Contents

Chapter 1

Introduction

It is difficult to say what is impossible,
for the dream of yesterday is the hope of today
and the reality of tomorrow.
Robert Goddard

That's one small step for a man, one giant leap for mankind.
Neil Armstrong

Space: the final frontier... to explore strange new worlds, to seek out new life and new civilizations, to boldly go where no man has gone before. This quotation from the opening of the original 'Star Trek' science fiction TV series summarizes in a few words the quest of humanity to discover and explore the unknown. For many centuries, exploration was confined to the Earth. Other planets and stars seemed totally out of reach. Yet, some visionary scientists patiently worked towards the realization of their dream. Then, in the middle of the 20th century, a new era begun: the first artificial satellite was launched into an orbit around the Earth. Nowadays, satellites have become essential ingredients of our everyday's life, robotic probes have explored remote places of our Solar System, and the presence of men and women in space has become commonplace. Yet, space exploration remains one of the most challenging and most fascinating undertakings of humanity.
The focus of the present course is mainly on space-borne **scientific** activities. Most of them are carried out either by robotic probes or unmanned satellites. The research activities that fall into these categories essentially deal with planetary science (including the Earth), astrophysics and fundamental physics. While the ultimate goal of many technological developments in the domain of spaceflight is of course to allow human beings to travel to remote places in our Solar System and beyond, we should stress that in many of the activities of interest here, the presence of human beings in space (such as on board the ISS) is not required and is not even desirable.
In this chapter, we will first start with a short (necessarily incomplete) overview of the history of space exploration. We will then consider the different reasons why to do science from space and, more generally, why to go to space. Next we will have a closer look at the life-cycle of a typical space mission and eventually introduce the European Space Agency (ESA) and its science programme.

1.1 A brief history of spaceflight

The history of spaceflight is tightly connected to the development of rockets which are the most-commonly used launch devices for space missions. The first documented usage of rockets dates back to the 13th century when

the Chinese army used small black powder rockets against the Mongols. The rockets were subsequently used as weapons by the Mongols, the Arabs, the French and the English. However, in the 16th century they lost their importance as weapons because of the development of new and more powerful cannons that were more precise and more successful in hitting their target. During several centuries rockets were mainly used for fireworks. It is during this period that the Austrian engineer Conrad Haas (1509 - 1576) made some remarkable developments including a four-stage rocket. At the end of the 18th century, rockets were again used as weapons, but also to fire rescue lines to ships in distress. However, applications of rockets for spaceflight were well beyond reach at that time.

1.1.1 The pioneers of the early 20th century

The history of spaceflight starts with pioneering theoretical considerations in the late 19th - early 20th century. In 1865, the French author **Jules Verne** (1828 - 1905) published his novel *De la Terre à la Lune* (followed in 1870 by the sequel *Autour de la Lune*). The story is about a projectile-like spaceship shot by an enormous cannon to reach a velocity of $11.2\,\mathrm{km\,s^{-1}}$ necessary to travel to the Moon. While Jules Verne's work is clearly a science-fiction novel, it is nevertheless the first book of its kind that attempts to account for most of the physical constraints of spaceflight known at that epoch.

In the early years of the 20th century, three pioneers, who were all heavily inspired by Jules Verne's novels, made substantial contributions to the theoretical conception of spaceflight. In 1903, the Russian scientist **Konstantin Tsiolkovsky** (1857 - 1935) published *The Exploration of Cosmic Space by Means of Reaction Devices* in which he theorized many aspects of space travel and rocket propulsion. For instance, Tsiolkovsky proposed to use multistage rockets fueled with liquid oxygen and hydrogen to reach the velocity necessary to put a spacecraft into a stable orbit around the Earth. Tsiolkovsky also derived a fundamental equation, known as the rocket equation (see Sect. 3.2), or Tsiolkovsky's equation[1].

Robert Goddard (1882 - 1945) was an American physicist who developed several techniques necessary to overcome the limitations of conventional powder rockets. In 1919, he published the book *A Method of Reaching Extreme Altitudes*. Goddard launched the first liquid-fueled rocket in March 1926. The various prototypes that he developed reached speeds up to $885\,\mathrm{km\,hr^{-1}}$. However, his work was not taken seriously by the American authorities. Goddard also worked on the concept of multi-stage rockets (1914) and the stabilization of rocket flights. He first used a nozzle to significantly increase the efficiency of his rocket engine. Many of his ideas were included by the Germans in the design of the V2 during the Second World War.

The Romanian-German scientist **Hermann Oberth** (1894 - 1989) was strongly influenced by Jules Verne's novels. In 1922, he submitted a PhD doctoral thesis work that was rejected as too utopian. Oberth published the work himself under the title *Die Rakete zu den Planetenräumen*. In 1929, he published a more extended version of his work entitled *Wege zur Raumschiffahrt*. In 1929, he launched his own liquid-fueled rocket. During the Nazi regime in Germany, Oberth became a collaborator of his former student **Wernher von Braun** in the development of the A4 (Aggregat 4, better known as the V2) combat rocket[2]. In 1950, Oberth published another visionary book, *Menschen im Weltraum* where he proposed a space-borne reflecting telescope for astronomical observations as well as a space station.

In 1936, the German Army established a rocket development and test centre in Peenemünde on the shores of the Baltic Sea, under the control of General Walter Dornberger. Wernher von Braun (1912 - 1977) acted as the system engineer, supervising the design and the development of the A4 rocket. The first successful launch took place on 3 October 1942. At first, these activities caught little interest from the Nazi regime, but this situation changed in

[1] Actually, this equation was already established 24 years earlier by the Belgian officer Casimir-Erasme Coquilhat (1811 - 1890) in a paper called *Trajectoires des fusées volantes dans le vide* published in the *Mémoires de la Société Royale des Sciences de Liège*.

