories in the table above, with exception of the numbers 3, 10 and 11. A trajectory correction maneuver, executed shortly before the final earth encounter, can adjust the trajectory in a way that Jupiter is passed with a different flyby distance; and then the spacecraft travels on to Saturn, Neptune and finally to Xena. The overall travel time depends mainly on the Saturn flyby distance and is typically 27 years, assuming that a Venus gravity assist maneuver is used to lower the injection energy of the mission. Xena is then encountered between 2038 and 2040.

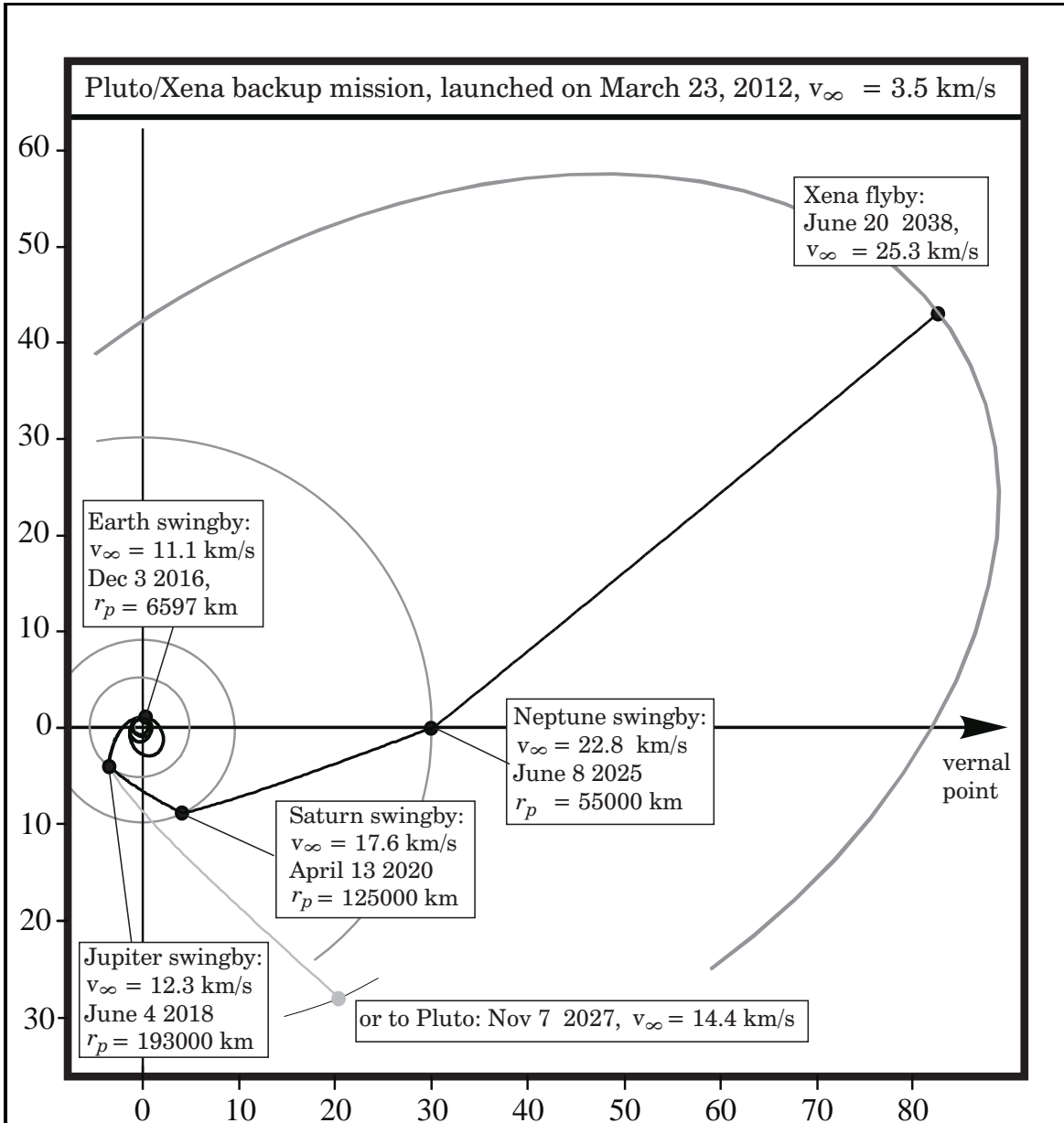
The trajectories number 8 and 9 are the best (or we consider Mars as target itself). Thus, the launch date for the “nominal mission” must be on September 15th 2010, the energy for the injection v_∞ is lower than 3.5 km/s. A “backup mission” exists for March 23rd 2012. When the spacecraft is directed to Pluto, it will arrive about 16 years after the injection (in October 2026 or in November 2027 respectively). We can reduce the flight time when we spend some Δv at the last earth flyby (about one year for 300 m/s). The nominal mission offers the opportunity for an asteroid flyby (177 Irma in Sept 2011, 1771 Makover in March 2012, 1046 Edwin in October 2012).



Example 8-11a: Pluto is the name of the most famous member of the Kuiper belt. A Pluto mission can utilize the gravitation of Venus, Earth and Jupiter to lower the injection energy, and the far target can be reached after a flight time of just 16 years. The gravity assist Pluto mission does not demand more launcher capacity than the positioning of a geostationary earth satellite; but Pluto's extreme distance in deep-space, where sunlight is not much brighter than moonlight on the earth, raises up challenging astronomical problems (data transmission, trajectory accuracy and energy supply). In case radioisotope thermoelectric generators are "not-available" for the mission, solar-cells in combination with lithium batteries can be a solution, provided that it is possible to let the spacecraft "sleep" on the long way from Jupiter to Pluto. The spacecraft "wakes up" before the Pluto encounter, finds orientation and starts data collection. The actual encounter takes just about 10 hours and requires approximately 1.7 kWh electrical energy. About 200 MBytes of data have to be transmitted to the earth after the flyby, requiring approximately 3 kWh of electric energy.



Example 8-11b: A low energy Pluto mission will not be launched within the next decades unless the American deep space probe “New Horizons” has either completed its mission or failed; however, a backup mission to the Kuiper belt object Xena is maybe sensible. “New Horizons” has been scheduled to pass Pluto on July 14, 2015. Assuming that the Pluto/Xena mission will be launched on March 23rd 2012 with a low injection energy to perform a Venus-Earth-Earth return trajectory, it will pass the earth on December 3rd 2016. A decision on the final target has to be made shortly before the earth encounter. In case the “New Horizons” spacecraft has failed and the Pluto flyby must be repeated, the backup probe is directed to a Jupiter-Pluto trajectory where it will encounter Pluto in November 2027 (after a flight time of 15.5 years). In case the “New Horizons” spacecraft has been successful, the backup mission can be directed to a Xena flyby in June 2038 (after a flight time of 26 years). The alternative trajectory includes visits of the planets Jupiter, Saturn and Neptune.



Example 8-11c: Object 2003 UB₃₁₃, discovered in 2003 and nicknamed “Xena”, is the largest object that has been found orbiting around the sun since the discovery of Neptune and its moon Triton. The planet (or planetoid) Xena has at least one moon. (orbital elements: $a = 67.6619$ AU; $\varepsilon = 0.442156$; $\vartheta = 44.1774^\circ$; $\psi = 35.875^\circ$; $\varphi_p = 151.334^\circ$; perihelion date: Nov. 9, 2256, perihelion distance $r_p = 37.745$ AU). Probably, in future it will be generally accepted that Xena is the 10th planet of the solar system, because Xena is even bigger than Pluto; even though its orbital data deviate much from what we know from the real planets: Xena needs 560 years to complete one orbital revolution, the object performs its perihelion passage at a distance of 38 AU from the sun. The orbit is with 44° highly inclined with respect to the ecliptic plane. Presently, the object is close to its aphelion position, at a distance of 97 AU from the sun (approximately 13.5 lighthours). However, a deep space probe can reach even this far distant target in a comparatively short time period, when swingby maneuvers at the outer planets accelerate the probe to an extraordinarily fast sun-escape velocity.

8.4. Manned Mars Missions

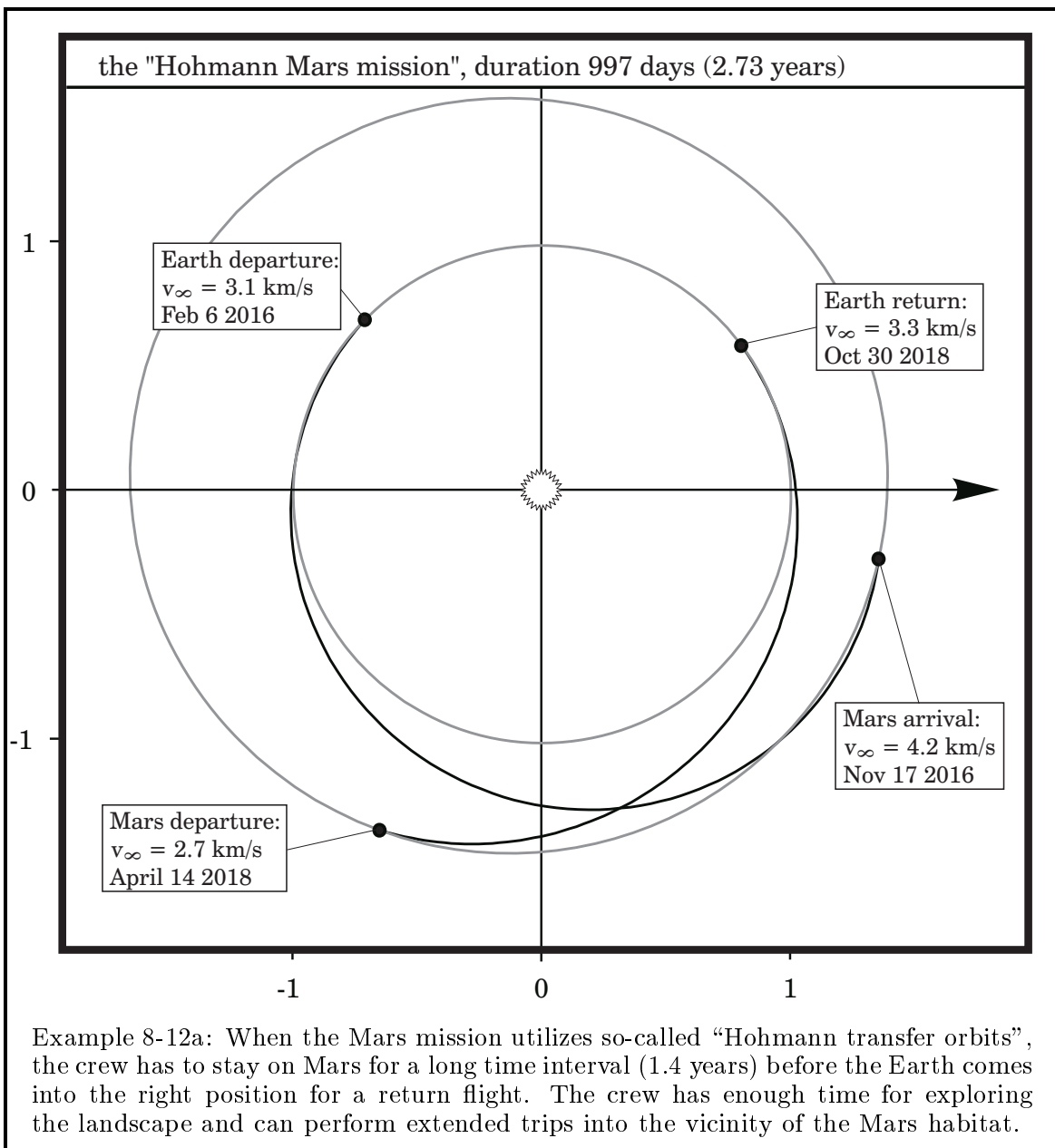
Today the surface of planet Mars is an uninhabitable desert, however, the planet may have carried life in a former period of its existence when parts of its surface were seemingly covered with water. A manned Mars mission is the most exciting project of future astronautics: it is the logical consequence of the American Apollo project and the international space station ISS; and it can help mankind to find an answer to the fundamental question about the origin of life. The surface of planet Mars, covered with sand and rocks, is accessible by humans with available technology; and maybe it is the only region in the universe that is still left over for a manned exploration. This section of the chapter shows that a manned Mars mission is feasible and not more expensive than the American Apollo project or the international space station.

8.4.1 Martian Trajectories

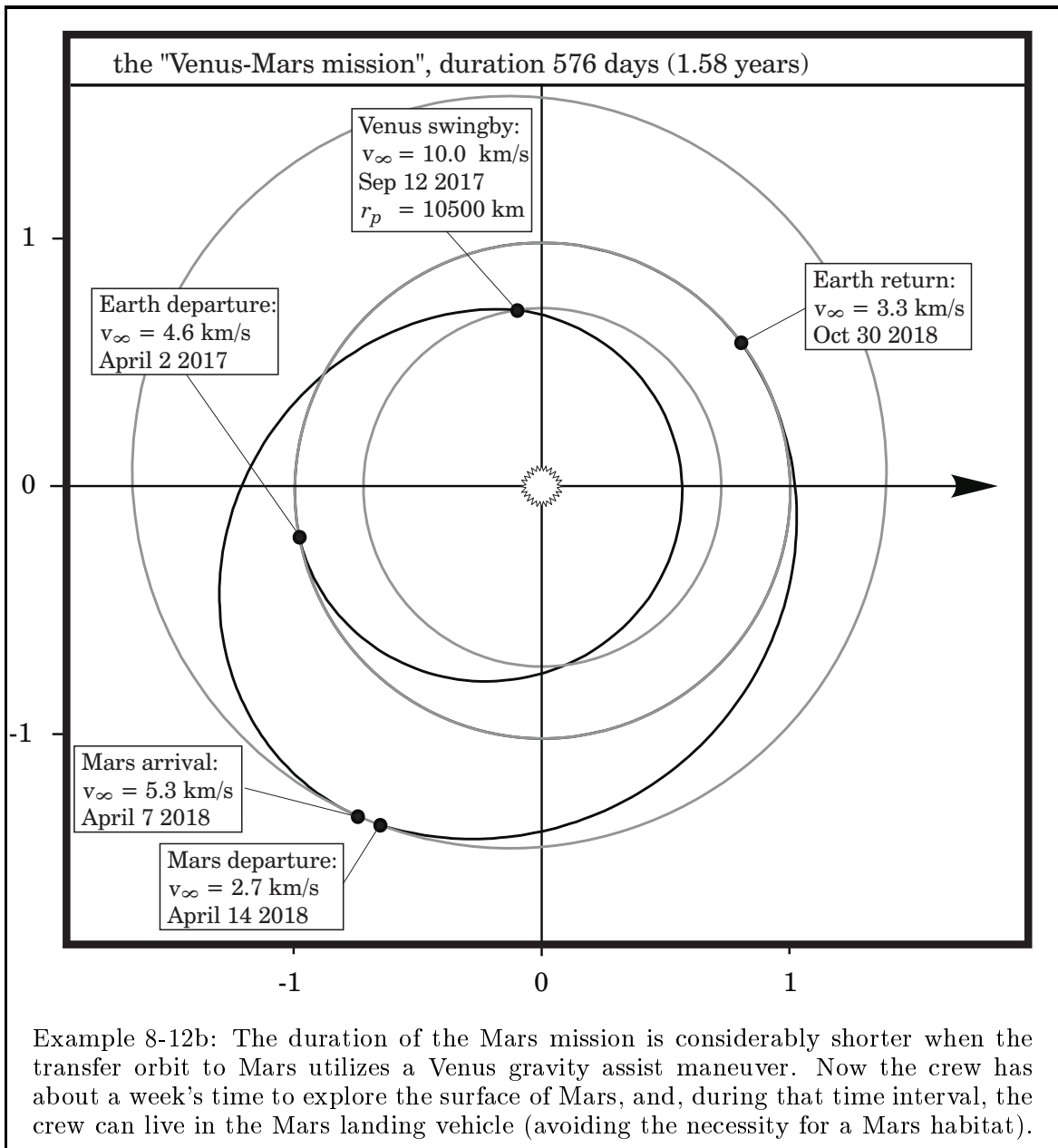
The target Mars and its orbit around the sun. Let us first have a look at Mars as a target for a planetary space mission. The planet encircles the sun at a distance between 1.381 AU and 1.666 AU (semiaxis $a = 1.524$ AU). The orbit is inclined with respect to the ecliptic and eccentric (inclination $\vartheta = 1.85^\circ$, eccentricity $\varepsilon = 0.09344$); and we can expect that the velocity requirement for a transfer trajectory from Earth to Mars varies substantially for the different launch opportunities. Mars has a carbon dioxide atmosphere with a surface density of 0.017 kg/m^3 and a scale height of 9 km; and, even though the atmosphere is extremely thin, it suffices for an aerodynamic braking maneuver. We can compare the density of the atmosphere of Mars with the density of the atmosphere of the earth at an altitude of 30 km above the surface. The landing of a spacecraft on the surface of Mars has to be supported and controlled by rocket motors, since parachutes are either “too large” or the landing vehicle is still “too fast” immediately before the final touch-down. The entry flight profile must take into consideration that the terrain of Mars is not plane, it contains surface obstacles (some mountains are quite high, for example Mons Olympus). Mars rotates on its axis with a period of 24.6 hours; the inclination angle between equator and ecliptic is 24° (Mars is similar to the Earth concerning rotation period and inclination of the axis). However, since the planet is considerably smaller than the earth, it is much easier to depart from its surface. The gravitational constant is $\gamma_{mars} = 4.28355 \cdot 10^{13} \text{ m}^3/\text{s}^2$ (the mass is 11% of the mass of the earth) and the equatorial radius is 3397 km; and we can use equation (8-24) to calculate for a flight altitude of 300 km above the surface a circular velocity of 3404 m/s and a parabolic escape velocity of 4814 m/s.

The “Hohmann Mars mission”. The so-called Hohmann transfer trajectory consists of two tangential thrust periods separated by a 180° coast arc; and a transfer orbit that is similar to the Hohmann trajectory it is the most economical way to reach the planet Mars. We can calculate a transfer time of $\Delta t = 259$ days for the flight from a circular orbit at 1 AU to a coplanar circular orbit at 1.524 AU; where the spacecraft departs from the initial orbit with a relative velocity of $v_\infty = 2942 \text{ m/s}$ and arrives at the target orbit with a relative velocity of $v_\infty = 2646 \text{ m/s}$.

A spacecraft that arrives at Mars on a Hohmann transfer trajectory cannot return back to the Earth immediately but it has to wait for the appropriate “return window”. The spacecraft needs 259 days for the flight to Mars and again 259 days for the return flight; and if it returned back immediately it would not encounter the Earth at the final instant of its mission (then the earth would be at another place on the orbit). The spacecraft has to wait approximately 454 days for the appropriate moment to return (on an average day Mars advances 0.524° while the Earth advances 0.985°). The entire mission takes about 972 days or 2.66 years. Even though we appreciate a long prevail time at Mars for extensive experiments, the long mission time is actually the disadvantage of the “Hohmann Mars mission”. The crew needs in any case a heavy container (“Mars habitat”) for staying more than a year on the surface of Mars.



The “Venus-Mars mission”. In order to reduce the mission time we have to include a gravity assist maneuver at Venus into the mission. Venus encircles the sun at a distance of about 0.723 AU. The flight times for transfers on Hohmann ellipses are: 146 days from Earth to Venus, 217 days from Venus to Mars and 259 days from Mars back to the Earth. Thus, the “Venus-Mars mission” takes all together 622 days or 1.7 years, typically. When we depart from the Earth a little faster than what we would need for a conventional Hohmann transfer, for example with a velocity of $v_\infty = 4.6$ km/s, then we can omit the propulsive maneuver at Venus. Now the spacecraft arrives at Mars coming from Venus; and therefore it approaches Mars with a faster relative velocity, $v_\infty = 5.2$ km/s. It stays at Mars for a short time period (less than a month) and returns to the Earth on a conventional Hohmann trajectory.



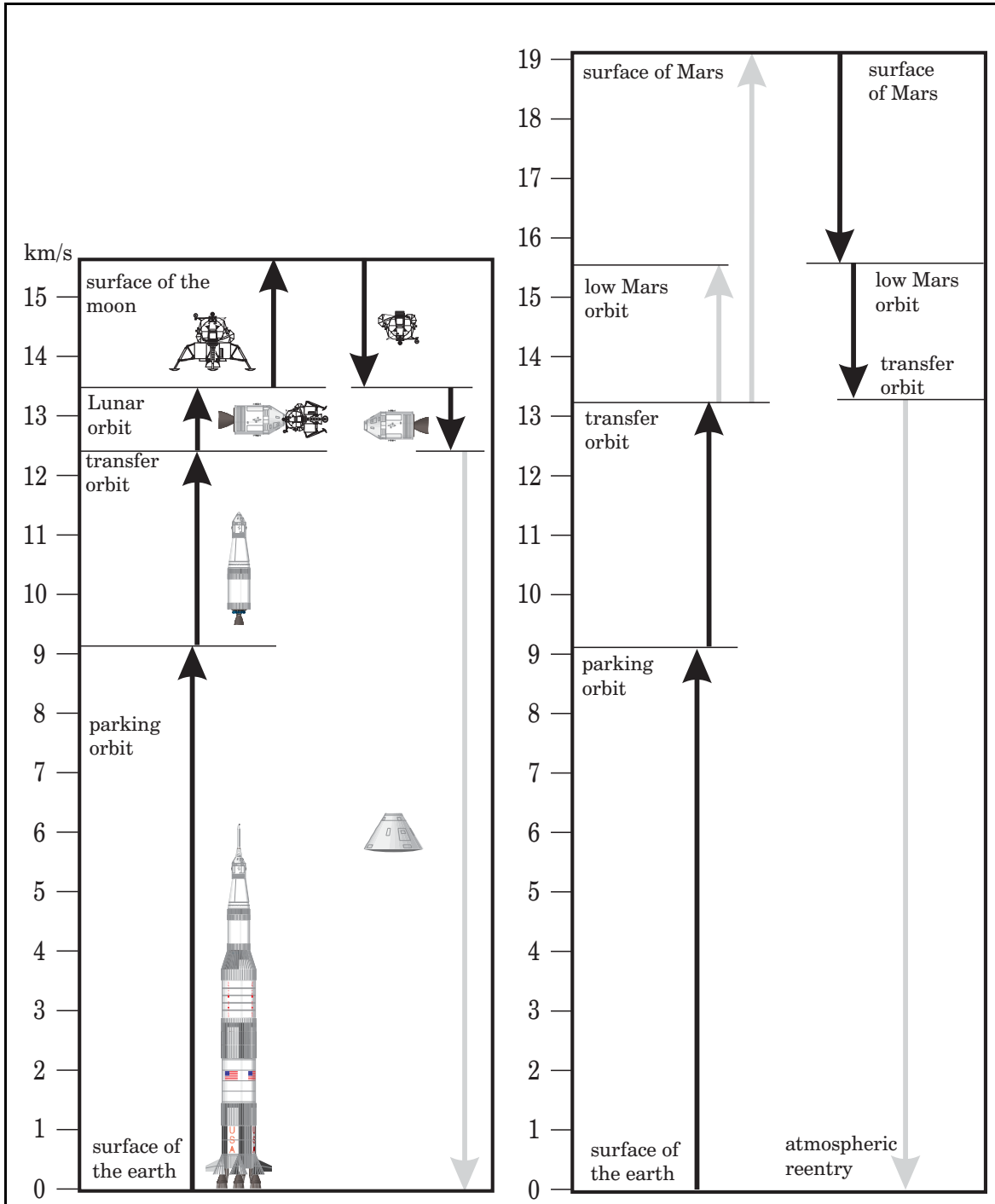
Example 8-12b: The duration of the Mars mission is considerably shorter when the transfer orbit to Mars utilizes a Venus gravity assist maneuver. Now the crew has about a week’s time to explore the surface of Mars, and, during that time interval, the crew can live in the Mars landing vehicle (avoiding the necessity for a Mars habitat).

The fast mission. An even shorter mission time is possible when we intercept Mars on its orbit around the sun with a spacecraft on a one year earth-return trajectory; however, this mission is not really an alternative because it requires an immense launch energy ($v_\infty \approx 10$ km/s). Remember that the orbital period T of a heliocentric orbit is exclusively a function of the major semiaxis: we can conclude that the semiaxis of the orbit for the transfer to Mars and back to the Earth must be: $a \approx 1$ AU. The spacecraft stays on Mars for a short period of several days before it returns to the Earth, where it enters the Earth's atmosphere with an extremely high reentry velocity.

Comparison of the mission alternatives. The "Venus-Mars mission" seems to be the best alternative: in contrast to the "Hohmann Mars mission" it is not necessary to bring a habitat to the surface of Mars, and the mission time is considerably shorter. Some disadvantages are involved. First, the thermal environment at a distance of 0.57 AU from the sun (near Venus) can mean a problem to a manned spacecraft. Then the mission needs more launch energy: when we assume the departure from a 200 km low earth orbit, the mission requires a Δv of approximately 4.3 km/s to achieve a velocity of $v_\infty \approx 4.6$ km/s (the Hohmann Mars mission: $\Delta v \approx 3.8$ km/s for $v_\infty \approx 3.3$ km/s). Another disadvantage is that the spacecraft arrives at Mars with a comparatively high relative velocity ($v_\infty \approx 5.2$ km/s); where an aerobraking maneuver must decelerate the vehicle. Finally we have to notice that launch windows for the Venus-Mars mission occur less frequently than launch windows for the Hohmann Mars mission. We can calculate the frequency of the appearance of launch windows when we compare the synodic periods of Earth, Mars and Venus. The Earth-Mars synodic period is 780 days, the Earth-Venus synodic period is 584 days. The next similar constellation for launching the Hohmann Mars mission appears already after 780 days (2.1 years); but the next launch opportunity for the Venus-Mars mission appears when four Earth-Venus synodic periods meet three Earth-Mars periods: after 2336 days or 6.4 years.

It is also possible to execute the Venus gravity assist maneuver not on the way to Mars but on the way back from Mars to Earth. Launch opportunities for such missions appear also in time intervals of 6.4 years. However, when we compare this alternative, we realize that then the velocity at Mars departure is comparatively high. A spacecraft that departs from Mars heading for Venus leaves with a relative velocity of $v_\infty = 5.2$ km/s; and this speed must be established by means of rocket propulsion (in contrast to the aerobraking maneuver that reduces the high entry speed at arrival). The comparatively high departure speed makes this alternative mission unattractive.

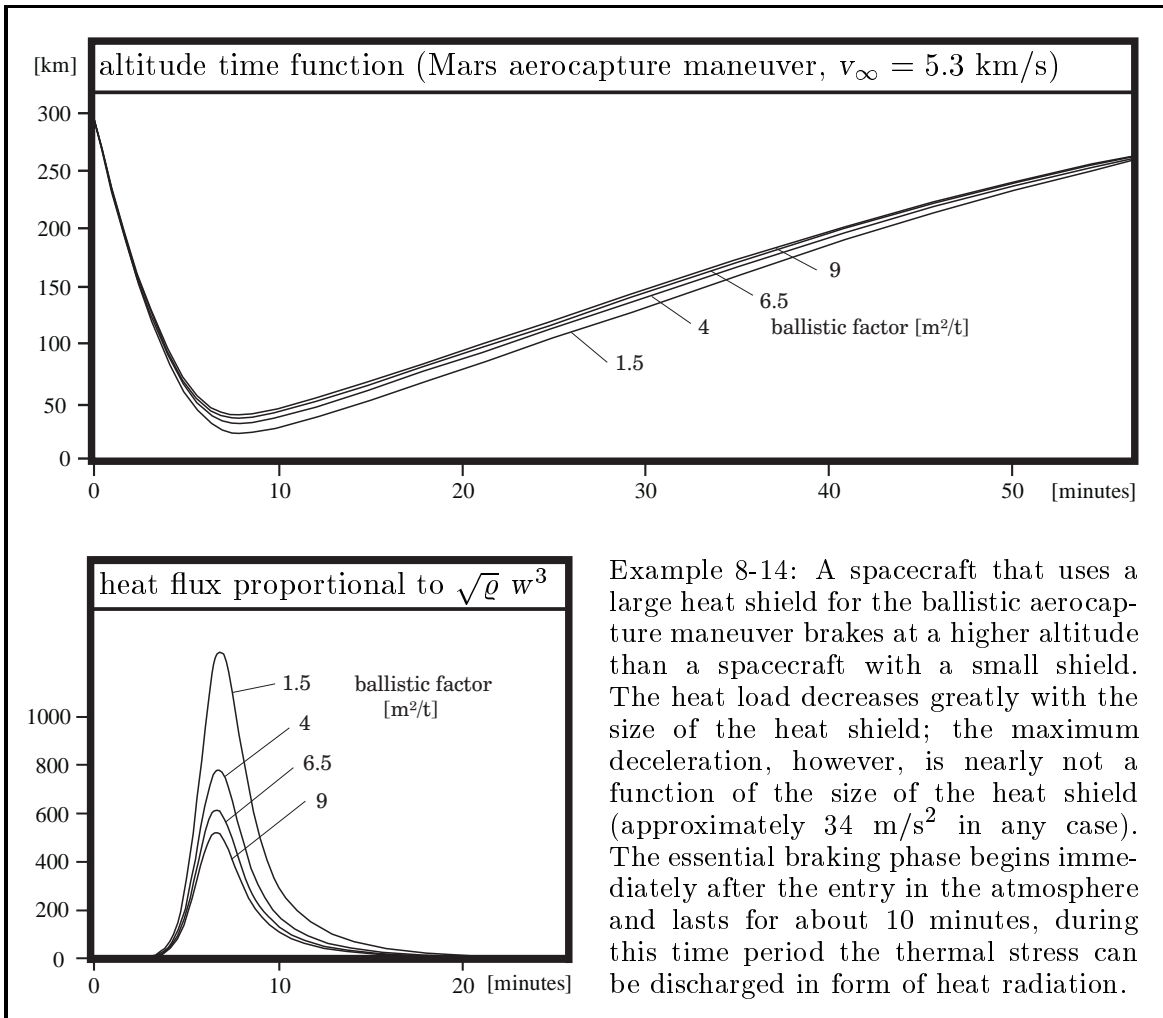
Velocity requirement for a manned Mars mission. When we compare the velocity requirement for a Mars mission with the velocity requirement for a moon mission we notice that it is easier to reach the surface of Mars than the surface of the moon: the reason for this is that the moon has no atmosphere which we could use for an aerodynamic braking maneuver. However, the return from Mars back to the Earth requires more rocket propulsion because Mars has a greater mass than the Moon. Thus, all together a manned Mars mission needs approximately the same propulsion as a manned Moon mission, but a mission to the planet Mars takes much longer: the actual problem is the duration of the mission rather than its velocity requirement.



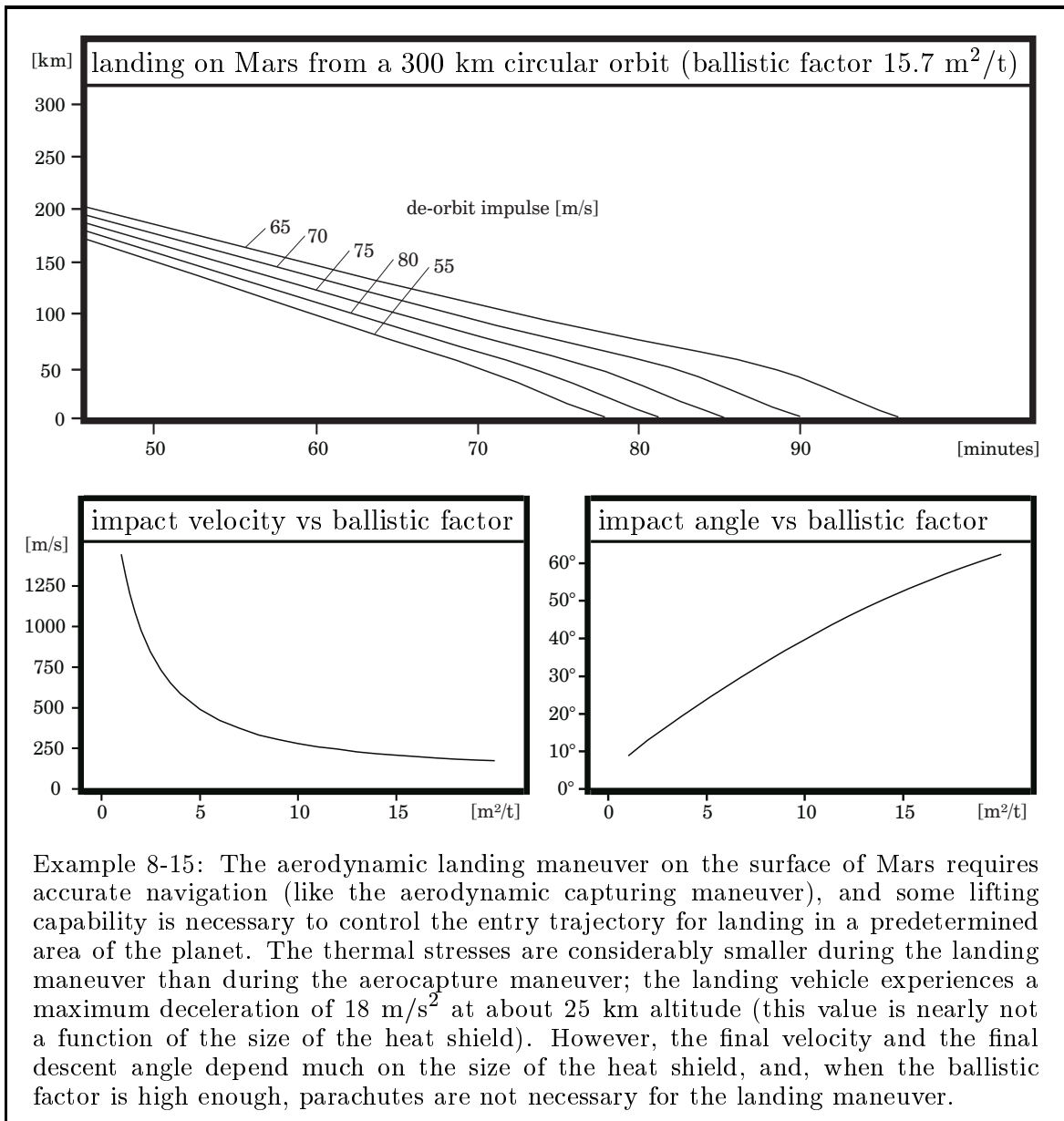
Example 8-13: Like a manned Moon mission, a Mars mission needs an enormous amount of rocket propulsion, where again the principal share is the launch of the mission ($\Delta v \approx 12.8 - 13.3$ km/s). The atmosphere of Mars can help to speed down the spacecraft at arrival, either to attain an orbit around Mars or to descend to the surface. However, then the mission has to use again rocket propulsion to depart for the Earth ($\Delta v \approx 3.6$ km/s for the ascent of the Mars lander and $\Delta v \approx 2.1$ for the return of the Mars orbiter from a 300 km circular orbit). The return to the surface of the Earth is again “cheap” when an atmospheric reentry maneuver decelerates the return cabin.

8.4.2 Flight Operations at Mars

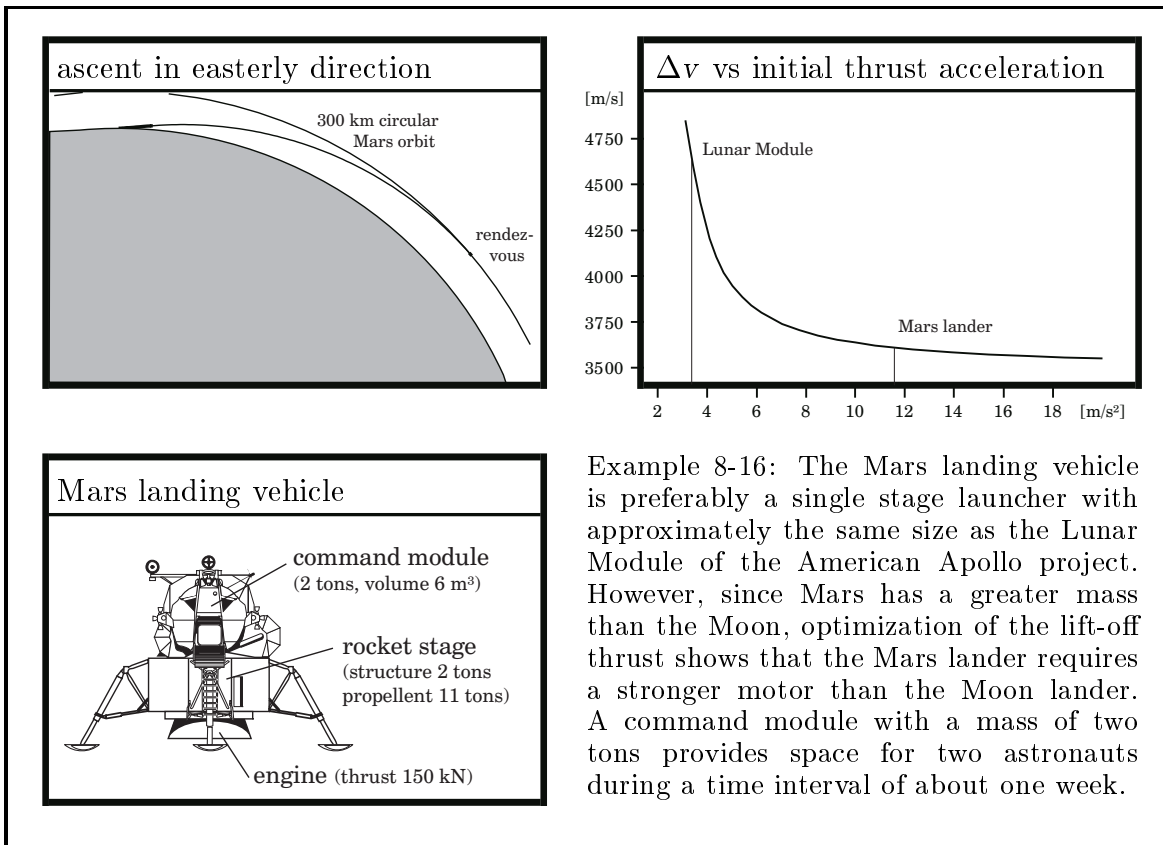
Aerocapture at arrival. The spacecraft arrives at Mars with a relative velocity $v_\infty = 4.2$ km/s (Hohmann Mars mission) or $v_\infty = 5.3$ km/s (Venus-Mars mission) and accelerates when it is attracted by the gravity of the planet. When we want to establish a circular orbit at a low altitude (for example 300 km above the surface) we have to decelerate the spacecraft; and it is much more economical to perform the velocity reduction by an aerocapture maneuver than by means of rocket propulsion ($\Delta v \approx 2.9$ to 3.8 km/s). The spacecraft enters the atmosphere of Mars on a hyperbolic trajectory with a velocity between 6.4 to 7.2 km/s; the pericenter of this orbit is just about 30 km higher than the surface of Mars. The ballistic entry maneuver requires very precise navigation: the entry angle has to be adjusted exactly in a way that, after braking, the spacecraft leaves again the atmosphere and climbs to the altitude of the 300 km where a small thrust maneuver establishes the circular velocity. The spacecraft needs heat protection to stand the loads of the braking maneuver. It is necessary that we use a heat shield with a large diameter, because the heat load reduces greatly when the vehicle decelerates in higher layers of the Mars atmosphere.



Aerodynamic landing on the surface. The landing vehicle is also exposed to aerodynamic heat loads during the landing maneuver, however, the stresses are much smaller now because the vehicle descends from a circular orbit at 300 km altitude and enters the Mars atmosphere with a relative velocity of just 3.3 km/s. The vehicle can use the same heat shield for the landing maneuver as for the aerocapture maneuver: heat protection is still necessary during the phase of aerodynamic braking, and, when the shield is large enough, the landing can be performed without parachutes. Impact velocity and impact angle for a ballistic reentry maneuver are mainly functions of the ballistic factor. A large shield with a high ballistic factor decelerates the vehicle down to a moderate velocity (for example 200 m/s for a ballistic factor of 15.7 m²/t). Shortly before the actual impact the Mars landing vehicle drops the shield, ignites its rocket motor and decelerates the velocity for a soft touch down on the surface of Mars.



Ascent to orbit. The Lunar Module of the American Apollo project was actually a two stage rocket: the first stage decelerated the velocity of the vehicle for landing on the moon; the astronauts returned to orbit with the propulsion of the command module (which was actually the second stage of the lander). The Lunar Module had a mass of 14.737 tons and an overall Δv capability of two times 2.3 km/s; and, when used as a two stage launcher, the vehicle had been able to reach an orbit from the surface of Mars. The velocity requirement for the ascent from the surface of Mars to a 300 km circular orbit amounts approximately to 3.6 km/s. However, the Lunar Module is not the ideal vehicle for the application as a Mars lander: the thrust of the first stage is too low (43 kN only), and the in-flight ignition of the engine of the second stage means a safety risk to the crew. The Mars lander must be a new construction, preferably a single stage vehicle with a stronger motor (for example with an initial thrust acceleration of a little more than 10 m/s^2). The vehicle needs a rocket stage that uses storable propellant (effective exhaust velocity $c = 3071 \text{ m/s}$), a controllable motor and a manned cabin as “payload”. When the cabin is made for two astronauts and has the same size and the same mass as the cabin of the Lunar Module (2 tons), the rocket stage must have a mass of 13 tons (charged with 11 tons usable propellant) and a rocket engine that generates a lift-off thrust of about 150 kN. We can use the Ciolkovskij equation to calculate a Δv capability of 4060 m/s for the vehicle. This includes some contingency and is enough for the final touch down on the surface after aerobraking ($\approx 300 \text{ m/s}$), the ascent back to orbit ($\approx 3600 \text{ m/s}$) and the rendezvous maneuver with the orbiter at an altitude of 300 km above the surface of Mars.



Example 8-16: The Mars landing vehicle is preferably a single stage launcher with approximately the same size as the Lunar Module of the American Apollo project. However, since Mars has a greater mass than the Moon, optimization of the lift-off thrust shows that the Mars lander requires a stronger motor than the Moon lander. A command module with a mass of two tons provides space for two astronauts during a time interval of about one week.

Departure for the earth. After having visited the surface of Mars the lander returns to orbit and performs a rendezvous maneuver with the orbiter. At an altitude of 300 km above the surface, for example, the spacecraft orbits with a circular velocity of $v_{circular} = 3404$ m/s. For a return flight to the earth the spacecraft has to increase its velocity to enter a hyperbolic escape orbit (therefore it has to accelerate to a velocity that is faster than the parabolic escape speed, $v_{parabolic} = 4814$ m/s). We can use the energy conservation law to find the relationship between the Δv (the velocity change on the orbit) and v_{∞} (the velocity in interplanetary space):

$$\Delta v = \sqrt{v_{\infty}^2 + 2 v_{circular}^2} - v_{circular} \quad (8 - 47)$$

For a velocity of $v_{\infty} = 2700$ m/s the equation above indicates that a tangential burn maneuver must raise the circular velocity by $\Delta v = 2115$ m/s. However, this value can be considered as the “minimum which is necessary”; the actual maneuver needs more Δv capacity because of the gravitational losses (approximately 100 m/s with contingency). Immediately after the burn maneuver the spacecraft departs from Mars with a relative velocity of 5519 m/s; it has still a velocity of 2700 m/s when it crosses the borders of the “sphere of gravitational influence” of the planet.