[2] The purpose of this introduction is to focus only on the technical contributions of Oberth and von Braun. We will not consider their ambiguous and much debated political opinions (in the case of Oberth) or their relationship with the Nazi regime (in the case of von Braun).

Figure 1.1: From left to right: Konstantin Tsiolkovsky, Robert Goddard, Hermann Oberth (sitting in the front) and Wernher von Braun (2nd from the right) along with other US space scientists and General Holger Toftoy who played a major role in operation 'Paperclip', Sergey Korolyov.

1943 when Hitler was looking for a 'Wunderwaffe' in a desperate attempt to prevent the German defeat. From 1943 on, the mass production of the 'Vergeltungswaffe 2' (V2) rocket was done by prisoners of the Mittelbau-Dora concentration camp near Nordhausen. The V2 rocket was used as a weapon mainly against Great Britain and Belgium. The V2 could deliver 900 kg to a target 322 km away. At the end of the war, von Braun and part of his team surrendered to the American troops and as part of operation 'Paperclip', he and several of his main collaborators of the V2 project moved to the US, where they played lateron a leading role in the development of the US rocket programme.

1.1.2 The Cold War

After the end of World War II, most of the highly-ranked engineers of the V2 project were working for the US army. The Soviets had captured only a few technicians that had been involved at a lower level in the Peenemünde activities. Still, it was the Soviet Union that made the most rapid progress in space exploration in the late 1950s. This was mainly the merit of the space engineer **Sergey Korolyov** (1907 - 1966), a former student of Andrei Tupolev. Although Korolyov had been temporarly arrested during Stalin's Great Purge in 1938, after the war, he became the mastermind behind most achievements of the early space programme of the USSR. To overcome the lack of reliability of rockets such as the V2, Korolyov developed the concept of automated gyroscope stabilisation systems for stable flights along a predefined trajectory. He designed the first intercontinental missile, the R-7 Semyorka which was subsequently modified to serve for space applications. On 4 October 1957, **Sputnik 1** became the very first artificial satellite launched into an orbit around the Earth. Only one month later (3 November 1957), Sputnik 2 was launched with the dog Laika on board. These satellites were each developed and built within less than one month. There was no time for testing nor quality checks!

After many fruitless attempts (mostly due to the lack of an efficient organization and a competition between different sections of the US defense department), it was eventually on 31 January 1958 that the USA launched their first satellite (Explorer 1) with a Jupiter-C rocket developed by von Braun and his team. To improve the coordination of the US space programme, NASA was established on 29 July 1958 and von Braun became the first director of NASA's Marshall Space Flight Centre in Huntsville, Alabama. Nevertheless, it was once more Korolyov and his team who managed to send the first man-made probes to the moon (Luna 1 flew by the Moon at

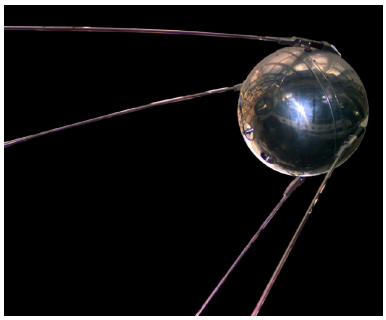
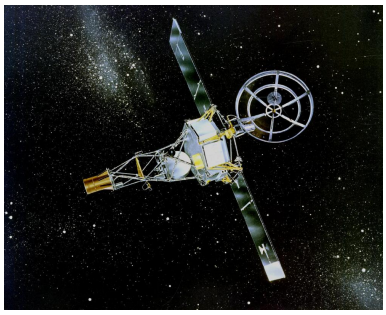
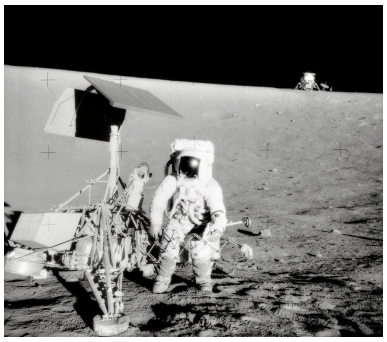




Figure 1.2: From left to right: Sputnik 1, the first man-made satellite of the Earth; Mariner 2, the first probe to fly by another planet; US astronaut Alan Bean near the landing site of the Surveyor 3 lunar probe.

a distance of 6400 km in January 1959, Luna 2 impacted the Moon in September 1959 and a few weeks later, Luna 3 provided the first images of the rear side of the Moon), and, most importantly, to send the first man into space and bring him safely[3] back to Earth (**Yuri Gagarin** on board Vostock 1 on 12 April 1961). The suborbital flight of the US astronaut Alan Shepard (5 May 1961) and the first orbital flight of US astronaut John Glenn (20 February 1962) followed several months later.

While some of the early satellites made some (serendipitous) scientific measurements (**Explorer 1** discovered the Van Allen radiation belts, Explorer 6 took the first images of the Earth from space), most of them were mainly built in a spirit of competition during the Cold War. It was in the same spirit that the race to the Moon started. This time, von Braun's Saturn V launcher turned out to be a better choice than the N-1 project on the Soviet side. Apollo 8 became the first manned mission that went into orbit around the Moon (December 1968) and on 20 July 1969, the crew of **Apollo 11** was the first to land on the Moon. While the USSR had lost the race to the Moon, Soviet scientists sent the **Lunokhod 1 and 2** rovers to the Moon (in 1970 and 1973) which were the first rovers on another body of the Solar System to be remotly operated from the ground.

Unlike von Braun in the USA, Korolyov never enjoyed the glory of being a national hero. In fact, to 'protect' him from foreign agents, the Soviet government actually never revealed his true identity until after his death. When the Nobel-prize committee wrote to the Soviet leader Nikita Khrushchev, asking him about the name of the designer of Sputnik 1, Khrushchev replied *'It's the people of the Soviet Union'*.