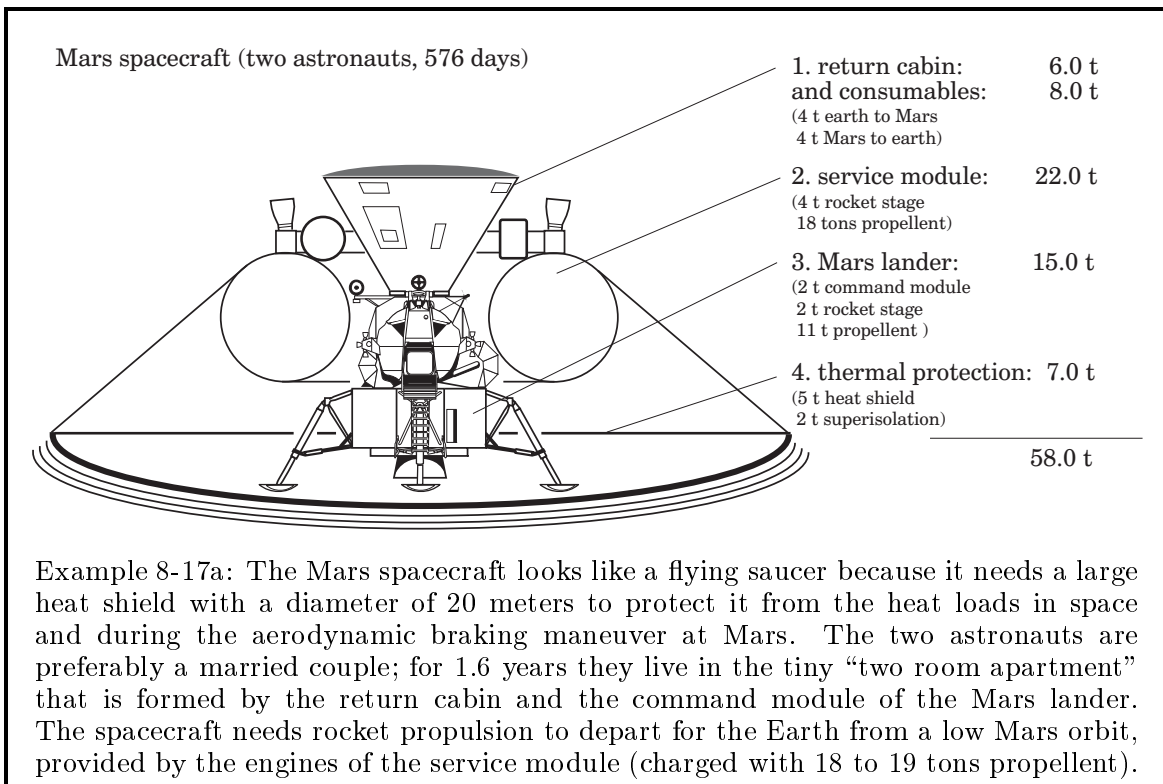
We assume that the flight operations at Mars involve that the spacecraft enters first a circular orbit around Mars before the lander descends to the surface and the orbiter stays on orbit. This mission concept has the disadvantage of a difficult rendezvous and docking maneuver between lander and orbiter, but the essential part of the spacecraft (that is necessary for the return to the earth) stays on orbit. The alternative mission concept is the landing of the entire spacecraft on the surface of Mars. Rendezvous and docking are avoided when the entire spacecraft lands on Mars; however, then the departure for the earth starts from the surface. The spacecraft needs for the return flight a Δv capability of approximately 5.8 km/s (3.6 km/s for the ascent to orbit and 2.2 km/s for the departure from orbit): either we have to bring a big two stage spacecraft with a large amount of propellant to the surface, or we have to produce the propellant for the return flight directly on Mars. The production of propellant from the “natural resources” of Mars via chemical processes operated by equipment with nuclear power or solar electric energy is apparently not completely impossible; however, this theoretical option is certainly not available technology now that could be used for the first manned Mars mission. We should bear in mind this theoretical option for later missions (for a permanent Mars base, for example).

The inclination of the orbit around Mars has to fit for the return flight to the Earth (otherwise the spacecraft has to perform an expensive inclination change maneuver). Therefore the spacecraft has to enter a low Mars orbit with the appropriate inclination against the ecliptic at the moment when it performs the aerocapture maneuver. However, the predetermined orbital inclination means another restriction concerning the possible landing areas on Mars. The mission has to land in one of the planes where there are no high mountain or deep valleys which could mean hazard to the landing maneuver (for example in “Mangala Vallis” 10° South and 150° West).

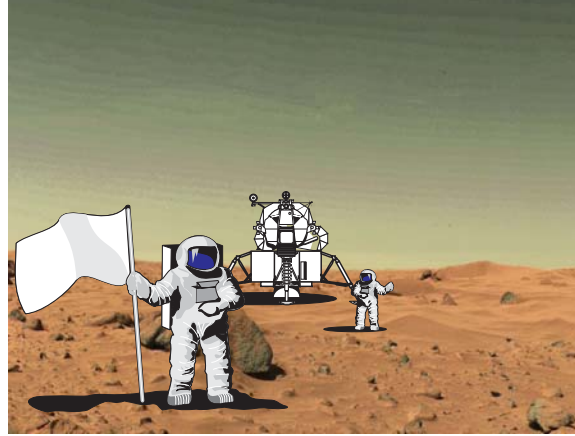
8.4.3 Expenditures for a Manned Mars Mission

The Mars spacecraft. Finally we are able to sketch out the example project of a manned Mars mission. The plausibility of the preliminary design increases when we take over approved components of the American Apollo project, even though these devices are not “available” anymore and not always “optimized” for our purpose.

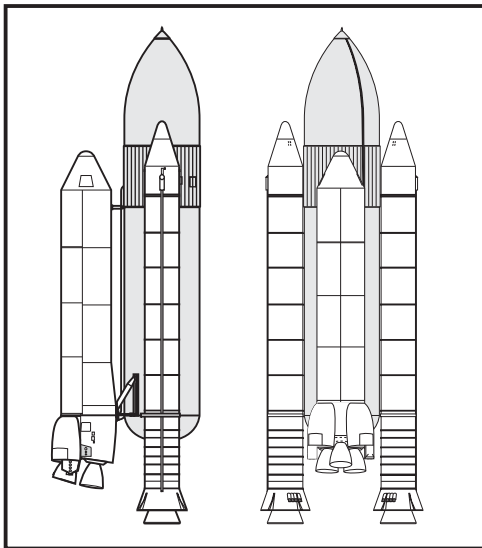
A certainly realistic assumption is that the budget is very tight. Therefore we consider that it is intended to send a crew of just two astronauts (preferably a married couple) on the fast “Venus-Mars Mission”. The mission lasts 576 days, where the crew stays for 7 days on the surface of Mars. The Mars-spacecraft consists mainly of four components: a return cabin similar to the Apollo capsule (6 tons), a service module (22 tons including 18 tons of usable propellant) a landing vehicle similar to the moon lander of the Apollo project (15 tons) and a large heat shield (7 tons). The life support system consumes every day 6.1 kg of material for every crew member: the spacecraft has to carry 7027 kg of consumables (or better 8 tons when we consider some contingency). More than half of this mass is drinking water (accurately 55%), the fresh water as well as the waste water is used for a radiation shielding to protect the crew against the radiation in space (the water for washing is continuously filtered and recycled). The spacecraft that departs from the Earth heading for Mars via Venus has a mass of 58 tons. About 5 tons of the material are consumed when the spacecraft arrives at Mars (where a part of this material can be discharged). Since the Venus swingby maneuver needs no propellant, a mass of nearly 58 tons arrives at Mars.



Example 8-17b: The two astronauts stay just for about a week on the surface of Mars; and during that short time interval they live in the small command module of their landing vehicle. They have just time to plant the flag, take pictures, collect sample material and visit interesting objects in the near vicinity of the landing place. The scenario resembles the moon landing mission of the American Apollo project, however, the Mars lander will use an aerodynamic braking maneuver for descending to the surface of Mars.



Injection of the mission. Since the Mars spacecraft is obviously too big for the payload bay of the American space shuttle (or any other available launcher), it has to be transported in pieces to a low earth orbit and assembled in space. The launch of the components of the spacecraft, their assembly and check-out can be done several months before the actual injection of the mission into interplanetary space

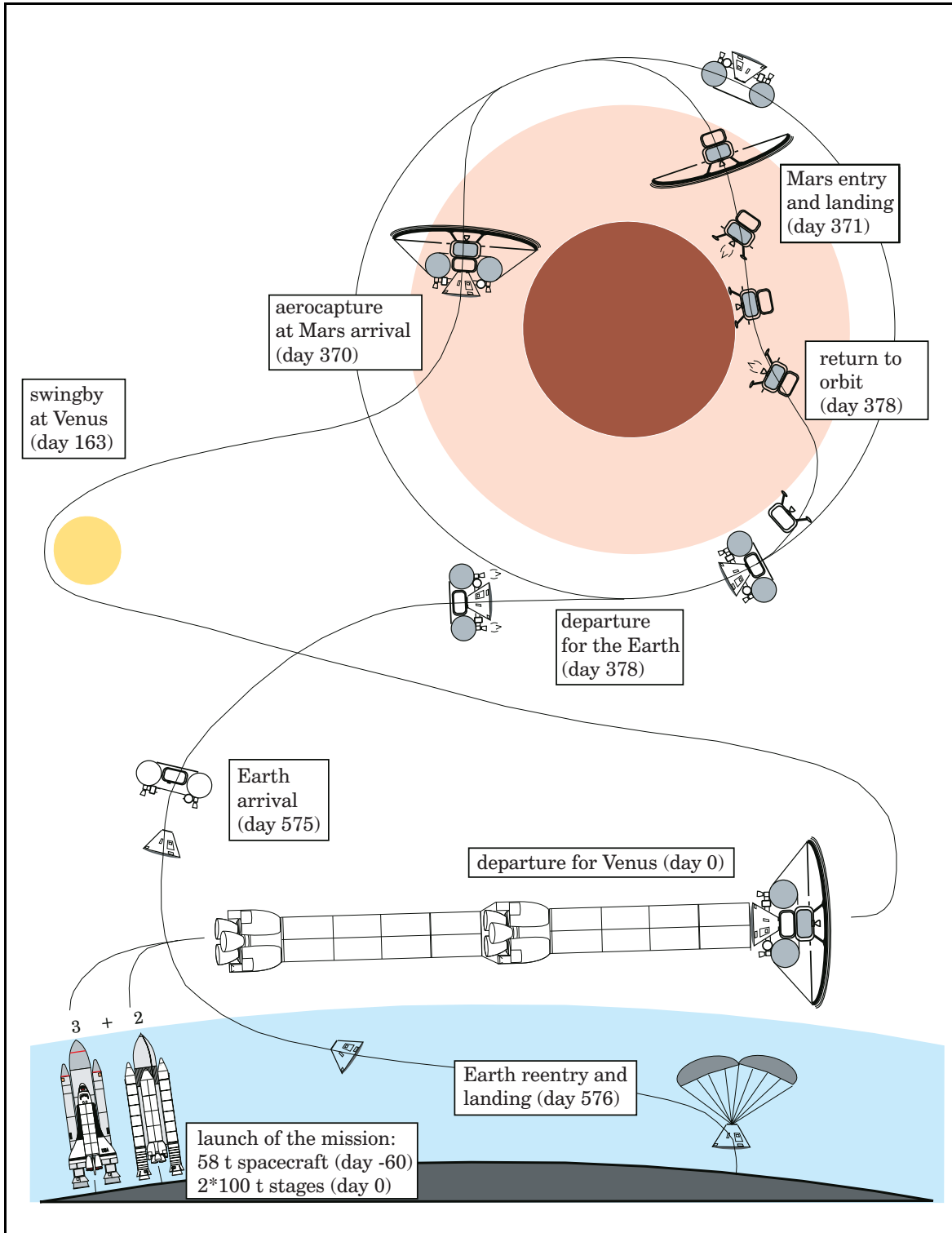


(before April 2nd 2017). Three shuttle flights are enough to bring the mass of 58 tons into a 200 km low earth orbit. The actual problem is the injection of the assembled spacecraft into interplanetary space: a “shuttle derived vehicle” can provide a solution to this problem. The proposal is to replace the reusable orbiter of the space shuttle system by an expendable cargo container for a payload mass of 77 tons. When we equip the expendable cargo shuttle with three SSME engines and design the payload container as a tank (with adapter) for LH₂/LOX propellant, then we are able to bring a complete filled-up high performance rocket stage to orbit (of about 100 tons mass).

The spacecraft has to depart from the earth with a velocity of $v_{\infty} \approx 4.6$ km/s; and, since the vehicle moves with a velocity of 7785 m/s on a 200 km low earth orbit, it has to raise its velocity by more than 4147 m/s (equation 8-47). Two “cargo shuttles” are necessary for the injection of the mission. Both vehicles are launched on April 2nd 2017, either simultaneously or with a short time interval in between (liquid hydrogen is not storable for an extended time period). In space the two stages are connected with the Mars spacecraft in a way that the flight configuration becomes a two stage rocket with an initial mass of approximately 258 tons. The first stage of the vehicle, ignited at the right moment, increases the velocity by 1511 m/s and brings the spacecraft on an elliptic orbit around the earth. With $\Delta v = 2830$ m/s, the second stage has enough capacity to bring the spacecraft on its way to Venus and Mars.

Mission execution. The mission is launched on April 2nd 2017 on a Venus-Mars swingby trajectory. For nearly a year the crew lives in the return cabin and the command module of the lander; their “two room apartment” has just a volume of 12 m^3 . Video entertainment, physical training and other activities serve as a way of passing the time. During the time of the flight to planet Venus the spacecraft approaches the sun as close as 0.57 AU, thermal control is provided via a super-isolation on the surface of the heat shield. Finally, on April 7th 2018, the spacecraft arrives at Mars. The crew checks out all subsystems of the spacecraft. In case of a serious malfunction the mission is terminated and the spacecraft returns back to the Earth directly with a propulsive maneuver at Mars. However, the spacecraft executes an aerocapture maneuver when all systems work normally and attains a circular orbit at 300 km altitude above the surface. The large heat shield has a diameter of 20 m; its ballistic factor ($6.5 \text{ m}^2/\text{ton}$) is high enough that the heat load of the braking maneuver can be emitted (radiation cooling). The heat shield is still usable when the spacecraft arrives at an altitude of 300 km above the surface of Mars, where it establishes circular velocity with a small impulse provided by its maneuver engines ($\Delta v = 90 \text{ m/s}$). Then the crew moves into the Mars lander; the vehicle separates from the capsule and initiates the aerodynamic landing with a small retro-thrust. The heat load of the landing maneuver is lower than the heat load of the aerocapture maneuver, because the smaller mass (only the lander) increases the ballistic factor for the aerobraking maneuver ($15.7 \text{ m}^2/\text{ton}$). Approximately one kilometer above the surface the vehicle falls with a velocity of 200 m/s. The crew ignites the engine of the lander, drops the heat shield and selects an appropriate place for the touch down on the surface of Mars.

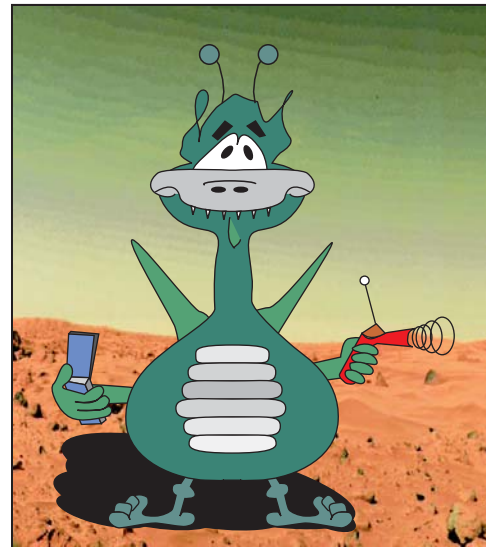
The astronauts stay for 7 days on the surface of Mars; during this time interval they live in the command module of the lander, but they get out of the vehicle to visit the surroundings of the landing area and to collect sample material. On April 14th 2018 the crew ignites the engine of the lander a second time for the return flight to orbit. The touch down on the surface with rocket propulsion has consumed not more than 2 tons of the propellant of the lander (equivalent to a Δv capability of 440 m/s). The ascent requires a Δv capability of 3610 m/s; and with the remaining 9 tons of propellant the lander is able to reach orbit and to perform a rendezvous maneuver with the orbiter. Then the crew discharges the empty rocket stage of the lander (however, the command module of the lander remains docked on the return cabin). The Mars spacecraft is smaller now because the heat shield and the rocket stage of the lander are missing (-20 tons). When we assume that from the initial 8 tons of consumables 4 tons suffice for the return flight (-4 tons), then the spacecraft weights all together 34 tons after the preparation for return. 18 tons of storable propellant provide a Δv capacity of 2.3 km/s to depart from the orbit. The spacecraft waits on orbit for the right time to ignite its engines for the return flight. 199 days later, on October 30 2018, it approaches the earth. The crew separates the service module and the command module of the lander from the return cabin and discharges the rest of the consumables. Finally the return cabin (6 tons) enters the earth’s atmosphere, either to perform a direct landing maneuver (with parachute watering on the ocean) or an aerocapture maneuver (with docking on the international space station).



Example 8-17c: The timeline for the Venus-Mars project study shows the chronological order of the most important events of the mission. The Mars spacecraft is launched about two months before the actual starting day and assembled in space; then the spacecraft is brought on a Venus-Mars swingby trajectory by two upper stages of the cargo shuttle. Mars is encountered after a flight time of nearly one year; then the crew stays for one week on Mars. The return flight back to the Earth lasts 198 days.

Mission cost. Our Mars mission has to be compared with other big space projects in order to estimate the costs. The comparison suggests that this mission is probably not more expensive but cheaper than the American Apollo project and the international space station ISS: the components which have to be constructed for the Mars mission are the “return cabin”, the “service module” and the “lander” (these components are in size and function similar to the devices of the Apollo project); only the “heat shield” for landing on Mars needs a completely new development. However, it is not necessary to build a large space launcher like America’s Saturn-5; the development of the cargo shuttle is certainly much cheaper (actually it requires just the development of a tank, the engines are available). The launch of our Mars mission requires three shuttle flights and two cargo shuttle flights: a mass of 258 tons has to be transported to a low earth orbit (and, in contrast to the space station ISS, this initial mass in low earth orbit consists mainly of propellant and consumables). Taking everything together we may conclude that the manned Mars mission is not necessarily more expensive than the Apollo project or the international space station.

Objective of the mission. The search for extraterrestrial life is certainly the most important reason for astronautical activities in deep space. Naturally, we do not expect to find an alien on Mars like the “martian” in the picture below; however, it is indeed possible that the material on the surface of Mars contains fossils. Even the discovery of fossil microorganisms would be an extraordinarily exciting result of the mission that would help finding an answer to the fundamental question about the origin of life. Some parts of the surface of Mars were apparently covered with water in a former period of its existence, and maybe the planet has lost its water in the consequence of a cosmic catastrophe. Today the surface of the Earth would be quite similar to the surface of Mars when the Earth had lost all its water; for example at the end of the age of the dinosaurs (about 60 to 70 million years ago). An unmanned space probe would probably not be able to discover the bones of the dinosaurs.



Political environment for a manned Mars mission. The technological basis for a manned Mars mission is already available for many years (such a mission could be performed either in the way that has been outlined in this book or in another way). Despite of its possibility the mission will not be performed within the next decades: the political basis that was present during the time of the American Apollo project (the “cold war”) is not present anymore. Even though we are happy that the cold war is over we regret that the political basis for scientific space activities has turned out to be so bad without it. Now we are looking forward to a peaceful political situation with a better environment for a manned Mars mission and, during that time, it is the important task for space engineers to show to the public how easy it would be.

9. Low Thrust

The expression “low thrust” refers to a variety of quite different propulsion concepts. They all have in common that the propulsion energy comes from the sun (or occasionally from a nuclear power generator, anyway the energy is not stored in chemical form in the propellant), and that the engines are too weak to propel heavy space launchers. Low thrust propulsion is limited to the application in satellites and space probes. Today, low thrust propulsion systems (resistojets, arcjets, ion-engines, magneto-electric plasma thrusters, and solar sails) live still in the shadow, but actually they offer a great potential for future astronautics: the reason is that the working fluid (propellant) is used with a higher efficiency when the energy comes from outside. Thus, when these unconventional engines are applied to satellite station keeping, they have the capability to increase the service life of earth satellites substantially; or, when they are applied as the main propulsion system to deep space probes, they can make planetary missions possible which are too difficult for conventional chemical propulsion. The ninth chapter of this book treats the great potential of low thrust propulsion and the difficult problems which are inherently involved with it.

There exists a variety of electric thrusters which utilize different working principles; the different working principles involve operational advantages and disadvantages. All these thrusters require electric energy, in space provided by large solar arrays or by radio-isotope generators. We will have a look at available thrusters and power supply systems in the first section of this chapter. Engines and power supply system necessary to propel an electric space probe are usually heavy in comparison with the overall propellant mass; and this leads to some apparently paradox effects: the theory of rocket propulsion applied to the optimization of electrically propelled space missions demonstrates that the staging principle deviates substantially from what we might expect knowing conventional rocket engines. The second section of this chapter analyzes the correct application of the staging principle for a “comet nucleus sample return” mission and for a “multi asteroid rendezvous and flyby” mission. Finally, the third section of this chapter treats the optimization of sailcraft propulsion: a small spacecraft, equipped with huge lightweight reflectors, could also utilize exclusively the pressure of sunlight as a propulsive force: the great advantage is that such a solar sailcraft would not require any propellant to perform its planetary mission.

The fundamental textbook on electric propulsion systems was written by E. Stuhlinger [“Ion Propulsion for Space Flight”, McGraw-Hill Book Company, New York, 1964]; however, also the books of R.G. Jahn [“Physics of Electric Propulsion”, McGraw-Hill Book Company, New York, 1968] and G.P. Brewer [“Ion Propulsion Technology and Applications”, Gordon and Breach Science Publishers, New York, 1970] are recommendable. Considerable advances have been attained since the time when these books were published, documented for example in many proceedings and papers of the annual “AIAA/DGLR/JSASS International Electric Propulsion Conference”. The technology of solar space sailing is summarized in the book of J.L. Wright [“Space Sailing”, Gordon and Breach Science Publishers, Philadelphia, 1992].

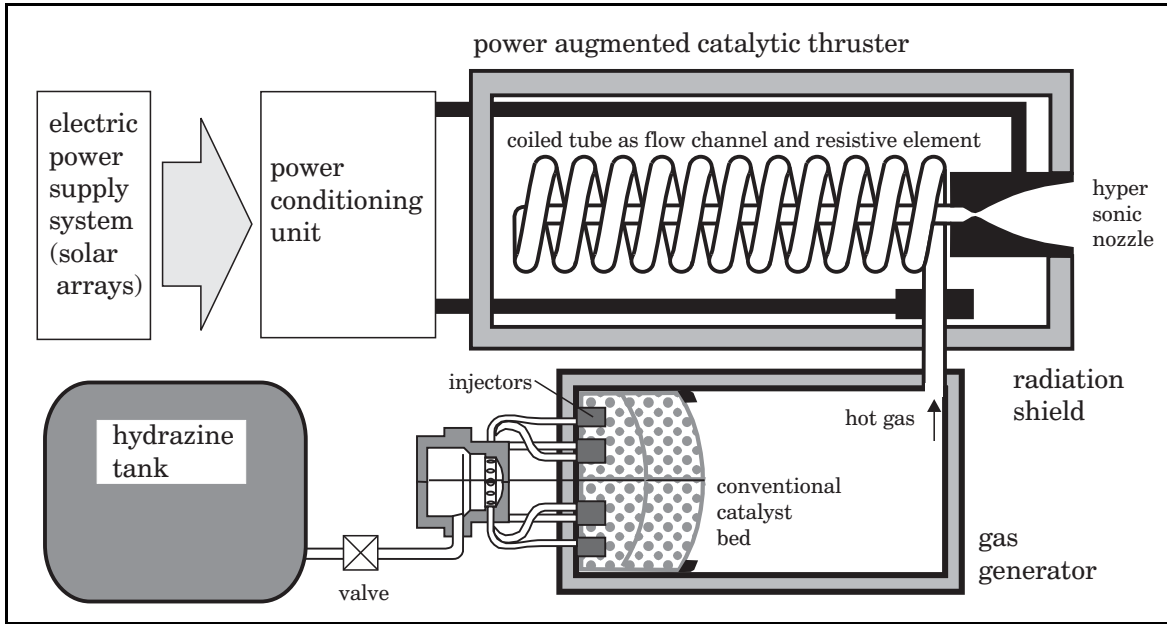
9.1. Low Thrust Propulsion Systems

Theoretically, there are many different ways how we can utilize the energy of the sun or the energy of a nuclear power generator for space propulsion. However, let us now concentrate our attention on really practicable ways: considerable industrial efforts have already been made in developing several different types of electric engines; and also the use of solar sails for some deep space missions seems to be promising.

9.1.1 Thermo-electric Engines

Working principle. Conventional chemical engines transform the heat energy that is released by the combustion of propellant via an expansion process in the rocket nozzle into kinetic energy. The hotter the combustion process the more kinetic energy can be generated. The idea is now to put more energy with a resistor or a light arc into the combustion chamber in order to improve the exhaust velocity of the engine, assuming that there is enough electric energy available in the spacecraft. Thermo-electric engines work like conventional chemical rocket engines; the only exception is that now electricity heats up the gas in the chamber. Expansion and exhaustion take place in a conventional hypersonic nozzle. It can be the case that the propellant contains no chemical energy and electricity is the only energy source; but it can also be the case that the propellant contains chemical energy and the combustion gas is additionally heated up by electricity. In practice, these engines can be used for station keeping of earth satellites (switching the energy for broadcasting every now and then for short time periods to the thrusters). Promising is also the application in manned spacecraft utilizing waste water or carbon dioxide as propellant for control maneuvers.

Resistojets. The use of an electric resistor is the easiest way to utilize electric energy for space propulsion: the energy is simply transferred to the working fluid via direct contact with an electrically heated coil of wire. Special high-temperature metal alloys (rhenium, platinum, molybdenum or tungsten) allow surface temperatures of maybe up to 2400°C. The exhaust velocity of the engine is a square root function of the specific chamber enthalpy; and, considering that the chamber temperature is limited, hydrogen gas with its low molecular mass would be the best propellant for a resistojet. However, other propellents are usually preferred because it is too difficult to store hydrogen in space. The use of a monopropellant (such as hydrazine) reduces the demand for electric energy, because the chemical reaction of the monopropellant provides additional heat energy. In practice it has been tried to improve the specific impulse of small hydrazine thrusters, for example from 230 seconds to 320 seconds. The problem is that the resistor has to be heated before the thrusting and that the hot decomposed gas has to stay for a certain time interval in contact with the resistor. Electric energy is anyway not cheap in space; the efficiency of resistojets ranges between 65% and 85%, depending on the propellant, on the pulsing time and on some other construction details. Resistojets have been developed for thrust levels between 10 mN and 25 mN, some versions have already flown in several experimental satellites.



Arcjets. A light arc is the alternative to a resistor for heating up the chamber gas. A light arc that flickers between the cathode (located in the center of the chamber) and the anode (the coaxial ring-shaped entry region of the hypersonic nozzle) transfers the electric energy directly to the chamber gas. Direct heating of the gas avoids the temperature limitation imposed by the melting point of the resistor material (the temperatures in a light arc range between 10000°C and 22000°C). However, the temperature of arcjet electrodes is limited to about 2700°C , when the electrode material is tungsten alloy with boron nitride as insulator; and in practice the arcjet introduces some other even more difficult problems: just a minor percentage of the gas experiences direct arc heating, the other part is heated indirectly by mixing with the hot gas, by heat radiation from the arc and by contact with the hot electrodes. The efficiency of an arcjet is usually low (below 50%); heat energy radiated to space and electric energy necessary to ionize the gas is lost for propulsion.

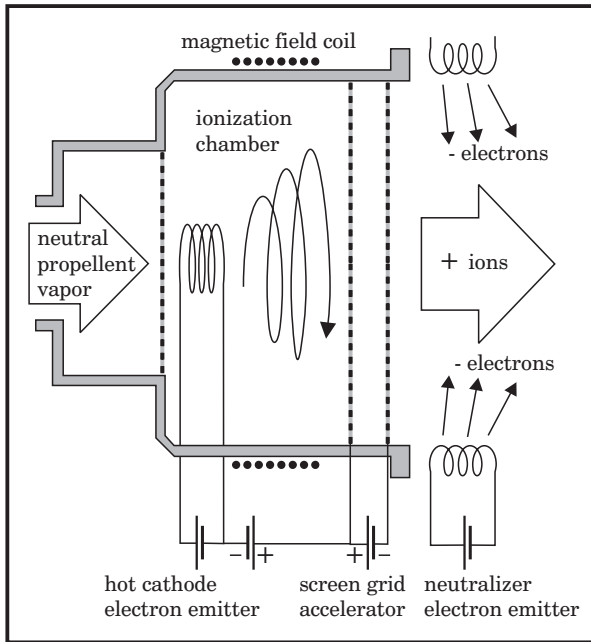
Arcjets need comparatively high electric currents which cause heavy conductors between the current transformer and the thruster; and the ignition of the arc requires other voltages and other currents than the steady state operation. The ignition of the arc is very destructive for the thruster: pulsed mode operation is difficult because the erosion of the electrodes is much higher (up to 100 times) when the thruster is operated in pulsed mode. Experimental thrusters have been developed which combine good performance (1400 seconds specific impulse) with high efficiency (between 60% and 85%), but the erosion is not acceptable then. Practical arcjet designs resemble the construction of resistojets: hydrazine monopropellant is catalytically decomposed inside a small gas generator and then conveyed to a tiny thrust chamber where a light arc (or a resistor) transforms the electricity into heat. For example, a small 100 mN hydrazine arcjet thruster (named HAJ1) for satellite stabilization is offered by the German space industry; the model is specified to operate with a specific impulse of 530 seconds at a power consumption of 250 W and a chamber pressure of 18 bar.

9.1.2 Ion Engines

Working principle. Ion engines use an electric field to accelerate the propellant, an ionized gas with an extremely low density (typical propellents for ion engines are mercury, xenon or cesium). In practice there are three different ways how ion thrusters can generate ions from the propellant: first, by the bombardment of the gas with energetic electrons; second, by direct contact of the gas with a hot metal electrode; and third, by the stimulation of a plasma with a radio-frequency field. The positively charged ions are extracted from the ionization chamber and accelerated in an electric field between two (or three) screen grids. Finally, a sub-assembly emits electrons to the ion beam, to neutralize it and to prevent a charging of the spacecraft.

The ion thrusters operate with thrust levels between 10 mN and 200 mN, typically, with electric efficiencies between 70% and 90%. The lower thrust versions are usually designed for earth satellite station keeping and the stronger versions for deep space missions. Typical specific impulses range between 3000 seconds and 6000 seconds. The working principle would make it possible to generate easily even higher specific impulses (simply by increasing the grid voltage); but then power consumption gets usually too high for the desired thrust level. Thrusters, power supply systems and electric converters are now heavy components of the propulsion system, compared with the weight of the propellant. Ion engines are the most promising low thrust propulsion systems; however, they suffer the problem of a deterioration with the operation time and have a limited lifespan. Experimental thrusters have flown several times on satellite platforms with the main objective to demonstrate the desired service lifespan (desired are at least two years of constant operation); and America's electrically propelled "Deep Space 1" probe flew by comet Borelly on September 22, 2001.

Electron bombardment thrusters. Most existing ion engines belong to the subclass of electron bombardment thrusters, also called Kaufman thrusters (named after H.R.Kaufman who developed in 1960 the first operational engine). The propellant, usually liquid mercury or xenon, is vaporized and conveyed to the ionization chamber where it is electrically activated to form a very low density plasma (≈ 0.001 torr). The ionization is initiated by the collision of high energetic "primary" electrons with neutral gas atoms. The electrons come either from a thermal filament or from a hot cathode discharge chamber. An electric current in a coil surrounding the chamber establishes a magnetic field in order to influence the motion of the electrons and thus to increase the probability of ionization by impact: lower energetic secondary electrons (byproducts of the ionization) contribute with about 50% to the ion production. It is desired that the plasma is nearly completely ionized (accumulation of neutral atoms can cause arcing and voltage breakdown) and that the plasma contains just singly charged ions. A strong electric field between the grid electrodes at the exit of the ionization chamber extracts the positively charged ions from the plasma and keeps the electrons in the chamber. The ions enter the field through the holes in the grids and accelerate by running through a potential between the grids (1000 V to 2000 V). The neutralizer device, another electron source, feeds electrons into the exhaust beam.



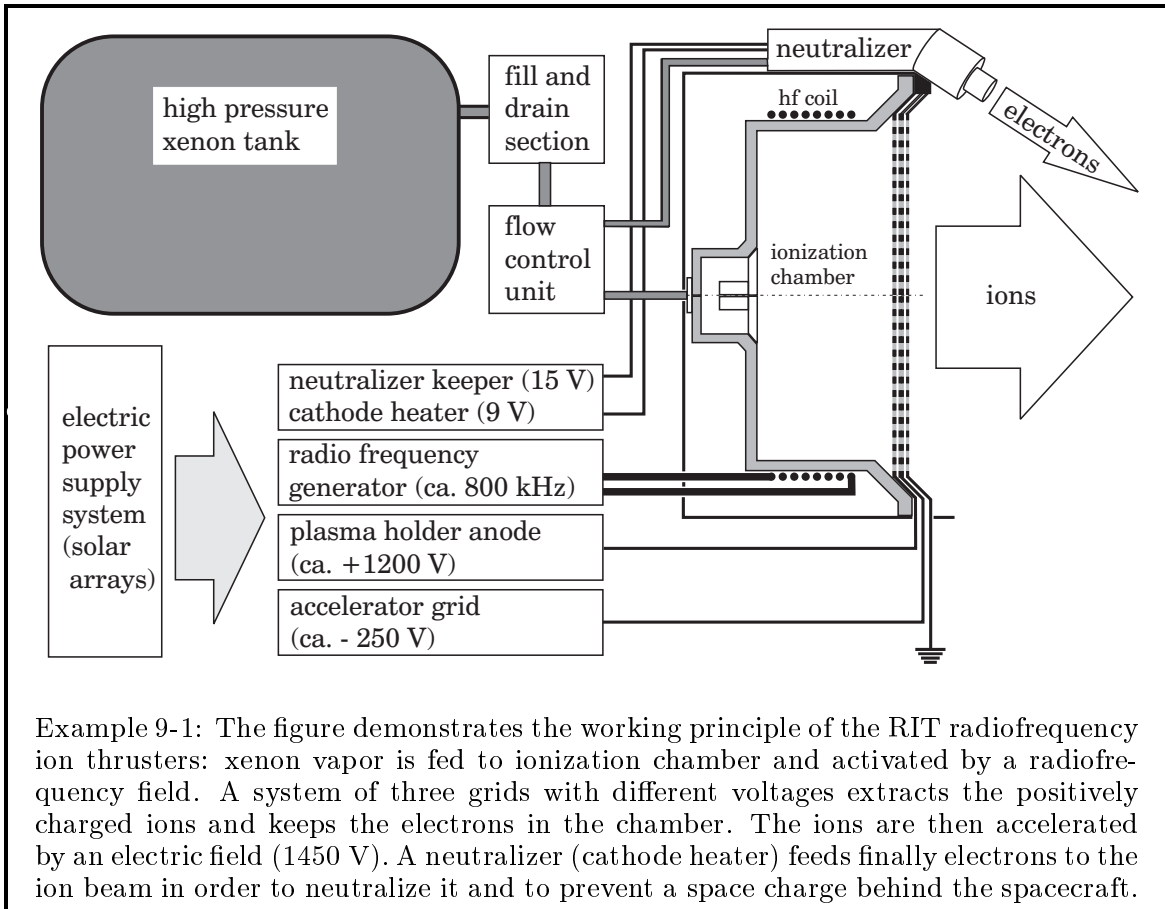
Ion engines have a limited service life: a small number of the ions extracted from the plasma is not expelled but hits the grids and causes an erosion of the grids (“spluttering”). This constant ion bombardment can be diminished by an appropriate “ion optic” (electro-magnetic fields). However, analytical and computational models for analyzing the quite complicated motion of the ions are still not really accurate; and thus the best extractor geometry with low spluttering rate must be determined by expensive long-time experiments. Special advanced grid materials can be a key to success (metal or conductive fiber material).

Today, the inert gas xenon is the preferred propellant for ion engines. Other inert gases have also been tested, but their suitability is not as good. Xenon has a high atomic mass (131.3 g/mol) and a comparatively low ionization potential (12.1 eV); and the exhaust plumes are not dangerous for the spacecraft (xenon cannot chemically react with surfaces or pollute instruments). The gas can be stored in its liquid phase in pressurized tanks (or close to the critical state, 500 kg/m³ at 60 bar and 35°C, a heater must be available to avoid freezing at lower temperatures). The source substance for production is air (xenon is a minor constituent of the atmosphere), but the production process is quite expensive. Earlier versions of ion thrusters used mercury vapor as propellant (atomic mass 200.6 g/mol, ionization potential 10.4 eV). Mercury is a liquid metal and several thousand times cheaper than xenon, but this propellant was abandoned because it was considered as dangerous for the spacecraft. Other early versions experimented with cesium vapor (atomic mass 132.9 g/mol, suitable because of its low ionization potential 3.9 eV), but this propellant ignites with air and reacts explosively with water and thus it is even more dangerous than mercury.

Ion contact thrusters. Ions can also be produced directly by the contact of mercury or cesium vapor with a hot tungsten electrode (“Langmuir-effect”). So-called ion contact thrusters (or Stuhlinger thrusters) work similar like other ion engines, except for using a different ionization method: low pressure mercury or cesium vapor flows through a hot porous tungsten ionizer which ionizes over 99% of the atoms. For contact ionization the metal tungsten must be hotter than a certain temperature (for example hotter than 1100°C); but in practice higher ionization potentials are selected to avoid the presence of neutral atoms which cause an erosion of the thruster. Heat radiation from the hot ionizer reduces the electric efficiency of the thruster. Experimental ion contact thrusters have been developed for different thrust levels, but it is apparently difficult to obtain a reasonable service life for this type of thruster.

Radio-frequency ion thrusters. In contrast to Kaufman thrusters and Stuhlinger thrusters which use hot electrodes for the initiation and activation of the plasma, the RIT-engine (or Loeb-thruster) uses an electromagnetic field with a frequency between 500 kHz and 1 MHz. The induction coil for the field is placed around the ionization chamber. The stimulation of the gas with a radio frequency field requires more energy but it avoids the lifespan problems of the electrodes and allows the generation of a plasma with a higher density. The thrusters RIT 10, RIT 15 and RIT 35 (the numbers refer to the chamber diameter) have been developed in a cooperation between the ESA, the German space industry and the university of Giessen; the thruster RIT-10 with 10 mN thrust (for satellite station keeping) was tested in 1992 on board of the retrievable European experimental satellite platform EURECA.

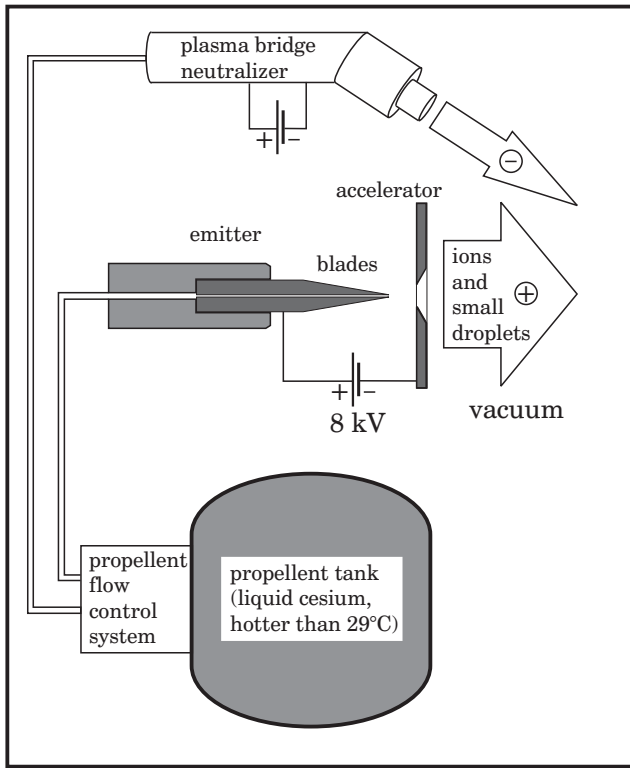
Presently, the RIT-10 device is offered to satellite manufacturers, specified for 15 mN thrust, 350 W electric power consumption and 3060 seconds specific impulse. Thus, the mass flow rate amounts only 1.8 mg per hour (xenon is used as propellant). The thruster weighs 1.8 kg and is with a nozzle diameter of 16 cm not really small. Until now practical problems have prevented the application in commercial satellites (incompatibilities in the electric interfaces, electro-magnetic interference problems, reliability concerns and so on), even though a reasonable lifetime has been experimentally demonstrated and the low propellant consumption rate is certainly appreciated.



Example 9-1: The figure demonstrates the working principle of the RIT radiofrequency ion thrusters: xenon vapor is fed to ionization chamber and activated by a radiofrequency field. A system of three grids with different voltages extracts the positively charged ions and keeps the electrons in the chamber. The ions are then accelerated by an electric field (1450 V). A neutralizer (cathode heater) feeds finally electrons to the ion beam in order to neutralize it and to prevent a space charge behind the spacecraft.

Field emission electric propulsion. So-called FEEP systems fall also into the category of electrostatic ion thrusters, however, their method of ion creation differs substantially from the ionization method of other ion engines. FEEP systems extract ions directly from the liquid phase of a conductive propellant by the action of a strong electric field (stronger than 3 kV). Under the influence of the strong electric field, the free surface of a liquid conductor is distorted into a series of “Taylor cones”, and ions depart from the tip of the cones apparently without intermediate gaseous phase.

The ion emitter is a needle with an extremely thin capillary tube inside ($\approx 10 \mu\text{m}$). In order to create more ions for higher thrust, an array of stacked needles can be used, or even better, the emitter is a wedge-shaped metal plate with a thin slit in its blade.

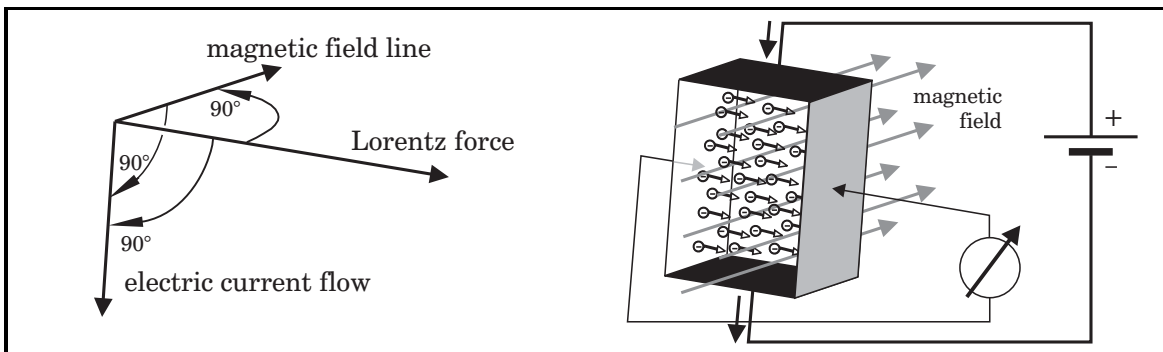


In a practical European design (a development in the French and Italian space industry, sponsored by ESA) the emitter is made of two perfectly polished metal blades (molybdenum) tightly clamped together, one of which is polluted with sputter-deposited metal spots (nickel) in order to give the desired slit width (“double knife edges” with a length of 8 cm and a slit width of $1.1 \mu\text{m}$). Capillary forces drive the liquid propellant through the slit; 8 kV between the front end of the emitter and the accelerator electrode extract and accelerate the ions. Then, as in all ion engines, the ion beam is discharged by a neutralizer device, to prevent that charged particles are later attracted by the spacecraft. The device develops 2.4 mN thrust.

The liquid metal cesium is the favorable propellant for the FEEP thruster, because it has an excellent wetting capability, a low ionization potential and a high atomic mass. The advantages of FEEP thrusters in comparison with other ion engines are the simple construction and a comparatively high electric power efficiency. A disadvantage is that the electric field for ion creation must be strong (high voltages are necessary). The high accelerator voltage gives the FEEP thruster a high specific impulse but it requires also comparatively much electric power per unit of thrust (for example 55 W/mN for a specific impulse of 6000 s). Even though several emitters can readily be clustered to form a stronger thruster, the demand for electric energy becomes excessive then. For satellite station keeping the FEEP thruster can be superior to other ion engines because it can instantaneously be switched on and off; precise attitude maneuvers with extremely low thrust are typical applications for FEEP thrusters.

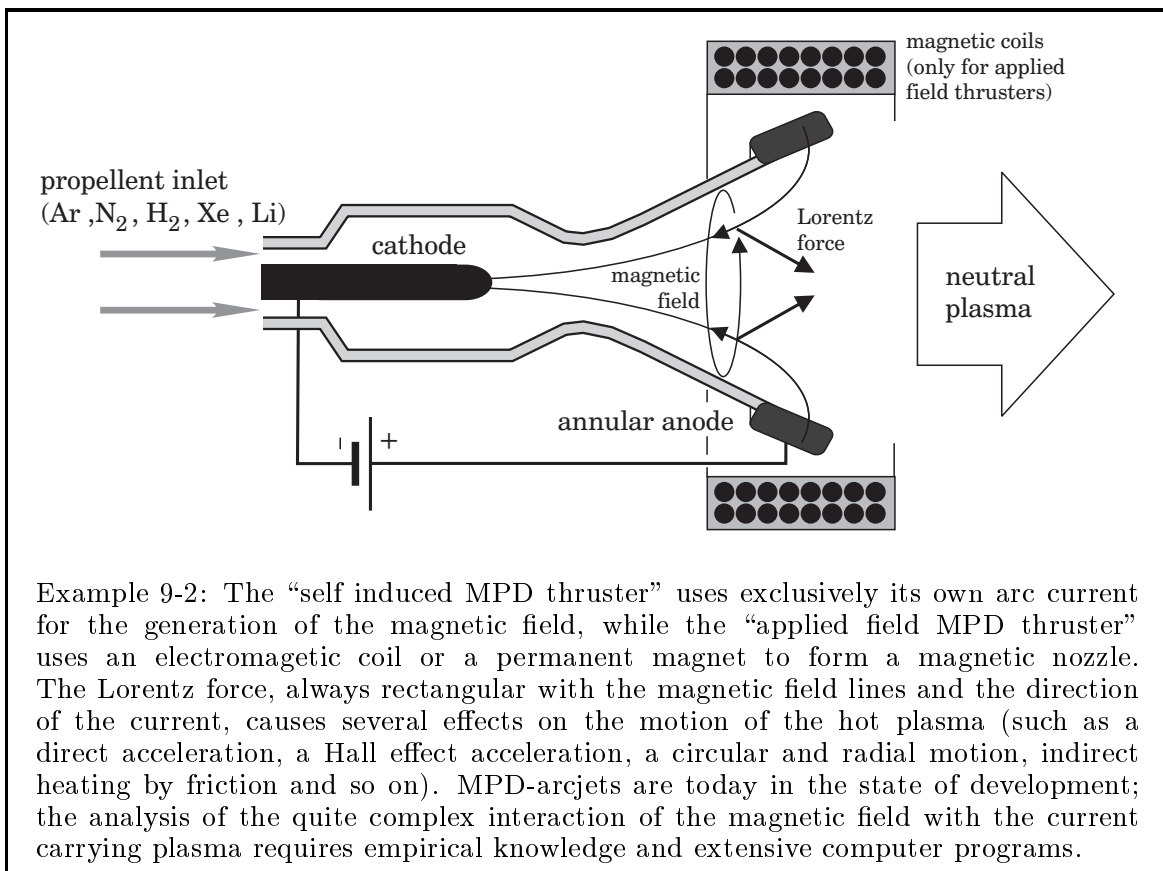
9.1.3 Magneto-plasma Dynamic Thrusters

Working principle. In contrast to ion engines which use an electric field for the acceleration of charged propellant, MPD thrusters use magnetic forces that act on a neutral plasma. A gas becomes an electrically conductive plasma when about 10% of its atoms are ionized, even though the plasma is still an essentially neutral gas (it contains ions, electrons and neutral molecules). Now an electric current can run through the plasma, and when the current carrying conductor is exposed to a magnetic field, the “Lorentz-force” (named after H.A.Lorentz, 1853-1928) appears which accelerates the conductive plasma (the Lorentz-force is proportional to current intensity and magnetic flux density, it aims in a direction rectangular to the current and rectangular to the magnetic field). The “Hall-effect” (E.H.Hall, 1855-1938) is the phenomenon that inside a magnetic field an electric potential can be measured between the side walls of an electric conductor, as the consequence of magnetism. The magnetic field can be “self induced”, referring to the fact that a current carrying conductor itself creates a magnetic field. The physical driving mechanism that propels magneto-plasma dynamic MPD thrusters is therefore under some aspects the same as the driving mechanism of conventional electro-motors: electromagnetism.



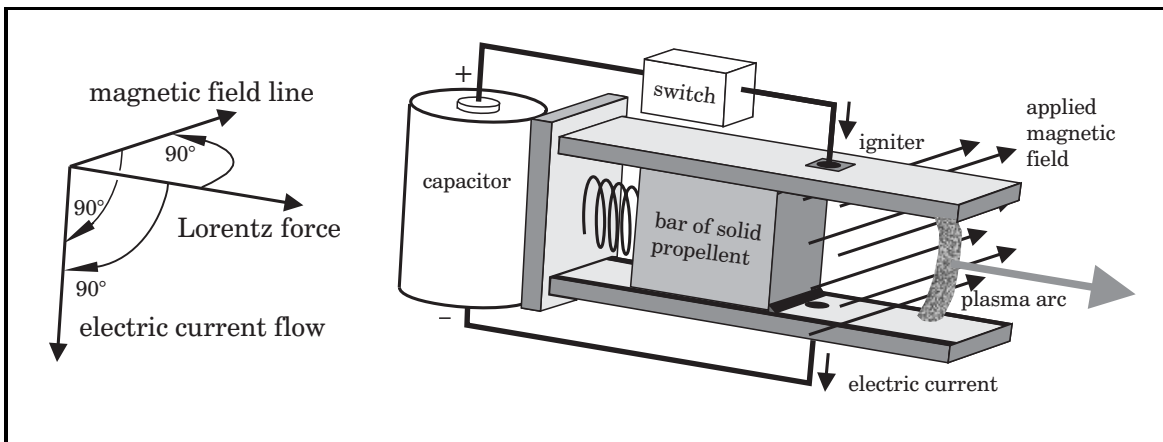
The MPD thrusters constitute another important class of electric engines; and, since about 1964, a wide variety of different types has been examined in conceptual studies, laboratory models, or industrial developments. Many universities and space agencies all over the world are presently experimenting with different concepts. Names and classification are quite confusing. Some thrusters operate in pulsed mode utilizing solid propellant, others operate in steady state; several types have flown successfully on American and Russian satellites. All these thrusters have in common that they generate thrust using a magnetic field that acts on a neutral plasma. Even though many thrusters have not reached the state of an advanced development jet, their operation principle is quite promising: while ion engines must use heavy ions and large beam widths because the charged particles in the electric field repel each other mutually, MPD thrusters can use light weight propellant particles and offer the principal advantage of a high specific impulse in combination with a high thrust capability (provided that a sufficient amount of electric energy is available in the spacecraft). Let us now have a look at those MPD thruster types which have been used for the station keeping of satellites, or which at least have undergone a concrete development.

MPD-arcjets. A thermal arcjet which incorporates a strong magnetic field into the nozzle for accelerating the working fluid is theoretically the easiest way to realize a “magneto-plasma-dynamic thruster” or “Lorentz-force accelerator”. Similar to a thermal arcjet, the MPD-arcjet creates a plasma (a hot electrically activated gas) by burning a light arc between the cathode in the center of the thrust chamber and the coaxial ring-shaped anode. However, this time the annular anode is placed at the exit of the hypersonic nozzle. It does not contribute much to the generation of thrust when the gas flow is heated in the divergent part of the nozzle; but there the gas is an electrically conductive plasma, and the current that runs from the anode to the central cathode is surrounded by a circular magnetic field (as every current carrying conductor). The Lorentz force, as the result of the interaction of the magnetic field with the current in the plasma, accelerates the plasma and contributes to the thrust generation.



MPD-arcjets have not been used in any spacecraft until now, but worldwide there is an ambitious research on this thruster type, because theoretically the concept allows high thrust densities (small engines with high thrust and low fuel consumption). Practical problems with this thruster type are excessive erosion of the hot electrodes and a comparatively low power efficiency. However, high thrust in combination with sparing fuel consumption is always at the expense of an enormous demand for power. Due to the non-availability of large power plants in space, it can be the case that high thrust MPD thrusters continue to be the space propulsion of the future.

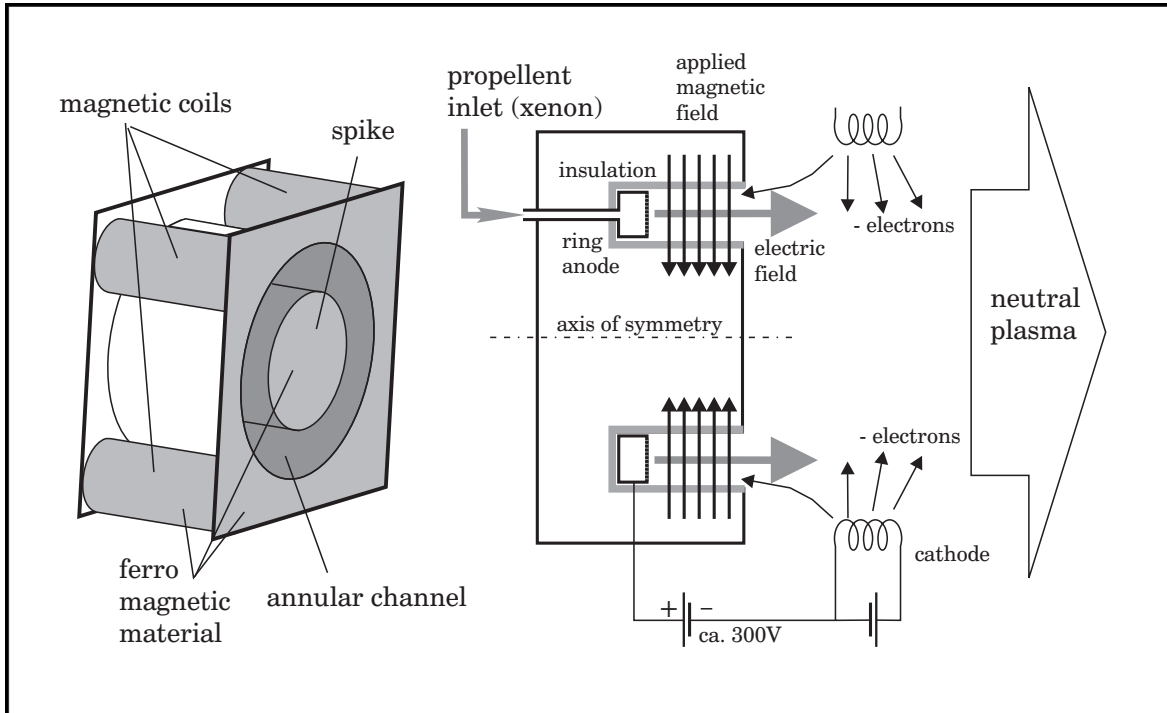
Solid pulsed plasma thrusters. In practice, the operation principle of the solid propellant plasma thruster is more promising for building a satellite control thruster for extremely small maneuvers. The “pulsed plasma thruster” uses electric power to ablate and electromagnetically accelerate molecules from a block of solid propellant. The propellant bar (usually teflon, polytetrafluorethylene) is essentially a part of the thruster itself and the “feed system” is very simple: a spring presses the front face of the propellant bar against the retaining lips between two electrodes. The spring feed mechanism is the only moving part of the thruster. A capacitor discharge, ignited by a spark, provides during a short time interval (microseconds) the current for the light arc (kiloamperes) between the two electrodes. When the light arc strikes it ablates several molecular layers of the propellant bar; the ablation products include a variety of fluorocarbons, such as HF, F₂, CF, CF₂ and CF₃ (all corrosive, toxic and cancer causing species). The plasma is then accelerated by the Lorentz force, in the simplest form established by an applied magnetic field between two “rail electrodes”:



However, also a self-induced magnetic field is able to accelerate the plasma, in case one rail electrode surrounds the other one (the pulsed coaxial ablative plasma cannon). The thruster can be fired several times per second producing a quasi steady thrust force, provided that the capacitor is constantly recharged by the power supply system of the spacecraft. The ability to create with a variable frequency many small impulses makes the solid pulsed plasma thruster ideal for some high precision pointing mission.

Even though the solid pulsed plasma thrusters has experienced a quite long development history (since 1964, early models have flown successfully on Russian and since 1970 on American satellites), the analytical understanding of its operation is very empirical and sophisticated computer programs are still missing. It was expected that propellant exhaust particles would pollute solar cells and instruments, but the experience has shown that this problem is not serious (nevertheless liquid xenon has also been tried as propellant). Today, some advanced concepts can nearly compete with the performance of ion thrusters (specific impulses between 800 and 1500 seconds, thrusts between 220 mN and 1100 mN, and power levels between 20 W and 160 W); but the electrical efficiency of the pulsed plasma thrusters is in any case quite poor.

Hall thrusters. The main component of this thruster type is a cylindrical chamber with a central spike, formed by a coaxial ring-shaped channel. Chamber and spike are made of ferromagnetic material; the annular channel is made of an electric insulating material. Chamber and spike incorporate electromagnets which establish a magnetic field with radial direction across the annular channel. The annular channel incorporates at its bottom the anode, a hollow ring with a channel for the propellant inside.



The propellant (usually xenon) can depart from the anode through small orifices and enter the annular channel. The hollow cathode, a thermal filament electron emitter, is an external device. During the operation about 95% of the propellant flows to the anode and 5% to the cathode. When during the operation a current flows through the plasma between the cathode and the anode, the lighter electrons travel from the cathode to the anode, while some heavier ions travel from the anode to the cathode. Now we might expect that most of the current is carried by the lighter electrons, but the same lightness makes it difficult for them to cross the magnetic field: the hall effect causes an azimuthal motion of the electrons, it forces them to move in spirals around the axis. Some of the electrons manage to hit the anode anyway and keep the current running, but most of them are trapped in a fast circular motion inside the annular channel. Thus, the magnetic field acts as an impedance to the electron flow and prevents the current from shorting. On their way the electrons collide with the propellant atoms and create ions. The surplus of electrons in the channel generates a strong electric field (about 100 V/cm) that attracts ions from the anode region. The ions are much heavier than the electrons, their motion is not much affected by the magnetic field; and thus the ions accelerate in the electric field. On their way they pick up electrons from the surplus in the channel or from the cathode emitter; and an essentially neutral plasma leaves the channel with a velocity of 15 to 25 km/s.

The Hall-effect thruster was originally invented during the 1960s in the United States, but later it was abandoned there. It was then perfected in the Soviet Union and applied many times in Russian satellites for attitude and orbit control purposes. Therefore this thruster is today the electric space engine with the highest state of development; presently the western space community tries to re-adopted this propulsion concept. The Hall-thruster (or Russian plasma thruster) is actually an ion engine because it uses an electric field for the acceleration of charged ions, but the electromagnetic “Hall-effect” plays a key role in the operation principle. Like ion engines the Hall-thruster uses inert gas as propellant (xenon or krypton), but it avoids erosion problem of screen grids and operates with more convenient voltages and currents.

Russia developed several thrusters with similar performance (specific impulses from 1600 seconds to 2000 seconds) for different power levels (between 0.3kW and 4.0 kW). The demonstrated efficiency of about 50% and the long service life (over 6000 hours) make these thrusters suitable for satellite control and for interplanetary missions.

9.1.4 Electric Energy Sources

Batteries. Electric propulsion systems consume comparatively much electric energy, a typical consumption rate for an ion engine is for example 30 W for every milliNewton thrust over a period of 10000 hours. Let us start the enumeration of space power systems with a type that is apparently not appropriate to supply electric propulsion systems with energy: the chemical battery. Today’s state of the art are rechargeable lithium carbon cells (“secondary batteries” for 5000 discharge/recharge cycles), they can store 1 kWh in 15 kg cell weight, equivalent to 67 Wh/kg (silver-cadmium accumulators can store about 40 Wh/kg). The capacity of non-rechargeable “primary” batteries is about five to ten times better, for example lithium cells with a weight of just 118 grams are specified to deliver 11 Ah at 3.6 V, equivalent to 336 Wh/kg. The performance depends on the temperature; batteries for space applications can operate at low temperatures (down to -80°C), they have a self discharge of less than 2% per year (the self discharge depends also on the ambient temperature, for example it amounts 2% per year for -40°C , 5% per year for $+21^{\circ}\text{C}$ and 9% per year for $+72^{\circ}\text{C}$).

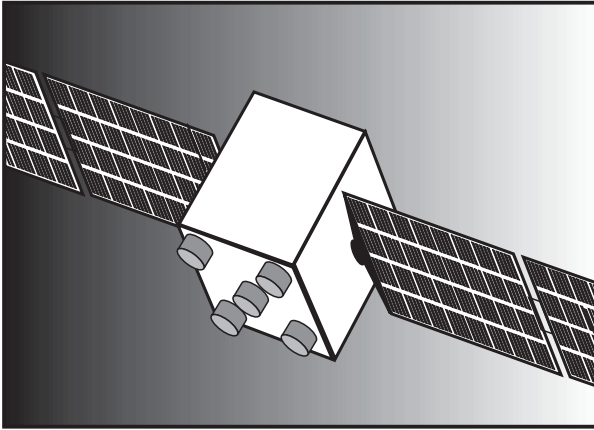
This is obviously not enough energy for long time propulsion, however, during short time periods secondary batteries can provide sufficiently energy, for example when the orbiting spacecraft travels through the shadow of the earth and the sunlight is not available (during a solar eclipse). Primary batteries are not an option, because their energy storing capacity is simply too small. We should bear in mind that otherwise primary batteries would be the ideal energy supply system for spacecraft: they are small, cheap, extremely reliable, shock resistant, not dangerous, need no orientation or deployment procedure, do not deteriorate in a radiation environment, they can deliver much power during short time intervals (high peak power levels), they do not disturb sensors or transmitters or cause any other interference with the spacecraft. Primary batteries are often not considered in astronautical engineering, even though they are actually the ideal power source when the demand for electric energy is small.

Fuel cells. Fuel cells are another type of primary batteries. They transform the energy generated by a chemical reaction (for example the oxygen-hydrogen reaction or the oxygen-methane reaction) with a remarkable efficiency of typically 50% directly into electric energy. Since they deliver drinking water as “waste product”, fuel cells have been used several times for manned missions, particularly with the intention to support solar cells (to cover peak loads). The fuel cells of the American Apollo spacecraft consumed 0.36 kg oxygen and 0.045 kg hydrogen for every kWh, equivalent to 2.47 kWh/kg (Apollo 13 has shown that the application of fuel cells is not always unproblematic). Liquid hydrogen cannot be stored in a spacecraft for a longer time period; thus another fuel has to be used, for example the combination of liquid methane and liquid oxygen with a smaller energy content of about 2 kWh/kg. Even though this is about seven times more capacity in comparison with the best conventional battery, it is still not enough power for an electrically propelled spacecraft.

Radioisotope generators. A nuclear process can liberate much more energy per unit of mass than a chemical process. Radioisotope thermoelectric generators (“RTGs”) transform the heat generated by the decay of radioactive material with appropriate half-life (plutonium 238) with thermoelectric converters into electric power, either by using the “Seebeck-effect” (a current is created between two junctions of different metal at different temperatures) or by using the “Langmuir-effect” (a hot tungsten electrode emits electrons which are then captured by a cooled collector electrode). These devices cannot be switched on and off, but they deliver electric power for decades, typically 100 W for a weight of 40 kg (the power diminishes slightly in time, about 8% in 9 years, due to a decay of the radioactive material). The electric efficiency of RTGs ranges between 5% and 10%; thus the unconverted heat must be discharged into space (utilizing heat pipes and radiators); but the device works without movable components. The additional heat can also be an advantage for a spacecraft that travels at a great distance from the sun, where the low temperatures provide a working problem for electronics.

RTG's operated successfully on American and Russian spacecraft, on earth satellites, on planetary spacecraft and even on the moon. All missions into deep space which went further away from sun than Mars used RTGs; so did Pioneer, Voyager, Ulysses, Galileo, Cassini and the Viking Mars landers. RTGs have indeed a fantastic performance, but their use involves two serious problems: first, non-acceptance of nuclear power by the taxpaying public; and second, very often in practice, the non-availability of the appropriate nuclear material. The material plutonium 238 (half-life 88 years) used in RTGs is not to same as the Plutonium 239 of nuclear bombs, but it is a byproduct of a certain bomb production (some touch of Plutonium 238 exists also in minerals). Like all Plutonium derivatives it is extremely toxic; and in case of a launch failure danger to man cannot be excluded completely. Careful attention has to be paid that a space probe with an RTG on board will not collide with the earth on its interplanetary trajectory. As substitutes there are not many elements available which have the appropriate half-life period of about 100 years. For shorter half-life periods the power decays too fast, for longer half-life periods the generated heat is insufficient.

Photo-voltaic generators. Radioisotope generators are well suitable as power providers for electrically propelled spacecraft, but more often than not photo-voltaic generators are better. Large solar arrays are the best electric power systems for spacecraft which do not travel too far away from the sun. Soon after their invention (about 50 years ago) photo-voltaic solar cells were used on the first operational earth satellites, even though that time their performance was poor in comparison with today. Solar cells have been constantly improved regarding efficiency, weight,



mechanical flexibility and radiation resistance; and the rapid development is still going on. Reference values today are 15% to 25% efficiency and a specific weight of 50 W/kg, typically. A reduction of the specific weight by the factor two or even three seems to be possible in the near future. The disadvantage of solar arrays in comparison with radioisotope generators is the demand for an always correct orientation with respect to the direction of the sunlight.

The sun provides at a distance of 1 AU a light flux of 1.395 kW/m²; thus depending on the efficiency, an area of 4 to 8 m² (weighing 20 kg) is required for every kilowatt electric power. Many photo-voltaic converters, thin flexible semiconductors with a cell area of about 2 times 4 cm² for 0.5 V and 130 mW, are integrated on the solar arrays (occasionally for lower power requirements also directly on the spacecraft's body), either connected in series for more volts or parallel for higher currents. The automatic deployment of solar arrays in space is not uncritical and the mechanism increases the weight of the system (large solar arrays are folded during the launch procedure).

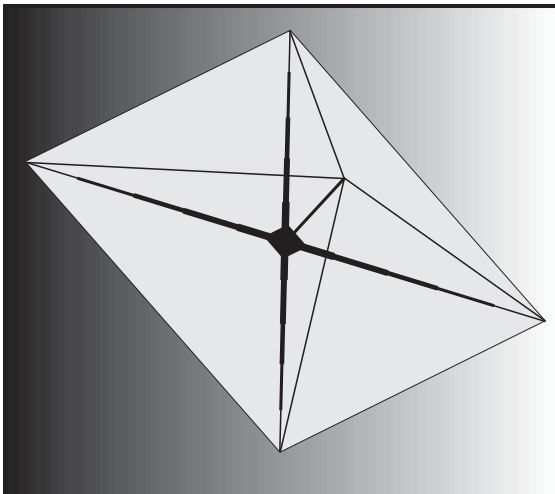
For more power larger solar arrays are necessary. The intensity of the sunlight diminishes inversely proportional to the square of the distance of the spacecraft from the sun, but the efficiency of solar cells improves when it gets colder: the exponent that characterizes the power degradation is not exactly 2 but approximately 1.7 (but solar cells refuse to work completely below a certain very cold temperature). Solar silicon cells have been developed for low intensity and low temperature conditions which would still work at Pluto's distance from the sun. When the spacecraft gets close to the sun, the solar arrays have to be turned away from the direction of the maximum intensity, otherwise the direct sunlight would destroy the sensitive cells. The energy which is not converted must be conducted away from the cells and radiated into space in order to keep the operation temperature in the limits. Radioactivity deteriorates solar cells and can even destroy them: a typical overperformance of 20% is given to the solar cells of earth satellites in order to cope with their degradation in time. Solar cells get damaged when a spacecraft passes the van-Allen radiation belt of the earth (or the radiation environment of another planet); and for many deep space missions it is still necessary to develop improved radiation resisting solar cells.

9.1.5 Solar Sails

Sunlight pressure. An area that is exposed to light (and other electromagnetic radiation) experiences a so-called light-pressure (discovered in 1884 by L. Boltzmann). We can readily compute the sunlight pressure on an illuminated area when we equate the incident light flux (on earth 1395 W/m^2) with the jet power (the expression $S \cdot c/2$, the thrust S in Newton and the velocity of light $c = 2.99793 \cdot 10^8 \text{ m/s}$). The result is a pressure of exactly $9.3064 \cdot 10^{-6} \text{ N/m}^2$, assuming that the entire flux is transformed into pressure (total rectangular reflection, no heating). In space the so-called “solar wind” acts additionally on the illuminated area, a plasma consisting of protons, electrons and some helium ions, but the pressure of this particle radiation is negligibly small: the solar wind has at the distance of the earth an average velocity of 500000 m/s and a density of approximately $1.5 \cdot 10^{-20} \text{ kg/m}^3$, thus the effect of the proton bombardment is about 1000 times smaller than the effect of the light pressure.

Sailcraft propulsion. The direct sunlight creates at the distance of the earth a pressure of about $9 \cdot 10^{-6} \text{ N/m}^2$ on a reflecting area. We are always excited by the idea that we might get something for free: a spacecraft equipped with huge light-weight reflectors could maybe utilize this pressure and generate a propulsive force directly from the sunlight without consuming any propellant. This great advantage would make it possible that the sailcraft could reach any planet from an orbit around the earth; and, when it is required, also return back to the earth without additional equipment.

Let us assume that a thrust acceleration of maybe 10% of the solar gravitation is sufficient for an interplanetary mission: then the thrust acceleration at the distance of the earth ($r \approx 1.5 \cdot 10^{11} \text{ m}$) would be $0.1 \cdot \gamma_{\odot}/r^2 = 0.587 \text{ mm/s}^2$; and every square meter of the spacecraft must not be heavier than 15.3 grams (approximately the weight of a letter). This seems to be possible for small probes; for example, a sailcraft with a mass of only 100 kg would then require sails with an entire area of 6536 m^2 . The sailcraft could be a small payload in the middle of the crossing of four booms which deploy huge reflectors with a total size of 81 times 81 meters. It is not allowed



to scale up these values proportionally for larger spacecraft, because the bending moments in the booms increase proportional to the square of the boom length, and the payload is not a surface specific load but concentrated on the center of the vehicle. The deployment of such fragile structures in space and the control of their attitude is certainly difficult but apparently not impossible. The deployment must occur at an orbital altitude where the remainder of drag will not cause a substantial deceleration of the lightweight vehicle: about 2000 km above the surface.

9.2. Electrically Propelled Missions

Let us return to electric propulsion. Presently the stabilization of earth satellites is considered as the main application for electric engines, but interplanetary missions with an extremely high demand for propulsion energy are actually more interesting.

9.2.1 Optimization of the Propulsion System

Electric propulsion for satellites. An electric propulsion system consumes propellant sparingly: the propellant mass, necessary to comply with a certain velocity requirement Δv , is very small in comparison with a chemical propulsion system, but thrusters and power supply system are now comparatively heavy. The advantage of a low fuel consumption plays a role in the mass balance only when the demand for velocity Δv is very high. Conventional earth satellites have today a typical service of eight years (after which the electronic equipment is out of date anyway), and the overall Δv requirement for position stabilization accumulated in this time interval (typically 1500 m/s) is simply too small to make the application of electric propulsion sensible. However, this result will change in favor of electric propulsion when the desired service life of earth satellites increases or when the weight of the thrusters decreases. More Δv capacity is required when the launch vehicle transports the satellite only to a low earth orbit: then an upper stage is not necessary and the satellite integrated propulsion system performs the initial positioning maneuver. The transport of a satellite from low earth orbit to geostationary position with the thrust of an electric propulsion system is a theoretical option but it is not advisable for two main reasons: first, the mission takes too much time for a commercial satellite service (additionally one year or more); second, the flight through the van Allen radiation belt is quite destructive for solar cells and other parts of the electronics. Until today these problems have obstructed the application of electric propulsion in commercial satellites. It seems to be more promising to use electric engines in scientific deep space probes.

The optimal exhaust velocity. Ion engines operate with exhaust velocities which are about ten times higher than the exhaust velocities of chemical engines, for example between 30 km/s and 60 km/s. The higher the exhaust velocity the more electric power is required. The exhaust velocity is a function of the grid voltage, usually it is constant for a certain engine (on request it would also be possible to construct ion engines with variable exhaust velocity). In order to find the optimal exhaust velocity for a certain mission we make now use of a simple model. Therefore we consider the spacecraft as composed of five sub-masses: the payload, the thrusters, the power supply system, the propellant and the tank. The payload, essentially the instrumentation of the probe and the structure which belongs to the instrumentation, is not a function of the other sub-masses. The thruster mass (and the concerning part of the structure) is a linear function of the thrust; the mass of the power supply system is a linear function of the power of the jet. The mass of the tank is approximately a linear function of the propellant mass which is determined by the Ciolkovskij equation.

In the second chapter of this book we have already used a similar model in order to analyze the performance of a chemical rocket stage. That time we used the expression $M_{ignition}$ for the initial mass of the stage; then we related all sub-masses (except for the payload mass) to the initial mass of the stage, and then we calculated the “mass ratio” of the rocket stage (the important factor $\mathcal{R} = M_{payload}/M_{ignition}$). The same procedure applied to the electrically propelled rocket stage yields now:

$$M_{ignition} = M_{payload} + M_{thrusters} + M_{powersupply} + M_{tank} + M_{propellent} \quad (9 - 1)$$

$$M_{burnout} = M_{ignition} - M_{propellent} \quad (9 - 2)$$

with: $M_{thrusters} = C_{thrusters} \cdot s \cdot M_{ignition} \quad (9 - 3)$

$$M_{powersupply} = C_{powersupply} \cdot 0.5 s c \cdot M_{ignition} \quad (9 - 4)$$

$$M_{tank} = C_{tank} \cdot M_{propellent} \quad (9 - 5)$$

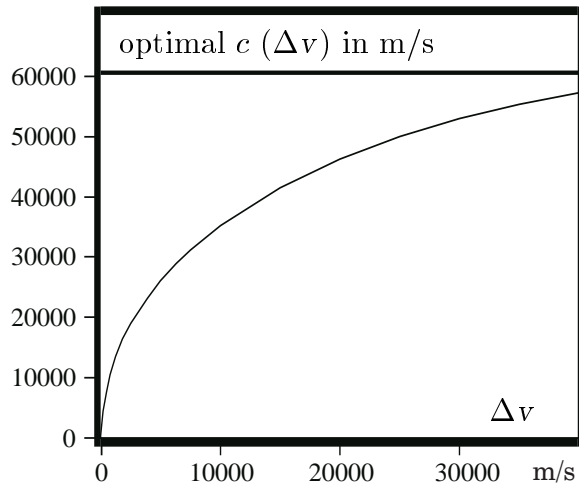
$$M_{propellent} = (1 - e^{-\Delta v/c}) \cdot M_{ignition} \quad (9 - 6)$$

Term $M_{payload}$ is the mass of the instrumentation, it is given by the scientific objective of the mission and it is not a function of the other subsystems when the initial mass of the spacecraft is predetermined. The other subsystems depend in a linear manner on the initial mass of the spacecraft: a bigger spacecraft requires proportional to its size more engines, more power and more propellent. Furthermore, the engine mass is a linear function of the initial thrust acceleration $s = cm/M_{ignition}$, and the mass of the power supply system is a linear function of the jet power $mc^2/2 = 0.5 s c M_{ignition}$ (m is the mass flow rate and c the effective exhaust velocity). We insert the equations (9-3) to (9-6) into (9-1) and transform. The result is the following relationship:

$$\mathcal{R} = \frac{M_{payload}}{M_{ignition}} = (1 + C_{tank}) \cdot e^{-\Delta v/c} - s (C_{thrusters} + \frac{c}{2} C_{powersupply}) - C_{tank} \quad (9 - 7)$$

The result is slightly different from the result for a chemical engine: now a power system is present with a mass fraction depending linearly on the exhaust velocity c . Differentiation $\partial\mathcal{R}/\partial c = 0$ brings us the conditional equation for the optimal exhaust velocity c . It is a transcendental equation that can be solved iteratively, thus:

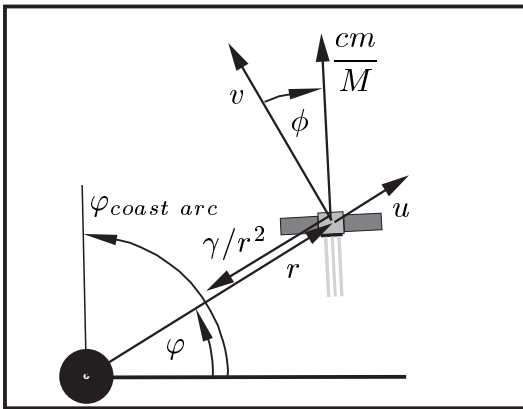
$$c = \sqrt{\frac{2(1 + C_{tank})}{s C_{powersupply}} \cdot \Delta v \cdot e^{-\Delta v/c}} \quad (9 - 8)$$



The plot on the left shows that the optimal exhaust velocity c of an electrically propelled spacecraft is approximately a square root function of the mission velocity requirement Δv ; assuming the values $C_{tank} = 0.1$ (10% tank weight), $s = 1/3000$ N/kg (or $3.33 \cdot 10^{-4}$ m/s²) and $C_{powersupply} = 0.04$ kg/W (this corresponds to solar arrays with 50 W/kg and thrusters with an efficiency of 50%).

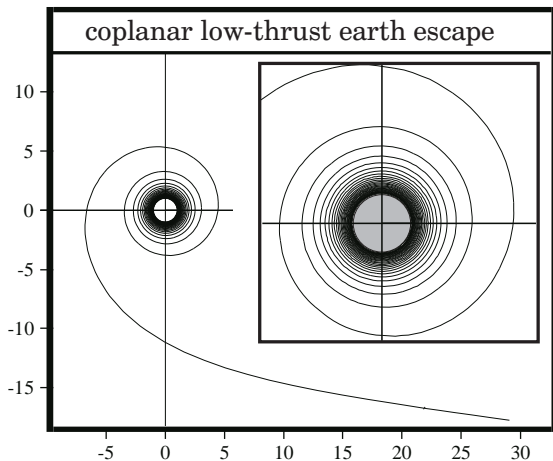
9.2.2 Low Thrust Earth Escape

The escape spiral. There are two ways how an electrically propelled deep space probe can enter the interplanetary transfer orbit: either the spacecraft is inserted directly by the chemical propulsion of the upper stage of the launch vehicle, or it uses its own propulsion system to spiral up from a low earth orbit. Let us now analyze the option of low thrust escape: in comparison with direct insertion it offers the advantage of the better efficiency of electric propulsion, but it involves the disadvantage of an extended mission time (and the undesirable prevail time in the van Allen radiation belt). To find out the expense of time and propellant of the low thrust earth escape mission we have to integrate the equations of motion numerically (the derivation of these differential equations is explained in detail in the chapter four of this book). Thus:



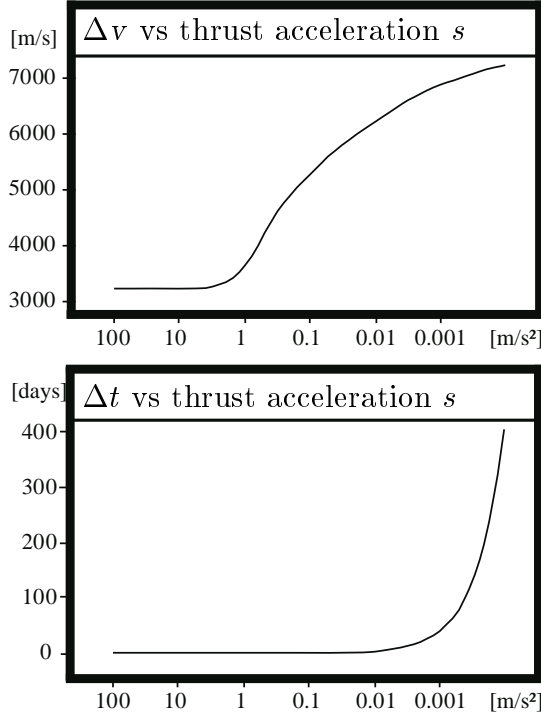
$$\begin{aligned} \dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} + \frac{c m}{M} \sin \phi \\ \dot{v} &= -\frac{u v}{r} + \frac{c m}{M} \cos \phi \\ \dot{r} &= u \\ \dot{\phi} &= \frac{v}{r} \\ \dot{M} &= -m \\ \tan \phi &= \frac{u}{v} \end{aligned} \tag{9 - 9}$$

The plane motion is defined by the polar coordinates r and ϕ (radius and path angle with pole at the center of the earth) and the velocity components $u = \dot{r}$ (vertical) and $v = r\dot{\phi}$ (horizontal). The trajectory is controlled by the thrust angle ϕ and the mass flow rate m (M is the mass of the spacecraft and c the constant exhaust velocity). Symbol γ represents the gravitational constant of the earth, $\gamma_{\oplus} = 3.9867 \cdot 10^{14} \text{ m}^3/\text{s}^2$. A control programme is called “tangential thrust” when the thrust vector is aligned with the velocity vector, $\tan \phi = u/v$. Tangential thrusting lets the mechanical energy and the angular momentum of the orbit grow simultaneously (the vertical velocity u remains small most of the time; the propulsive power is $dE/dt = \vec{v} \cdot \vec{S}$,



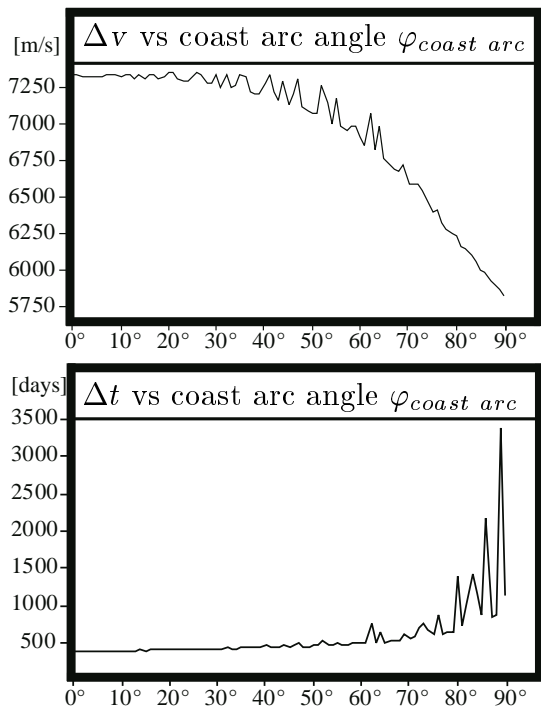
with the mechanical energy on orbit E , the thrust vector \vec{S} and the velocity vector \vec{v}). Thus, a simple steering programme with tangential thrust all the time is nearly an optimum escape strategy (actually, the really optimum escape strategy requires tangential thrust only at the final time): tangential thrust maximizes the increase of the mechanical energy of the orbit at every instant of the entire flight time; and it is easy to integrate the trajectory even when the thrust is extremely low.

Constant thrusting. A spacecraft that orbits with a velocity of $v = 7785$ m/s at an altitude of 200 km above the earth ($r_{earth} = 6378$ km) requires a Δv capacity of 3225 m/s to enter with a single impulsive maneuver a parabolic escape trajectory. Initially the Δv requirement remains nearly constant when we start to reduce the



thrust level of the maneuver, but it grows considerably when we reduce the thrust acceleration $s = cm/M$ below 1 m/s². However, the Δv requirement does also not grow without limits when the thrust level gets really low (below 1 mm/s²); but it remains always below a certain value (the circular velocity of the initial orbit). The time interval of the escape maneuver of a space probe equipped with electric engines last about one year ($s \approx 0.2$ mm/s²). The numerical simulation of the trajectory shows that the spiral deforms considerably shortly before the spacecraft has reached the parabolic escape velocity. In order to avoid that the spacecraft collides with the earth, it is necessary to replace the tangential thrust programme finally by more sophisticated navigation strategy.

Coast arcs. A thrust acceleration $s = 0.2$ mm/s² and an exhaust velocity $c = 35240$ m/s are typical values for an electrically propelled spacecraft: several thousand orbital



revolutions are necessary before the probe reaches the parabolic escape velocity. It is possible to reduce the Δv requirement of the mission at the expense of more mission time Δt . The figures on the left assume that the thrust is switched off always when the spacecraft travels through a certain sector of the escape spiral (determined by the angle $\varphi_{coast arc}$). Suspending the thrust ($m = 0$) every orbital revolution in a certain sector makes the escape spiral more eccentric and reduces greatly the gravitational losses. The computed plots show irregular behaviour because the duration of the last coast arc (immediately before the spacecraft reaches escape velocity) has a main influence on Δv and Δt . For a short mission time it is advisable that we omit the last coast arcs.

Hyperbolic insertion. In practice however, it is probably better that we insert the electrically propelled spacecraft directly with the upper stage of the launcher into the interplanetary transfer orbit. This avoids a long prevail time in the radiation belts of the earth and reduces the mission time by approximately one year. The question arises how fast the spacecraft should depart from the earth at the beginning of its mission: the electric propulsion of the probe is much more efficient than the chemical propulsion of the upper stage; but the propellant of the upper stage is very efficiently used because it is expended near the gravitational center. The chapter four of this book explains why: the mechanical energy of the escape orbit with respect to the gravitational center remains constant, and thus the hyperbolic escape velocity v_∞ (far away from the earth) is much higher than the velocity $\Delta v_{superparabolic}$ (what the upper stage has to expend for changing from parabolic escape velocity to hyperbolic escape velocity). The energy relationship is responsible for the following equation:

$$\begin{aligned}\Delta v_{superparabolic} &= \sqrt{v_\infty^2 + v_{parabolic}^2} - v_{parabolic} \\ &= c_{chemical} \cdot \ln(M_{initial}/M_{final})\end{aligned}\quad (9 - 10)$$

The right hand side of the equation above is the Ciolkovskij equation applied to the maneuver that brings the vehicle from parabolic velocity to hyperbolic velocity: $M_{initial}$ is the mass of the spacecraft before the maneuver and M_{final} the mass after the maneuver. The index “chemical” in the expression for the exhaust velocity c indicates that the upper stage of the launcher performs the maneuver. To establish the same velocity v_∞ with the help of electric propulsion we have to switch the propulsion system on when the probe has already left the gravitational influence of the earth and moves on a parabolic escape orbit. The Ciolkovskij equation becomes then:

$$v_\infty = c_{electrical} \cdot \ln(M_{initial}/M_{final})\quad (9 - 11)$$

When the velocity v_∞ is small, the use of chemical propulsion is preferable: the non-linear energy relationship (9-10) shows that we get a small hyperbolic escape velocity v_∞ nearly for free, without expending much propulsion (the parabolic escape velocity on a circular orbit 200 km above the surface of the earth amounts 11010 m/s). However, when the velocity v_∞ is not small, the use of electric propulsion is preferable because its efficiency is higher ($c_{electrical}$ is typically ten times higher than $c_{chemical}$). The optimal injection velocity is determined by the margin when the electric propulsion starts to become better than the chemical propulsion. To find it we can assume that the ratio $M_{initial}/M_{final}$ must be approximately the same for both maneuvers:

$$(c_{chemical}/c_{electrical}) \cdot v_\infty = \sqrt{v_\infty^2 + v_{parabolic}^2} - v_{parabolic}$$

or:

$$v_\infty = \frac{2 (c_{chemical}/c_{electrical})}{1 - (c_{chemical}/c_{electrical})^2} \cdot v_{parabolic} \approx 2 \frac{c_{chemical}}{c_{electrical}} \cdot v_{parabolic}\quad (9 - 12)$$

The parabolic escape velocity is approximately 11 km/s, and when the exhaust velocity of the electric propulsion system is ten times better than the exhaust velocity of the chemical engine of the upper stage, the optimal hyperbolic escape speed is 2200 m/s. Actually it is better to select a value slightly above this margin, because, in contrast to electric propulsion, chemical propulsion adds no time to the mission.

9.2.3 The Comet Nucleus Sample Return Mission

The staging principle for electric propulsion. We have already analyzed the staging principle for conventional chemical rockets at the beginning of the second chapter of this book. When a chemical rocket accelerates to a high final velocity, the staging principle is applied to get rid of superfluous tank weight, essentially. The tank of an electrically propelled spacecraft is small and its weight is not important in comparison with the other components of the propulsion system, and thus the staging principle is apparently not sensible in electric propulsion. This is, however, not true for high energetic missions (like the comet nucleus sample return mission).