After the Moon landing, the focus of manned spaceflight moved mainly towards space stations in low-Earth orbits. In the US, the development of the **Space Shuttle** was approved in February 1972. The Shuttle system featured two re-usable solid fuel boosters plus an external fuel reservoir. The first successful launch (Space Shuttle *Columbia*) took place on 12 April 1981 (astronauts John Young and Robert Crippen). The Space Shuttle program did not fulfill its expectations though. It turned out to be extremely expensive and most of all, was marked by the accidents that led to the loss of the *Challenger* (explosion at launch on 28 January 1986) and *Columbia* (disintegration during atmospheric re-entry on 1st February 2003) shuttles and the deaths of fourteen astronauts. Beside the USSR (nowadays Russia) and the USA, China has also developed its own means to send humans into space. The first taikonaut Yan Li Wei reached the orbit on 15 October 2003 on board of a Shenzhou spacecraft.

Over the last decade, private companies have progressively taken over some of the space activities that were previously under the responsibilities of national agencies. The best-known examples are certainly **Space-X** and **Blue Origin**. Both companies have developed a procedure to land and re-use the first stage of their launcher which could possibly lead to a reduction in the costs for access to space. In May 2020, Space-X launched the **Crew Dragon** spacecraft with two astronauts to the ISS. This was the first manned space mission, fully operated by a

[3] Actually, because of a risk of malfunction of the retro-rockets, Gagarin ejected himself from his Vostock capsule at an altitude of 7 km and landed with a parachute.

private company. On the longer term, the presence of private companies as forefront actors in the space business will certainly lead to a re-definition of the objectives of space activities. We will discuss some aspects related to this development in Chapter 9.

1.1.3 The exploration of the Solar System

After visiting the Moon with robotic probes and manned missions, the next logical destinations in the exploration of the Solar System became the planets Mars and Venus. Following a number of unsuccessful attempts for a Venus flyby, in December 1962 the US spacecraft Mariner 2 became the first man-made probe that succeeded to fly by another planet. The first probe that impacted on the surface of Venus was Venera 3 (USSR) in March 1966 and the first orbiter around the planet was Venera 9 in October 1975. More recently, the European orbiter **Venus Express** studied the properties of the planet between April 2006 and January 2015.

The first attempts to reach Mars date back to 1960, but the first successful flyby of the red planet was done in July 1965 by the US probe Mariner 4. In July and September 1976, the **Viking 1 and 2** missions (both consisting of an orbiter and a lander) reached Mars. The Viking missions lasted until 1980 (Viking 2) and 1982 (Viking 1) and allowed to make in-situ measurements of the Martian soil and atmosphere to search for traces of biological activity. Over the last decade a number of missions have reached the red planet and are still exploring it; the best known ones are probably the European orbiter **Mars Express** (in orbit around Mars since December 2003), the US rovers **Spirit** and **Opportunity** (on the planet's surface since January 2004[4]), the US orbiter **Mars Reconnaissance Orbiter** (in orbit around Mars since March 2006), the US lander **Phoenix** (May – November 2008), and most recently, the US rover **Curiosity** (since August 2012).

The first missions to the outer Solar System were the American probes **Pioneer 10 and 11** that flew by Jupiter (December 1973) as well as Jupiter (December 1974) and Saturn (September 1979), respectively. In 1977, **Voyager 1 and 2** were launched to take advantage of a favorable planetary alignment. Each of the two spacecraft were equipped with colour TV cameras as well as a suit of instruments to measure and record magnetic and atmospheric data for each of the planets and their moons. These missions performed a Grand Tour of the outer Solar System and sent back spectacular close-up colour images leading to many important discoveries concerning these planets (auroral activities on Saturn, the magnetic field of Uranus and Neptune tilted with respect to their rotation axes and offset with respect to the centres of the planets), their moons (volcanic activity on Io, the possibility of a liquid ocean beneath the frozen crust of Europa, the atmosphere of Titan, cryo-volcanism on Triton) and rings. Both spacecraft are now heading towards the outer edges of the Solar System. Subsequently, Jupiter and Saturn have been visited by the **Galileo** and **Cassini-Huygens** missions respectively. These missions performed a detailed exploration of these giant planets and their moons by means of orbiters and landers: Galileo orbited around Jupiter between December 1995 and September 2003, Cassini orbited Saturn between July 2004 and September 2017, Huygens landed on Titan on 14 January 2005. In July 2016, the NASA mission **Juno** got into orbit around Jupiter with the goal to study the planet's structure and the properties of its atmosphere.

Minor bodies of the Solar System have also been visited. In July 2015, the US probe **New Horizons** was the first spacecraft to fly-by the dwarf planet Pluto. A major effort was dedicated to the study of comets. Here the most spectacular results were obtained by the European probe **Giotto** that obtained the first-ever images of the nucleus of Halley's comet during its perihelion in March 1986, as well as by ESA's **Rosetta** probe that remained in orbit about comet 67P Churyumov-Gerasimenko between September 2014 and September 2016.

[4]NASA lost contact with Spirit (trapped in deep sand) in the spring of 2010 and with Opportunity in June 2018.

1.2 Why go to space? (Part 1)

There are a number of obvious motivations why to do science from space (usually defined as beginning at an altitude of more than 100 km). Here we give a non-exhaustive list of such reasons.

1.2.1 Get out of the Earth's atmosphere

Astronomy and astrophysics rely mostly on the measurement (imaging, photometry, spectroscopy, polarimetry,...) of the electromagnetic radiation from remote sources. However, whilst the spectrum of electromagnetic radiation ranges from the shortest wavelengths (highest energies) found in γ-rays up to the longest wavelengths (lowest energies) typical of radio waves, the Earth's atmosphere is largely transparent over only tiny windows of the electromagnetic spectrum. These windows correspond to the visible light (roughly 4000 – 7000 Å) and to the radio domain (from a few centimeters to about 20 meters), as well as to several narrow, partially transparent, infrared windows. The absorption by the atmosphere over the rest of the electromagnetic spectrum is so heavy that essentially no radiation from cosmic sources reaches the ground. To study the cosmos over energy domains that are not accessible from the ground, it is therefore necessary to go to space. This is especially true since there are a number of phenomena that manifest themselves only in specific energy domains.