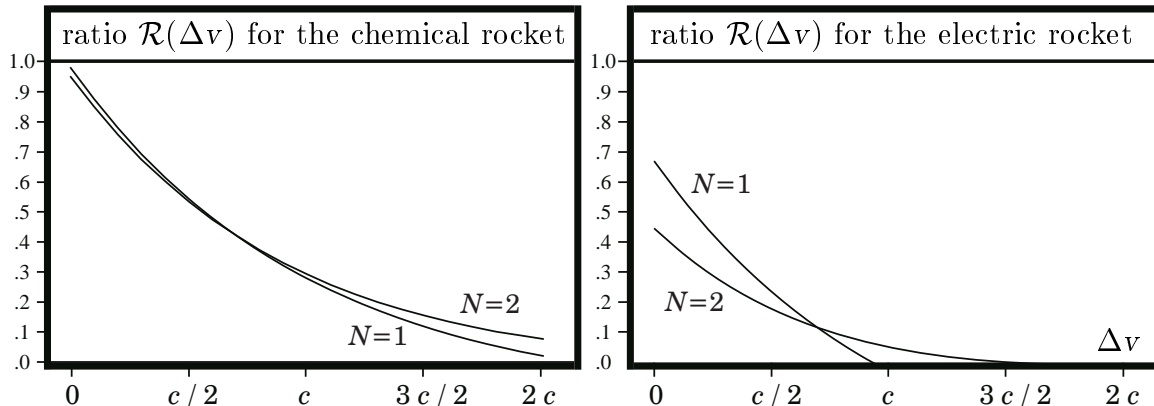
To analyze the staging principle for electric propulsion we assume that the optimal exhaust velocity c for a certain mission is determined and that the thrust acceleration s cannot become smaller than a certain limit. With these assumptions it is possible to replace in equation (9-7) the expression $s(C_{thrusters} + 0.5 c C_{powersupply})$ by the single constant $C_{engine} = M_{engine}/M_{ignition}$. The constant C_{engine} describes the weight of the thrusters, the power supply system, the cables, converters, structure and thermal control system which belong to the electric propulsion system. Then we have:

$$\mathcal{R} = \frac{M_{payload}}{M_{ignition}} = (1 + C_{tank}) \cdot e^{-\Delta v/c} - C_{engine} - C_{tank} \tag{9-13}$$

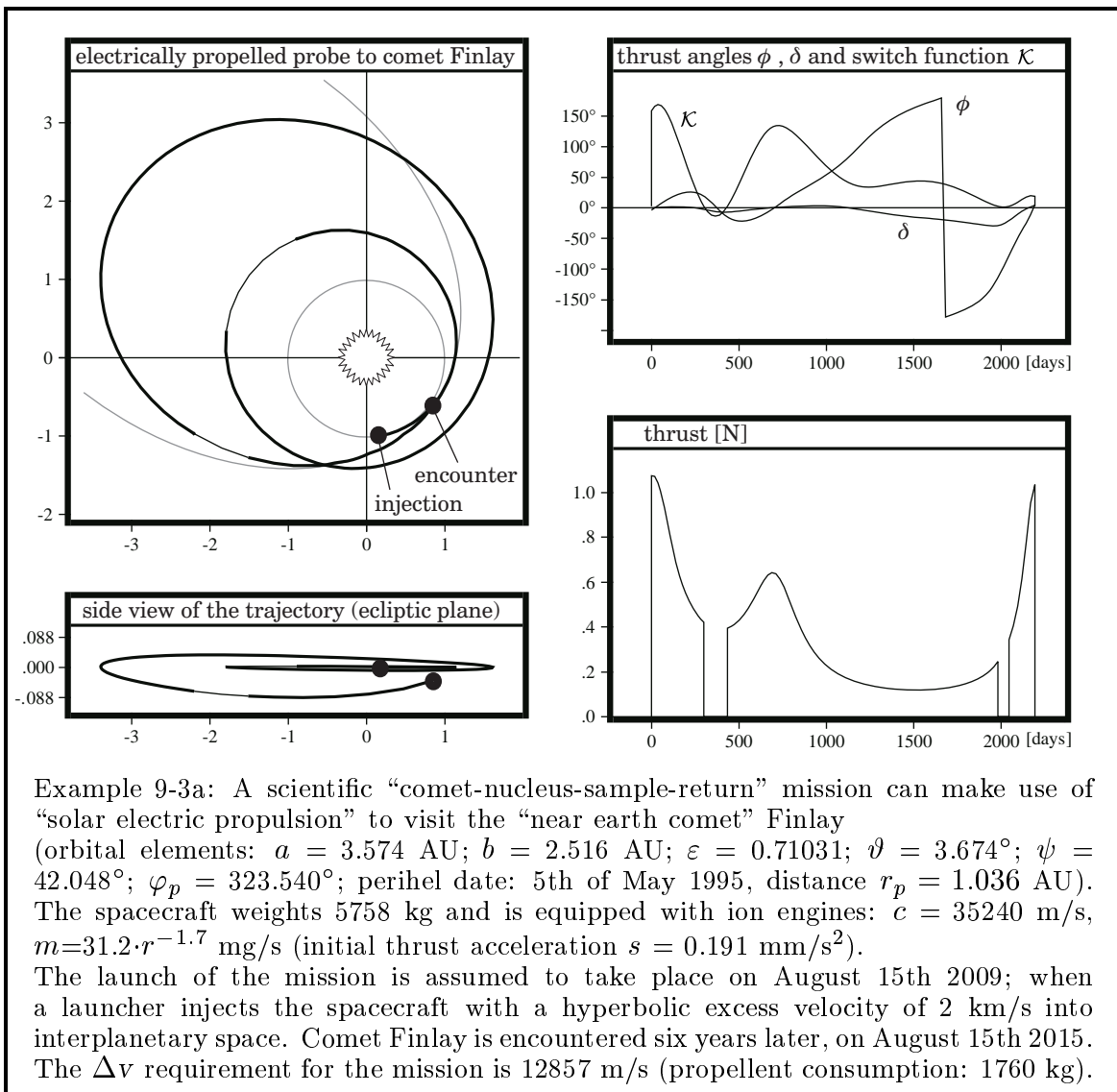
Now the relationship is valid for both, electrical and chemical propulsion, however the actual value of the constant C_{engine} is quite different. The ignition mass of the second stage is exactly the payload mass of the first stage, and, assuming that the stages operate with the same kind of propulsion system and that the Δv requirement of the mission is distributed in equal portions among both stages, we can denote:

$$\mathcal{R} = \frac{M_{payload}}{M_{ignition}} = \left[(1 + C_{tank}) \cdot e^{-\Delta v/(2c)} - C_{engine} - C_{tank} \right]^2 \tag{9-14}$$

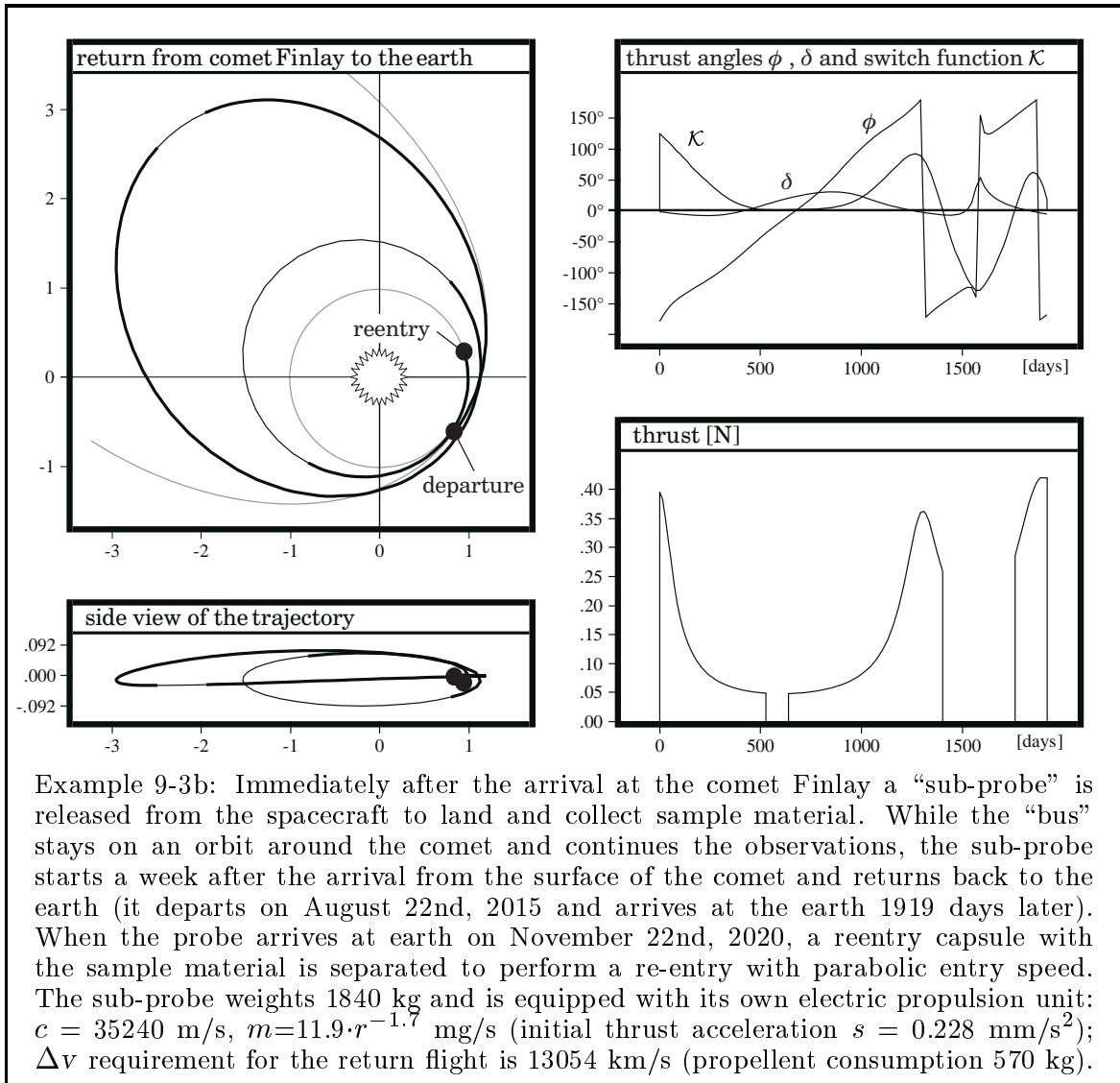
Equation (9-14) is valid for the two stage vehicle. Let us evaluate these equations for chemical and an electric propulsion, assuming that $C_{tank} = 0.1$ in both cases but that $C_{engine} = 0.03$ for the chemical rocket stage and $C_{engine} = 0.33$ for the electric stage. The plots compare single stage vehicles ($N = 1$) with two stage vehicles ($N = 2$):



The plots on the preceding page show that in case of a high energetic mission the staging principle is also sensible for electrically propelled spacecraft. The reason, however, is not the same as the reason for staging conventional chemical rockets. Here the mass of the strong rocket engine is usually small in comparison with the other components of the chemical stage (particularly in comparison with the propellant mass), but this is not the case for an electrically propelled spacecraft: the thrust acceleration cannot be lowered under a certain level because the mission must be accomplished in a reasonable time interval; and the electric engines and the power supply system are now heavy components of the spacecraft. When the spacecraft has expended a certain amount of propellant it is sensible to discharge thruster mass and solar cells with the intention to reduce the thrust acceleration. The staging of an electrically propelled mission should always be considered when the propellant load for the single stage configuration would be about 50% of the total spacecraft mass. For example, this condition is fulfilled for the comet nucleus sample return mission.



Mission alternatives. Comets consist of the “original material” of the solar system; probably it is only a sort of dirty snow, but the actual composition is very interesting: the material is obviously identical or very similar to the material which was present at the beginning of the existence of our solar system. It would be very interesting if we had samples of the material to analyze it in laboratories. The Δv requirement for the comet nucleus sample return mission is with about 26 km/s extremely high (13 km/s for the outward journey and 13 km/s for the return trip); but an electrically propelled spacecraft equipped with ion engines has the capability to perform such a mission. After having been injected into the interplanetary transfer orbit the spacecraft deploys its solar arrays, ignites the electric engines and travels to the comet. At arrival the spacecraft enters an orbit around the comet and releases a sub-probe (“lander”) for collecting the sample material. Now there are two possibilities how the material can be brought back to the earth: either the lander performs a rendezvous with the orbiter before the orbiter returns to the earth, or the lander is equipped with its own electric propulsion system and returns autonomously back to the earth.



Let us now compare the two mission alternatives a little more in detail. In any case a re-entry capsule with a parachute is necessary to bring the sample material to the surface of the earth. Such a system has never been constructed, maybe it weights 200 kg for 50 kg of material. We have also hardly more than an idea what instrumentation is necessary for the mission. A planetary spacecraft needs in any case a complicated control system: computer, transmitter and receiver, equipment for navigation, thermal control, and so on; and it would also be nice to have on board some contingency for scientific experiments and a wide angle camera. Let us assume that the instrumentation of the planetary spacecraft weights 400 kg (including structure and equipment necessary to support the payload). This is certainly not accurate, but without a detailed study we have no better information on the weight of the instrumentation.

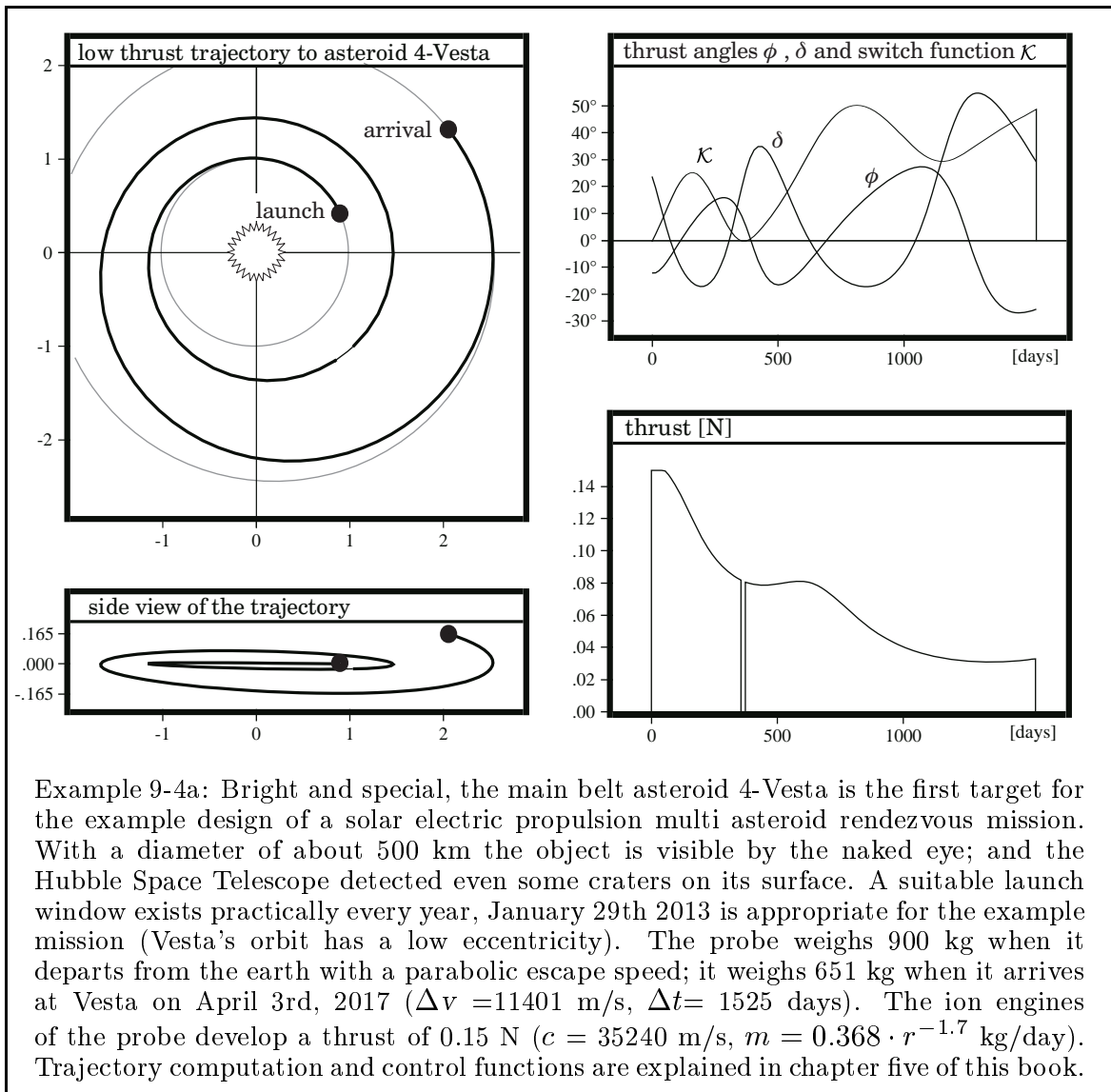
The construction of the sub-probe that is released to land on the comet and collect sample material is not easy. The vehicle requires also a complex navigation system, with transceiver, camera, electric energy supply system and so on; and, in case this is required, also the capability to perform a rendezvous and docking maneuver with the orbiting spacecraft (“bus”). The subprobe is discharged after it has transferred the material to the bus. Again, in the absence of better data, let us assume that the subprobe weights 400 kg. To avoid the risky rendezvous maneuver on the orbit around the comet, we can alternatively equip the subprobe with an own electric propulsion system (solar arrays, thrusters, propellant and tank) and let the subprobe return to the earth while the bus continues to orbit around the comet. Concept 1, the return of the bus, represents the single stage vehicle; concept 2, the return of the subprobe, represents the two stage vehicle. The electric propulsion system weighs about 33% of the initial stage mass ($C_{engine} \approx 0.33$); propellant tanks weigh about 10% of their filling capacity ($C_{tank} \approx 0.1$). When we assume that the exhaust velocity c of the thrusters is $c = 40$ km/s, we can use the Ciolkovskij equation (9-6) to calculate a propellant mass fraction of 27% for a Δv requirement of approximately 13 km/s.

outbound:	concept 1 (6000 kg)		concept 2 (5960 kg)	
	bus	subprobe	bus	subprobe
instrumentation	400 kg	400 kg	400 kg	400 kg
propulsion system	2000 kg		2000 kg	622 kg
propellant	1620 kg + 1108 kg		1620 kg	505 kg
tank	272 kg		162 kg	51 kg
re-entry capsule	200 kg			200 kg
homebound:	bus (4030 kg)		subprobe (1828 kg)	
instrumentation	400 kg		400 kg	
propulsion system	2000 kg		622 kg	
propellant	1108 kg		505 kg	
tank	272 kg		51 kg	
re-entry capsule	200 kg		200 kg	
sample material	50 kg		50 kg	

The table on the previous page shows that in the considered case the initial mass of the two stage configuration is with 5960 kg slightly better than the initial mass of the single stage configuration (6000 kg). This advantage, however, is negligibly small in comparison with the advantage that the rendezvous and docking maneuver between the two spacecraft is avoided. The comet nucleus sample return mission should use an autonomous return voyage of the electrically propelled subprobe, because this alternative improves greatly the probability that the difficult mission is finally successful.

9.2.4 The Multi Asteroid Rendezvous Mission

Mission objective. The Multi-Asteroid-Rendezvous and Flyby Mission is the other big future project where electric propulsion could demonstrate its fantastic capability.



These asteroids (or strictly speaking planetoids), smaller celestial bodies with a few dozen kilometers diameter, are interesting for the same reason as comets: they consist also of the original material of the solar system. Several thousand asteroids encircle the sun between the orbit of Mars and the orbit of Jupiter; a classification is necessary because albedo and spectrum of these “main belt asteroids” are different. A spacecraft equipped with ion engines could visit one particularly interesting object, establish an orbit around it and examine it for a certain time interval (“rendezvous”). Then, since the spacecraft stays in the asteroid main belt anyway, it can travel on to the next interesting object nearby. In case another asteroid comes close to the interplanetary transfer trajectory of the spacecraft, the trajectory can be deformed in a way that the probe flies past this asteroid and takes some photos (“flyby”). The mission is continued until finally the spacecraft has expended all its propellant.

The staging principle. In the multi asteroid mission the high Δv capability of electric propulsion is utilized to visit several objects: now the optimization objective is not a maximization of the scientific payload anymore, as usual, but a maximization of the number of target objects. This optimization criterion has a staging principle as a consequence which is quite different from the well-known payload maximization principle and which is seemingly paradox at the first glance. Under certain circumstances it is sensible to inject with the same launcher several electrically propelled probes into interplanetary space. A distribution of the mission among several independent probes can increase the number of targets and reduce the mission time and the mission costs. To analyze the staging principle we make again use of the simple model described by the equations (9-1) to (9-6); and again we assume that the specific weight of the electric propulsion system is predetermined (equation 9-13). This time, however, we take into consideration that several spacecraft (number N) are launched together with the same launch vehicle. The equations take then the form of:

$$\begin{aligned} M_{spacecraft} &= M_{launcher\ capacity}/N & (9-15) \\ &= M_{instrumentation} + M_{engine} + M_{tank} + M_{propellant} \end{aligned}$$

with: $M_{launcher\ capacity} = \text{predetermined}$ (9-16)

$$M_{instrumentation} = \text{predetermined} \quad (9-17)$$

$$M_{engine} = C_{engine} \cdot M_{spacecraft} \quad (9-18)$$

$$M_{tank} = C_{tank} \cdot M_{propellant} \quad (9-19)$$

$$M_{propellant} = (1 - e^{-\Delta v/c}) \cdot M_{spacecraft} \quad (9-20)$$

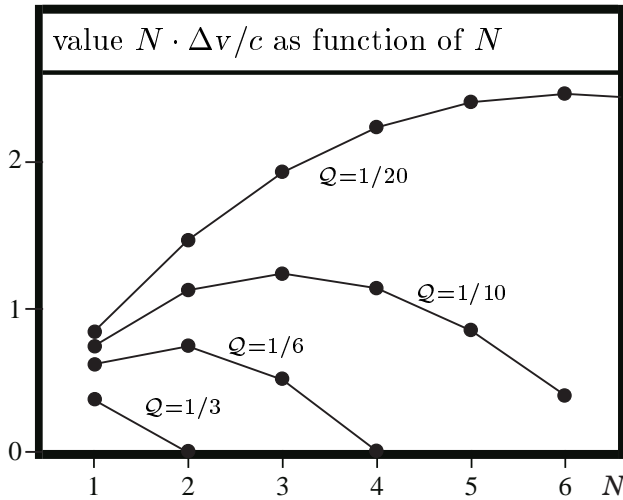
Let us consider now that the launch vehicle for the mission is predefined and that its capacity for the injection of a spacecraft on an interplanetary transfer orbit is given, thus that the expression $M_{launcher\ capacity}$ is not a variable. The mass of the instrumentation of the spacecraft $M_{instrumentation}$ is exclusively defined by its purpose; the instrumentation mass is not a function of the mass of the propulsion system or the capacity of the launcher. For an electrically propelled spacecraft the constant C_{engine} is approximately 33%, the constant C_{tank} is approximately 10%. The propellant mass required for a certain mission Δv is determined by the Ciolkovskij equation.

Our intention is now to find to a given launcher capacity and a given spacecraft instrumentation the Δv capacity of every individual probe and the Δv capacity of all probes together as a function of the probe number N . Therefore we introduce the factor \mathcal{Q} as:

$$\mathcal{Q} = \frac{M_{instrumentation}}{M_{launcher capacity}} = \frac{M_{instrumentation}}{N \cdot M_{spacecraft}} \tag{9 - 21}$$

Factor \mathcal{Q} is the instrumentation mass related to the launcher capacity, this value must be much smaller than 1 (for example $\mathcal{Q} = 1/15$ when the instrumentation weighs 400 kg and the launcher capacity is 6000 kg); and as agreed, this value is predetermined for a certain mission and not open for optimization. Now we insert the equations (9-16) to (9-20) into the equation (9-15) and transform the result using \mathcal{Q} . Thus:

$$\frac{\Delta v}{c} = - \ln \left(\frac{N \cdot \mathcal{Q} + C_{engine} + C_{tank}}{1 + C_{tank}} \right) \tag{9 - 22}$$



The left hand side of the equation above is the Δv capacity of one spacecraft, divided by the exhaust velocity c of its propulsion system. When we plot the overall flight performance of all spacecraft together $N \cdot \Delta v/c$ as function of the number N with \mathcal{Q} as parameter (assuming $C_{engine} = 0.33$ and $C_{tank} = 0.1$), we can observe that there exist a maximum for the overall flight performance that depends on N , the integer number of probes.

The smaller the value \mathcal{Q} the more probes are necessary for the $N \cdot \Delta v/c$ maximum. The physical interpretation of this effect is that a single spacecraft carries a lot of propellant inefficiently from one asteroid to another one. Let us now use the equation (9-22) to estimate for a certain mission the number of targets and the mission time t . Therefore we have to consider that the Δv requirement and the flight time requirement as well are different for the first target and for the other following targets:

$$\begin{aligned} \Delta v &= \Delta v_{first\ target} + \Delta v_{other\ targets} \\ \Delta t &= \Delta t_{first\ target} + \Delta t_{other\ targets} \end{aligned}$$

with: $\Delta v_{first\ target} \approx 10$ km/s; $\Delta v_{other\ targets} \approx 3$ km/s per target
 $\Delta t_{first\ target} \approx 4$ years; $\Delta t_{other\ targets} \approx 3$ years per target

These numbers are “typical values” and not at all accurate: they are the result of many experimental trajectory computations; and they will only help us to figure out the consequences of the equation (9-22) for a certain example mission. Maybe it is not necessary that we equip the spacecraft with many instruments; and when the capacity of a satellite launcher is large in comparison with the size of this instrumentation, then it can be better to send several spacecraft on their way to study many asteroids.

For example, when the instrumentation weighs 600 kg and a launcher with the capacity of 6000 kg is selected for the mission, then $Q = 0.1$ and the optimal spacecraft number is $N = 3$. The value $c = 40$ km/s is a typical value for the exhaust velocity of electric engines. The following table shows the evaluation of the equation (9-22):

spacecraft number N	1	2	3	4
individual Δv capacity [km/s]	29.2	22.3	16.4	11.2
overall Δv capacity [km/s]	29.2	44.6	49.2	45.0
number of target objects	$1 \cdot (1+6)=7$	$2 \cdot (1+4)=10$	$3 \cdot (1+2)=9$	$4 \cdot (1+0)=4$
mission duration Δt [years]	$4+6 \cdot 3=22$	$4+4 \cdot 3=16$	$4+2 \cdot 3=10$	$4+0 \cdot 3=4$

We can learn from the table that the distribution of the mission among several spacecraft reduces the available Δv capacity of every spacecraft, as it can be expected for a fixed instrumentation mass and a fixed launcher capacity. However, the overall Δv capacity of all spacecraft together can increase when several spacecraft are used, and this can have the consequence that the multi-asteroid-mission can visit more target objects. Several parallel operating spacecraft reduce in any case the overall mission time; and the probability of a total failure of the mission is also reduced with more than one spacecraft. The results of the comparison depend much on the underlying assumptions, important is here the ratio Q of instrumentation mass and launcher capacity: the bigger the launch vehicle the more spacecraft should be used.

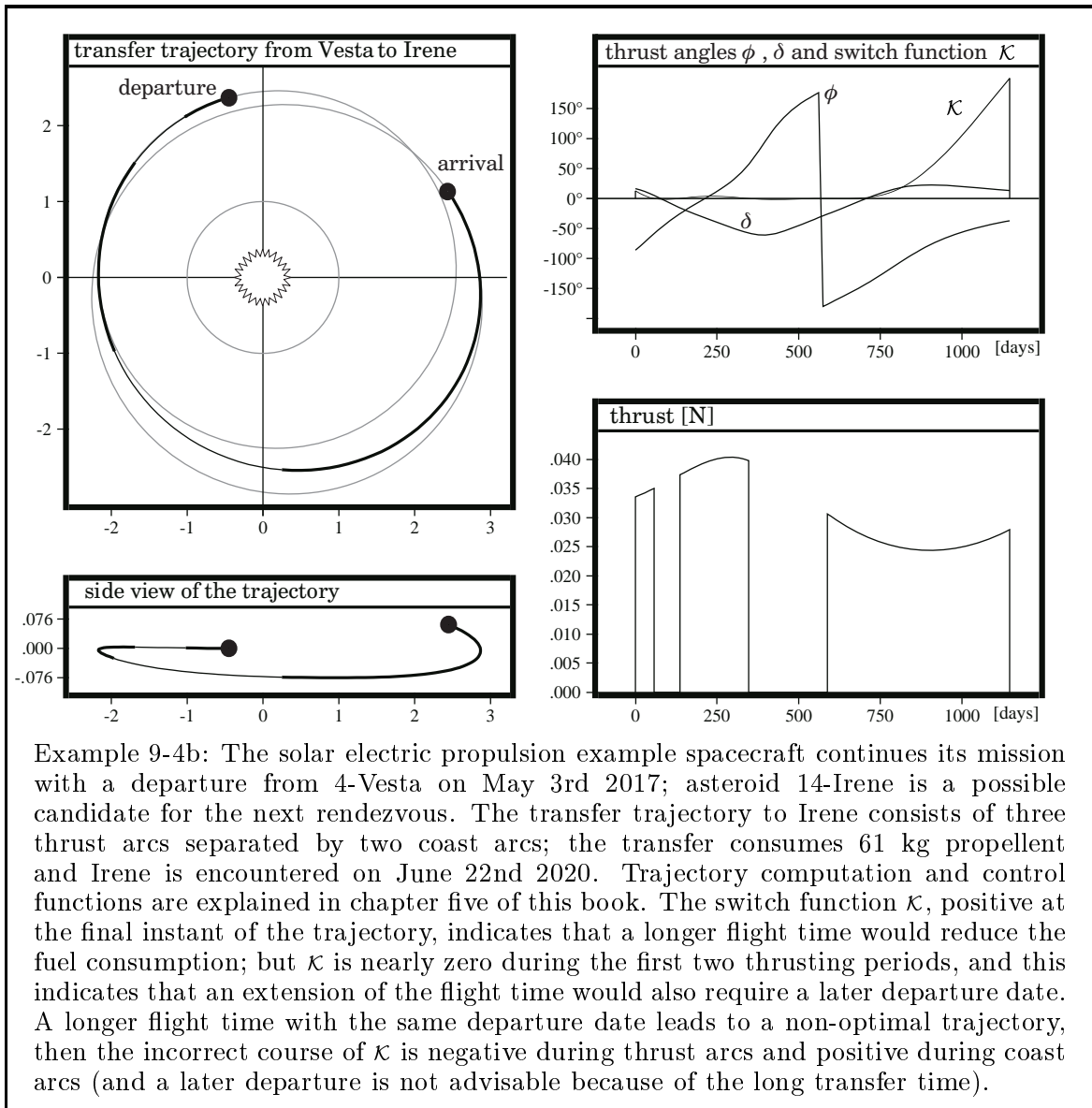
Mission costs. Now some people might expect that the multi asteroid mission would become more expensive when several small spacecraft are launched together with the same launcher, in comparison with a mission that uses only one big spacecraft. This is, however, not necessarily true. The decisive aspect is the capacity of the launch vehicle: obviously a reduction of the project costs is possible simply by using a smaller space launcher. However, when a decision concerning the launcher for the mission has been made, then a distribution of the launcher capacity among several small spacecraft does not raise the mission costs: the mass that is injected into interplanetary space is essentially the same; and the production of several identical smaller probes is maybe even cheaper than the production of a single big spacecraft.

Interplanetary low thrust transfer trajectories. When the launcher capacity is small, however, for example only 900 kg for a parabolic earth escape mission, then the “optimal” probe number is $N = 1$. A small electrically propelled spacecraft with an instrumentation mass of 200 kg might then consist of the following subsystems:

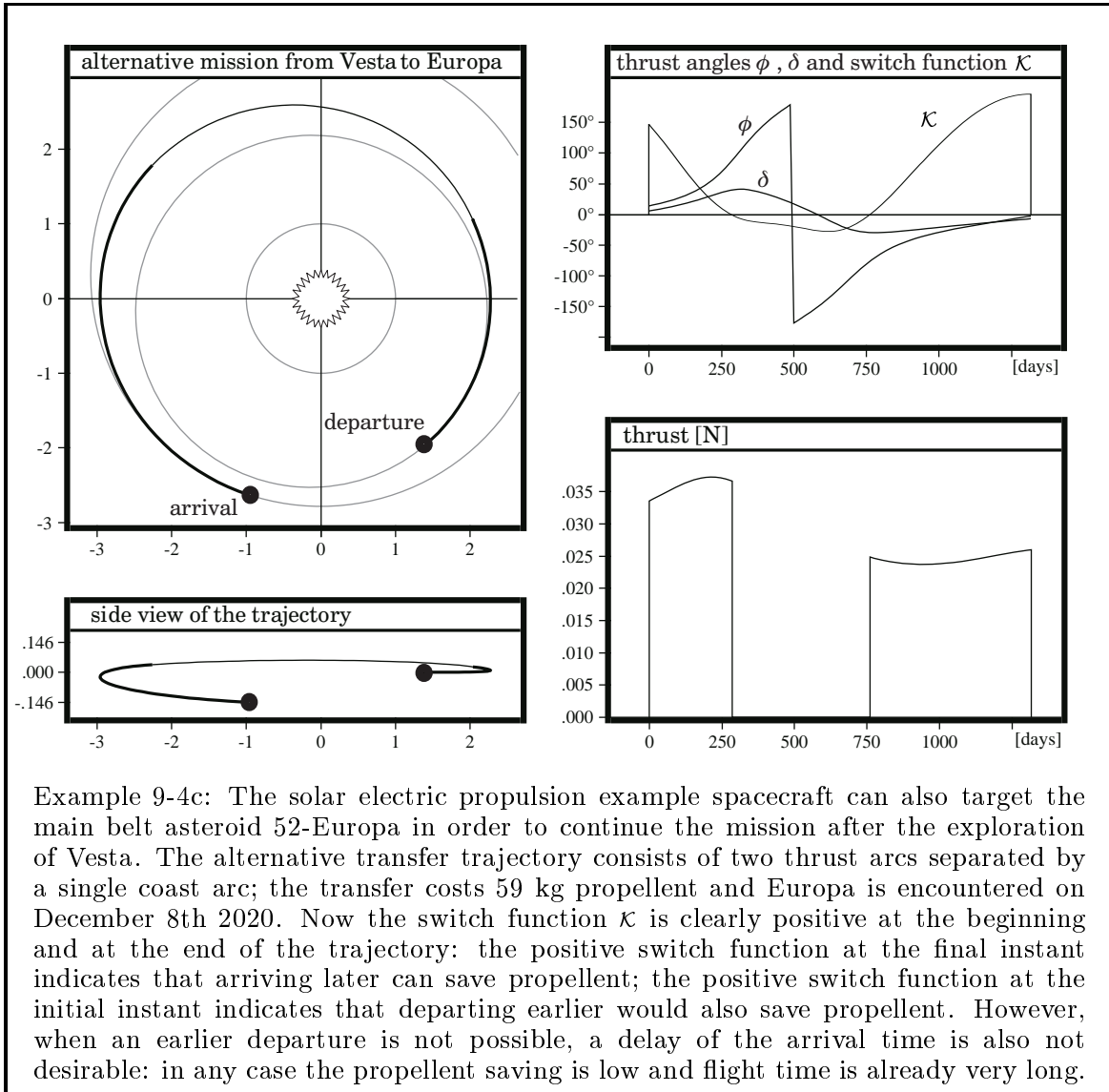
scientific payload and wide angle camera:	66 kg
transceiver with memory and parabolic antenna:	66 kg
attitude control system and navigation computer:	66 kg
main structure, tank and thermal control system:	90 kg
electric engines assembly (for example ten RIT 10):	100 kg
solar arrays for 5 kW power (at 1 AU distance):	200 kg
xenon propellant:	310 kg

Radiofrequency ion engines of the German RIT-10 type weigh 2 kg, they develop a thrust of 15 mN, and operate with an exhaust velocity of 30 km/s and consume 350 W electric power (the mass flow rate is then 5 micrograms per second only). For an interplanetary mission it is sensible to increase the exhaust velocity slightly, for example to 35240 m/s, and we can calculate a Δv capacity of 14.9 km/s for the probe.

Asteroid 4-Vesta is the most interesting candidate for an asteroid mission and thus it should be the first target. The transfer trajectory from the earth to Vesta consumes about 11.4 km/s of the Δv capacity of the spacecraft. When the spacecraft approaches this object it establishes an orbit around the asteroid and stays there for some time (in order to take photos and execute experiments). The relative velocity is small on an orbit around Vesta, about 10 to 100 m/s, depending on the altitude. Now the spacecraft has a Δv capacity of about 3.5 km/s left over for a second rendezvous with another asteroid from the catalogue.



Example 9-4b: The solar electric propulsion example spacecraft continues its mission with a departure from 4-Vesta on May 3rd 2017; asteroid 14-Irene is a possible candidate for the next rendezvous. The transfer trajectory to Irene consists of three thrust arcs separated by two coast arcs; the transfer consumes 61 kg propellant and Irene is encountered on June 22nd 2020. Trajectory computation and control functions are explained in chapter five of this book. The switch function κ , positive at the final instant of the trajectory, indicates that a longer flight time would reduce the fuel consumption; but κ is nearly zero during the first two thrusting periods, and this indicates that an extension of the flight time would also require a later departure date. A longer flight time with the same departure date leads to a non-optimal trajectory, then the incorrect course of κ is negative during thrust arcs and positive during coast arcs (and a later departure is not advisable because of the long transfer time).



However, there are many candidates because the catalogue contains more than 35000 objects, even when the search for the consecutive target is restricted to the 4000 candidates with accurately known orbital elements. Trajectory simulation is very important to determine which objects are in an appropriate position. For example, when the spacecraft stays on orbit around Vesta in the beginning of May of the year 2017 and the search is concentrated only upon the first 100 objects of the catalogue, then the following four main belt asteroids seem to be in an appropriate position:

asteroid	9-Metis	14-Irene	26-Proserpina	52-Europa
Δv [m/s]	4137	3471	3310	3330
Δt [days]	882	1146	1022	1315

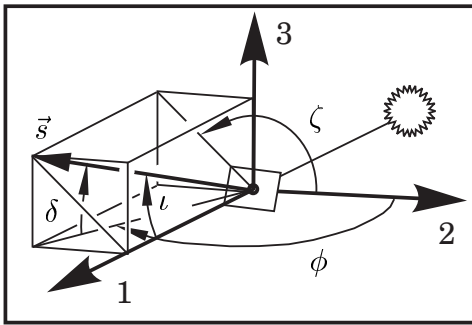
The orbital elements of the asteroid Vesta and the four possible candidates for a consecutive rendezvous are listed in a table in the first chapter of this book; and how to calculate the optimal transfer trajectories is explained in the chapter five.

9.3. Solar Sailing

Sailing in the radiation of the sun is the other exotic “low thrust” propulsion concept and like solar electric propulsion, it promises a great potential for future astronautics. Photon-radiation from the sun generates on earth a pressure of about $9.1 \cdot 10^{-6} \text{ N/m}^2$. A spacecraft that is equipped with huge light-weight reflectors could utilize this sunlight pressure and generate a propulsive force within the magnitude of maybe 10% of the solar gravity, equivalent to a thrust acceleration of $0.1 \cdot \gamma_{\odot}/r^2 = 0.587 \text{ mm/s}^2$ at a distance of 1 AU (1 astronomical unit $\approx 1.5 \cdot 10^{11} \text{ m}$). Naturally, the control of extremely light structures and their deployment in space is difficult, but the solar sails do not consume propellant. This only great advantage makes it possible that the sailcraft can reach any planetary destination from an orbit around the earth; and, when it is required, also return back to the earth without additional equipment.

9.3.1 Earth Escape with Solar Sails

Thrust vector of a sailcraft. When we analyze the thrust vector of a solar sailcraft we realize that now the thrust magnitude is a function of the actual thrust direction. We get the maximal thrust force when the sail area is exposed to the full sunlight, and this is obviously the case when the normal vector of the sail area is aligned with the direction of the sunbeams. Term ι is the angle between the sunbeams and the normal vector, and we resolve the normal vector using another angle called ζ . Thus:

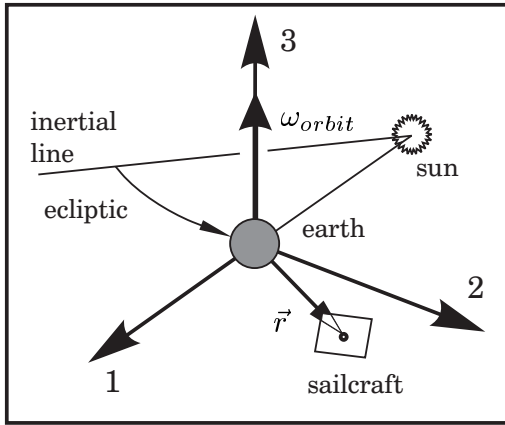


$$\vec{s} = s \cdot \cos^2 \iota \begin{pmatrix} \cos \zeta \\ \cos \zeta \sin \iota \\ \sin \zeta \sin \iota \end{pmatrix} \quad (9 - 23)$$

$$= s \cdot (\sin \phi \cos \delta)^2 \cdot \begin{pmatrix} \sin \phi \cos \delta \\ \cos \phi \cos \delta \\ \sin \delta \end{pmatrix}$$

For this notation we use a coordinate system that contains the ecliptic in its 1-2-plane; the 1-direction is always aligned with the direction of the sunbeams. The sails cannot generate a thrust force in the negative 1-direction; the thrust force vanishes when the normal vector is perpendicular to the sunbeam direction. We assume that the thrust acceleration \vec{s} is aligned with the normal vector of the sails. Its magnitude is proportional to the exposed area (the “number of photons” which hit the sails) and proportional to the change of the linear momentum of the impacting photons. The constant s characterizes the maximum thrust acceleration when the thrust vector is aligned with the direction of the sunbeams. When the thrust vector is inclined, however, the thrust is smaller and s must be multiplied two times by the factor $\cos \iota$. Alternatively we could also use the thrust angles ϕ and δ , but particularly in the calculation of sailcraft escape trajectories the angles ι and ζ are better conditioned.

Equations of motion. The equations of motion for the earth escape trajectory of a sailcraft take a simple form when we established them using a moving geocentric coordinate system: the 1-axis is always aligned with the sunbeams and the 3-axis is the normal vector of the ecliptic. The coordinate system rotates slowly on the 3-axis with the angular velocity $\omega_{orbit} = 360^\circ/\text{year}$ or $1.991 \cdot 10^{-7}$ rad/s. We cannot neglect this slow rotation because the flight time of a sailcraft on an earth escape trajectory can last several years. However, we may ignore the gravitational force of the sun which acts in a direction opposite to the sunbeams, because the gravity of the sun is neutralized by a counteracting centrifugal force with the same magnitude. Let \vec{r} be the vector that locates the sailcraft with respect to the earth. When we differentiate \vec{r} to find the velocity of the sailcraft \vec{v} we have to observe the laws for vector differentiation in a rotating coordinate system. The second total time derivative of the location vector is the acceleration vector with respect to inertial space:



$$\begin{aligned}\vec{r} &= \begin{pmatrix} x \\ y \\ z \end{pmatrix} & (9-24) \\ \dot{\vec{r}} &= \begin{pmatrix} \dot{x} - y \omega_{orbit} \\ \dot{y} + x \omega_{orbit} \\ \dot{z} \end{pmatrix} \\ \ddot{\vec{r}} &= \begin{pmatrix} \ddot{x} - 2\dot{y} \omega_{orbit} - x \omega_{orbit}^2 \\ \ddot{y} + 2\dot{x} \omega_{orbit} - y \omega_{orbit}^2 \\ \ddot{z} \end{pmatrix}\end{aligned}$$

The equations of motion follow when we equate the acceleration $\ddot{\vec{r}}$ of the location with the accelerations of gravity and thrust (we ignore the drag deceleration and assume that the travel starts from an orbit around the earth with a high altitude, for example 2000 km above the surface of the earth). The earth attracts the sailcraft with a force that is inversely proportional to the square of distance \vec{r} , where the gravitational constant is $\gamma_{\oplus} \approx 3.9866625 \cdot 10^{14} \text{ m}^3/\text{s}^2$. Then, in order to get the vector of the gravitational acceleration, we multiply the magnitude of the gravitational acceleration with the inverted direction of the unit vector that locates the sailcraft with respect to the center of the earth. Finally we can write down the equations of motion as:

$$\begin{pmatrix} \ddot{x} - 2\dot{y}\omega_{orbit} - x\omega_{orbit}^2 \\ \ddot{y} + 2\dot{x}\omega_{orbit} - y\omega_{orbit}^2 \\ \ddot{z} \end{pmatrix} = -\frac{\gamma_{\oplus}}{\sqrt{x^2 + y^2 + z^2}^3} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + s \cos^2 \iota \begin{pmatrix} \cos \zeta \\ \cos \zeta \sin \iota \\ \sin \zeta \sin \iota \end{pmatrix} \quad (9-25)$$

Equation (9-25) is the vector form of the equations of motion for the earth escape trajectory of a sailcraft; ω_{orbit} is the angular velocity of the earth orbit around the sun. When we want we can neglect the terms $x\omega_{orbit}^2$ and $y\omega_{orbit}^2$; but we better consider the terms $-2\dot{y}\omega_{orbit}$ and $+2\dot{x}\omega_{orbit}$: they describe the effects caused by the rotation of the coordinate system and, at a velocity of 6 km/s, for example, their values are with 2.4 mm/s^2 greater than the thrust acceleration of the sailcraft ($\approx 0.6 \text{ mm/s}^2$).