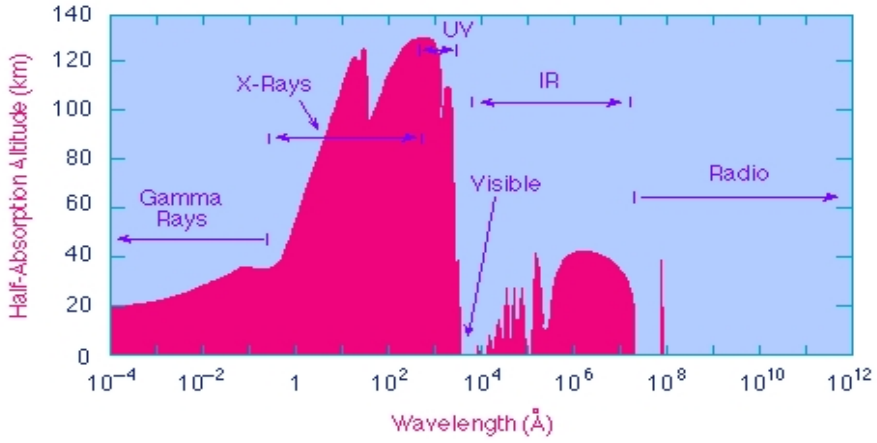


Figure 1.3: Atmospheric transmissivity windows as a function of wavelength.

But even in the visible domain, there are good reasons to do observations from space. In fact, the Earth's atmosphere not only absorbs part of the radiation from cosmic sources (this is the case at any wavelength even over atmospheric windows), but it also blurs the images due to the atmospheric turbulence. For a ground-based telescope, the image of a point-like source will be a disk with a FWHM equal to the seeing which results from the combination of the capabilities of the telescope itself and the atmospheric blurring. To reach a sub-arcsec resolution with a single telescope on the ground, one needs to use active and adaptive optics even with the largest existing telescopes. The situation is much better in space. For instance, the Hubble Space Telescope with its 2.4 m diameter primary mirror achieves an angular resolution of 0.05 arcseconds.
Another problem is the variability of the atmospheric transparency. To obtain high-precision photometric data (e.g. photometric light curves to search for the transit of an exoplanet in front of its parent star), one needs to have incredibly stable conditions. From the ground, this can only be done during exceptionally clear and stable (so-called photometric) nights. Even then, the precision that can be achieved is not comparable to what can be done from space with satellites such as CoRot or Kepler. In its most extreme version, the variability of the atmosphere's

transparency can even prevent observations (cloudy sky) although one should be aware that for certain kinds of observations, there is also an issue with what is called the *space weather*.

Finally, depending on the orbit selected for the mission, space-borne observatories are able to perform un-interrupted observations of time-dependent phenomena (e.g. observations of the Sun from L_1). This is extremely important since it eliminates the aliasing problem that arises from the daily cycle for ground-based observatories.

1.2.2 In-situ measurements

For research related to our Solar System, it is (at least theoretically) possible to travel to the various places where one wishes to perform the relevant measurements. The most obvious examples of these missions are orbiters around planets (such as Mars Express) that dress a detailed map of the planet in different wavelengths, measure its electric and magnetic environment and perform observations of its climate. The advantages of an orbiter are obvious. To illustrate this point, let us consider the diffraction limit of a telescope, i.e. the smallest angular detail that can be resolved. For a perfect telescope of diameter D, this is λ/D. The actual size of the objects that can be resolved is obtained by multiplying the diffraction limit with the distance between the telescope and its target. Orbiters can therefore resolve much smaller details at the surface of a planet. For instance, a telescope observing Mars with a mirror of diameter 10 cm from an orbit around Earth can resolve details of about 5.5 km under the best conditions, whilst the same instrument in orbit around Mars can resolve details of 3 m. Orbiters further allow to monitor the climate of the planet.

Much more sophisticated concepts are of course possible. One can imagine landers that explore a specific site on the planet's surface and perform an in-situ study of its atmosphere (chemical composition, atmospheric pressure, temperature,...), its geological properties,... Even more ambitious is the concept of rovers that move at the surface of the planet, thereby allowing to visit a number of different places with different properties. In some cases, it is also interesting to consider sample return to perform a detailed analysis in a laboratory on the ground, using more sophisticated (hence more massive and more energy-consuming) instrumentation, although there is a major issue with contamination here and the pros and cons of a sample return mission need to be evaluated very carefully. In-situ measurements can also be done to study the environment in which the Earth and the other planets are evolving. This environment is dominated by the Sun and its solar wind and the interactions between the latter and the magnetosphere and atmosphere of a planet are best studied in-situ.

1.2.3 Earth observations

Whilst astronomical satellites look up to the sky, a major application of satellites is the observation of the Earth itself. There are many reasons to do this (apart from military reasons and commercial tele-communication or navigation satellites that we will not consider here). Satellites are unique in their ability to constantly monitor the entire Earth for weather forecast and to monitor the health of our planet by studying natural phenomena (floods, hurricanes,...) as well as the impact of human activities on the climate and the environment in general (e.g. forest fires, measurements of pollutants, changes in the ice cover).

1.2.4 Microgravity

Some specific research topics require the absence of significant acceleration and/or the absence of vibrations that are generated by seismic and human activities on the ground. This is the case for instance in experiments (such as eLISA) aiming at the detection of gravitational waves of certain frequencies.

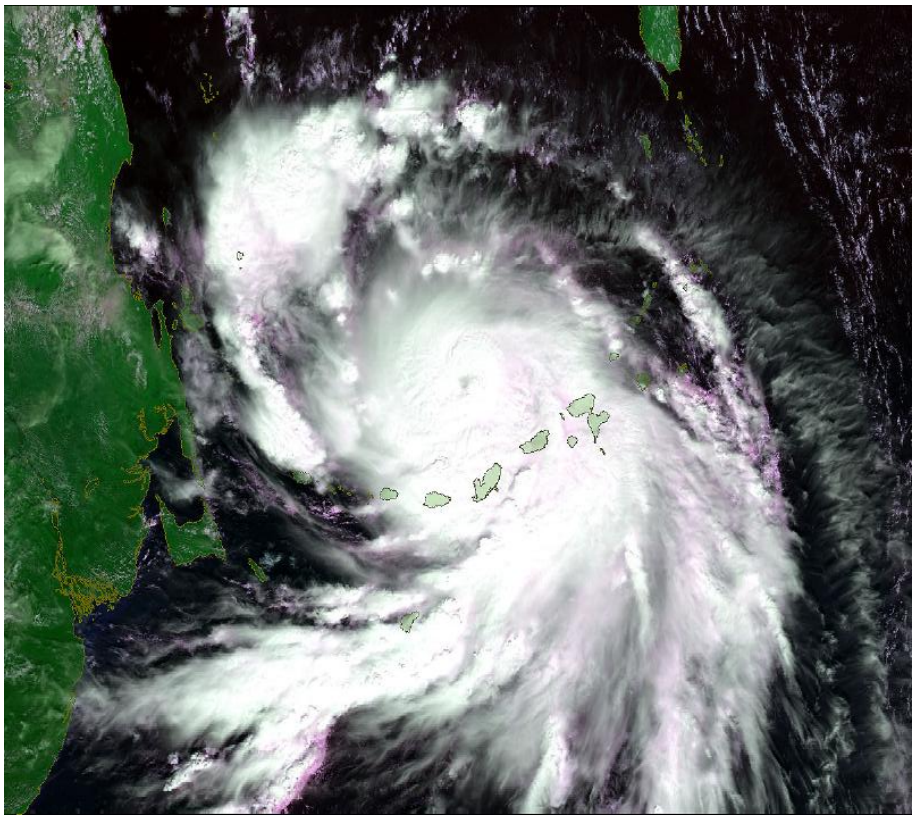


Figure 1.4: ENVISAT image of Hurricane Dean leaving Martinique (August 2007).

1.3 Why go to space? (Part 2)

Space-borne scientific research has led to immense progress in our understanding of the Solar System and the wider Universe (see the examples above and in the forthcoming chapters of these lectures). Ambitious scientific projects often trigger the development of new technologies some of which eventually find an application in our everyday's life. However, one should not take these side effects as a justification for space research. *In fact, fundamental research does not need an economic justification. It rather reflects the attempt of any civilized society to reduce its degree of ignorance. Giving up such an activity for economic reasons would eventually be equivalent to declaring the cultural and scientific bankruptcy of our society.*

Yet, the bulk of the space activity is not motivated by scientific research. Indeed, there are a number of non-scientific reasons for space activities. These activities generally belong to either of the following categories: commercial or political applications (the latter being either of civil or military nature). The most obvious examples of such activities are found in the domains of telecommunications, microgravity, Earth observation and manned spaceflight.

However, after six decades of spaceflight, some of these applications have actually failed to fulfil their promise whilst others have an enormous success. Microgravity, for instance, falls into the first category. In principle, microgravity offers the possibility to manufacture alloys that cannot be obtained on the ground. Still, in order for a material manufactured in space to be commercially competitive, and thus to overcome the enormous transportation costs, its value ought to exceed a hundred times that of gold and there ought to exist a sufficient demand

on the market. These two conditions have not been met so far and parabolic flights offer a much cheaper solution than manufacturing aboard the ISS for instance. As far as telecommunications are concerned, transatlantic telecommunications via satellite (Intelsat) are nowadays outdated by the new generations of intercontinental fiber optics cables. However, there is of course still a broad market for TV broadcasting via satellite and mobile-phone telecommunications via satellite.
A very successful spin-off of an initially military application of space is the navigation system GPS (Global Positioning System) which has currently a tremendous commercial success.
Earth observation from space is the first satellite-based activity worldwide and provides another discipline which has a huge impact on our society. For instance, the images of the hole in the ozone layer and other global views of the impact of human activities on our planet's ecosystem that can be obtained from space contribute to the awareness of the general public and the politicians for ecological questions. Of course, Earth observations are also of strategic importance, even though the importance of this domain is quite different among the various actors in space exploration.
Manned spaceflight, finally, was used in the early days of spaceflight, and still is to some extent, as a political tool to demonstrate the capabilities of the various space nations. Beyond these propaganda effects and the fact that manned spaceflight contributes of course to make the general public (hence the tax payers) dream about space exploration, the scientific interest of a permanent presence of men in space, such as on board the ISS, is however rather limited. Indeed, the ISS itself is too unstable for astronomy observations, manufacturing of high-tech products is not commercially interesting (see above) and the most obvious scientific interest concerns the studies of the impact of a long space mission on the human body, which are relevant in the preparation of manned missions to other planets.