A nearly optimum sailcraft escape strategy. The next step is that we find a steering programme that brings the sailcraft in a minimal time from a low earth orbit to interplanetary space. Remember that a tangential thrusting programme provides a good (but suboptimal) solution to the problem of low thrust earth escape. However, for sailcraft the thrust magnitude is a function of the thrust direction, and therefore we must replace the tangential thrusting programme by another navigation strategy. The intention is to let the mechanical energy of the sailcraft grow rapidly. A fundamental law of mechanics states that the change in time of the specific energy e of the orbit is the dot vector product of velocity and thrust acceleration; thus:

$$\dot{e} = \dot{\vec{r}} \cdot \vec{s} = s \cos^2 \iota \cdot ((\dot{x} - y \omega_{orbit}) \cos \iota + (\dot{y} + x \omega_{orbit}) \cos \zeta \sin \iota + \dot{z} \sin \zeta \sin \iota) \tag{9 - 26}$$

Now we want to control the thrust direction in a way that the orbit accepts at any instant the maximum increase of mechanical energy. We know that such a steering strategy is actually not the real optimum, but the principle leads to a nearly optimum escape trajectory that can easily be integrated by a numerical integration procedure (and we can say that in any case the calculated trajectory complies with the reality, however, a more sophisticated control strategy would bring us even better results).

Let us first consider the control of the thrust angle ζ . When we restrict the angle ι as $0 \leq \iota \leq 90^\circ$ for a unique determination of the thrust direction, the angle ζ may assume any value between 0° and 360° . Since $\sin \iota$ is then always positive, the following control of the angle ζ maximizes the growing of the mechanical energy e of the orbit:

$$\sin \zeta = \frac{\dot{z}}{\sqrt{(\dot{y} + x \omega_{orbit})^2 + \dot{z}^2}} ; \quad \cos \zeta = \frac{\dot{y} + x \omega_{orbit}}{\sqrt{(\dot{y} + x \omega_{orbit})^2 + \dot{z}^2}} \tag{9 - 27}$$

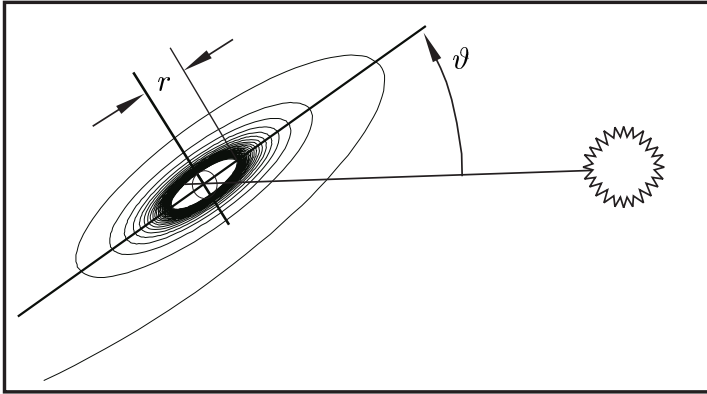
Derivation of a corresponding law to control the angle ι requires more complicated transformations, because the thrust magnitude is now a function of this angle (the factor $\cos^2 \iota$). When we form $\partial \dot{e} / \partial \iota = 0$ and divide the result by $\cos^2 \iota$, then we find a quadratic equation for $\tan \iota$. Fortunately, since the angle ι is positive by definition, the quadratic equation has just one positive solution. The result is:

$$\tan \iota = \frac{-3(\dot{x} - y \omega_{orbit}) + \sqrt{9(\dot{x} - y \omega_{orbit})^2 + 8(\dot{y} + x \omega_{orbit})^2 + 8\dot{z}^2}}{4\sqrt{(\dot{y} + x \omega_{orbit})^2 + \dot{z}^2}} \tag{9 - 28}$$

This equation serves for the numerical integration of the sailcraft escape trajectory.

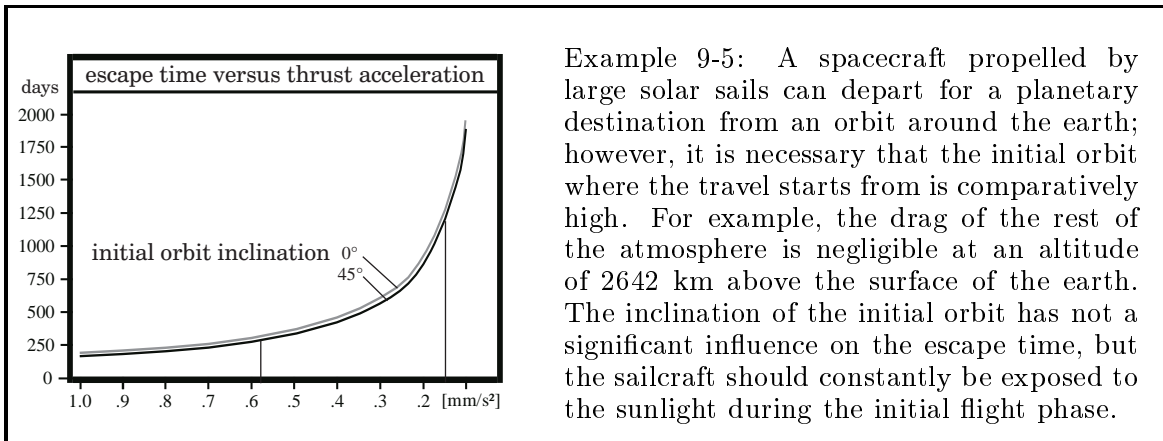
Initial altitude and inclination. The optimization objective for a sailcraft escape trajectory is the minimization of the flight time; and we may conclude that a control strategy is nearly the optimum when it increases the mechanical energy rapidly. However, the mission duration of the earth escape maneuver is obviously also a function of the altitude r of the initial orbit and its inclination ϑ . The change of energy which is necessary to escape from a circular low earth orbit into interplanetary space is given by the expression $\Delta e = \gamma_{\oplus} / r$; this states trivially that Δe gets smaller for greater values of r . The inclination ϑ between initial orbit and the ecliptic has an influence on the time the sails can push the spacecraft into the favorable direction.

Let us first consider the altitude r of the initial orbit. A space station orbit would allow that astronauts deploy the fragile structures of the solar sails. Unfortunately,



the drag deceleration of the rest of the atmosphere is still too high on a low earth orbit; a typical value for the aerodynamic pressure on a circular orbit at 500 km altitude is $2 \cdot 10^{-5} \text{ N/m}^2$ (about twice the value of the light pressure). The sail deployment must take place automatically at a considerably higher altitude.

It is obviously advantageous to avoid flight phases with a solar eclipse or with a motion in a direction towards the sun, because during these flight phases the sails cannot accelerate the sailcraft. However, the earth encircles the sun, and even when we choose an initial orbit with the sunbeams perpendicular to the orbital plane, the sun comes every year twice into an unfavourable position (the sailcraft needs several years to escape from the earth). We have to select a high inclination angle ϑ to ensure that at least initially the sailcraft is always exposed to the sunlight, for example $\vartheta > 45^\circ$.

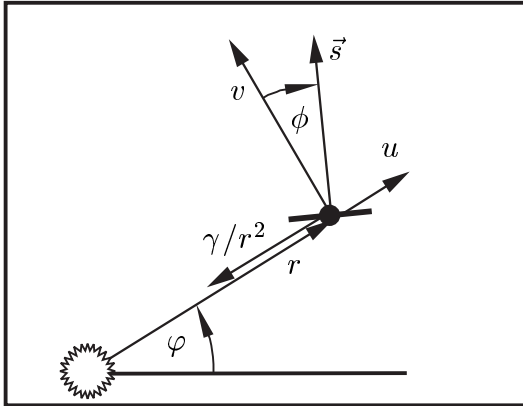


Example 9-5: A spacecraft propelled by large solar sails can depart for a planetary destination from an orbit around the earth; however, it is necessary that the initial orbit where the travel starts from is comparatively high. For example, the drag of the rest of the atmosphere is negligible at an altitude of 2642 km above the surface of the earth. The inclination of the initial orbit has not a significant influence on the escape time, but the sailcraft should constantly be exposed to the sunlight during the initial flight phase.

Let us choose a circular orbit at an altitude of 2642 km above the surface with an inclination of 45° as a basis for the departure of the sailcraft. To escape from this orbit the sailcraft needs about ten months, assuming that extremely light weight sails develop a maximal thrust force that is equivalent to 10% of the solar gravity ($s=0.587 \text{ mm/s}^2$). However, when we assume a more realistic value for the thrust acceleration, for example $s=0.147 \text{ mm/s}^2$ only (four times smaller), then the sailcraft needs more than three years to escape. The actual problem is that the sailcraft has to assume the appropriate orientation with respect to the sunlight during the entire mission, and this means that the vehicle must rotate on its axis while it orbits around the earth. The attitude control system must constantly establish and keep the right orientation (either by the application of gyroscopic actuators or by a shift of the center of mass).

9.3.2 Sailing in Interplanetary Space

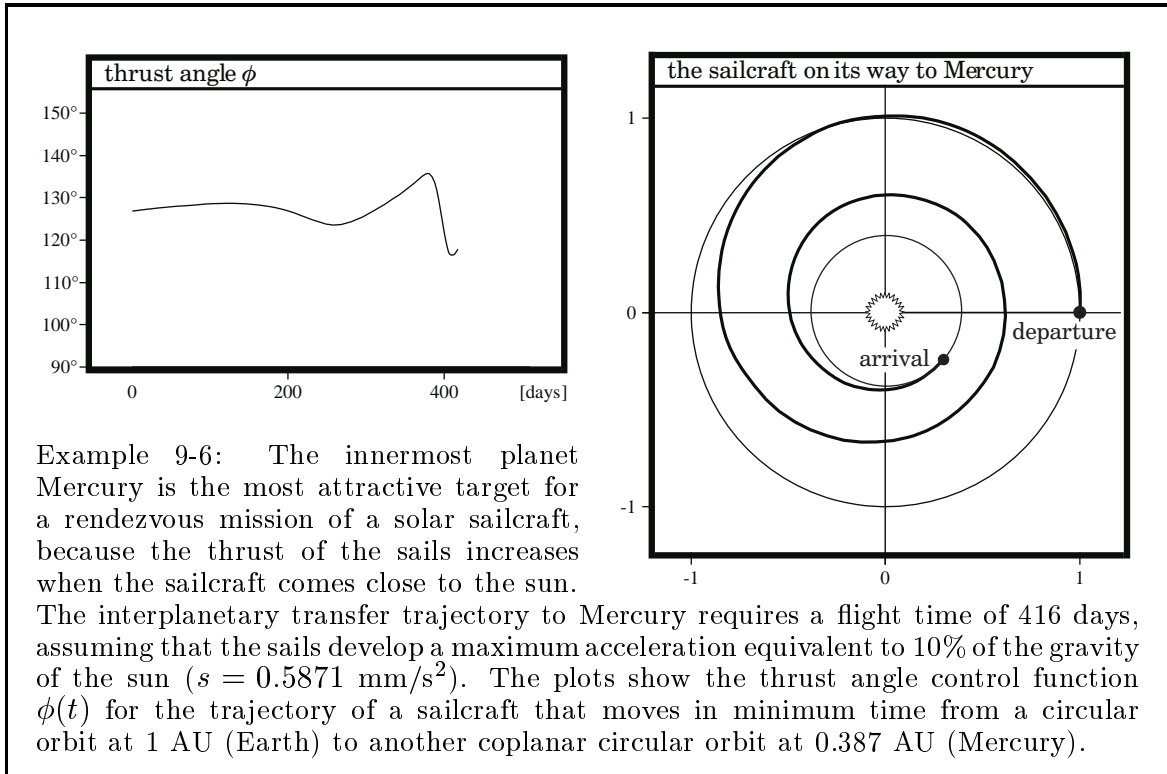
Equations of motion. In interplanetary space the sailcraft demonstrates a better performance than on earth orbit: the travel to other planets needs time anyway, and the sails can take advantage of the long mission duration to accelerate the sailcraft. However, the thrust acceleration s of the sails diminishes inversely proportional to the square of the distance r from the sun; the factor is $(r_{\oplus}/r)^2$ with $r_{\oplus}=1$ AU. The thrust is also a nonlinear function of the inclination of the sails with respect to the incident light (equation 9-23). Let us concentrate our attention only on the plane motion.



The equations of motion for the plane sailcraft trajectory can be written as:

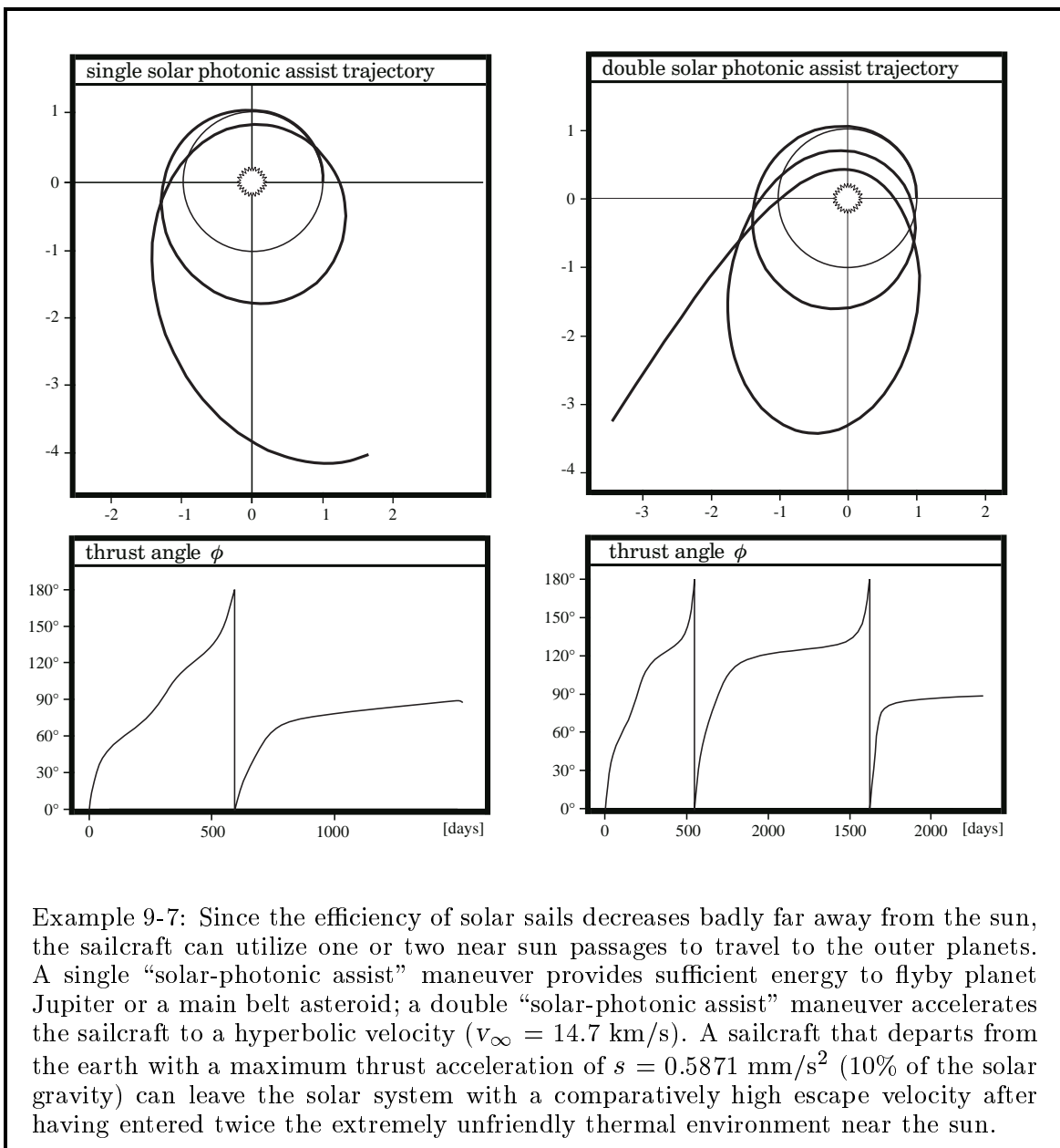
$$\begin{aligned} \dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} + s \left(\frac{r_{\oplus}}{r}\right)^2 \sin^3 \phi \\ \dot{v} &= -\frac{u v}{r} + s \left(\frac{r_{\oplus}}{r}\right)^2 \sin^2 \phi \cos \phi \\ \dot{r} &= u \\ \dot{\phi} &= \frac{v}{r} \end{aligned} \tag{9 - 29}$$

The derivation of the equations of motion is explained in the chapter four of this book. The trajectory is controlled by an appropriate time function for the thrust angle ϕ (with $0 \leq \phi \leq 180^\circ$), r and ϕ are polar coordinates; u and v velocity components.



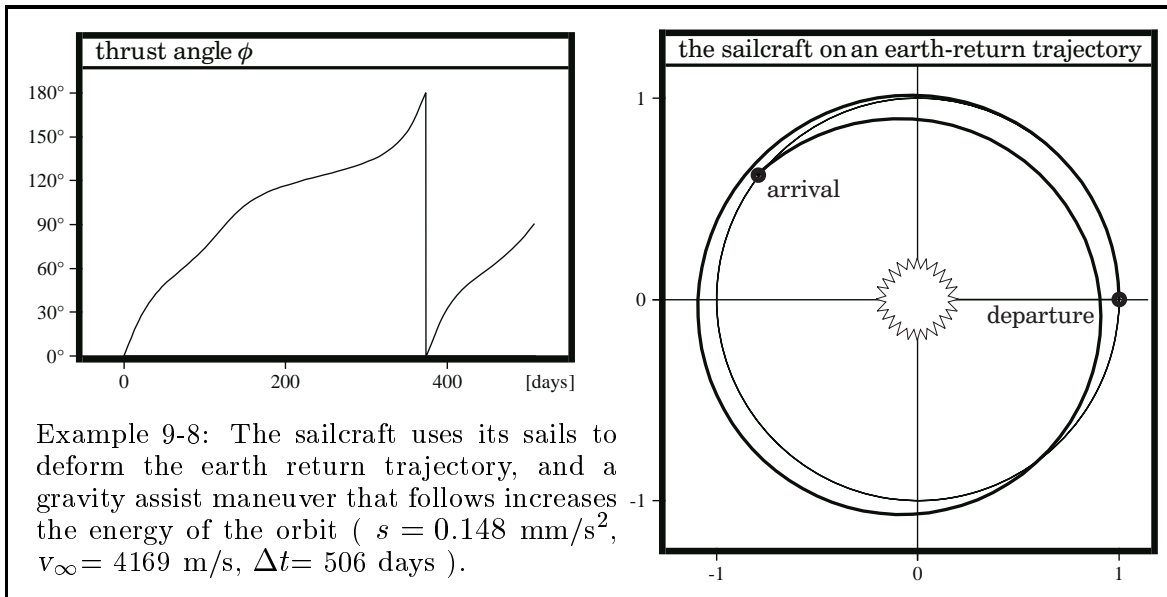
Sailcraft trajectory optimization. It is easy to integrate the sailcraft trajectory numerically when the thrust angle time function $\phi(t)$ is well-known (equation 9-29). However, the mission should be performed in a minimal time interval and the trajectory must satisfy certain final conditions, and the concerning time function for the angle ϕ is usually not well-known. The appropriate tool for the optimization of sailcraft trajectories is the Hamilton-Lagrange theory (also called “calculus of variations”). The application of this mathematical method transforms the nonlinear optimization problem into the numerical solution of a “two-point-boundary-value” problem.

The application of the Hamilton Lagrange theory is not quoted here but demonstrated in the chapter five of this book for various trajectory optimization problems.

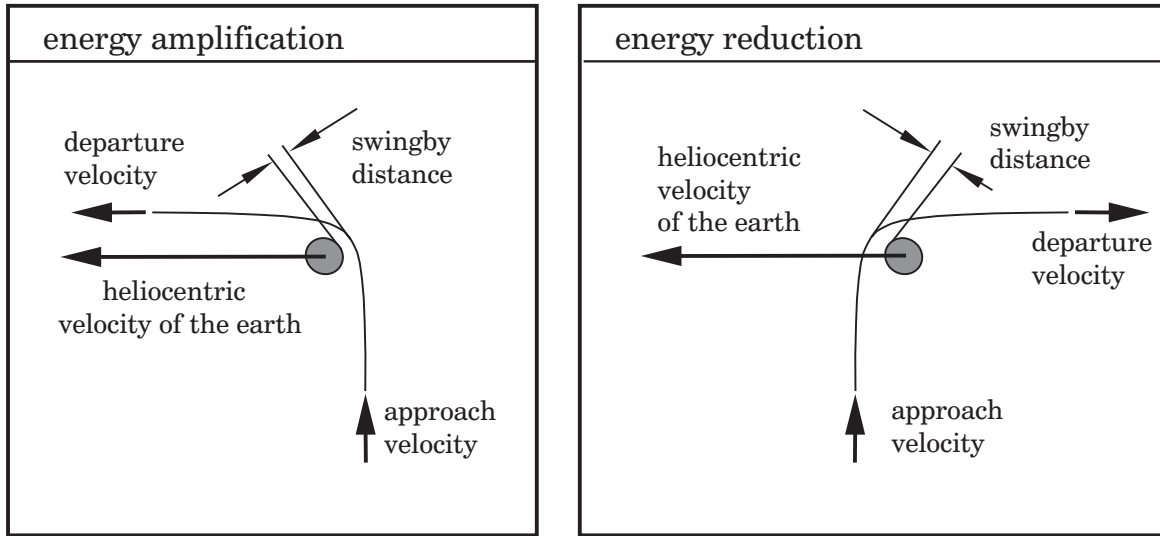


9.3.3 The Sailcraft on an Earth-Return Trajectory

Gravity swingbys. The mission of a sailcraft to a distant planet improves greatly when we include a gravity swingby maneuver; and therefore the departure planet is the most attractive candidate. Assume that the sailcraft has already left the sphere of gravity influence of the earth and now, however, instead of heading for the distant planet, the sailcraft re-targets the earth for a gravity assist maneuver. During a time period of roughly 1.4 years the sails are effective to deform the earth return orbit. By this the sails do not put energy into the trajectory but they amplify the velocity of the sailcraft in relation to the earth (the orbital period of the return orbit remains approximately one year). The capacity of solar sails to deform the elliptic return orbit is much better than their capacity to increase the heliocentric energy of the orbit. When the spacecraft returns to the earth after some time it executes a gravity swingby maneuver which increases the heliocentric energy of the trajectory considerably.



The example above demonstrates the principle: now we can equip the sailcraft with smaller sails; a maximum thrust acceleration of 2.5% of the solar gravity suffices to execute the maneuver (only $s=0.148 \text{ mm/s}^2$ at 1 AU). The sailcraft escapes from the earth with a parabolic speed and drifts later in “near earth space” with a small relative velocity. During the drifting period of 506 days the sailcraft uses its sails to deviate the heliocentric orbit; the optimization objective is the maximization of the relative flyby velocity at the earth encounter that follows. Then the sailcraft approaches our planet with a relative velocity of $u= 4169 \text{ m/s}$, (coming directly from the sun for an observer on the earth) and passes the earth at an altitude of 3123 km. The gravity turns the relative velocity in a direction parallel with the heliocentric velocity of the earth (deflection angle 90°); either in the same direction for a journey to Mars or in the opposite direction to reduce the velocity for a journey to Venus (it depends on the side where the sailcraft passes the earth whether the relative velocity must be added to or subtracted from the heliocentric velocity of the earth).



Journeys to the outer planets. A gravity swingby maneuver at the earth can utilize the relative velocity of the sailcraft at earth encounter, and a relative velocity of 4169 m/s suffices for a mission to Mars. The sailcraft can also execute the swingby maneuver in a way that it reduces the heliocentric energy of the orbit, and enter the near sun space with the intention to perform a “solar-photonic assist maneuver”. However, when the objective is an amplification of the heliocentric energy of the orbit, the best alternative is that the sailcraft utilizes the swingby maneuver to re-target the earth once more on another earth return trajectory for another swingby maneuver.

The second earth swingby maneuver, which takes place about two or three years after the first swingby maneuver, puts enough energy into the trajectory for a mission to a main belt asteroid or to planet Jupiter. The method of repeatedly returning to the earth for a gravity swingby maneuver can be continued, however, the flight time on the earth return orbit becomes long while the energy of the orbit grows. When the sailcraft should flyby a far distant planet like Neptune or Pluto, a third earth gravity assist maneuver on a five-year (or six-year) earth return trajectory becomes necessary.

Solar sails in comparison with gravitational maneuvering. We have seen that the return for a gravity swingby maneuver at the earth can greatly improve the maneuvering capability of a sailcraft in interplanetary space. However, when we intend to incorporate also other planets like Mars and Venus into a sailcraft trajectory, then the main disadvantage of the sail propulsion concept becomes obvious: a spacecraft that is capable to flyby Mars coming from the earth can use exclusively the gravity of the planets. For most missions which are interesting in practice the sails are not necessary at all. Thus, in order to be an attractive propulsion concept, the sailcraft must at least save the propellant that is usually required for the interplanetary injection; and this means that it must depart from a low earth orbit to interplanetary space. The departure from a low earth orbit involves an automatic deployment procedure, a long prevail time in the radiation belt of the earth and many attitude maneuvers. Until today no realistic attempt has been made to realize an interplanetary sailcraft.

10. Reentry Maneuvers

What we call an atmospheric reentry maneuver is the return of a spacecraft back from space down to the surface of the earth. This return flight capability with a soft landing on the surface is only occasionally necessary for unmanned spacecraft, however, it is an essential premise in manned space flight. It became obvious that on principle such a recovery was feasible when the extraterrestrial nature of meteorites was recognized. During the second world war Germany developed the first operational space rocket: the V2 missile. The V2 climbed on a suborbital trajectory up to space and returned. Even when its “payload” (a bomb) did not explode, the impact of the vehicle with an enormous kinetic energy caused a big crater in the terrain. Later Russia and America experimented the recovery of spacecraft, first unmanned and then manned vehicles. On August 10th, 1960, sea divers recovered parts of the American Discovery 13 satellite from the ground of the Pacific Ocean near Hawaii; and on April 12th, 1961, the Russian Vostok spacecraft returned from orbit with Yuri Gagarin on board. Today, return of America’s space shuttles and Russia’s capsules seems to be routine, but the accident of the Space Shuttle Columbia has demonstrated that the maneuver is still dangerous. Forces which act on a reentry vehicle are strong, and not everything what happens in the flow field around a reentry vehicle is actually well-known. Many thermodynamic phenomena do still resist a numerical computation. The design of a reentry vehicle is an iterative process, it relies on inaccurate computations, wind channel experiments and on the comparison of the results with measurements from test flights. It is normal that in some flight phases the computed loads deviate more than by factor two from the values which are then encountered during the real mission.

A body that returns from space must dissipate on its reentry trajectory an enormous amount of kinetic energy. An essential part of the heat load must be transferred to the air flow that surrounds the body, otherwise the body would melt and vaporize. Therefore the reentry vehicle must be a blunt body; the vehicle should not have a slender fuselage or wings with sharp leading edges exposed to the incident air flow. This fundamental design law for reentry vehicles was first recognized and studied by H.J Allen [“Hypersonic Flight and the Re-Entry Problem”] and J.A.Fay and F.R.Riddell [“Theory of Stagnation Point Heat Transfer in Dissociated Air”], both articles were published in the Journal of the Aeronautical Sciences, Vol.25, 1958. In this chapter we will discuss the thermodynamic phenomena that can be encountered in the different flow regimes on a reentry trajectory. However, we will not study the way how to solve the complicated partial differential equations which describe the flow field around a reentry vehicle. Computational fluid dynamics applied to reentry vehicles you can find, for example, in the text books of J.D.Anderson [“Hypersonic and High Temperature Gas Dynamics”, McGraw-Hill, New York, 1989] and H.Oertel [“Aerothermodynamic”, Springer Verlag, Berlin Heidelberg, 1994]. Flight mechanics and attitude stabilization during the reentry maneuver is treated in the comprehensive book of F.J.Regan and S.M.Anandakrishnan [“Dynamics of Atmospheric Re-Entry”, AIAA Education Series, Washington DC, 1993]. The numerical results demonstrated in this book were generated using special computer programs.

10.1. High-Speed Gasdynamics

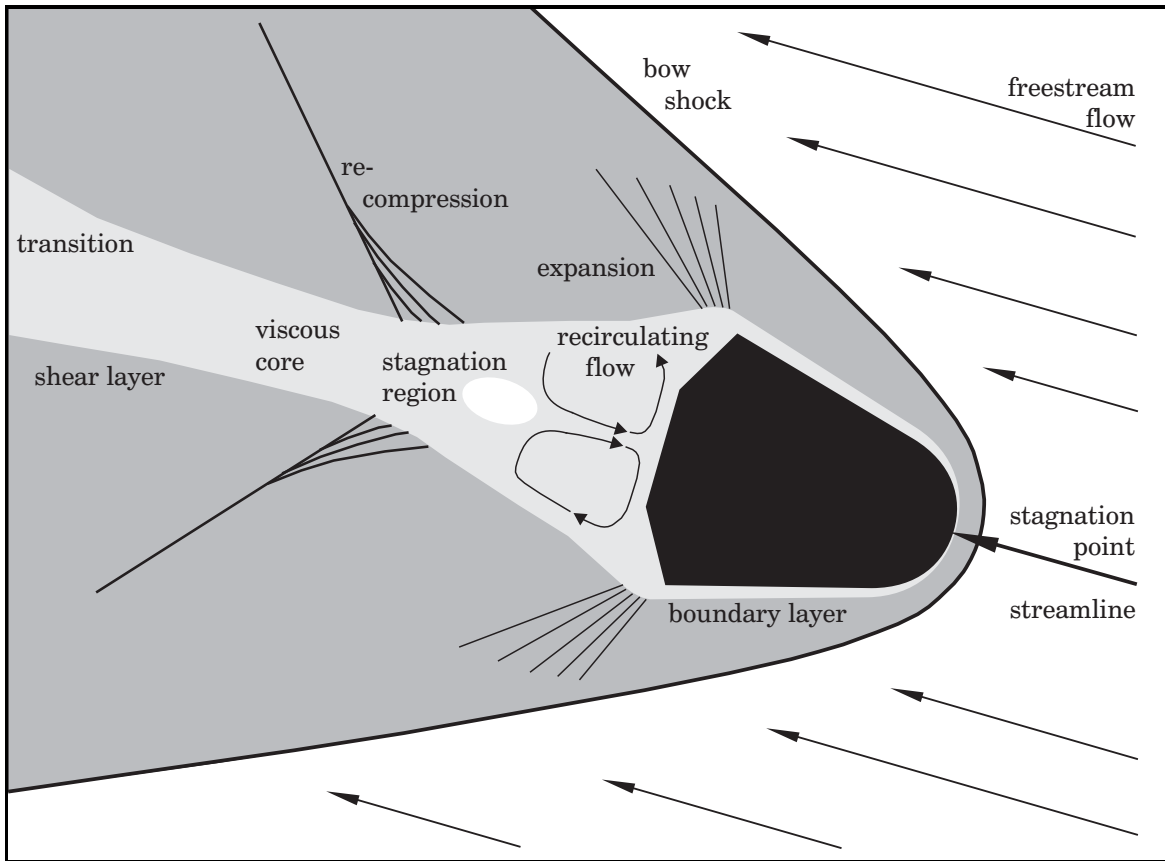
Gasdynamics is a comparatively new field of mechanics, particularly when we consider the numerical computation of complicated three dimensional flow fields. The analysis of the flow field that surrounds a reentering space vehicle involves some effects which make the problem even more complicated: the energies involved are extraordinarily high so that we cannot treat the gas as a “perfect gas” anymore. Chemical reactions take place and “real gas” effects have to be taken into consideration. The speed on a reentry trajectory ranges, for example, from Mach 27 down to zero; and the ambient density ranges from vacuum to the normal ground level pressure. The wide range of different velocities and different densities makes it necessary that we divide the trajectory into sections with quite different conditions for the analysis of the flow field.

10.1.1 Flow Regimes

Free molecular flow regime. Consider a spacecraft that moves on a circular low earth orbit. The reentry maneuver is initiated with a small retro-thrust. During the initial descent the vehicle gets more and more into contact with gas molecules of the upper atmosphere. The vehicle is quite fast and the atmosphere is extremely thin. Collisions take place between gas molecules and the surface of the vehicle, however, the “mean free path length” of the rarefied gas is actually much longer than the length of the vehicle. Thus, collisions of gas particles reflected from the surface of the vehicle with gas molecules from the “free flow” take place far behind the vehicle, and essentially these collisions do not influence the flow field in the near vicinity of the vehicle. The flow field in the vicinity of the reentry vehicle is dominated by particle-surface collisions; and the incident flow can be treated as a collisionless flow. A small part of the hot gas that surrounds the vehicle is ionized; the ionization makes the air to an electrically conductive plasma and prevents transmission of radio signals to and from the vehicle (for the US-Space Shuttle, the 20 minutes “radio blackout” begins at an altitude of 134 km and lasts until an altitude of approximately 54.8 km).

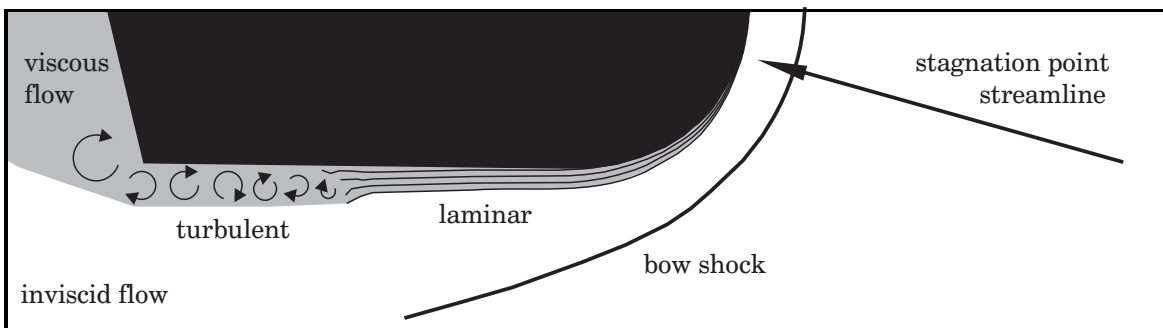
Transition from free molecular to continuum flow regime. At an altitude of about 90 km above the surface of the earth the vehicle enters the transition regime. The vehicle moves extremely fast with a speed of about Mach 27 and the gas density is low. Mutual interactions between gas particles take place, these interactions start to form a shock wave at a distance of about 10 free path lengths in front of the vehicle (however, the characteristic phenomenon of a bow shock is not really present now). The compression of the air behind the shock wave causes an enormous increase of the gas temperature, and real gas effects occur: essential portions of the energy liberated in the collisions between gas molecules stimulate molecular vibrations, break molecules into atoms and activate chemical reactions. The flow field in front of the vehicle is not in equilibrium, because interactions between gas particles with exchange of energy are rare and the gas velocity is extraordinarily high.

At the opposite (rear) side of the vehicle the gas flow expands rapidly. Lighter atoms congregate behind protecting surfaces, they form a stagnation region with frozen chemical equilibrium nearly as hot as the stagnation region in front of the vehicle. A shear layer separates the hot rear stagnation region from the cold outer flow field.



Continuum flow regime. The regime begins with a velocity of about 6 km/s or Mach 19 at an altitude of about 75 km above the surface of the earth. Now the mean free path length of the gas particle motion is short enough so that the flow field can be treated as a “continuum”, considering the state of the gas (velocity, density, temperature, chemical composition and so on) at every location in the flow field, but ignoring that actually the gas consists of individual particles with individual behaviour. The characteristic bow shock wave is present in front of leading edges of wings and in front of the nose of the fuselage (at a distance of about 6% to 12% of the nose radius). The gas experiences a considerable increase in temperature and density when it travels through the shock wave, however, the stagnation point is not as hot as it would be in comparison with the compression of a perfect gas. Real gas effects absorb essential parts of the energy (about 50%, depending on the velocity). The vibrational modes of the oxygen and nitrogen molecules are excited, the gas is partly dissociated and chemical reactions take place (which form nitrogen monoxide molecules). Behind the bow shock wave the gas comes to its equilibrium only in a small region near the stagnation point, at other locations the velocity is so extraordinarily high and the density so low that the gas remains in a non-equilibrium state.

Frictional effects in the gas flow form the boundary layer behind the bow shock wave. The boundary layer is a thin layer attached to the walls of the vehicle. Its thickness is essentially a function of the Mach number and the wall temperature, it equals zero at the stagnation point and grows approximately proportional to the square root of the distance from the stagnation point. The gas velocity at the inner edge of the boundary layer corresponds to the velocity of the vehicle; the velocity at the outer edge corresponds to almost 99% of the undisturbed “freestream” velocity. The temperature profile in the boundary layer is dominated by the thermal and catalytic behaviour of the surface material of the vehicle (its influence on chemical reactions, as well as its heat conductivity and radiation coefficient). The heat flux from the gas flow to the walls increases considerably when the boundary layer changes form “laminar” to “turbulent”.



Such a transition can be expected at an altitude of about 50 km, and it is advisable to design the vehicle in a way that the transition from laminar to turbulent takes place at an altitude where the maximum heat flux has already been passed through. The turbulent boundary layer is much thicker than the laminar layer. The transition begins at the rear end of the vehicle and moves to the front while the vehicle descends. Shock waves from other parts of the vehicle cause hot spots when they interact with the boundary layer: for example such a situation is present on the surface of wings which intersect the cone that is formed by the bow shock of the nose of the vehicle.

Perfect gas flow regime. The real gas effects disappear when the vehicle decelerates from hypersonic to supersonic velocity. Air dissociation, chemical reactions and the vibrational modes of the molecules play no significant role anymore at flight altitudes below 30 km and velocities below Mach 2. The temperature in the stagnation point is now cold enough so that the air may be considered as a perfect gas. The gas in the boundary layer at the side walls of the vehicle’s fuselage is now in equilibrium. The dynamic pressure is the dominant load for the vehicle in the supersonic flight regime. The flow field changes once more when the vehicle crosses the sonic point during the deceleration from supersonic to subsonic speed: the bow shock moves forward and gets more detached from the nose of the vehicle; and the cone angle grows. When the vehicle moves with the velocity of sound, weak shock waves accept nearly rectangular cone angles before they finally disappear. When the vehicle moves slower than Mach 0.5, it can also be ignored that the air is compressible. In case the vehicle is a capsule it can now deploy its parachutes. In case the vehicle is a space shuttle, it can now establish the final glide slope for the landing approach.

10.1.2 Aero-Thermodynamic Phenomena

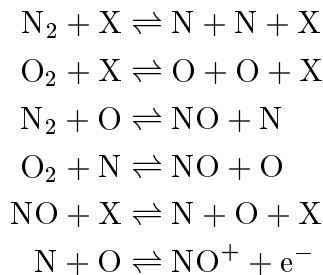
Translational, rotational and vibrational excitation of molecules. Let us now discuss the real gas phenomena which appear in a hypersonic flow field and which make the accurate computation of the flow field so difficult. The so-called “inner energy” of a gas is a thermodynamic state variable, it is actually the sum of the energies of all the gas particles. For a perfect gas the inner energy is exclusively a function of the temperature, for a real gas it is also a function of the pressure. Single-atomic molecules (inert gases, for example) can just accept kinetic energy in form of translational velocity. Double-atomic molecules however, for example nitrogen N_2 or oxygen O_2 , can also accept kinetic energy in form of rotation. The probability that after an impact a gas molecule accepts rotational energy is described by the Boltzmann distribution, considering the physical effect that the molecules can accept continuous portions of translational energy but only discrete portions of rotational energy (quantum physics). When the gas gets hot, impacts between the particles excite also molecular vibrations, where again energy is transferred in form of quantized portions. To excite the vibrational modes at the normal ambient pressure of 1 bar, for example, oxygen must be hotter than 400 K and nitrogen hotter than 600 K.

We say that the gas is in its “thermodynamic equilibrium” when the distribution of the inner energy among translational, rotational and vibrational form corresponds to the Boltzmann distribution (or Maxwell distribution). To come to this state the gas requires a finite time interval, called relaxation time (the gas needs a sufficient number of impacts between the gas molecules). In the upper layers of the atmosphere the gas density is extraordinarily low. Impacts between the gas molecules are seldom in the rarefied gas, and consequently the gas is not in its thermodynamic equilibrium state at many locations in the flow field that surrounds the reentry vehicle.

Dissociation. The collisions start to break the molecules apart when the temperature grows (for example above 2000 K for oxygen and above 4000 K for nitrogen); and the gas changes its chemical composition. It consists now of more particles, and the new single-atomic molecules can just accept kinetic energy in form of translational velocity. The concentration of atomic nitrogen N and atomic oxygen O in the hot gas is mainly a function of the temperature but also a function of the pressure: the concentration increases with growing temperature and with dropping pressure. Oxygen is nearly completely dissociated at temperatures higher than 4000 K, nitrogen is nearly completely dissociated at temperatures above 9000 K.

Ionization. The production of electrons and positively charged ions plays not a significant role in the absorption of energy on a reentry trajectory, however, it is responsible for the radio transmission blackout. Atomic oxygen and atomic nitrogen get ionized at high temperatures and low densities, but most of the ionization of the hot air is the consequence of the generation of ions from nitrogen monoxide (NO^+). However, only a small part of the nitrogen monoxide gas ionizes (approximately 1%).

Chemical reactions. The earth's atmosphere consists mainly of 78% nitrogen N_2 , 21% oxygen O_2 (and about 1% other gases). This composition does nearly not change with the altitude. However, chemical reactions take place in the hot air that surrounds a reentering space vehicle, the molecules break apart and nitrogen monoxide is formed. Neglecting an insignificant amount of other species, the plasma in the flow field around the vehicle can be considered as composed of seven major constituents: N_2 , O_2 , N , O , NO , NO^+ and electrons e^- . Then the chemical reactions responsible for the gas composition can be described by the following set of equations. The term X is used to specify always a different member of the seven species. Thus:



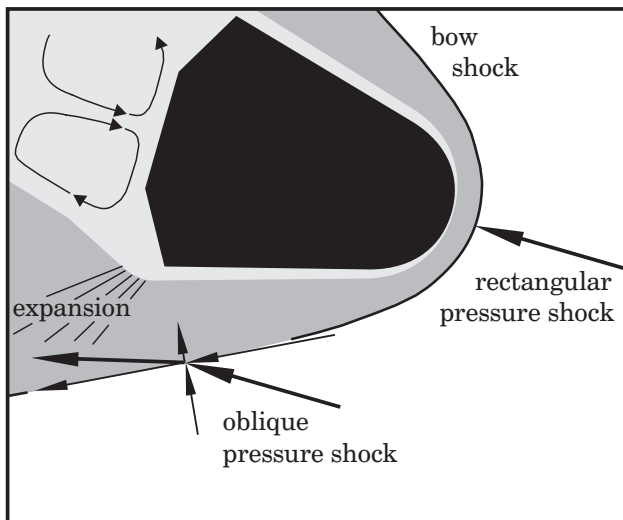
The first two equations describe the dissociation of nitrogen and oxygen, the third and fourth equation describe the generation of nitrogen monoxide, the fifth equation describes the dissociation and the sixth equation the ionization of nitrogen monoxide. The last equation does nearly not contribute to the energy balance of the reactions, however, it should be considered in order to determine the ionization state of the gas.

When a reversible chemical reaction is in its equilibrium, an "equilibrium constant" can be devised that relates the partial pressures of the participants of the reaction (the actual value of this equilibrium constant depends sensitively on the temperature). Together with the principle of the conservation of matter (no atomic species gets lost during the reaction) the equilibrium constant enables us to compute the equilibrium composition of the gas as a function of temperature and pressure. Once we have found the composition of the gas we can readily calculate the "heat of reaction". The heat of reaction follows from the "heat of formation" (tabulated in textbooks) and from the concentration of all the species which participate in the reaction. Knowing the heat of reaction we know finally the amount of energy that must be transferred to the gas in order to reach the corresponding thermochemical state.

The chemical non-equilibrium. It is comparatively easy to compute the equilibrium composition of hot air as a function of temperature and pressure (the computation requires only the solution of a nonlinear equation system). In textbooks on aerothermodynamics we find the molar composition of the air as a function of the temperature plotted for several different pressure levels. However, the problem gets much more complicated when also the velocity of the chemical reactions must be considered. These reactions are fast but they require finite time intervals. The velocity of the air in a flow field surrounding a reentry vehicle is so extremely high that often the chemical reactions do not reach at all positions in the field the equilibrium state.