1.4 The ESA science programme

In 1962 ESRO, the European Space Research Organization, was founded by several European Countries (Belgium, Switzerland, Germany, Denmark, Spain, France, Italy, the Netherlands, the United Kingdom and Sweden). The first satellite launched successfully by ESRO was ESRO2B a satellite studying cosmic rays and the solar UV radiation (1968). In May 1975, the European Space Agency (ESA) was created by the ESRO member states. Lateron, Austria, Norway, Finland, Ireland, Portugal, Greece, Luxembourg, Czech Republic, Romania, Poland, Estonia and Hungary have joined ESA. The first successful flight of an Ariane 1 launcher took place on 24 December 1979. Since then, ESA has undertaken a very successful science programme, although its financial resources are much more modest than those of its American cousin NASA.
ESA takes care of scientific space missions that are too ambitious for a national European space agency. The ESA science programme is the only mandatory element of the ESA programme: each member state contributes with a share proportional to its gross national product. ESA has built and operated many successful space science missions, some of them in collaboration with other space agencies. Past missions include Giotto (encounter with Halley's Comet), ISO (Infrared Space Observatory), Hipparcos (astrometric mission), IUE (International Ultraviolet Explorer), Herschel (far-IR observatory), Planck (measurements of the Cosmic microwave background), Venus Express (orbiter around Venus), Rosetta (mission to comet 67P/Churyumov-Gerasimenko), and Cassini-Huygens (mission to Saturn). Currently, the ESA science directorate is operating a number of spacecraft: SOHO (Solar Heliospheric Observatory), Proba-2 (space weather), Mars Express (a Martian orbiter), Double Star and Cluster (study of the Earth's magnetosphere), INTEGRAL (γ-ray astronomy), XMM-Newton (X-ray astronomy), Hubble Space Telescope (with NASAS, observations from UV to near-IR wavelengths), Gaia (high-precision astrometry), LISA Pathfinder (a technology mission to test components for an experiment to detect gravitational waves), ExoMars 2016 (a Martian orbiter), BepiColombo (on its way to planet Mercury), CHEOPS (detection of extrasolar planets), Solar Orbiter (observations of the Sun from a distance of 0.25 AU). New missions are being implemented. These

are the James Webb Space Telescope (in collaboration with NASA, 2021), Euclid (map the distribution of dark energy, second M-class mission, 2022), PLATO (detect extrasolar planets, third M-class mission, 2026), JUICE (Jupiter ICy moons Explorer, first L-class mission of the Cosmic Vision plan, 2022), Athena (next generation X-ray observatory, second L-class mission of Cosmic Vision, 2030) and ARIEL (a mission to study atmospheres of exoplanets in the near-IR, fourth M-class mission, 2028).

Nowadays, ESA employs just over 2000 people[5]. The ESA headquarters are located in Paris. The European Space Research and Technology Centre (ESTEC) in Noordwijk (NL) is the place where most of the technological developments are coordinated and where satellites are tested prior to launch. The European Astronaut Centre (EAC) in Cologne (Germany) is the training centre of the ESA astronauts who visit the International Space Station. ESOC (European Space Operations Centre) in Darmstadt (Germany) is the control centre for the real-time operation of space missions: the health of the spacecraft is monitored and all commands for orbital manoeuvres and pointings are issued by the ESOC spacecraft controlers. The European Space Research Institute (ESRIN) in Frascati (Italy) is responsible for the coordination and scientific operation of the Earth observation satellites, whilst the European Space Astronomy Centre (ESAC, Villafranca del Castillo, Spain) assumes the same responsibilities for the astronomy and Solar System missions.

In the ESA ministerial council, each member state has a single vote. This gives relatively more weight to the smaller member states such as Belgium. However, ESA also applies the principle of *fair return* which means that each country recovers industrial contracts with a value proportional to its contribution to the ESA budget. The decisional process within ESA is admittedly rather slow and rigid. However, this inertia contributes to some extent to make the ESA space program more stable than the NASA one. Indeed, in the US, the NASA budget is voted every year by the Congress who can reverse its decisions at any time.

The global yearly budget for space activities worldwide is around 300 billion €. About three quarters of this sum are related to commercial activities. The largest chunks correspond to telecommunications, navigation and Earth observation. Among all civil space agencies, NASA has by far the largest yearly budget (about 19.5 billion USD vs. about 5.8 billion € for ESA in 2016). About half of NASA's budget is spent for the human spaceflight programme. ESA is second in absolute terms. However, the (dis)proportion between NASA and ESA is even larger for the space science budget: 5.6 billion USD for NASA versus roughly 500 million € for ESA. Yet, even NASA's budget is much lower than what it used to be in the past. Funding resources were plentiful in the Moon race: at that time the NASA budget amounted to about 4.4% of the US federal budget. Within half a year after the Moon landing, NASA workforce dropped from 400 000 employees and contractors to less than 200 000. Currently, the NASA budget corresponds to 0.5% of the US federal budget.

In Europe, France is clearly the leader in space activities with 0.18% of its gross national product (GNP) being spent on space applications. Belgium actually comes second with 0.085% of its GNP, well before Germany (0.048%), Italy (0.048%) and the UK (0.024%). Belgium contributes on average 180 million€ per year into ESA activities, which makes it the sixth contributor to the ESA budget in absolute terms.

Currently, momentum is shifting away from national space agencies towards private companies and ambitious businessmen. For scientific missions, given the increasing costs of space missions, international cooperation is becoming more and more important. However, some space-qualified technologies fall under US ITAR (International Traffic in Arms Regulation) regulations and their import or export or even share with the partners of the project hence become political issues that prevent notably simultaneous cooperation for instance between ESA and NASA and ESA and China. At the same time, the reduced funding for science missions threatens the implementation of new scientific space missions. Even flaggship missions such as the *James Webb Space Telescope*, under construction by NASA with an important ESA contribution, are suffering recurrent launch delays and others might

[5] In Europe, space-related industries represent about 40 000 jobs.