The velocity of a reversible chemical reaction depends on the temperature and on the concentration of the participants. At the beginning the reaction proceeds rapidly towards the equilibrium state, but when it gets near to the equilibrium it slows down. The individual values of the rate coefficients for reversible reactions can be determined experimentally (and sometimes theoretically). However, many of these values are not accurately known for gas under extraordinary conditions, and thus the computation of a reentry flow field depends usually on simplified chemical reaction models.

Shock waves and expansion waves. A shock wave is a phenomenon that occurs in every supersonic flow field. In the supersonic flow field that surrounds a reentering space vehicle a distinct “bow shock wave” can be observed, a cone with rounded tip. The thermodynamic change in state is usually considered as a rectangular shock wave in the rounded front section immediately in front of the nose of the vehicle, and as an oblique shock wave in the conic section. The shock wave increases instantaneously the thermodynamic state variables temperature, pressure and density, while the energy (or total temperature) of the gas remains constant. The entropy grows when the gas travels through the shock wave, indicating that total pressure (or stagnation pressure) is lost in the shock wave, and that only an instantaneous compression is possible (and not an instantaneous expansion). Behind the vehicle the gas expands rapidly, however, in contrast to the compression the expansion of the gas is a continuous function and, essentially, the entropy remains constant. Temperature, pressure and density drop while the gas velocity increases. The magnitude of the thermodynamic change in state depends mainly on the Mach number of the incident air flow.

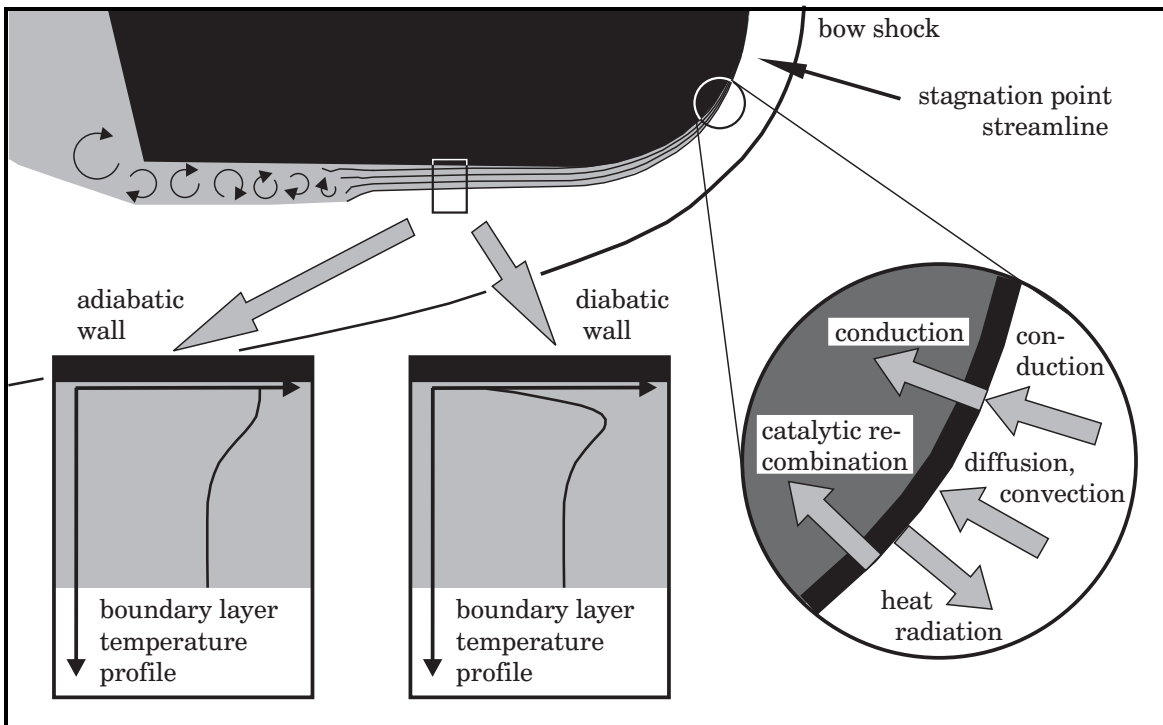


Behind the vehicle the gas expands rapidly, however, in contrast to the compression the expansion of the gas is a continuous function and, essentially, the entropy remains constant. Temperature, pressure and density drop while the gas velocity increases. The magnitude of the thermodynamic change in state depends mainly on the Mach number of the incident air flow.

When in the supersonic flow regime the air is considered as a perfect gas, then the thermodynamic state of the gas immediately behind a rectangular shock wave can be readily calculated from the state of the gas before the shock wave (its temperature, pressure, density and velocity). The concerning relationships are not quoted here, they can be written exclusively as functions of the Mach number of the incident flow and the ratio of the specific heats of the gas κ . The behaviour of the gas that travels through an oblique shock wave can be calculated in a similar manner. Then the velocity vector is resolved into two components, one of which is parallel with the shock wave and the other one is rectangular. The parallel component remains unchanged when the gas travels through the shock wave; and for the rectangular component of the incident velocity the relationships for a rectangular pressure shock are valid. When the air is considered as a perfect gas, it is also possible to calculate the thermodynamic situation for every location in the expanding flow field analytically.

Even though the shock wave itself is not a “real gas effect”, the appearance of the shock in a hypersonic flow field (where real gas effects are considered) is different from the appearance in a supersonic flow field (where real gas effects are ignored). For example, for the velocity Mach 15 and $\kappa = 1.4$ we can calculate a factor 44.7 for the temperature raise of a perfect gas that flows through a rectangular shock wave. However, the stagnation point temperature is actually much colder than 10000 K. The equations valid for shock waves have to be modified in order to consider the real gas effects; and in the hypersonic flow regime the relationships are more complicated.

Boundary layers and heat transfer. Temperature profile and thickness of the boundary layer are both considerably influenced by the properties of the wall material (its surface roughness and chemistry). Inside the boundary layer the hot compressed air accepts heat input from the outer flow field by friction, and thus the temperature in the boundary layer can be hotter than in the outer “freestream” flow field.



The hot gas transfers heat to the skin of the vehicle via convection and conduction. In case the wall material acts as a catalyst, the recombination of dissociated molecules means an additional heat load for the walls. The skin of the reentry vehicle, however, is usually much colder than the gas in the boundary layer (or in the stagnation point): energy is discharged from the walls to outside in form of heat radiation and to inside in form of heat conduction, sometimes the walls are also cooled by the ablation of wall material. Today, the heat transfer models used for the calculation of reentry flow fields are hardly more than a rough estimate. The problem is complicated by the fact that the results of the calculations, the surface temperatures, influence the temperature profile in the boundary layer, its thickness and therefore the complete flow field (the unknown wall temperature is a boundary condition for the flow field).

10.1.3 Computation of Flow Fields

Dimensionless numbers in aerodynamics. In order to assess the validity of wind channel experiments or computer simulations, experimental and computational aerodynamics make use of dimensionless numbers. The intention is to determine whether two different flow fields are “similar” (or comparable) regarding certain aspects, in practice whether experimental or computational results comply with the reality. The problem is that it is usually not possible to achieve comparability of two different flow fields under all the important aspects: the experiments depend usually on the assumption that at least only one characteristic aspect of the experimental flow field is similar to the reality. Three of these numbers play a role in reentry vehicle aerodynamics: the Knudsen number, the Mach number and the Reynolds number.

The Knudsen number relates the “mean free path length between particle collisions” to the “characteristic length of the vehicle”. This number is used to determine what kind of flow regime is present when the vehicle descends through the upper layers of the atmosphere: a free molecular flow regime can be expected when the Knudsen number is greater than 10, and a continuum flow regime can be expected when the Knudsen number is smaller than 0.01. In between the transition regime is present.

The Mach number is the ratio of the “actual velocity” to the “local velocity of sound”. Two different flow fields are comparable regarding the gas compressibility when the Mach number of the flow is the same. The occurrence and form of shock waves depends sensitively on the Mach number: shock waves can occur only in supersonic fields where the Mach number is greater than 1. Shock waves cannot occur in subsonic flow fields where the Mach number is smaller than 1, except for the “transsonic” situation when the vehicle moves nearly with sonic speed and when the actual velocity at some locations is greater than the velocity of sound. The compressibility of the air can be neglected at velocities smaller than Mach 0.5, and real gas effects have to be considered in the hypersonic regime when the Mach number is greater than 3. The Mach number is the most important parameter for supersonic wind channel experiments.

The Reynolds number relates in the flow field “inert forces” to “viscous forces”; this number is used to determine whether a boundary layer is laminar or turbulent. It is assumed that the transition from laminar to turbulent takes place at the moment when the Reynolds number accepts a certain value, called critical Reynolds number. The critical Reynolds number in a compressible boundary layer is influenced by many factors (for example by the shape and size of the vehicle, surface roughness and temperature); however, the theory for the unstable transition from laminar to turbulent is today not satisfactorily developed and thus the critical Reynolds number has to be determined experimentally. The Reynolds number influences the aerodynamic drag of a vehicle and is therefore the most important key parameter for subsonic wind channel experiments. Wind channel experiments where both, Reynolds number and Mach number, comply with the reality are not state of the art yet, but such experiments would be very helpful in developing new economical aeroplanes.

Computation of the free molecular flow regime. Methods of the so-called “kinetic gas theory” are used to evaluate the flow field in the free molecular regime (and in the transition regime): the gas is considered as composed of many individual particles; each particle has its individual position, velocity and energetic state; energy and linear momentum are exchanged between the particles via mutual collisions and impacts on surfaces. The Boltzmann equation can be used to describe the spatial distribution of the particles and the change of this distribution with respect to the time and the location. However, finding a solution to the Boltzmann equation is numerically very difficult and actually just possible in certain simplified situations. Thus, the flow field in the free molecular regime is often evaluated using more direct methods, for example via the individual calculation of trajectories for a great number of representative particles. The occurrence (and the effect) of collisions between particles is either directly calculated or assumed with the help of statistical models.

Computation of the continuum flow regime. A simplification of the Boltzmann equation is possible when the gas can be treated as a continuum, and the result of this simplification is the so-called Navier Stokes equation. The Navier-Stokes equation is valid for the evaluation of flow fields with friction but without shock waves; however, it still suffers the disadvantage that numerical solutions are only exceptionally possible. The Navier-Stokes equation can further be simplified when the friction is neglected, and the result is the so-called Euler equation for frictionless continuous flow fields. The Euler equation can be solved numerically by standard computer programs (under certain conditions it is even possible to solve the Euler equation analytically, for example for the motion of a perfect gas in a plane or axisymmetric flow field). Therefore, a flow field that contains shock waves and boundary layer friction can be intersected and solved, provided that the thickness of the boundary layers and the orientation of the shock waves is well-known. The problem is that exactly this information is not present at the beginning of the simulations; because finding this information is the main result of the computations. The final result should be the accurate pressure and heat load for every location on the surface of the vehicle; but the iterative numerical solution process can be cumbersome and costly, and the quality of the simulation results depends much on the underlying models for the thermodynamic behaviour of the gas and the catalytic behaviour of the wall material.

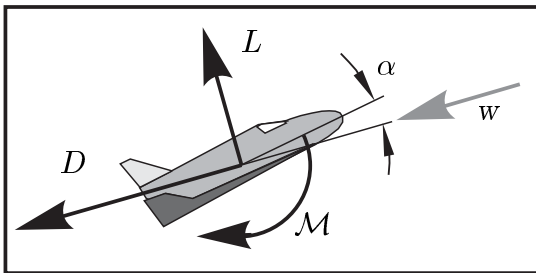
Newton’s approximation. The costly numerical simulation is often not done in preliminary design studies, usually approximation methods are used for the determination of the aerodynamic coefficients. The most important method is the Newton method, postulated by Sir Isaac Newton for the calculation of forces on bodies in incompressible flow fields. It assumes simply that the moving fluid particles hit the surfaces of an obstacle all with the same freestream velocity vector, and that then the linear momentum of a fluid particle associated with the velocity component normal to a surface is transferred to the obstacle, while the linear momentum associated with the velocity component tangential to the surface remains unchanged. The change of momentum corresponds to a pressure force; and the aerodynamic coefficients can be found by an integration of all pressure forces along the entire surface of the obstacle.

The results of the Newton theory comply quite good with the reality, particularly for supersonic flight with high velocity. However, the Newton theory fails in the subsonic velocity regime for which it was initially postulated, because it ignores the random motion of the gas particles (responsible for the static pressure and thus for the forces on shadowed surfaces at the rear end of the vehicle). The Newton theory can be modified to take the influence of the Mach number under consideration (“modified Newton theory”). Beside the Newton method, there exist other impact methods useful for the preliminary aerodynamic design of reentry vehicles, which are based on experimental data in combination with empirical and theoretical knowledge.

10.1.4 Aerodynamic Loads

Aerodynamic forces and moments. Let us now consider a certain position on the reentry trajectory of a vehicle that returns from space. The vehicle experiences at all locations of its surface normal and tangential forces caused by the air flow. The static effect of all these small forces is equivalent to the effect of a single strong force that acts on a certain point of application, called the “center of pressure”. The equivalent aerodynamic force can be resolved into two rectangular components, the one parallel with the incident air flow is called drag, the other one is called lift.

Let us change the attitude of the vehicle is such a way that the lift force vanishes. For example, in case the vehicle has a body with an axis of symmetry, the lift force vanishes when the symmetry axis is exactly parallel with the direction of the incident flow; but the lift force appears when we turn the vehicle so that the vehicle’s axis is inclined with respect to the incident flow. The angle between the axis and the air flow is called α (“angle of attack”). It is obvious that even for slender bodies the drag does not disappear when $\alpha = 0$, however, the drag will accept a minimum then. Experience shows that for small angles of attack the lift force L is approximately a linear function of α , and the drag force D is a nonlinear (quadratic) function of α . Experience shows also that the aerodynamic forces lift and drag are approximately proportional to the dynamic pressure, defined as $Q = \frac{1}{2} \rho w^2$, where w is the velocity of the incident wind and ρ the density of the atmosphere at the actual flight altitude.

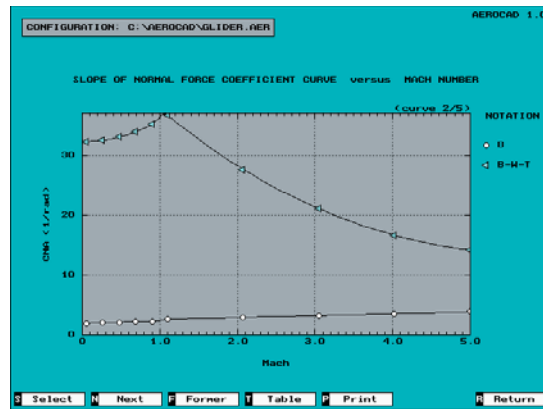
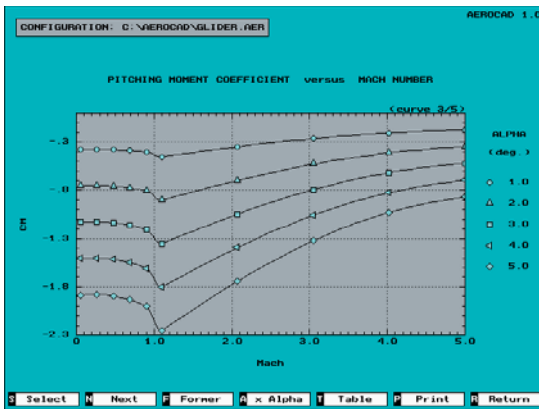
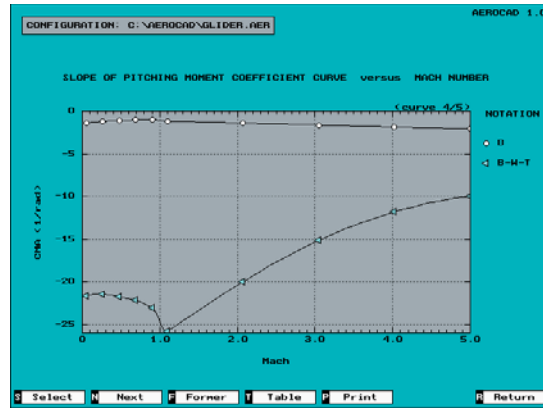
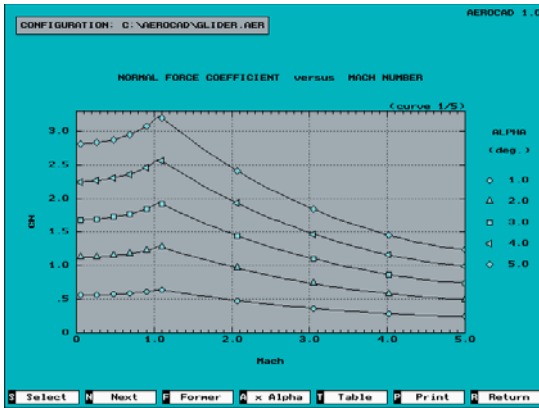
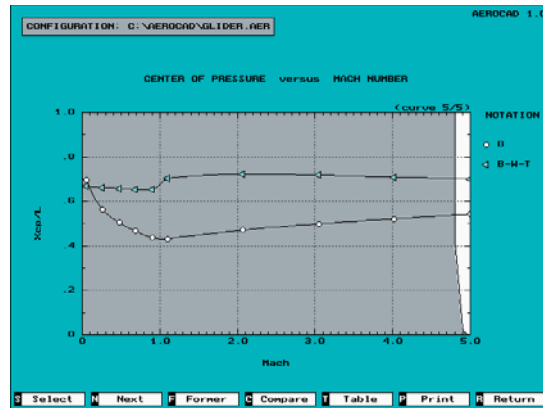
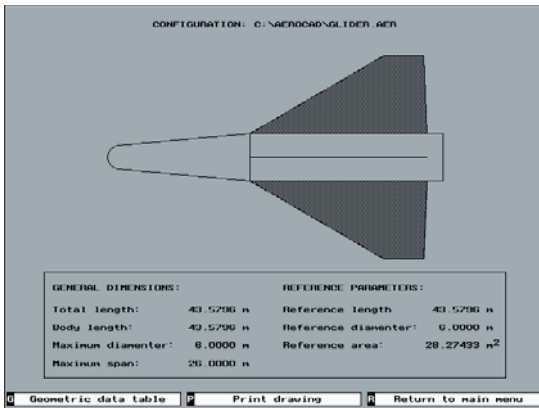


Thus, we can write for the aerodynamic forces lift and drag the following equations:

$$L = \frac{1}{2} \rho w^2 (C_{L\alpha} \cdot \alpha) A \quad (10 - 1)$$

$$D = \frac{1}{2} \rho w^2 (C_{D0} + C_{D\alpha} \cdot \alpha^2) A \quad (10 - 2)$$

Term A is a reference area used to make the aerodynamic coefficients dimensionless. The three coefficients $C_{L\alpha}$, C_{D0} and $C_{D\alpha}$ are only approximately constants; particularly, when a wide range of different velocities is considered, a considerable dependency of these coefficients on the Mach number is noticeable.



Example 10-1: The aerodynamic behaviour of a winged vehicle can be estimated during the preliminary design phase with approximation methods based on theoretical and empirical knowledge. The figures show some of the results of a design study for a winged space glider, performed with the computer program Aerocad: center of pressure, normal force coefficient and pitching moment coefficient as function of the Mach number in the supersonic flight regime.

By definition the aerodynamic force causes no torque in the center of pressure; and, also by definition, the gravitational force causes no torque in the center of mass. The center of mass is the point of application for gravity and for inert forces, and we can conclude that the aerodynamic force may not cause a torque in the center of mass unless it is our intention to change the attitude of the vehicle. Therefore it is necessary that the line of action of the aerodynamic force runs through center of mass. (at the moment lift, drag and gravitation are the only forces that act on the vehicle).

However, when the vehicle's orientation deviates slightly from the nominal attitude, the line of action of the aerodynamic force does not run through the center of mass anymore, and the aerodynamic force causes a torque in the center of mass (the center of pressure is usually not coincident with the center of mass). For stability it is required that the aerodynamic torque is a retaining moment, this means that the torque must turn the vehicle back to its nominal attitude (where it disappears). It can easily be verified that the aerodynamic torque is only then a retaining moment when the center of pressure is always located behind the center of mass. Then we can shift lift and drag to the center of mass and consider the aerodynamic torque as:

$$\mathcal{M} = \frac{1}{2} \varrho w^2 C_{M\alpha} \cdot (\alpha - \alpha_{nominal}) A R \quad (10-3)$$

Equation (10-3) assumes that the aerodynamic moment \mathcal{M} with respect to the center of mass is a linear function of the deviation of the angle α from the nominal "trimmed" value $\alpha_{nominal}$ (and proportional to dynamic pressure $Q = \frac{1}{2}\varrho w^2$). Term A is a reference area and R a reference distance (the actual location of the center of pressure is not constant but changes with the value of the angle of attack α). For aerodynamic stability the pitching moment coefficient $C_{M\alpha}$ with respect to the center of mass must be a quantity that accepts under all conditions a negative sign. When the vehicle is a capsule, its center of mass must be placed close to the heat shield in order to ensure sufficient aerodynamic stability. The heat shield itself is usually a heavy component of the vehicle (it has a mass fraction of 10% to 15%, typically), but a part of the material ablates during the main braking phase. When the vehicle is a glider, large control surfaces are necessary because the vehicle has to trim during the initial entry phase great angles of attack ($\alpha_{nominal}$ between 30° to 40° , for example).

The aerodynamic coefficients $C_{L\alpha}$, $C_{D\alpha}$, $C_{D\alpha}$ and $C_{M\alpha}$ (equations 10-1 to 10-3) vary their actual values substantially during the different flight phases on the reentry trajectory. The accurate evaluation of these coefficients is one of the main objectives of the difficult computation of the flow field that surrounds a reentering space vehicle.

Heat flux in the aerodynamic stagnation point. The other main objective of the numerical simulation is the attempt to calculate the heat load for all locations on the surface. However, as we have seen, due to a missing accuracy of chemical models, the results of these simulations are usually uncertain. Considering that the drag force is approximately proportional to the square of the incident wind velocity w^2 , we can expect that the power associated with the deceleration will be proportional to the expression w^3 (power is the scalar vector product of force and velocity).

Frequently used in preliminary design studies is a simplified equation for the heat flux in the stagnation point. This relationship is the result of early works on the reentry problem, it is based on the solution of the Navier-Stokes equation for the stagnation point streamline in the boundary layer and it considers the heat transfer from the hot reacting gas to the walls of the vehicle. The solution is then greatly simplified in order to come to a practically useful result. It can be shown that the stagnation point heat flux is inversely proportional to the square root of the radius of the (spherical) “nose” of the vehicle. This demands that the vehicle must be a blunt body and that wings and fins must have well-rounded leading edges. Assuming a nose radius of one meter, for example, the heat flux transferred from the gas to the nose is approximately:

$$\dot{q} = 1.705 \cdot 10^{-7} \sqrt{\rho} w^3 \quad (10 - 4)$$

Term \dot{q} is the specific heat flux in the aerodynamic stagnation point in kW/m² (provided that the wind velocity w is specified in m/s and the air density in kg/m³). The equation (10-4) is not at all accurate, it can also be found in literature denoted as $\dot{q} = 1.24 \cdot 10^{-7} \sqrt{\rho} w^3$, or with the term $\sqrt{\rho} w^{3.15}$ and a completely different constant.

10.2. Reentry Trajectories

We have seen that the accurate computation of the flow field that surrounds a reentry vehicle is very difficult. In comparison, the computation of reentry trajectories is easy. The theory distinguishes between ballistic reentry and gliding reentry trajectories.

10.2.1 Ballistic Reentry

The ballistic factor. A trajectory with vanishing lift is called a ballistic trajectory. Let us first study ballistic reentry maneuvers where the angle of attack α equals zero. The entire influence of the vehicle on the maneuver is then characterized by a single value, the important factor $C_{D_0} \cdot A/M$ (the drag coefficient at zero angle of attack C_{D_0} multiplied by the reference area A and divided by the mass M of the vehicle). We will designate the symbol B to this factor and call it “ballistic factor”. Therefore:

$$B = C_{D_0} \cdot A/M \quad (10 - 5)$$

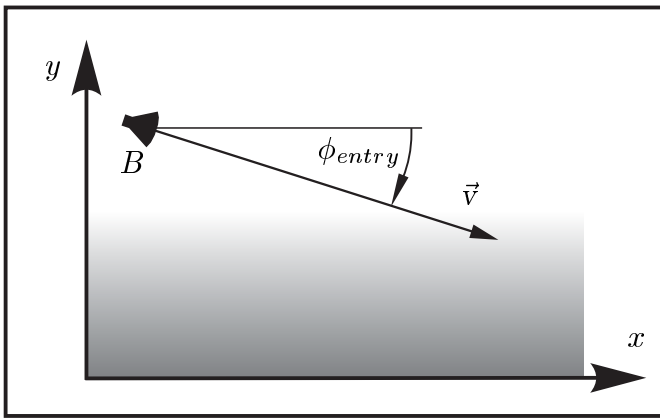
Note that often in other literature the reciprocal value is called “ballistic factor”.

The density of the atmosphere. The atmospheric density ρ is known to be approximately an exponentially declining function of the altitude. It can be denoted as:

$$\rho = \rho_0 \cdot e^{-(r - R_{earth})/h_{scale}} \quad (10 - 6)$$

In this “barometric representation” $(r - R_{earth})$ is the flight altitude above sea-level ($R_{earth} \approx 6378$ km), term $\rho_0 \approx 1.225$ kg/m³ is the density at sea-level, and $h_{scale} \approx 6700$ m is the atmospheric scale height. Equation (10-6) is only an approximation, for accurate computations the barometric representation has to be replaced by a better model (for example, the density can be specified in the form of a table).

Straight line entry. Let us now concentrate the attention on a simplified scenario: the direct entry on a straight line. The intention is not to replace an accurate computer solution but to find some analytical relationships. Therefore it is necessary that we define the underlying assumptions clearly: here the ballistic entry into the atmosphere takes place above a flat non-rotating earth; the air density is given as a function of the altitude (coordinate y) by the barometric representation (10-6) (no use is made of the range coordinate x); and the gravitational force is neglected. Drag (10-2) is the only force that acts on the vehicle, and consequently the trajectory is a straight line. Term B (equation 10-5) is used as the ballistic factor of the vehicle. Incident wind w and vehicle velocity v have the same value but opposite directions; and, making use of the entry angle ϕ_{entry} , the velocity components can be written as:



$$\begin{aligned} \dot{x} &= v \cos \phi_{entry} \\ \dot{y} &= -v \sin \phi_{entry} \end{aligned} \quad (10 - 7)$$

$$(v = \sqrt{\dot{x}^2 + \dot{y}^2} = w)$$

The drag deceleration is then:

$$\dot{v} = -\frac{B \rho_0}{2} e^{(-y/h_{scale})} v^2 \quad (10 - 8)$$

In order to integrate this equation we divide it by the vertical velocity component \dot{y} (equation 10-7), replacing the time coordinate t by the altitude coordinate y . Thus:

$$\frac{dv}{dy} = \frac{B \rho_0}{2 \sin \phi_{entry}} e^{(-y/h_{scale})} v, \quad \rightarrow \quad \frac{dv}{v} = \frac{B \rho_0}{2 \sin \phi_{entry}} e^{(-y/h_{scale})} dy \quad (10 - 9)$$

Integration of the equation (10-9) yields the velocity v as a function of the altitude:

$$v(y) = v_{entry} \cdot e \left[-\frac{B \rho_0 h_{scale}}{2 \sin \phi_{entry}} e^{(-y/h_{scale})} \right] \quad (10 - 10)$$

Term v_{entry} is the velocity at infinite altitude ($y \rightarrow \infty$). Note that the argument of the exponential function (in square brackets) is a linear function of the air density $\rho(y)$.

Maximum deceleration. The next point in question is the maximum deceleration and the altitude where the maximum occurs. We insert (10-10) into (10-8) to find:

$$\dot{v}(y) = \frac{-B \rho_0 v_{entry}^2}{2} \cdot e \left[-\frac{B \rho_0 h_{scale}}{\sin \phi_{entry}} e^{(-y/h_{scale})} - \frac{y}{h_{scale}} \right] \quad (10 - 11)$$

The maximum deceleration takes place where the differential $d\dot{v}/dy$ vanishes; to find this point we differentiate the expression in square brackets and put the result to zero.

The equation that defines the altitude $y(\dot{v}_{max})$ of the maximum deceleration is then:

$$e^{(-y/h_{scale})} = \frac{\sin \phi_{entry}}{B \rho_0 h_{scale}}$$

or transformed:

$$y(\dot{v}_{max}) = h_{scale} \cdot \ln \frac{B \rho_0 h_{scale}}{\sin \phi_{entry}} \quad (10 - 12)$$

This inserted into the equation (10-8) yields $v(\dot{v}_{max}) = v_{entry} \cdot e^{-1/2} = 0.606 \cdot v_{entry}$, indicating that the maximum deceleration occurs at about 60% of the entry speed. To calculate the actual value of \dot{v}_{max} we use the result (10-12) for the equation (10-9):

$$\dot{v}_{max} = -\frac{\sin \phi_{entry} \cdot v_{entry}^2}{2e \cdot h_{scale}} \quad (10 - 13)$$

For example, for $\phi_{entry} = 3^\circ$, $v_{entry} = 7.8$ km/s and $h_{scale} = 6.7$ km we can calculate a maximum deceleration of $\dot{v}_{max} = -87.4$ m/s². Surprisingly this value is only a function of the entry conditions (speed and angle), the ballistic factor B has no influence.

Maximum heat flux. The next point in question is the maximum heat flux and the altitude where the vehicle experiences the maximum heat flux. To find it we have to insert the velocity (10-10) into the equation (10-4), considering that $w = v(y)$:

$$\dot{q} = 1.705 \cdot 10^{-7} \sqrt{\rho_0} v_{entry}^3 \cdot e \left[-\frac{3 B \rho_0 h_{scale}}{2 \sin \phi_{entry}} e^{(-y/h_{scale})} - \frac{y}{2 h_{scale}} \right] \quad (10 - 14)$$

Term \dot{q} is specific stagnation point heat flux at the “unit nose” (with 1 meter radius). The maximum heat flux takes place where the differential $d\dot{q}/dy$ vanishes; therefore:

$$e^{(-y/h_{scale})} = \frac{\sin \phi_{entry}}{3 B \rho_0 h_{scale}}$$

or:

$$y(\dot{q}_{max}) = h_{scale} \cdot \ln \frac{3 B \rho_0 h_{scale}}{\sin \phi_{entry}} \quad (10 - 15)$$

This inserted into the equation (10-10) yields $v(\dot{q}_{max}) = v_{entry} \cdot e^{-1/6} = 0.846 \cdot v_{entry}$; and we can observe that maximum heating occurs earlier than maximum deceleration. The altitude difference is $h_{scale} \cdot \ln 3$, or approximately one scale height ($\ln 3 \approx 1.098$):

$$y(\dot{q}_{max}) - y(\dot{v}_{max}) = h_{scale} \cdot \left(\ln \frac{3 B \rho_0 h_{scale}}{\sin \phi_{entry}} - \ln \frac{B \rho_0 h_{scale}}{\sin \phi_{entry}} \right) \quad (10 - 16)$$

Finally we use the maximum heating altitude (10-15) for the equation (10-14) to find:

$$\dot{q}_{max} = 1.705 \cdot 10^{-7} v_{entry}^3 \cdot \sqrt{\frac{\sin \phi_{entry}}{3e B \cdot h_{scale}}} \quad (10 - 17)$$

Obviously, the maximum heat flux \dot{q}_{max} is not independent of the ballistic factor B , in contrast to the maximum deceleration \dot{v}_{max} . The heat flux has a direct influence on the surface temperature. Since the heat flux is smaller for vehicles with a large B factor, we may conclude that increasing the drag coefficient reduces the surface temperatures. A large heat shield is obviously desirable for a ballistic reentry vehicle.

Integral heat flux. The actual heat flux \dot{q} is the important quantity which determines the surface temperatures. However, the integral heat flux during the entire maneuver, $q = \int \dot{q} dt$, is the other important quantity, it determines the total heat load for the reentry vehicle, and thus size and capacity of its thermal protection system. In order to integrate the conditional equation (10-4) we divide it by the velocity component \dot{y} and replace again the time coordinate t by the altitude coordinate y . Then we insert the velocity v according to the equation (10-10). The result is:

$$\frac{dq}{dy} = \frac{-1.705 \cdot 10^{-7} \sqrt{\rho_0} v_{entry}^2}{\sin \phi_{entry}} \cdot e^{\left[\frac{-y}{2 h_{scale}} \right]} \cdot e^{\left[-\frac{B \rho_0 h_{scale}}{\sin \phi_{entry}} e^{(-y/h_{scale})} \right]} \quad (10-18)$$

We can transform (10-18) into a more convenient representation by introducing z^2 as:

$$z^2(y) = \left[\frac{B \rho_0 h_{scale}}{\sin \phi_{entry}} e^{(-y/h_{scale})} \right] \quad \left(= 2 \ln \frac{v_{entry}}{v} \right) \quad (10-19)$$

Hence $2zdz = -z^2 \cdot dy/h_{scale}$; and, replacing the terms y and dy in equation (10-18) by the corresponding functions of the new variables z and dz , the equation becomes:

$$dq = 1.705 \cdot 10^{-7} \cdot v_{entry}^2 \sqrt{\frac{h_{scale}}{B \sin \phi_{entry}}} \cdot 2 e^{-z^2} dz \quad (10-20)$$

The expression $\int e^{-z^2} dz$ is well-known in the mathematical field of error calculations (it is the Gaussian error distribution); and $\sqrt{\pi}/2$ is the value of the integral in the range between $z = 0$ and $z \rightarrow \infty$. This integral expression accepts the value $\sqrt{\pi}/2$ when we integrate the equation (10-20), because the boundaries for the integration are "space level" (where $z = 0$ for $y \rightarrow \infty$) and "ground level" ($z \rightarrow \infty$ for $v \rightarrow 0$). Then we can write the integral stagnation point heat flux on the reentry trajectory as:

$$q = 1.705 \cdot 10^{-7} \cdot v_{entry}^2 \sqrt{\frac{h_{scale} \pi}{B \sin \phi_{entry}}} \quad (10-21)$$

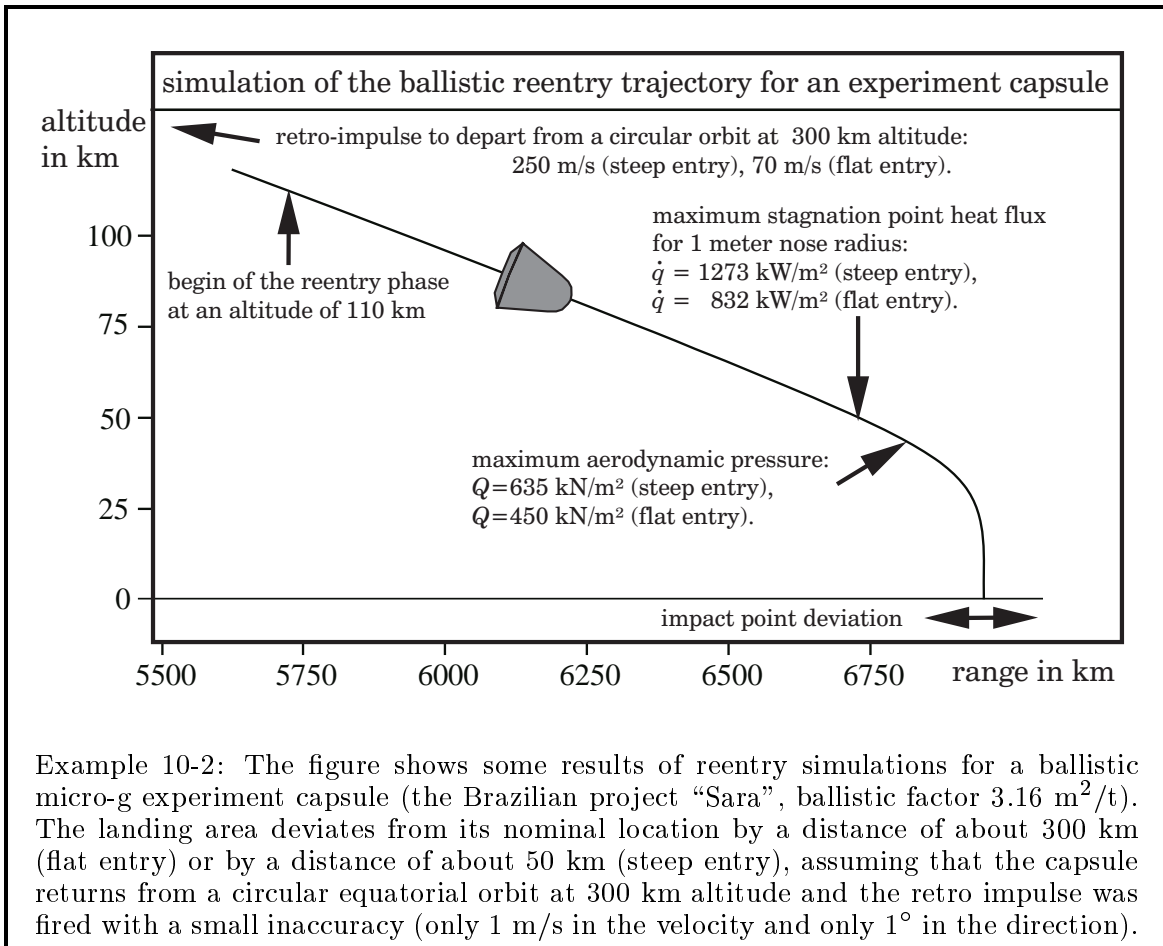
Two items are remarkable when we consider this result: first, as well as the maximum heat flux \dot{q} the integral heat flux q diminishes when we increase the ballistic factor B (enlarge the specific size of the vehicle). Second, in contrast to the maximum heat flux that grows with ϕ_{entry} , the integral heat flux gets smaller when we increase this angle. Descending faster means higher temperatures, but the entire heating is smaller then.

Uncertainty of the impact point. In practice there is another reason that can make steeper descends interesting: the accurate predictability of the range variable x . When a ballistic reentry capsule returns from orbit, its final impact location can only be determined with a typical precision of ± 50 km, for example. The atmosphere is a thin gas layer above the surface of the earth, and the air density is an exponential function of the altitude. Usually it would be desirable to know the landing (or watering) area with a high accuracy before the reentry maneuver is initiated, but small uncertainties in the climate can cause a substantial deviation of the landing location. The predictability of the landing area improves when a steeper descent is performed.

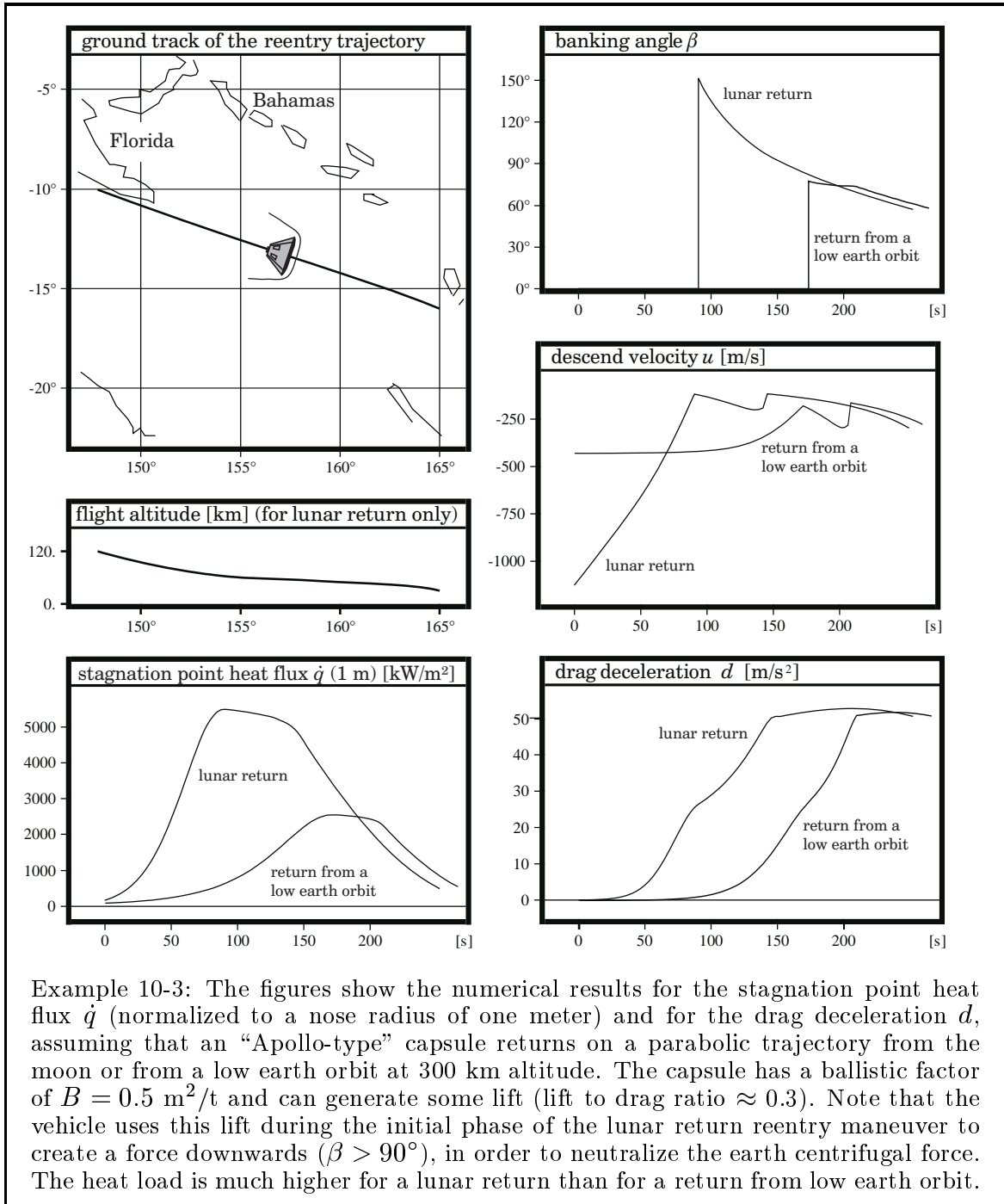
When we want to simulate the ballistic reentry maneuver in order to analyze the precision of the impact point, the model of the straight line entry is not accurate enough. For example, we can assume that the reentry takes place in the equatorial plane of the central gravitational field of the earth; and we can write the equations of motion using polar coordinates r and φ . Velocity components are vertical u and horizontal v ; however, since the earth rotates once a day with the angular velocity ω_{earth} , the components of the incident wind are vertical $(-u)$ and horizontal $(r\omega_{earth} - v)$. Thus:

$$\begin{aligned}\dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} - \frac{\varrho(r)B}{2} u \sqrt{u^2 + (v - r\omega_{earth})^2} \\ \dot{v} &= -\frac{uv}{r} - \frac{\varrho(r)B}{2} (v - r\omega_{earth}) \sqrt{u^2 + (v - r\omega_{earth})^2} \\ \dot{r} &= u \\ \dot{\varphi} &= \frac{v}{r}\end{aligned}\quad (10 - 22)$$

Term γ is the gravitational constant of the earth ($\gamma = 3.9866625 \cdot 10^{14} \text{ m}^3/\text{s}^2$). When we integrate the equation system (10-22) using a numerical standard routine, we can easily change the atmospheric density $\varrho(r)$, the ballistic factor B or the initial conditions, in order to check the influence of these modifications on the landing area.



Quasi ballistic reentry trajectories. The use of a small lift force during certain time intervals of the maneuver can greatly reduce the maximum heat flux and the maximum deceleration as well. Reentry capsules can usually generate a small lift force, either by using control surfaces (fins) or by shifting the center of mass away from the symmetry axis. Since the vehicle is still extremely fast during the time periods of the maximum loads, a small control force suffices to control the descent velocity u . The substantial deceleration is performed at an altitude of about 60 km.



Example 10-3: The figures show the numerical results for the stagnation point heat flux \dot{q} (normalized to a nose radius of one meter) and for the drag deceleration d , assuming that an “Apollo-type” capsule returns on a parabolic trajectory from the moon or from a low earth orbit at 300 km altitude. The capsule has a ballistic factor of $B = 0.5 \text{ m}^2/\text{t}$ and can generate some lift (lift to drag ratio ≈ 0.3). Note that the vehicle uses this lift during the initial phase of the lunar return reentry maneuver to create a force downwards ($\beta > 90^\circ$), in order to neutralize the earth centrifugal force. The heat load is much higher for a lunar return than for a return from low earth orbit.

10.2.2 Gliding Reentry

Trajectory optimization. The problem in trajectory optimization is the determination of appropriate control functions; and, as we have seen in other chapters of this book, the Hamilton-Lagrange theory (or calculus of variations) is usually the right tool for doing this. However, the Hamilton-Lagrange theory is quite complicated, and fortunately its application is not necessary for the optimization of reentry trajectories. Reentry trajectory optimization is not a problem of the Hamilton-Lagrange theory.

To understand these circumstances we have to look at two important equations: the one for the drag (10-2) and the other one for the stagnation point heat flux (10-4). When we divide the equation (10-4) by the equation (10-2), then we find the stagnation point heat flux related to the drag deceleration of the vehicle. The relationship is:

$$\frac{dq}{dv} \approx \frac{\dot{q}}{D/M} = \frac{(-2 \cdot 1.705 \cdot 10^{-7}) w}{\sqrt{\rho} (C_{D0} + C_{D\alpha} \cdot \alpha^2) A} \quad (10 - 23)$$

Since drag is the dominant force during the reentry maneuver, the time derivative of the velocity dv/dt accepts approximately the same value as the drag deceleration d ($d = D/M$). Term dq in the equation (10-23) is the amount of heating which the thermal protection system has to stand in order to reduce the velocity of the vehicle by dv . Integration of the quantity dq yields the integral heat flux q of the whole maneuver. Obviously it is important that we minimize the total heat flux. When we look at the right hand side of the equation (10-23) we see that we can influence two quantities in order to minimize q during the essential braking phase: we can brake at a low altitude where the air density ρ is high, and we can brake with a large the angle of attack α . Term A in equation (10-23) is the constant reference area and w the wind velocity.

When we look a second time at the two important equations (10-2) and (10-4), we see that both, the actual heat flux \dot{q} and the actual drag deceleration $d = D/M$, increase sensitively when we brake at a lower flight altitude. The actual heat flux determines the surface temperatures; the actual drag deceleration determines the structural loads for the vehicle. We cannot select a flight altitude for the essential braking phase where an extremely high stagnation point temperature would burn the thermal protection system, or where an extremely high deceleration would break the fragile structure of the vehicle. Thus, we must select the braking altitude as low as possible, but still high enough so that surface temperatures and structural loads are bearable.

Reentry flight phases. Under these conditions the optimal control of a reentry trajectory is entirely determined. The trajectory must consist of the following phases:

Initially, in order to pass the higher layers of the atmosphere as quickly as possible, the reentry vehicle enters the atmosphere with maximum lift and maximum drag. In the upper atmosphere the deceleration is not significant but the heating is high. The entry angle ϕ_{entry} must be selected in a way that the vertical velocity u is very small at the moment when the stagnation point temperature reaches its maximum.

The stagnation point temperature may not increase anymore in the flight phase that begins now, the phase of the flight with maximum stagnation point temperature. The vehicle flies still with the maximum angle of attack, but the lift force is not used to let the vehicle climb but to let the vehicle fly a curve. The descent velocity u is exclusively controlled by an appropriate adjustment of the banking angle; and it is controlled in a way that the stagnation point heat flux remains approximately constant.

The drag deceleration grows while the vehicle descends at the limitation imposed by the maximum stagnation point heating rate. The phase of maximum deceleration begins at the moment when deceleration of the vehicle reaches its allowed limitation. The descent velocity is reduced once more in order to keep the dynamic pressure approximately constant. The angle of attack can now be reduced if the landing site is far distant from the ground track of the orbit (if the vehicle must have a large cross range capability). The heating rate \dot{q} reduces rapidly while the vehicle descends at the limitation imposed by the maximum deceleration. The flight phase ends when the reentry vehicle flies with a velocity where the heating rate is not significant anymore.

Now the vehicle is slow enough to enter the final flight phase. When the vehicle is a space shuttle it can now open ventilation holes to cool the hot inner structure and establish the glide slope for the landing site approach. When the vehicle is a capsule it can now discharge the heat shield and start with the parachute deployment sequence.

The type of heat protection system determines the altitude of the braking maneuver: surface temperatures are lower at higher altitudes but all together more heat is transferred to the vehicle. Space shuttles use tiles as heat protection system and discharge heat mainly in the form of radiation; capsules use heat shields and discharge heat mainly in the form of material ablation. Consequently, space shuttles begin the essential deceleration phase at higher altitudes than capsules, because low surface temperatures are more important than the entire heat load. Capsules decelerate at lower altitudes because the entire heat load is more important than the surface temperature.

Banking the vehicle makes it possible to obtain a “cross-range capability” for landing areas distant from the geographical ground track of the orbit. A substantial cross range demands during the essential braking phase a constant banking to one side. When the desired cross range is smaller, the vehicle must be banked alternately to one side and the other. To reach the maximum cross range distance the angle of attack must be reduced slightly (at the expense of a higher total heating rate).

Equations of motion. Let us now write down the equations of motion for a reentering space glider (the derivations of these equations is demonstrated in the chapter four of this book). These second order differential equations for the powered motion in the atmosphere are essentially the same for all flying objects, for space gliders we can ignore the thrust force and assume that the mass of the vehicle is constant. However, for accurate numerical computations we cannot ignore that the earth is rotating and that the atmosphere is carried along with the rotation of the earth.

The equations of motion which we use here are based on a moving coordinate system. The origin is placed on the center of the earth, but the attitude of the system turns with the displacement of the vehicle: the 1-axis aims always at the vehicle, and the 2-axis is always aligned with the inertial velocity vector of the vehicle (inertial means in relation to a non-rotating earth). The actual position of the vehicle with respect to the North pole and the earth equator is then determined by the three Euler angles ψ , ϑ and φ . We use the symbol r ("radius") to specify the flight altitude ($r = 6378$ km at the surface of the earth) and the inertial velocity components u (vertical) and v (horizontal). The earth rotates on the South-North axis once a day with the angular velocity ω_{earth} . The incident wind velocity is the difference of the inertial velocity of the atmosphere and the inertial velocity of the vehicle:

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} -u \\ r \omega_{earth} \cos \vartheta - v \\ -r \omega_{earth} \sin \vartheta \cos \varphi \end{pmatrix}$$

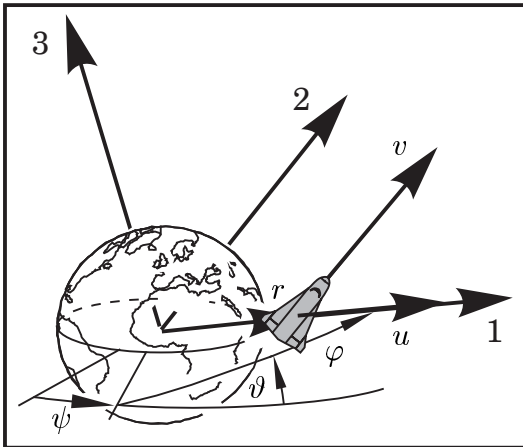
$$w = \sqrt{w_1^2 + w_2^2 + w_3^2} \quad (10 - 24)$$

Drag deceleration \vec{d} and lift acceleration \vec{l} are functions of the incident wind w , the air density ρ and the angle of attack α . The aerodynamic drag coefficient is $C_D = C_{D_0} + C_{D_\alpha} \cdot \alpha^2$ and the lift coefficient is $C_L = C_{L_\alpha} \cdot \alpha$. Therefore:

$$\vec{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \frac{(C_{D_0} + C_{D_\alpha} \cdot \alpha^2) A \rho}{2 M} w \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \quad (10 - 25)$$

$$\vec{l} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \frac{(C_{L_\alpha} \cdot \alpha) A \rho w}{2 M \sqrt{w_1^2 + w_2^2}} \begin{pmatrix} ww_2 \cos \beta - w_1 w_3 \sin \beta \\ -ww_1 \cos \beta - w_2 w_3 \sin \beta \\ (w_1^2 + w_2^2) \sin \beta \end{pmatrix} \quad (10 - 26)$$

Term β is the banking angle, A is a reference area and M the mass of the vehicle. Like always, γ is the gravitational constant. The equations of motion are then:



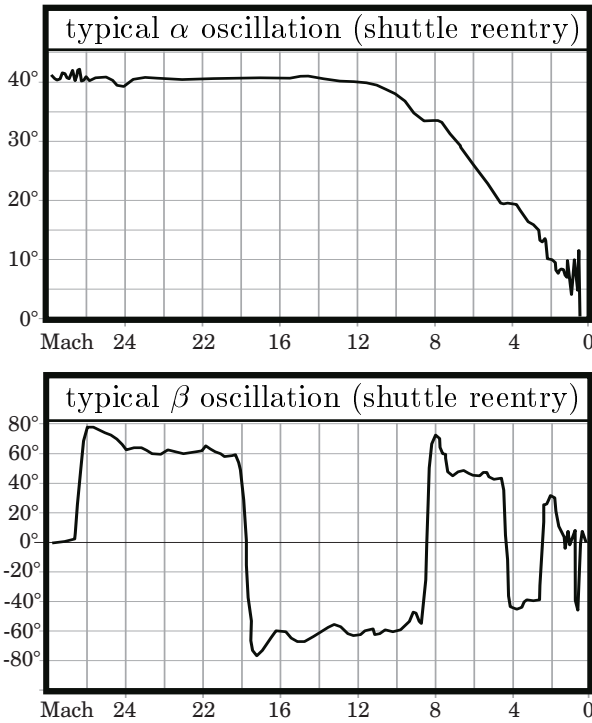
$$\begin{aligned} \dot{u} &= \frac{v^2}{r} - \frac{\gamma}{r^2} + d_1 + l_1 \\ \dot{v} &= -\frac{u v}{r} + d_2 + l_2 \\ \dot{r} &= u \\ \dot{\varphi} &= \frac{v}{r} - \frac{d_3 + l_3}{v} \sin \varphi \cot \vartheta \\ \dot{\vartheta} &= \frac{d_3 + l_3}{v} \cos \varphi \\ \dot{\psi} &= \frac{d_3 + l_3}{v} \cdot \frac{\sin \varphi}{\sin \vartheta} \end{aligned} \quad (10 - 27)$$

These differential equations can easily be integrated numerically using a standard integration routine, provided that the control functions for α and β are well-known.

Control of the initial entry phase. During the initial reentry phase the nose of the vehicle is up and the angle of attack is high. The control functions for the vehicle are then $\alpha = maximum$ and $\beta = 0$. The magnitude of the de-orbiting maneuver has to be adjusted in a way that the descent velocity is small at the end of this phase.

Control of the essential braking phase. When the maximum heating has been reached the vehicle is banked to one side with the intention to control the descent velocity. The descent velocity is very small during this phase, and neglecting u and its time derivative \dot{u} we can follow the so-called “equilibrium glide condition” from the first equation of the system (10-27). The result is a condition for the banking angle β :

$$\cos \beta \approx \left(\frac{v^2}{r} - \frac{\gamma}{r^2} \right) / \left(\frac{\rho C_{L\alpha} \alpha A}{2M} w w_2 \right) \tag{10 - 28}$$



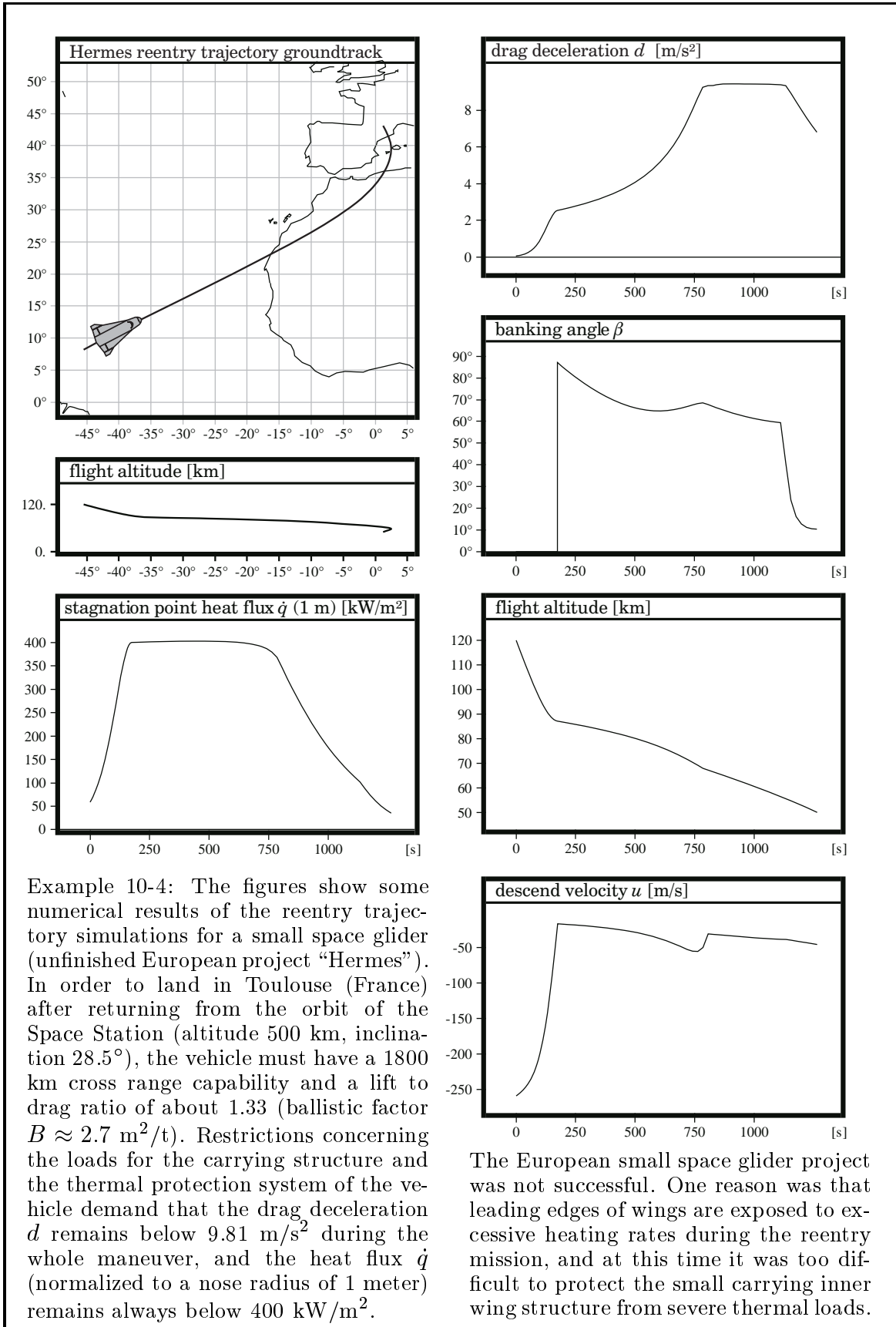
Note that also the third component of the incident wind vector is negligibly small ($w_3 \approx 0$), and that the vehicle must have a sufficient lifting capability to satisfy the condition (10-28). However, it is usually possible to satisfy this condition also when the lift coefficient $C_{L\alpha}$ is very small, because the centrifugal acceleration v^2/r neutralizes a part of the gravitational acceleration γ/r^2 and the velocity of the incident wind w is high. Space gliders perform reentry maneuvers with high banking angles ($\beta \approx 80^\circ$); gliders have considerable cross range capabilities (the cross range distance can be reduced by banking the vehicle several times alternately to both sides). In contrast capsules have nearly no cross range capability.

During the essential braking phase, the descent velocity u of the vehicle must be controlled in order to keep initially the heat flux $\dot{q} = 1.705 \cdot 10^{-7} \sqrt{\rho} w^3$ and later the aerodynamic pressure $Q = \frac{1}{2} \rho w^2$ approximately constant. Both quantities are mainly functions of the incident wind velocity w and the air density ρ . Therefore:

$$\frac{d\dot{q}(w, \rho)}{dt} = \frac{\partial \dot{q}}{\partial w} \cdot \dot{w} + \frac{\partial \dot{q}}{\partial \rho} \cdot \dot{\rho} = 0 \quad \rightarrow \quad u \approx -3 \rho w \frac{C_D A}{M} h_{scale} \tag{10 - 29}$$

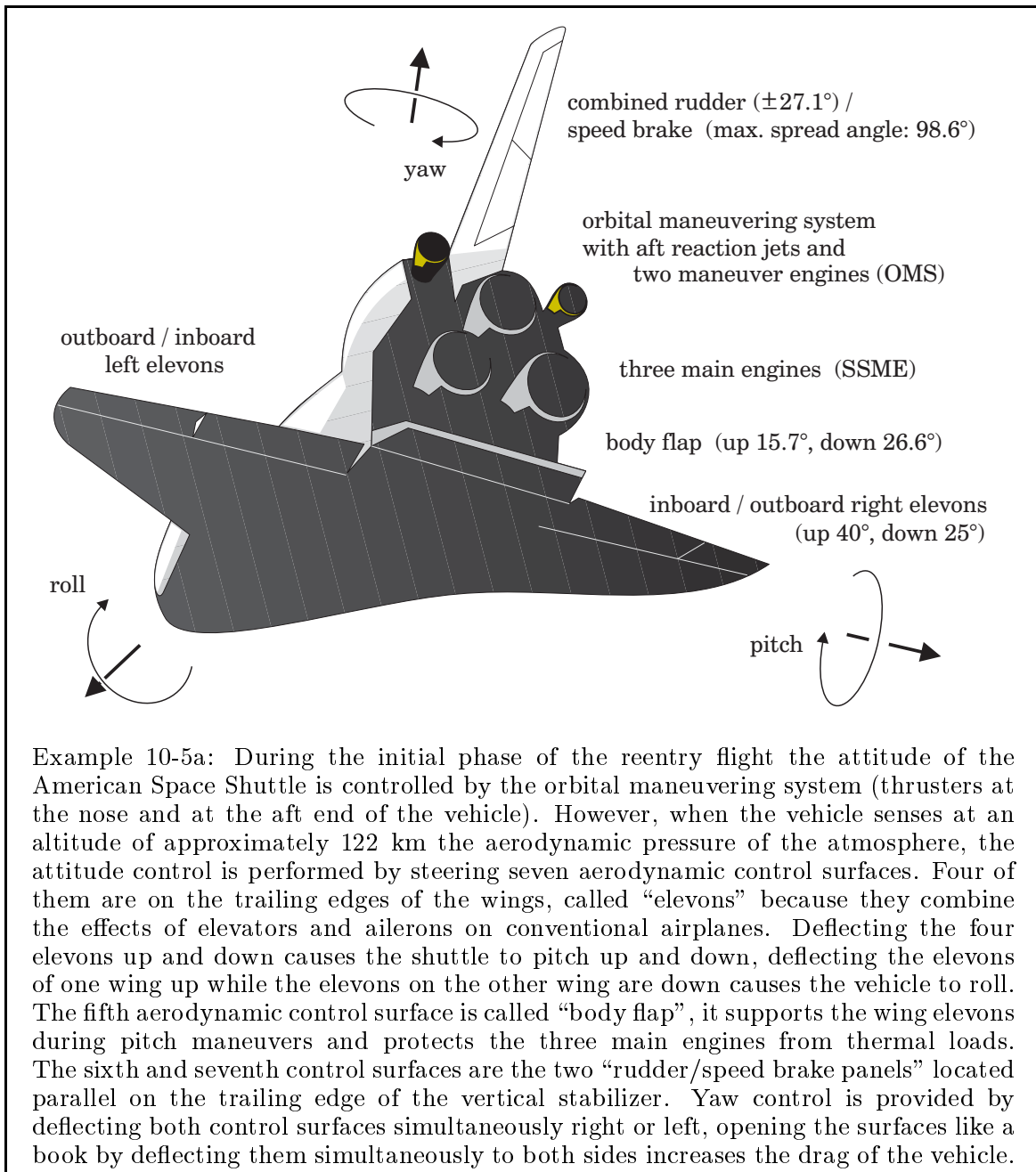
$$\frac{dQ(w, \rho)}{dt} = \frac{\partial Q}{\partial w} \cdot \dot{w} + \frac{\partial Q}{\partial \rho} \cdot \dot{\rho} = 0 \quad \rightarrow \quad u \approx -\rho w \frac{C_D A}{M} h_{scale} \tag{10 - 30}$$

(Note that $\dot{\rho} = -\rho u/h_{scale}$ and $\dot{w} \approx -Q C_D A/M$.) Thus, the descent velocity u has to be reduced approximately by the factor 3 when the vehicle transits from the phase with maximum temperature to the phase with the maximum dynamic pressure.



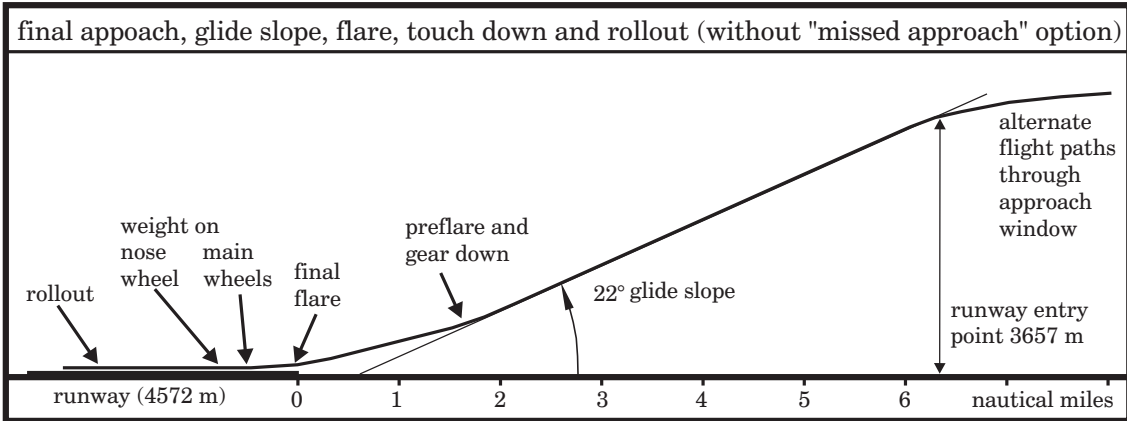
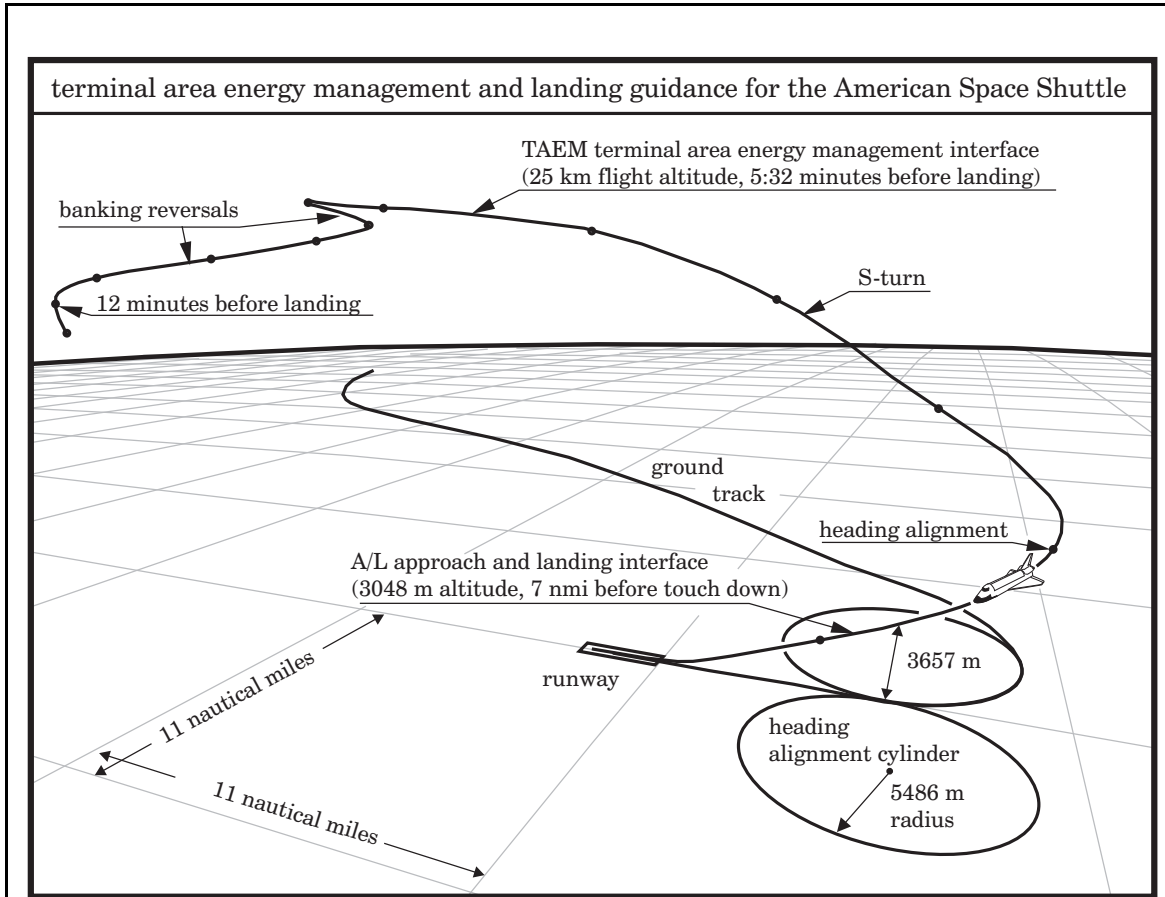
10.2.3 Control and Navigation

Aerodynamic surfaces. We have seen that the objective in finding an optimized space glider reentry trajectory is simply to bring the vehicle safely down to the surface of the earth. Therefore the vehicle has to fly through a certain tight entry corridor: as low as possible, but not exceeding the thermal and the mechanical load limitations. After the essential braking maneuver the glider must land on a predefined airport. Like an aircraft it is equipped with aerodynamic control surfaces, in order to be able to control the braking maneuver and to perform the precise terminal navigation.



Guidance during the braking maneuver. On orbit the reentry maneuver of the glider is initiated by firing a small retro-impulse. Immediately after this the attitude control system establishes a vanishing banking angle ($\beta \approx 0$) and a high angle of attack ($\alpha \approx 30^\circ$ to 40°); and the vehicle waits for the aerodynamic surfaces to become effective. At the moment when aerodynamic pressure is sensed (altitude ≈ 122 km), the control is handed over the aerodynamic control surfaces and the orbital maneuvering system is deactivated. The reentry navigation begins. During the following aerodynamic braking phase the control surfaces keep a predetermined time function for the angle of attack α (for example α can be kept constantly 30° , or α decreases slowly from 40° to 30°). The navigation is exclusively performed by an appropriate control of the banking angle β . The temperature control phase begins when the vehicle senses the maximum permissible surface temperature; the drag control phase begins later when the vehicle senses the maximum deceleration. The vehicle executes roll maneuvers and adjusts the banking angle, first for a constant heating rate and then for a constant deceleration rate (for example, during the essential braking phase the vehicle banks initially with $\beta \approx 80^\circ$ and finally with $\beta \approx 60^\circ$). A larger banking angle means a steeper descent and therefore a higher energy dissipation rate, conversely, a smaller banking angle means a smaller energy dissipation rate. The cross range distance is measured while the vehicle decelerates; and when the actual cross range distance exceeds a certain computer-loaded value, the vehicle performs a roll maneuver in the opposite direction in order to reverse the banking angle.

Terminal guidance for the final landing maneuver. The essential braking phase is finished when the glider enters the transition phase to terminal approach. During the transition phase the angle of attack is substantially reduced (until $\alpha \approx 15^\circ$) and the banking angle β is turned until the vehicle has an upright attitude for gliding. The radio transmission blackout is over and the navigation system of the vehicle is now able to receive signal from terrestrial navigation stations. The glider can accurately measure its position and find out whether it must still dissipate excess energy (this can be done either by flying a certain "S-curve" or by utilizing the speed brakes). In case of the American space shuttle, the "terminal area energy management guidance" begins at 83000 feet altitude with a speed of 2500 feet per second. The Shuttle attains subsonic velocity flying at an altitude of approximately 49000 feet, when the vehicle is still 22 nautical miles distant from the landing site. The "approach and landing guidance" begins at an altitude of 10000 feet with an airspeed of 290 ± 12 knots, 6.9 nautical miles distant from the airport. The vehicle is then aligned with the centerline of the runway and guided to descend with a 22° glide slope (seven times more than the approach glide slope of a civil jet airplane) aiming at a point 0.86 nautical miles in front of the runway. It descends now with a rate of 10000 ft/minute, about twenty times the sink rate of an airliner during a standard instrument approach. A "pre-flare" maneuver is executed at an altitude of 1750 ft above ground in order to reduce the sink rate to 9 ft/s and to attain a glide slope of 1.5° . At 500 ft altitude the landing gear is deployed and the speedbrakes are adjusted; the runway threshold is crossed at 100 ft. The Shuttle flares once more and touches down with a speed of 184 to 196 knots, ca.2500 ft past the threshold of the runway (total length 15000ft).



Example 10-5b: The American Space Shuttle is famous for landing like an airliner, however, in one important aspect the comparison fails completely. The Shuttle is not equipped with jet engines, and thus it has to land safely like a glider on the first attempt. Precise navigation performed by the “TAEM”-system ensures that the vehicle reaches the landing area; and then the landing maneuver is guided by the automatic “A/L”-system. Since the gliding properties of the Shuttle are quite bad compared with a conventional jet airplane, the Space Shuttle descends with a much steeper glide slope and flies much faster when it approaches the runway of the airport.

10.3. Thermal Protection Systems

A returning space vehicle is exposed to high thermal loads on its reentry trajectory. We have seen the heat flux in the aerodynamic stagnation point of the nose or the leading edges of wings is mainly a function of the velocity, the flight altitude and the radius of the exposed surface. The greater the radius the lower the stagnation point temperature. A large space shuttle is usually decelerated at high altitudes and protected with a reusable heat protection system that consists of a hot structure at the nose and at the leading edges of the wings, tiles (or shingles) at the highly stressed bottom surfaces and an insulation at the lower stressed back surface. A space capsule is usually decelerated at lower altitudes and protected with an ablative heat shield.

10.3.1 High Temperature Materials for the Reusable Space Shuttle

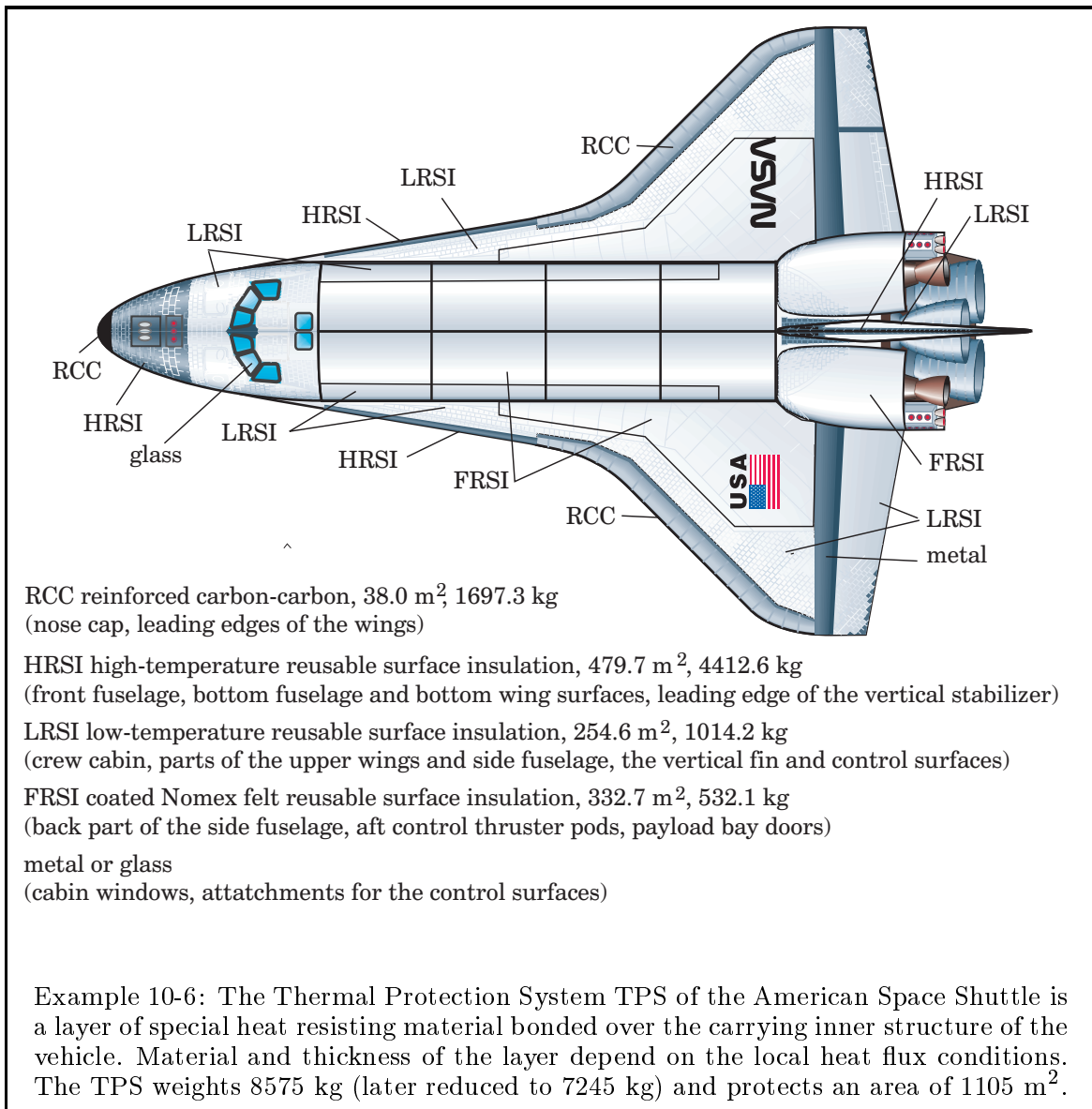
The load carrying aluminum substructure. The carrying structure of a space vehicle is primarily made of light metal (aluminum alloy) or graphite epoxy, materials which lose their strength at temperatures above 180° Celsius, typically. The surface temperatures during the reentry maneuver are much higher (for example at the nose of the American Space Shuttle they reach values as high as 1650° Celsius); and thus the carrying structure must be protected against the excessive heating.

Fiber reinforced composites. The nose and the leading edges of the wings are made of carbon fibers imbedded in a matrix of carbonized phenolic resin material. A special production process transforms the outer layers of these “reinforced carbon carbon” panels into a silicon-carbide coating that protects the carbon fibers from oxidation. Parts made of this material are highly resistant to fatigue loading and can operate between temperatures of -121° Celsius (in space) and +1650° Celsius (during reentry). Today ceramic materials as matrix for the carbon fibers (or for silicon carbide fibers) play a key role in the development of new materials for hot structures. These “fiber reinforced ceramic composites” allow the realization of lightweight load carrying structures for short time operation up to temperatures of 1800° Celsius, for longtime operation up to maybe 1000° Celsius.

Tiles. The tiles for the protection of the bottom surface of the American Space Shuttle consist of special glass fibers that became rigid by a ceramic bonding process. They are coated with a thin black glass layer. Tiles for upper parts of the surface consist essentially of the same material, but they have a white color because aluminum oxide was added during the coating process for a better moisture resistance. The tiles with a typical area of 15 times 15 centimeters are manufactured with different shapes and different densities to protect the individual locations on the surface, the thickness varies from 2.5 cm to 12 cm. The volume specific weight of the tiles is comparatively low, because only a small part of the volume contains material (most of the volume is void). The tiles can stand temperatures up to 1260° Celsius, but they are fragile (sensitive to deformations and vibrations), and a flexible stress insulation is necessary between the tiles and the carrying structure of the Shuttle.

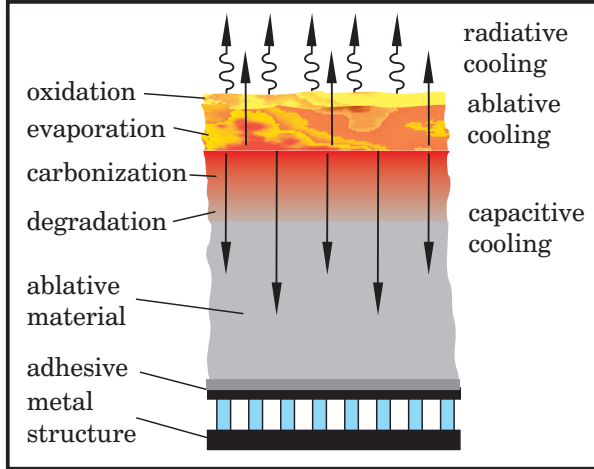
Missing tiles on the lower surface of the left wing (which broke off during the launch procedure when they were impacted by a piece of ice from the external tank) caused later during the reentry maneuver the fatal accident of the Space Shuttle Columbia.

Nomex felt surface insulation. About 50% of the surface of a space shuttle are not exposed to the incident air flow and therefore not excessively heated during the reentry maneuver. The surfaces are parts of the back fuselage, parts of the upper wings and the side fuselage that is located in the shadow of the wings. These lower stressed areas are exposed only to temperatures below 371° Celsius, they are protected by the so-called “coated Nomex felt reusable surface insulation”, a textile layer with a thickness between 0.4 and 1 cm bonded to the surface of the shuttle by a silicon adhesive (the “Nomex” felt is a double layer cross-lapped aramid fiber web; pressed, sewn, and thermally stabilized during the production process).



10.3.2 Heat Shields

The ablative heat shield. The classical heat shield is essentially a honeycomb structure filled with phenolic resin, reinforced with fibers (carbon, glass or asbestos).



The thickness of the shield is not constant for all locations on the surface, but it depends on the local heat flux. For example, the thickness of the heat shield of the retired American Apollo capsule varied from 2.3 cm to 7.6 cm, resulting in an area specific mass of 98 kg/m². Cooling mechanism is the discharge of energy in the form of heat radiation to outside, heat conduction to inside (so-called “capacitive cooling”), and in form of chemical decomposition, evaporation and ablation of material.

The inflatable heat shield. We have seen that the ballistic factor B plays a key role in the design of heat protection system. The actual heat flux \dot{q} (equation 10-17) and the integral heat flux q (equation 10-21) diminish when B grows, indicating that it is advisable to use a heat shield with a large diameter (however, the maximum deceleration is essentially not a function of B). When the heat shield is large enough, radiation cooling suffices to discharge the incident heat load to outside. An inflatable heat shield could be the best solution, since a large solid heat shield is heavy and does usually not fit in the payload bay of a conventional space launcher. The manned Mars mission might be an application for inflatable heat shields, to provide heat protection during aerocapture and landing on Mars, and during the reentry maneuver on earth.

Weight of heat protection systems. The American Space Shuttle has obviously not satisfied the high expectations during its development phase regarding operational safety and low costs. Even though the Space Shuttle is a fantastic tool for manned space activities on a large scale, it is quite expensive and not free of failures. Several times it has been tried to develop a smaller winged reentry vehicle (for example Dynasoar, Hermes and Hope), another reason that these projects were unsuccessful is that heat protection problems increase when the size of the vehicle is reduced. The American Space Shuttle system will be phased out in some years (or probably immediately when another fatal accident occurs); and, due to the low budget situation in the present time, the new American manned return vehicle may be again a capsule.

	vehicle mass	thermal protection system	
Mercury (return from LEO)	1179 kg	143 kg	12.0%
Apollo (parabolic reentry)	4309 kg	590 kg	13.7%
US-Space Shuttle	68000 kg	7245 kg	10.8%
European Hermes (never made)	12940 kg	1135 kg	8.8%

List of Symbols

Greek letters:

- α The angle of attack (equation 4-25). The angle α describes together with the banking angle β the attitude of the launch vehicle with respect to the direction of the incident wind. In chapter 2 the same symbol α is used to determine the angle between the actual flow direction of a gas molecule and the centerline of the nozzle (equation 2-118).
- β The banking angle (equation 4-25). The angle β describes together with the angle of attack α the attitude of the launch vehicle with respect to the incident wind. The “nose” of the vehicle is “up” for $\beta = 0$; full banking to one side means $\beta = \pm 90^\circ$.
- γ The gravitational constant (equation 1-3 or 4-16). The individual gravitational constant of a celestial body is universal gravitational constant $\gamma_o = 6.672 \cdot 10^{-11} \text{ m}^3/(\text{kg s}^2)$ multiplied with the mass of the body.
- δ The thrust angle out of the flight plane (equation 4-33). We use the thrust angles ϕ (in flight plane) and δ (out of flight plane) to resolve the thrust vector into a component notation of the moving coordinate system. For a unique determination δ is limited to $\pm 90^\circ$ and ϕ is limited to $\pm 180^\circ$.
- Δ Generally a difference between two similar values. For example, we use the symbol Δx to describe a displacement or Δr to describe the difference between two orbital altitudes.
- Δv The velocity requirement (equation 2-6 or 2-7). The capability of the rocket propulsion system, determined by the Ciolcovskij equation 2-7, must comply with the velocity requirement for a certain mission, determined by the calculation of the concerning trajectory.
- Δt Generally a time interval, for example the flight time between two locations on a conic orbit (equations 4-57 to 4-60).
- ε The numerical eccentricity of a conic orbit (equation 4-52 or 4-53). The dimensionless quantity ε is always positive, frequently used as an “orbital element” to determine the type of conic orbit: $\varepsilon = 0$ for a circular orbit, $\varepsilon < 1$ for an elliptic orbit, $\varepsilon = 1$ for a parabolic orbit and $\varepsilon > 1$ for a hyperbolic orbit.
- ϵ The expansion ratio (equation 2-88); ϵ characterizes the geometry of the rocket nozzle, it is the exit area of the nozzle divided by the throat area.
- ζ The thrust angle of a solar sailcraft (equation 9-23). For the computation of sailcraft trajectories it can be advantageous to determine the thrust direction with respect to the sunbeams, using the angles ι and ζ .

- η The right running characteristic (equation 2-130). The characteristics η and ξ are curves in a two-dimensional (or axisymmetric) flow field, where partial derivatives can be composed to total derivatives.
- θ Kepler's eccentric anomaly, important for the calculation of the flight time on a conic orbit (equation 4-58).
- ϑ The Euler angle that characterizes the orbital inclination (the "inclination angle", equation 4-2). The actual attitude of a coordinate system (or a rigid body) with respect to reference attitude is determined by a set of three angles, very often used are the Euler angles φ , ϑ , ψ .
- Θ The Cardan angle that characterizes the geographical latitude ("latitude angle", equation 4-3). The actual attitude of a coordinate system (or a rigid body) with respect to reference attitude is determined by a set of three angles, very often used are the Euler angles φ , ϑ , ψ or the Cardan angles Φ , Θ , Ω .
- ι The thrust angle of a solar sailcraft (equation 9-23). For the computation of sailcraft trajectories it can be advantageous to determine the thrust direction with respect to the sunbeams, using the angles ι and ζ . Angle *iota* is the inclination of the sail area.
- κ The ratio of the specific heats and also the exponent for an isentropic change in state (equation 2-31 or 2-59). The value κ (between 1.1 and 1.3 for the fire gases in a rocket chamber) is a typical propellant characteristic, it depends mainly on the number of atoms in the gas molecules.
- λ Generally a multiplier, very often a Lagrange multiplier (equation 5-3). A Lagrange multiplier has usually an index that indicates the state variable to which the multiplier belongs.
- μ Generally a multiplier. This term is used with three different meanings in three different chapters. In chapter 2 "Rocket Propulsion" term μ is used as the angle between the flow direction and the characteristic curves, thus it is the "Mach angle" in case of an isentropic flow field (equation 2-131). In chapter 5 "Orbit Optimization" the term μ is used as a Lagrange multiplier that is adjoined to a restriction in equation form (equation 5-9 or 5-13). Finally, in chapter 7 "Launcher Dynamics" term μ is used as an oscillation frequency (equation 7-95 or 7-114).
- ν This term is used with two different meanings in two different chapters. In chapter 2 "Rocket Propulsion" term $\nu(\mu)$ is the "Prandtl-Meyer" angle for the computation of a flow field (equation 2-134). In chapter 7 "Launcher Dynamics", however, the term ν is used as an oscillation frequency (equation 7-93 or 7-114).