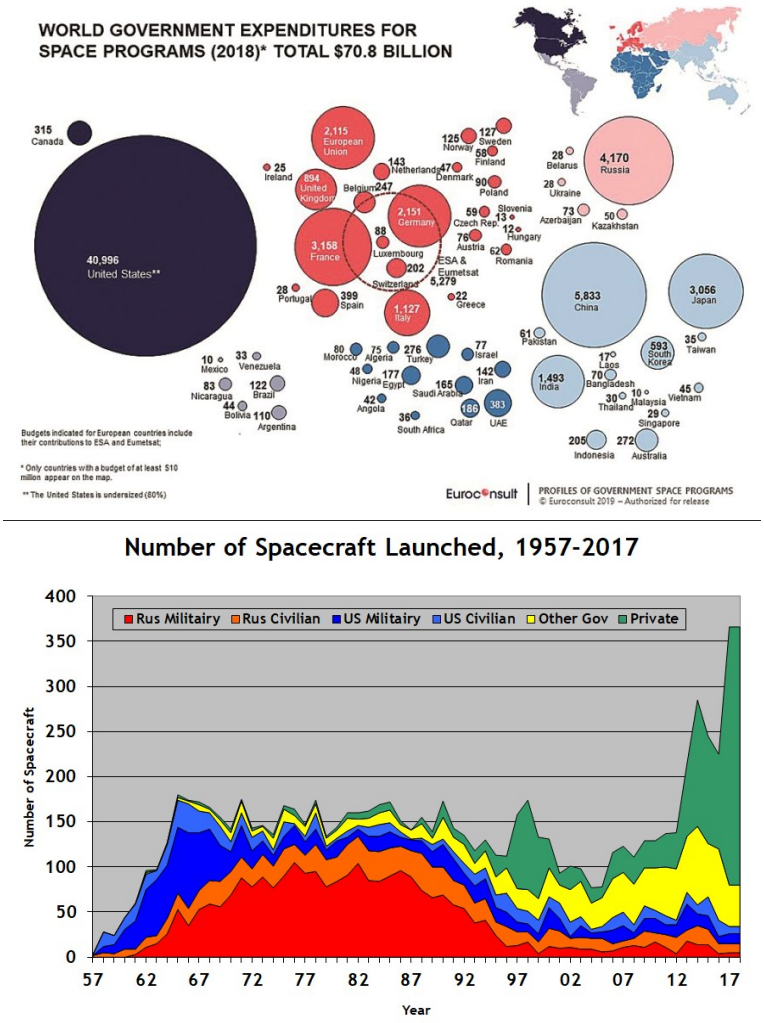


Figure 1.5: Top panel: government expenditures for space activities in 2016. ©Euroconsult. Bottom panel: number of spacecraft launched by country.©claudelafleur.qc.ca.

eventually be cancelled as a result of exploding costs and budget cuts.

1.5 The life cycle of a space mission

In Europe, ESA acts as the main driver for space science and space exploration. The ESA science programme contains missions that address scientific questions in three distinct fields: astrophysics, Solar System science and fundamental physics. The various satellites, probes and missions are selected based on the input from the European scientific community.

The very first step in the life cycle of a science mission is a call for ideas. ESA issues such calls roughly every ten or fifteen years; the most recent one is *Voyage 2050* which opened in March 2019. Before that the *Cosmic Vision* call was issued in 2004. The aim is to identify paramount scientific questions and themes that will require specific technology developments and to allow for long-term planning. Some time after the call for ideas, a call

for proposals is issued where the scientific community actually submits proposals for new missions. For the first flight opportunities of the Cosmic Vision plan (originally foreseen for 2017-2018), this happened in the spring 2007. The proposals are evaluated by ESA's advisory structures: the Astronomy Working Group (AWG), the Solar System Working Group (SSWG) and the Fundamental Physics Advisory Group (FPAG). In 2010, as a result of the merger between the science and robotic exploration directorates, the advisory structure was slightly reorganized: the SSWG became the Solar System Exploration Working Group (SSEWG). Lateron, the FPAG was absorbed into the AWG.

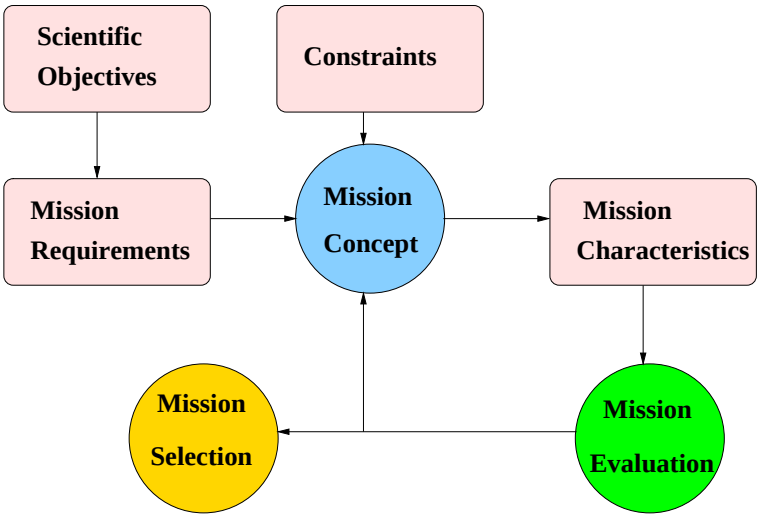


Figure 1.6: Schematic summary of the selection process of new space missions.

A mission is selected based on its science objectives, its payload (can this payload meet the science objectives?), the complexity of the overall mission concept (launch vehicle, the spacecraft bus, the orbit, a preliminary operation profile, the complexity of the model payload,...), the technical feasibility and re-use of existing technology. For many Solar System exploration missions, the timing is affected by the planetary configuration, leading to specific, often rather narrow, launch windows. Another aspect that comes into play is the competition with other missions for ground segment antenna time. Missions that travel far away from the Earth require the use of large antennas for telecommunications such as the Deep Space Network (DSN). If possible, it is advantageous to fly towards a direction where the spacecraft faces little competition with other missions for DSN time. During the so-called Pre-Phase A conceptual study, these different aspects are critically reviewed and at the end of this phase, a decision is taken whether or not to proceed to the preliminary analysis of the mission, which is called Phase A. During this phase, the project team establishes a preliminary design of the spacecraft, a roadmap for the different steps of the development, building, testing and operation of the mission as well as a preliminary assessment of the mission budget. Indeed, understanding the purpose, scope, schedule and cost of a mission is mandatory before a realistic spacecraft design becomes possible.
Once a mission has been selected, it goes into the definition phase (Phase B) where requirements and schedules are defined. Two parallel industrial studies are initiated through an Invitation to Tender. This definition phase lasts about 2 to 3 years and includes also an announcement of opportunity for proposals for instrumentation for the spacecraft payload. The definition phase is followed by an implementation phase (Phase C/D) that includes the selection of a prime contractor, the final design of the mission, the development of critical technologies, the integration and verification of the various components of the mission and the spacecraft as a whole, as well as the launch and commissioning of the mission. It is also during this phase that the ground segment of the mission is designed, developed and tested. This phase is coordinated by a project manager who has to deal most of the