- ξ The left running characteristic (equation 2-130). The characteristics η and ξ are curves in a two-dimensional (or axisymmetric) flow field, where partial derivatives can be composed to total derivatives.
- π The number for the circumference of a circle, $\pi = 3.141592653589793\dots$
- ϱ Generally the density, usually the density of the atmosphere (equation 1-14), for example $\varrho_o = 1.225 \text{ kg/m}^3$ is the standard density at sea-level. In chapter 2 ϱ is the density of the gas in a rocket thrust chamber.
- ρ Generally a length or or distance (equations 1-5, 4-61 or 6-71 for example).
- ς This symbol defines an angle only important for the flight time calculation on a conic orbit using the Lambert theorem (equations 4-62 to 4-64).
- ϕ The thrust angle in the flight plane (equation 4-33). We use the thrust angles ϕ (in flight plane) and δ (out of flight plane) to resolve the thrust vector into a component notation of the moving coordinate system.
- ϕ_{entry} The entry angle of the ballistic maneuver (equation 10-7).
- φ One of the three Euler angles (equation 4-2). The actual attitude with respect to reference attitude is generally determined by a set of three angles, the Euler angles φ , ϑ , ψ are very often used for this. φ becomes the path angle in case of a plane trajectory, then it is a “polar coordinate” that determines together with the distance r the position of the spacecraft (equation 4-49).
- χ The thrust deviation angle, the angle between the actual thrust direction and the centerline of the launch vehicle (equations 7-82 to 7-84).
- ψ The Euler angle that determines the orientation of the line of nodes (the “node angle”, equation 4-2). The actual attitude of a coordinate system (or a rigid body) with respect to reference attitude is determined by a set of three angles, very often used are the Euler angles φ , ϑ , ψ .
- ψ The Cardan angle Φ that determines the geographical longitude (“longitude angle”, equation 4-3). The actual attitude of a coordinate system (or a rigid body) with respect to reference attitude is determined by a set of three angles, very often used are the Euler angles φ , ϑ , ψ or the Cardan angles Φ , Θ , Ω .
- ω The angular velocity. In orbital mechanics the angular velocity is important to calculate the attitude of the rotating coordinate system (see equation 4-7). In attitude mechanics ω is important to determine the motion of the body (equation 6-20).
- Ω The rotation angle $\Omega = \int \omega dt$ and also the Cardan angle that determines the rotation on the body fixed 1-axis (equation 4-3).

Calligraphic letters:

- $\mathcal{A}, \mathcal{B}, \mathcal{C}$ The diagonal elements of the matrix of inertia \mathfrak{S} (equation 6-22), they define the moments of inertia of a rigid body that rotates on the 1-axis (\mathcal{A}), on the 2-axis (\mathcal{B}) and on the 3-axis (\mathcal{C}).
- $\mathcal{D}, \mathcal{E}, \mathcal{F}$ The off-diagonal elements of the matrix of inertia \mathfrak{S} (equation 6-22).
- $\vec{\mathcal{F}}$ Term \mathcal{F} with an arrow on top symbolizes a general force (equation 6-7). \mathcal{F} is also the shear force that acts on a cross-sectional area (equation 7-100).
- \mathcal{G} The objective function (equation 5-11).
- \mathcal{H} The Hamilton function or Hamiltonian (equation 5-2 or 5-9).
- \mathfrak{S} The matrix of inertia of a rigid body (equation 5-22); or, in case of a simple rotation, just the moment of inertia for the rotation on this axis (equation 7-82). The index indicates the reference point, for example M means “with respect to the center of mass”.
- \mathcal{I} The vector function that contains numerical integrations (equation 5-20).
- \mathcal{J} A functional, function of the final conditions (equation 5-13).
- \mathcal{K} The switch function (equations 5-30, 5-64, 5-77 or 7-14).
- \mathcal{M} The molecular mass of the gas in gram/mol (equation 2-28).
- $\vec{\mathcal{M}}$ A general torque, often used with a vector arrow on top (equations 6-8). \mathcal{M} can also be the a disturbing torque that acts on a launcher (equation 7-85) or the bending moment inside the vehicle (equation 7-103). When \mathcal{M}_α is used with the index α , then it is the unit aerodynamic torque, related to the the angle of attack α (equation 7-85).
- \mathcal{N} The normal force that acts on a cross-sectional area inside the launcher (equation 7-101).
- \mathcal{Q} The instrumentation mass of an electrically propelled space probe related to the capacity of the launch vehicle (equation 9-21).
- \mathfrak{R} The individual gas constant, \mathfrak{R} is the universal gas constant $\mathfrak{R}_o = 8314.5$ J/kmol K divided by the mean molecular mass \mathcal{M} (equation 1-9 or 2-26).
- \mathcal{R} The payload mass of a rocket stage (or space launcher) related to the ignition mass of the vehicle (equation 2-14 or 9-13).
- \mathcal{S}, s The entropy of the gas in the thrust chamber (equation 2-44 and 2-60).
- u, v, w The elements of the velocity vector of a flowing gas molecule inside the thrust chamber (equation 2-118).

Roman letters:

- a* In chapter 2 “Rocket Propulsion”: the speed of sound (equation 2-72); in all other chapters: the major semiaxis of the elliptic orbit (equation 4-53).
- A* Generally an area; for example the reference area for the aerodynamic coefficients (equations 4-24 and 4-25) or the duct area (equation 2-80 or 2-108).
- b* The minor semiaxis of the elliptic orbit (equation 4-53).
- B* The ballistic factor of a re-entering space vehicle, with the dimension m^2/kg (equation 10-5).
- c* The effective exhaust velocity, the velocity of the exhausted gases with respect to the final area of the rocket nozzle (equation 2-3 or 4-29). In chapter 2 the term *c* is also used to determine the actual velocity of a certain gas molecule inside the rocket nozzle before the exhaustion (equation 2-118).
- c_p, c_v* The specific heats at constant pressure or at constant volume (equation 2-30 or equations 2-56, 2-57).
- C* Generally a constant. Terms *C* are often used throughout the entire book, but usually these different constants are characterized by an individual index. For example, the aerodynamic coefficients for lift and drag are called *C_L* and *C_D* (equations 4-26 and 4-27).
- d, D* The drag deceleration *d* and the drag force *D*, sometimes with a vector array on top when the vector property is important (equations 4-24, 7-51, 7-70, or 10-2).
- e, E* Generally the energy *E* or the specific (mass related) energy *e*. Term *E* is used for the inner energy of a gas (equation 2-25 and 2-54); as well as for the mechanical energy on an orbit (equation 4-42) or the kinetic energy of a spinning satellite (equation 6-38).
- EI* The flexural stiffness (equation 7-103).
- e* The Euler number, $e = 2.718281828459045\dots$
- f* The derivatives of the state variables (equation 5-1).
- g* Functions that determine the constraints (equation 5-8).
- g* The gravitational acceleration (equation 1-3 or 4-16).
- G* The gain factor (equation 7-87). *G_α* is the gain factor that belongs to the rotation angle and *G_{α̇}* is the gain factor that belongs to the rotation rate.

- h, H In chapter 2 the chamber enthalpy H or the specific (mass related) enthalpy h (equation 2-25 or 2-55).
- h, H The specific (mass related) angular momentum h of the orbiting spacecraft with respect to the gravitational center (equation 4-44); or the total angular momentum H of a spinning satellite with respect to the reference point that is given as the index (equation 6-6). For example, H_M is the angular momentum with respect to the center of mass of the satellite.
- h_{scale} The atmospheric scale height (equation 1-12), for the standard conditions on the surface of the earth approximately 8.431 km.
- $I_{specific}$ The specific impulse (equation 2-5).
- J, J_o The Jacobi integral for a spacecraft that coasts in the gravitational field of two celestial bodies (equation 8-14). J_o is the value of the Jacobi integral at a certain position with “zero relative velocity”.
- $J_2, J_3...$ The dimensionless quantities which take into consideration that the gravitational field of a real celestial body has not exactly a spherical form (equation 1-8). For example, J_2 characterizes the oblateness of the body.
- k The oscillation mode (equation 7-121).
- $K(T)$ The chemical equilibrium constant (equation 2-41).
- l, L The lift acceleration l and the lift force L , with a vector array on top when the vector property is important (equations 4-25, 7-69, or 10-1).
- m The mass flow rate of the rocket engine (the rate of expenditure of propellant, equation 2-1).
- M Generally a mass, for example the actual rocket mass (equation 2-2). M is also the mass of a planet (equation 1-1); or, with an appropriate index, M is also used to symbolize the mass of an engine or launcher subsystem (for example the payload mass, the propellant mass or the nozzle mass).
- $Mach$ The Mach number, quotient of velocity and sonic speed (equation 2-73).
- n The amount of substance of a constituent of the fire gas in the combustion chamber (equations 2-27).
- N The number of stages of a launch vehicle (equation 2-15) or the navigation coefficients (equations 7-73, 7-74).
- p Generally the pressure, for example the actual pressure of the gas in the thrust chamber (equation 2-29) or the actual air pressure (equation 1-9).
- P An abbreviation (“primer”) according to the definition (5-32) or the primer vector (equation 5-61).

- q The specific (mass related) heat that enters the hot flowing gas in the rocket nozzle (equation 2-60), or the specific (area related) aerodynamic stagnation point heat load that the re-entering space vehicle experiences on its entry trajectory (equation 10-4).
- q The control variables (equation 5-1).
- Q The aerodynamic stagnation point pressure (equation 7-66).
- r The distance between the gravitational center and the space vehicle, or, using a vector arrow on top, the vector that locates the space vehicle in relation to the gravitational center (equation 4-1).
- R Generally a displacement or a distance (“radius”), individualized by its index. For example R_{earth} is the equatorial radius of the earth, approximately 6378 km. The radius of gyration is determined by the symbol $R_{\mathfrak{S}}$, or, with reference to the principal body axis frame, by $R_{\mathcal{A}}$, $R_{\mathcal{B}}$ and $R_{\mathcal{C}}$ (equation 6-25 and 6-26).
- s, S The thrust acceleration s and the thrust force S , sometimes with a vector arrow on top (equation 2-4 or 4-29).
- t, T The time coordinate t and the orbital period T (equation 4-55).
- T In chapter 2 the temperature of the gas in the thrust chamber.
- u, v The components of the velocity vector in 1-direction u and in 2-direction v with respect to the moving coordinate system. The component in 3-direction equals zero by definition (equation 4-1).
- v The velocity, often with a vector arrow on top (equation 4-1).
- \vec{V} The linear momentum vector (equation 6-5).
- v The chamber volume (equation 2-25).
- w The velocity of the incident wind in relation to the moving vehicle (equation 4-22).
- W The mechanical work stored in form of potential energy (equation 4-17).
- x, X Generally a variable or a vector (with an arrow on top). The elements of the vector \vec{x} are x_1 , x_2 and x_3 are used to locate a mass element with respect to the center of a Cartesian coordinate system (equation 6-21).
- x, y, z The three axes of a general Cartesian coordinate system. The coordinate x aims usually in a horizontal (or axial) direction, the coordinate y aims in vertical (or radial) direction.

References

- [1] Agrawal, B.N.; *Design of Geosynchronous Spacecraft*; Prentice-Hall Inc., Englewood Cliffs, New York, 1986.
- [2] Allen,H.J.; *Hypersonic Flight and the Re-Entry Problem*; Journal of the Aeronautical Sciences, Vol.25, 1958.
- [3] Anderson,J.D.; *Hypersonic and High Temperature Gas Dynamics*; McGraw-Hill, New York, 1989.
- [4] Büdeler,W.; *Geschichte der Raumfahrt*; Stürtz Verlag, Würzburg, 1982.
- [5] Barrère,M.; Jaumotte,A.; Fraeijs de Veubeke,B.; Vandenkerckhove,J.; *Rocket Propulsion*; Elsevier Publishing Company, Amsterdam, 1960.
- [6] Beatty,J.K.; Petersen,C.C.; Chaikin,A.; *The New Solar System*; Sky Publishing Co., Cambridge, Massachusetts, 1999.
- [7] Bello-Mora, M.; Hechler, M.; *DeltaV-EGA Comet Rendezvous and Earth Return*; ESOC-MAO Working paper, No. 237, 1986.
- [8] Berlin, P.; *Satellite Platform Design*; Kiruna Space and Environment Campus, Kiruna Sweden, 2003.
- [9] Brüning,G.; Hafer,X.; *Flugleistungen*; Springer-Verlag, Berlin Heidelberg New York, 1978.
- [10] Brewer,G.P.; *Ion Propulsion Technology and Applications*; Gordon and Breach Science Publishers, New York, 1970.
- [11] Bryson,A.E.; Ho,Y.C.; *Applied Optimal Control*; Hemisphere Publishing Corporation, New York, 1975.
- [12] Danby,J.M.A.; *Celestial Mechanics*; Willmann-Bell, Richmond, Virginia, 1988.
- [13] Engelhardt,W.; *Planeten Monde Ringsysteme*; Birkhäuser Verlag, Basel, 1984.
- [14] Farquhar,R.W.; *Mission design for a halo orbiter of the Earth*; Journal of Spacecraft, Vol.14, pp. 170-177, 1977.
- [15] Fay,J.A.; Riddell,F.R.; *Theory of Stagnation Point Heat Transfer in Dissociated Air*; Journal of the Aeronautical Sciences, Vol.25, 1958.
- [16] Fischer,H-M.; *Europas Trägerrakete Ariane*; Stediner Verlag Lemwerder, 2002.
- [17] Gürtler,J.; Dorschner,J.; *Das Sonnensystem*; Johann Ambrosius Verlag Leipzig, 1993.

- [18] Goetzel, C.G.; Rittenhouse, J.B.; Singletary, J.B.; *Space Materials Handbook*; Addison-Wesley Publishing Company, Reading Massachusetts, 1965.
- [19] Groepper, P. and others; *Solar sails - an innovative and enabling technology for gossamer space structures*; 54th International Astronautical Congress, IAC-03-U.2.a.08, 2003.
- [20] Hallmann, W.; Ley, W.; *Handbuch Raumfahrttechnik*; Carl Hanser Verlag, Munich Wien, 1999.
- [21] Igenbergs, E.; Kuzcera, H.; Weishaupt, U.; *Results of Impact Simulation for the Giotto Mission*; Acta Astronautica, Vol.10, No.12, 1983.
- [22] Isakovits, S.J.; *International Reference Guide to Space Launch Systems*; AIAA, Washington D.C., 1991 and 1999.
- [23] Jahn, R.G.; *Physics of Electric Propulsion*; McGraw-Hill Book Company, New York, 1968.
- [24] Kaplan, M.H.; *Modern Spacecraft Dynamics and Control*; John Wiley and Sons, New York, 1976.
- [25] King-Hele, D.; *Satellite Orbits in an Atmosphere*; Blackie and Son Ltd., Glasgow, London, 1987.
- [26] Kozak, D.V.; Moraes, P.; *AEROCAD - A Design Tool for Aerospace Vehicle Aerodynamics Preliminary Design*; Ingenieurbüro Dr.Schlingloff, Munich, 1993.
- [27] Krafft A.Ehrlicke; *Space Flight*; D.Van Nostrand Company, Inc., Princeton, New Jersey, 1962.
- [28] Lawden, D.F.; *Optimal Trajectories for Space Navigation*; Butterworth, London, 1963.
- [29] Lawden, D.F.; *Analytical Methods of Optimization*; Scottish Academic Press, Edinburgh and London, 1975.
- [30] Leitmann, G.; *Foundations of Optimal Control Theory*; McGraw Hill, New York, 1966.
- [31] Magnus, K.; *Kreisel*; Springer Verlag, Berlin-Heidelberg-New York, 1971.
- [32] Marek, J.P.; *Optimal Space Trajectories*; Elsevier, Amsterdam, 1979.
- [33] Martin, J.A.; Manski, D; *Variable-Mixture-Ratio and Other Rocket Engines for Advanced Shuttle*; AIAA-89-2282, 25th Propulsion Conference, Monterey, 1989.

- [34] Marty,D.; *Conception des Véhicules Spatiaux*; Masson, Paris, N.Y., Barcelone Milan, Mexico, São Paulo, 1986.
- [35] Messerschmid,E.; Bertrand,R.; Pohlemann,F.; *Raumstationen*; Springer Verlag, Berlin Heidelberg New York, 1997.
- [36] Minovitch, M.A.; *Fast Missions to Pluto Using Jupiter Gravity-Assist and Small Launch Vehicles*; Journal of Spacecraft and Rockets, Vol.31, No.6, 1994.
- [37] Minovitch, M.A.; *A Method for Determining Interplanetary Free-Fall Reconnaissance Trajectories*; JPL, TM 312-130, pp.38-44, 1961.
- [38] Oertel,H.; *Aerothermodynamic*; Springer Verlag, Berlin Heidelberg, 1994.
- [39] Pontryagin,L.S.; *The Mathematical Theory of Optimal Processes*; Interscience, 1962.
- [40] Regan,F.J.; Anandakrishnan,S.M.; *Dynamics of Atmospheric Re-Entry*; AIAA Education Series, Washington DC, 1993.
- [41] Richardson, D.L.; *A note on a Lagrangian formulation for motion about the collinear points*; Celestial Mechanics, Vol.22, pp. 231-253 (a and b), 1980.
- [42] Ruppe,H.O.; Schlingloff,H.; *Assessment of a European Pluto Flyby Mission*; Jahrbuch der DGLR, Bremen, DGLR-JT98-130, 1998.
- [43] Ruppe,H.O.; Schlingloff,H.; *Manned Mars Mission Revisited*; Jahrbuch der DGLR, Munich, DGLR-JT2003-012, 2003.
- [44] Ruppe,H.O.; *Introduction to Astronautics*; Vol.I and II., Academic Press, New York, 1966..
- [45] Ruppe,H.O.; *Die grenzenlose Dimension- Raumfahrt*; Volume I and II, ECON-Verlag, Düsseldorf-Wien, 1981.
- [46] Ruppe,H.O.; *Expedition to Mars - a baseline mission now*; 3.IAC, Lausanne, IAF-84-198, 1984.
- [47] Ruppe,H.O.; *Design Considerations for Future Space Launchers*; Acta Astronautica Vol.29, No.9, pp. 705-722, 1993.
- [48] Sanders, H.M.; Koelle, D.E.; Schöyer, H.F.R.; *Cost Optimum Launcher Design by Cost and Performance Normalization*; 49th International Astronautical Congress, IAA-98-IAA.1.1.02, 2002.
- [49] Sanders, H.M.; *Analysis of the Technical Characteristics and Performance of the Small Cosmos Launcher*; JBIS, Vol 54 pp 377 to 388, 2001.

- [50] Schlingloff,H.; *SKYNAV User's Documentation*; Ingenieurbüro Dr.Schlingloff, Munich, 1991.
- [51] Schlingloff,H.; *Control Laws for Optimal Spacecraft Trajectories*; AIAA-Journal of Spacecraft and Rockets, V 24, No.1, pp.48-51, Jan.-Feb.1987.
- [52] Schlingloff,H.; *Bahn- und Kostenoptimierung wiederverwendbarer Raumfährenoberstufen*; Dissertation, TU-Munich, 1983.
- [53] Schlingloff,H.; *Ein Optimierungsparadoxon bei Multi-Asteroiden-Missionen*; Jahrbuch der DGLR, Bonn, 1985.
- [54] Schlingloff,H.; *Optimierung von Multi-Asteroidenmissionen mit elektrischem Antrieb durch den Einsatz von mehreren Einzelsonden*; Zeitschrift für Flugwissenschaften und Weltraumforschung 9, Heft 6, pp. 349-355, 1985.
- [55] Schlingloff,H.; *Stufungsprinzipien bei elektrisch angetriebenen Raumfahrzeugen*; Jahrbuch der DGLR 86, pp. 86-103, 1986.
- [56] Schlingloff,H.; *A Trajectory Optimization Program for Spacecraft Navigation*; Second International Symposium on Spacecraft Flight Dynamics, ESA SP-255, pp.103-108, Darmstadt, 1986.
- [57] Schlingloff,H.; *Trajectory Optimization Techniques with Application to Mission Design of Solar-Electric Propulsion Spacecraft*; 19th AIAA/DGLR/JSAASS International Electric Propulsion Conference, Colorado Springs, 1987.
- [58] Schlingloff,H.; *Preliminary Design Techniques for Interplanetary Solar-Electric Propulsion Spacecraft*; Computational Mechanics Publications, Springer-Verlag, Berlin-Heidelberg-New York, pp.1-8, 1987.
- [59] Schlingloff,H.; *Fundamental Design Principles for Orbital Transfer Systems to Future Space Launchers*; Proceedings of COBEM-87, Universidade Federal de Santa Catarina, Florianopolis-SC, Brasil, pp. 887-890, 1987.
- [60] Schlingloff,H.; *Beiträge zur Optimierung der Flugleistung des europäischen Raumtransport-Systems ARIANE-5P/HERMES*; Jahrbuch der DGLR, Darmstadt, 1988.
- [61] Schlingloff,H.; *European Airborne Launch Capacity using ARIANE Components*; AIAA/DGLR Fifth International Aerospace Planes and Hypersonic Technologies Conference, Munich, 1993.
- [62] Schlingloff,H.; *Moderne CAD/CAE-Systeme für die Entwicklung von Raumtransportsystemen*; VDE-VDI Zeitschrift Mikroelektronik und Mikrosystemtechnik, Bd.8, Heft 4, pp. 230-233, 1994.
- [63] Schmucker,R.; *Hybridraketen*; Wilhelm Goldmann Verlag, Munich, 1973.

- [64] Sharpe,M.R.; *Living in Space*; Aldus Books London, 1969.
- [65] Sidi,M.J.; *Spacecraft Dynamics and Control*; Cambridge University Press, 1997.
- [66] Stuhlinger,E.; *Ion Propulsion for Space Flight*; McGraw-Hill Book Company, New York, 1964.
- [67] Sutton,G.P.; *Rocket Propulsion Elements*; John Wiley & Sons Inc., New York, sixth edition, 1992.
- [68] Szebehely,V.; *Theory of Orbits*; Academic Press, New York, 1967.
- [69] Szebehely,V.; Mark,H.; *Adventures in Celestial Mechanics*; John Wiley & Sons Inc., New York, 1998.
- [70] Thomsom,W.T.; *Introduction to Space Dynamics*; Dover Publications Inc., New York, 1986.
- [71] Vinh,N.X.; *Optimal Trajectories in Atmospheric Flight*; Elsevier Publishing Co., Amsterdam-Oxford-New York, 1981.
- [72] Wiesel,W.E.; *Spaceflight Dynamics*; McGraw-Hill, New York, 1989.
- [73] Wilson,A.; *Jane's Space Directory*; Jane's Information Group, London, 1997-1998.
- [74] Wright,J.L.; *Space Sailing*; Gordon and Breach Science Publishers, Philadelphia, 1992.
- [75] Yoshino,T.; Schlingloff,H.; Boscov,J.; *Improvement of Payload Capability of the Brazilian VLS Satellite Launcher*; 17th International Symposium on Space Technology and Sience ISTS, Tokyo, Mai 1990.

Index

- Ablative heat shield, 486
- Absolute changes, 171
- Accelerations, 22
- Acceleration
 - with respect to inertial space, 172
 - by the lift force, 177
 - control, 340
- Accessible regions, 372
- Accurate representation of the field of gravity, 13
- Advance of the perigee, 196
- Aerocapture, 384
- Aerocapture at arrival, 410
- Aerodynamic
 - coefficients, 177
 - landing on the surface, 411
 - forces and moments, 467
 - surfaces, 481
- Air-breathing engines for space transportation, 160
- An airborne launcher composed of Ariane-4 stages, 149
- Analysis of the control equations, 220
- Angular
 - displacement of the reference system, 260
 - momentum with respect to the gravitational center, 183
- Angular momentum vector
 - in a moving coordinate system, 262
 - of a dual-spin satellite, 275
- Angular momentum and energy, 381
- Application of
 - Newton's law, 181
 - the optimization theory, 307
- Arcjets, 421
- Ariane-5 design concept, 145
- Ascent to orbit, 412
- Asteroids of the Kuiper belt, 7
- Atmosphere
 - for breathing, 23
 - of Saturn's moon Titan, 19
- Atmospheric
 - pressure, 14
 - density, 176
- Attitude
 - control, 335
 - controller, 336
 - control, 346
 - reference gyro, 283
- Attraction
 - by the gravity of a celestial body, 173
 - in the central gravitational field, 289
- Axisymmetric
 - spacecraft, 265
 - space station, 297
- Balancing of wheels, 278
- Ballistic
 - factor, 470
 - missiles, 114
- Barometric representation, 14
- Basis of
 - anticipated technology, 27
 - available technology, 27
 - science fiction, 27
- Batteries, 430
- Behaviour of the switch function for impulsive transfers, 222
- Bending vibrations, 341

- Benefits of
 - airborne launching, 149
 - using an aircraft as a launch platform, 323
- Big solid strap-on booster motors, 100
- Biggest nuclear reactor of our solar system, 2
- Bipropellant thrusters, 110
- Booster motors for selection, 119
- Boundary layers and heat transfer, 464
- Breakwell diagram, 383
- Burning rate and
 - chamber pressure, 102
 - grain temperature, 102
- Burning solid propellant grain, 48
- Burnout control and reignition, 107
- Cabin temperature, 23
- Cartesian coordinate systems, 164
- Cascade control, 333
- Center of mass of an extended object, 252
- Chamber
 - enthalpy, 51
 - size, 90
- Change of the inclination angle, 238
- Changing of
 - the velocity, 199
 - orbital elements, 378
- Characteristic equation, 294
- Charon, 7
- Chemical
 - equilibrium, 58
 - non-equilibrium, 462
 - reaction process, 54
 - reactions, 462
- China's Long March launcher family, 132
- Ciolkovskij's equation, 35
- Circular low earth orbit LEO, 196
- Circularization in GEO position, 117
- Classification of rocket propulsion systems, 32
- Classification, 109
- Classification of transfer maneuvers, 224
- Closed cycle engines, 94
- Coast arcs, 185 , 223 , 437
- Cold gas thrusters, 109
- Combustion instabilities, 49
- Comets, 5
- Commercial LEO launch vehicles, 133
- Comparison of
 - mixed mode and single mode operation, 309
 - Venus-Earth and Mars-Earth maneuvers, 396
 - the mission alternatives, 408
- Composite
 - propellant, 45
 - solid rocket propellant, 50
- Composite solid propellant, 100
- Computation of the
 - free molecular flow regime, 466
 - continuum flow regime, 466
- Conditions for
 - having an atmosphere, 18
 - the occurrence of a pressure shock, 73
- Conditions for the stability of a dual-spin satellite, 276

- Conservation
 - laws for a diabatic flow process, 77
 - of matter, 82
- Conservation of
 - linear momentum, 82
 - the angular momentum, 83
- Constant thrusting, 437
- Construction of the flow field, 87
- Continuation of the mission, 392
- Continuous control of the mass flow rate, 303
- Continuum flow regime, 459
- Contours of zero relative velocity, 371
- Control
 - equations for the flight on circular orbits, 224
 - of impulsive escape trajectories, 242
- Control of the
 - thrust direction, 229
 - banking angle, 317
 - angle of attack, 318
 - thrust magnitude, 319
 - initial pull-up maneuver, 323
- Control of the
 - initial entry phase, 479
 - essential braking phase, 479
- Cooling of the thrust chamber wall, 91
- Coplanar
 - return, 385
 - transfer with intermediate impulse, 203
- Cost models in space transportation, 136
- Costs of
 - a space launcher, 139
 - liquid rocket stages, 137
 - solid rocket stages, 138
- Deceleration by the drag force, 176
- Definition of conic orbits, 191
- Definitions, 252
- Deflection of the
 - heliocentric orbit, 376
 - earth-return trajectory, 391
- Density of the atmosphere, 470
- Departure
 - date and flight time, 392
 - for the earth, 413
- Differential vector equations for optimal maneuvers, 230
- Dimensionless numbers in aerodynamics, 465
- Direction of the principal axis, 259
- Disadvantages of a Jupiter swingby maneuver, 399
- Discontinuous control of the mass flow rate, 303
- Dissipation of energy, 270
- Dissociation, 461
- Docking, 250
- Double-base propellant, 45
- Double earth swingby, 394
- Drag deceleration, 313
- Duct area, 70 , 79
- Dynamical Euler equations, 263
- Earth-moon gravitational system, 358
- Earth, 3
- Earth
 - departure seen from the heliocentric system, 375
 - escape trajectories, 112

- Eccentricity and pericenter location, 379
- Effect of
 - rest atmosphere, 193
 - the flyby distance restriction, 389
- Ejection principle, 31
- Electric propulsion for satellites, 434
- Electron bombardment thrusters, 422
- Ellipsoid of inertia, 258
- Energy and
 - angular momentum of conic orbits, 185
 - work, 262
- Entropy change, 74
- Environment for the payload, 144
- Equations of
 - motion for orbital flight, 184
 - motion, 216
 - motion for a transfer trajectory in vector form, 228
- Equations of motion for
 - the gravity gradient satellite, 292
 - small perturbations, 294 , 297
- Equations of motion, 313 , 335 , 450 , 453 , 477
- Erosive burning and surface cracks, 102
- ESA's
 - Ariane-4 launcher family, 122
 - Ariane-5 launcher, 122
- Escape
 - spiral, 436
 - utilizing a lunar gravity assist maneuver, 372
- Estimated performance, 78
- Euler and Cardan angles in the spherical triangle, 168
- Execution of the lunar missions, 356
- Exhaust
 - temperature, 77
 - velocity and chamber enthalpy, 53
 - velocity, 78
- Exosphere, 16
- Expansion ratio, 72
- Exploring the sun, 2
- Fast mission, 408
- Fiber
 - material cases, 105
 - reinforced composites, 484
- Field emission electric propulsion, 425
- Final conditions for
 - flyby maneuvers, 226
 - escape maneuvers, 227
- Flight
 - precision, 143
 - safety, 143
- Flyby velocity, 376
- Food, water, oxygen and hygiene, 24
- Four different airborne launcher concepts, 150
- Free molecular flow regime, 458
- Fuel cells, 431
- Future
 - aerospace planes, 161
 - of the public interest in space-planes, 161
- Gain factors, 339
- Gas generators, 94
- GEO capacity, 140
- Geometry of
 - conic orbits, 186
 - the escape orbit, 374

- Geostationary
 - transfer orbit GTO, 112
 - earth orbit GEO and geostationary transfer orbit GTO, 19
- Gimbal frame of reference, 280
- Gimbaled actuator gyro, 285
- Global players, 126
- Grain stresses and failures, 104
- Gravitational
 - acceleration, 11
 - acceleration of a spherical celestial body, 12
- Gravity assistance from
 - Jupiter and Saturn, 399
 - Jupiter, Saturn and Neptune, 400
- Gravity gradient
 - vector, 290
 - torque, 290
- Gravity
 - swingby at the earth, 387
 - swingbys, 455
- Growth factor, 115
- GTO market, 113
- Guidance during the braking maneuver, 482
- Gyrostabilizers, 285
- Hall thrusters, 429
- Halo orbits, 368
- Hamilton function and Lagrange equations, 315
- Hardware, 354
- Heat
 - flux in the aerodynamic stagnation point, 469
 - input, 76
 - of reaction and chamber enthalpy, 55
- Heliocentric view of the flyby maneuver, 377
- High energetic liquid propellant LH2/LOX, 117
- Hohmann
 - Mars mission, 405
 - transfer with inclination change, 204
- Human health factor, 26
- Hydrazine, 43
- Hydrazine monopropellant thrusters, 110
- Hyperbolic
 - earth escape trajectories, 198
 - escape from circular low earth orbit, 206
 - earth escape velocity, 374
 - insertion, 438
- Ignition
 - and control, 97
 - systems for solid rocket motors, 107
- Improvement of the Ariane-5, 146
- Impulsive
 - inclination change maneuvers, 201
 - transfer trajectories, 221
 - thrust arcs, 221
- Impulsive thrust attitude maneuver, 268
- In-flight handling, 41
- Inaccuracy of aerodynamic coefficients, 178
- Inclination of the transfer orbit, 378
- Inertial platform, 284

- Inflatable heat shield, 486
- Initial
 - altitude and inclination, 451
 - conditions, 318
- Injection
 - of the mission, 415
 - zone, 46
- Injectors, 90
- Instabilities, 47
- Integral heat flux, 473
- Integrating rate gyro, 282
- Interaction of solar wind and magnetic field, 21
- Interplanetary escape from
 - LEO, 207
 - GTO, 207
- Interplanetary low thrust transfer trajectories, 446
- Intersection of the characteristics, 86
- Ion contact thrusters, 423
- Ionization, 461
- Isentropic change in state, 66
- Iterative path adaptive guidance, 328
- Jacobi's integral, 370
- Japan's
 - H2 launch vehicle, 130
 - H1 space launch vehicle, 130
 - M-V launch vehicle, 130
- Jet damping, 335
- Journeys to the outer planets, 456
- Jupiter, 6
- Jupiter gravity assist on the way to Pluto, 398
- Kepler's
 - flight time equation, 188
 - laws, 187
- Kerosene/LOX, 117
- Kerosene as fuel, 43
- Kinematical Euler equations, 264
- Kinetic energy, 182
- Lagrange's equations for the perturbation of orbital elements, 192
- Lagrange equations, 218
- Lambert's theorem, 190
- Landing on the moon, 362
- Laser-gyro, 284
- Launch
 - contract, 144
 - location loss factor, 141
- Launcher
 - as a multi-body system, 352
 - families, 119
- LEO
 - capacity, 140
 - market, 114
- Lift limitation, 314
- Linear momentum and angular momentum of
 - a mass particle, 253
 - an extended object, 253
- Linear proportional terminal control, 329
- Lines and valves, 89
- Liquefied gases, 42
- Liquid
 - hydrogen LH2 as fuel, 42
 - oxygen LOX as oxidizer, 42
 - propellant rocket motors, 33
 - propellents, 88
 - rocket propellant, 50

- Load carrying aluminum substructure, 484
- Location and velocity of a flying spacecraft, 165
- Low
 - earth orbit LEO, 112
 - thrust escape trajectories, 242
- Lunar rendezvous, 360
- Magnetic
 - dipole moment, 20
 - field lines, 20
- Main belt asteroids, 5
- Manned
 - interstellar missions, 30
 - missions to planet earth, 28
- Manned missions to other planets, 28
- Market for earth observation and navigation, 114
- Mars, 4
- Mars
 - atmosphere, 19
 - spacecraft, 414
- Mass
 - flow rate profile, 307
 - ratio of multi-stage rockets, 37
- Mathematical formalism to solve trajectory optimization problems, 211
- Maximum
 - deceleration, 471
 - heat flux, 472
- Medium earth orbit MEO, 112
- Mercury, 3
- Mesosphere, 16
- Metal cases, 104
- Metallic hydrogen, 162
- Meteorites and space debris, 26
- Mission
 - alternatives, 441
 - cost, 418
 - costs, 446
 - execution, 416
 - objective, 443
- Mixed mode, 306
- Modeling the sloshing motion, 351
- Molniya orbits, 198
- Molnya orbits, 112
- Moment of
 - constant precession, 281
 - inertia matrix, 256
 - inertia, 277
- Momentum wheel, 288
- Moon, 4
- Moon landing maneuvers, 244
- More detailed engine mass model, 99
- Motion of
 - charged particles, 21
 - liquids in tanks, 351
- Motor ignition, 48
- MPD-arcjets, 427
- Multiple launching, 119
- Natural frequencies of the bending vibration, 344
- Navigation, 326
- Navigation accuracy, 329
- Near earth asteroids, 5
- Nearly optimum sailcraft escape strategy, 451
- Neptune, 6

- Newton's
 - approximation, 466
 - law, 253
- Nitrogen tetroxide as oxidizer, 44
- Node angle, inclination angle and pericenter angle, 191
- Noise and vibrations, 22
- Nomex felt surface insulation, 485
- Nozzle
 - contour, 90
 - shape, 105
- Nuclear propulsion, 162
- Numerical
 - eccentricity of conic orbits, 185
 - solution to a trajectory optimization problem, 214
- Nutation
 - dampers, 272
 - frequency, 267
- Objective of the mission, 418
- Oblateness of celestial bodies, 13
- Open cycle engines, 94
- Optimal control using
 - inertial coordinates, 231
 - polar coordinates, 232
 - energy and angular momentum as coordinates, 234
- Optimal
 - exhaust velocity, 434
 - final pressure, 72
 - number of stages, 38 , 116
 - size of the stages, 38
 - thrust magnitude control, 230
- Optimization of
 - functionals, 210
 - the final path angle, 225
- Optimization of the
 - flight time, 226
 - thrust direction, 300
 - mass flow rate, 302
- Optimum size of the stages, 116
- Orbital tower satellite, 162
- Orientation
 - of the magnetic dipole, 20
 - angles, 164
- Origin of atmospheres, 18
- Parabolic earth escape trajectories, 198
- Parallel
 - axis displacement, 261
 - staging, 40 , 116
- Partial differential equation for the bending vibrations, 342
- Payload mass ratio, 36
- Perfect gas flow regime, 460
- Performance
 - optimization, 62
 - as a quality factor of a launcher, 140
- Pericenter and apocenter, 186
- Periodic orbits in the earth-moon system, 362
- Phobos and Deimos, 4
- Photo-voltaic generators, 432
- Plane
 - impulsive maneuvers, 200
 - thrust angle control, 236
 - transfer between two circular orbits, 202
- Planetary capture, 384
- Pluto's atmosphere, 19
- Pluto, 7
- Point of application, 164
- Polar earth orbit PEO, 112

- Political environment for a manned Mars mission, 418
- Possibilities to support the lift-off phase, 117
- Potential energy, 174
- Precession of a spinning satellite, 267
- Pressure
 - gas, 108
 - inside the chamber, 51
 - ratios in the overexpanding nozzle, 74
 - shocks, 73
- Principal body axis frame, 257
- Principle of conservation of matter, 57
- Progress by precaution and redundancy, 354
- Propellant accommodation, 88
- Propellant, 108
- Propellant
 - for control thrusters, 50
 - feeding cycles, 94
 - geometry, 103
 - performance, 41
 - production and ground handling, 42
 - performance analysis, 51
- Pumps and turbines, 92
- Qualification and quality, 120
- Quasi ballistic reentry trajectories, 475
- Radiation, 25
- Radiation belts, 20
- Radio-frequency ion thrusters, 424
- Radioisotope generators, 431
- Rapid combustion zone, 46
- Rate gyro, 281
- Reacting expansion process, 75
- Reaction wheel, 287
- Redundant
 - core stage engines, 121
 - upper stage engines, 121
- Reentry flight phases, 476
- Regression of the node, 194
- Regulators and terminal controllers, 327
- Relative
 - changes, 171
 - velocity and inertial velocity, 172
- Relative velocity of the atmosphere, 176
- Rendezvous and flyby, 246
- Replacement of satellites, 114
- Resistojets, 420
- Resistor jets and arc jets, 109
- Resonant return, 386
- restricted three body problem, 359
- Restrictions or constraints, 212
- Return
 - from space, 22
 - to the same point in space, 385
- Reusability of space launcher components, 154
- Reusable
 - ballistic SSTO vehicle, 158
 - space launchers, 152
- Roll control, 335
- Rotational dynamics of the vehicle, 334
- Rotations on the principal
 - body axes, 257
 - axis of inertia, 269

- Russia's
 - Proton launcher, 128
 - Soyuz launcher, 128
- Safety by using a parallel staged launcher concept, 120
- Sailcraft
 - propulsion, 433
 - trajectory optimization, 454
- Satellite
 - attitude maneuver, 286
 - position stabilization, 250
- Saturn, 6
- Scale height, 15
- Search for extraterrestrial life, 30
- Semi-rigid spacecraft, 271
- Semimajor axis and eccentricity, 191
- Several compatible upper stages, 119
- Shape of the satellite and the payload bay, 274
- Shock waves and expansion waves, 463
- Simple stage mass model, 98
- Simplified equations, 311
- Situation in the throat, 79
- Size of the cabin, 23
- Small
 - all solid launchers, 101
 - LEO launchers, 134
 - solid strap-on booster motors, 100
- Solar sails in comparison with gravitational maneuvering, 456
- Solid propellant
 - rocket motors, 33
 - for space applications, 44
 - ingredients, 45
- Solid
 - propellant, 118
 - pulsed plasma thrusters, 428
 - upper stages, 100
- Space
 - cannon, 162
 - railroad, 162
 - shuttle, 114
- Spaceplane technology, 160
- Specific
 - costs, 142
 - energy of the trajectory, 182
 - enthalpy, 63
 - entropy, 65
 - heat, 52
 - Impulse, 35
 - mechanical energy, 370
- Speed of sound and the Mach number, 68
- Sphere of lunar influence, 372
- SSTO vehicles based on conventional technology, 156
- St-Vernant-Wantzel equation, 69
- Stability
 - diagram, 296
 - of the rotating gravity gradient satellite, 298
- Stability of the liberation points, 364
- Stabilization, 326
- Stages of equal size, 39
- Staging
 - principle for electric propulsion, 439
 - principle, 444
- Stationary attitude, 294
- Steady flow
 - process, 66
 - through a volume element, 81
- Steady state performance, 337

- Storability and stability, 41
- Storable liquid propellant, 118
- Straight line entry, 471
- Stratosphere, 16
- Stream-tube zone, 46
- Sun
 - as variable star, 2
 - synchronous orbit SSO, 197
- Sunlight, 26
- Sunlight pressure, 433
- Superposition of bending vibrations, 346
- Swingby at earth encounter compared with direct injection, 389
- Switch function for a Hohmann transfer, 239
- Symmetrical gyroscopic wheel, 279
- Synodic period, 382
- Target Mars and its orbit around the sun, 405
- Temperatures in the rocket nozzle, 69
- Terminal guidance for the final landing maneuver, 482
- Terraforming, 29
- Thermosphere, 16
- Three collinear liberation points L1, L2 and L3, 364
- Thrust
 - acceleration, 34 , 314
 - angle control for intercept maneuvers, 300
- Thrust
 - angle control for the ascent to orbit, 301
 - chambers, 108
 - direction, 179
 - direction aligned with the attitude of the vehicle, 180
- Thrust
 - direction before and after the impulse, 201
 - magnitude, 178
 - magnitude control, 237
 - of parallel operating rocket engines, 179
 - vectoring, 106
 - vector of a sailcraft, 449
- Tiles, 484
- Time
 - derivatives of Euler and Cardan angles, 170
 - derivative of the angular momentum vector, 183
 - interval of stable rotation, 274
 - of pericenter passage, 192
- Trajectories
 - with several burn periods, 240
 - of electrically propelled spacecraft, 249
- Trajectory
 - control with the Hamilton-Lagrange theory, 330
 - integration, 182
 - optimization, 476
 - profile, 310
 - re-optimization, 327
- Transfer
 - time, 382
 - with several pericenter or apocenter burn maneuvers, 204
- Transformation of
 - the fundamental equations of gas dynamics, 84
 - vector notations, 166
 - Euler angles to Cardan angles, 168
 - Cardan angles to Euler angles, 168
 - the moment vector, 254

- Transformation of the angular momentum
 - vector, 254
 - law, 255
- Transient performance, 337
- Transition from free molecular to continuum flow regime, 458
- Translational and rotational kinetic energy, 262
- Translational, rotational and vibrational excitation of molecules, 461
- Transversality conditions, 213
- Triangular liberation points L4 and L5, 364
- Troposphere, 15
- Two-dimensional axisymmetric flow field, 81
- Typical annual launch frequency, 113
- Ukraine's Zenit launcher, 128
- Uncertainty of the impact point, 473
- Uniform gravitational field without drag, 299
- Unloading of momentum wheels, 288
- Unmanned
 - interstellar missions, 30
 - planetary missions, 29
- Unsymmetrical dimethylhydrazine and monomethylhydrazine, 43
- Uranus, 6
- US-America's
 - Delta-II launch vehicle, 124
 - Atlas-IIA launch vehicle, 124
 - launchers Titan IV and Titan II, 124
- Using
 - additionally Mars or Venus swingby maneuvers, 400
 - an aircraft as a first stage, 158
- Variable mixture ratio, 304
- Variation of state variables and control variables, 330
- Vector
 - representation of the rotation velocity, 169
 - resolution, 164
 - transformation into a notation of the moving system, 167
- Velocity
 - and pressure in the throat area, 71
 - of the atmosphere, 175
 - reduction at aphelion position, 388
 - requirement for the ascent to orbit, 322
- Velocity requirement for a manned Mars mission, 408
- Venus-Mars mission, 407
- Venus, 3
- Venus
 - atmosphere, 19
 - swingbys on earth-return trajectories, 395
- Vertical flight through the atmosphere, 301
- Visits of extra-terrestrial creatures on earth, 30
- Waveforms of the bending vibrations, 345
- Weight of heat protection systems, 486
- Weightlessness, 24
- Wings for the initial pull-up maneuver, 323
- Working principle, 420 , 422 , 426

The purpose of this book is to develop an understanding of the main engineering problems involved in spaceflight. Mathematics has not been avoided, and a graduate level is necessary to follow the derivations. The book gives insight into the engineering of rocket engines, launch vehicles, flight mechanics, system optimization and mission design; it can be recommended to students in advanced lessons as well as to all spacecraft system engineers.

ISBN 3-00-016289-5