time with an overconstrained problem: a budget or schedule that may be inadequate, instruments that require more resources than allocated, subcontractors who would prefer deliver their hardware as late as possible while simultaneously extorting as much money as possible out of the project... Sometimes the only way out of this dilemma is to actually descope the mission (i.e. suppress one instrument or reduce the requirements). Unfortunately, descoping is always done at the expense of the scientific return of the mission. This phase lasts about 4 to 5 years.
The in-orbit commissioning and performance verification phases are crucial steps for each scientific space mission. Despite the careful design, integration, tests and calibrations on the ground, it is quite normal to discover and mitigate problems during these phases. Space is indeed a hostile environment, difficult to fully simulate on the ground. Moreover, the launch itself is a critical event that can also affect the performances of a spacecraft. Some examples of problems discovered and mitigated during a commissioning campaign are described in section 7.1.2, when we discuss the Gaia mission.
The exploitation phase (Phase E) starts after the launch when the actual scientific observations/operations start. The life of a mission is most often limited by the quantity of fuel on board, the aging of the electronics, the power production (accounting for the degradation of the solar panels, batteries, generators,...), the quantity of cooling liquid (for cryogenic devices),... Most space missions are built to last much longer than their nominal lifetime and many of the currently active spacecraft are well beyond their initially foreseen lifetime. Actually, the annual costs of operations are much less than the costs for designing, building and launching the spacecraft (see Table 1.1).

Table 1.1: Cost at launch and cost for annual operations of some past missions.

Mission	Cost at launch	Annual cost for exploitation
Hyabusa	240 million USD	5 million USD
Ulysses	400 million USD	2.5 million USD
Spirit & Opportunity	820 million USD	20 million USD
Galileo	1.9 billion USD	15 million USD
Cassini	2.6 billion USD	80 million USD

However, operations of a space mission are not automatically extended by the relevant space agency. The mission first needs to undergo a mission extension review where the health of the spacecraft is checked and the prospectives of future operations are investigated. The final decision about the extension is often taken based on purely budgetary considerations.

1.6 Various categories of space missions

One can distinguish various types of space missions in astrophysics:

- pioneer missions that open up a new wavelength range and perform a low sensitivity all-sky survey (e.g. the first X-ray satellite Uhuru),
- second generation missions that perform pointed observations and imaging (e.g. the first space-borne X-ray telescope EINSTEIN),
- observatory class missions that do both spectroscopic and imaging observations (e.g. Chandra and XMM-Newton, today's X-ray observatories).

A similar classification can be done for planetary missions:

- first exploration mission (flybys as done by the Voyager probes),

- comprehensive studies with an orbiter (Galileo, Cassini, Mars Express),
- in situ studies with landers (Huygens, Spirit & Opportunity, Curiosity,...),
- sample return missions (Apollo, Hyabusa,...).

A scientific space mission can be either of observatory or Principal Investigator (PI)-type. A scientist can be involved in a mission at different levels: either as PI or CoI (these people have access to guaranteed observing time), as mission scientists (this is a committee where the scientific performances of the mission are monitored), or as guest observers (for an observatory-type mission).

Chapter 2

Our Solar System and its ingredients

Before we discuss the constraints on space travel, it is important to recall some of the properties of the Solar System. In fact, it is the Sun and its planets that set the (gravitational, radiative,...) environment in which the spacecraft evolves.

2.1 Our Solar System

Our Solar System consists of the Sun (a G2 V star) and all the bodies that are gravitationally bound to it. These are the eight planets (Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus and Neptune) and their moons, the dwarf planets (such as Pluto, Ceres and Eris) with their moons as well as a multitude of small bodies (asteroids, Kuiper belt objects, comets, interplanetary dust...).
Most of the mass (99.86%) of the Solar System is concentrated in the Sun itself. The planets and all minor bodies of the Solar System move around the Sun on orbits that are ruled by Newton's law of gravity. In the particular case of planetary motion, this translates into three laws first discovered by Johannes Kepler (1571 - 1630):

- *The orbit of a planet is an ellipse with the Sun at one of the foci.* An ellipse is mainly characterized by its semi-major axis a and its eccentricity e. In heliocentric polar coordinates, the position of a planet is thus given by its (r, ϕ) coordinates that are related to each other:

$$r = \frac{a\,(1 - e^2)}{1 + e\,\cos\phi} \tag{2.1}$$

  The point of closest approach to the Sun is the perihelion, the point of maximum separation is called aphelion.

- The orbital velocity of a planet is not constant. *The line joining the planet and the Sun sweeps out equal areas over equal intervals of time.* Therefore, the planet moves faster while it is close to perihelion than around aphelion.

- *The cube of the semi-major axis is proportional to the square of the orbital period.*

$$a = \mu^{1/3}\,P^{2/3}\,(2\pi)^{-2/3} \tag{2.2}$$

  where a is the semi-major axis, P the orbital period and $\mu = G\,(M_\odot + m)$ with G the gravitational constant, $M_\odot$ the mass of the Sun and m the mass of the planet.

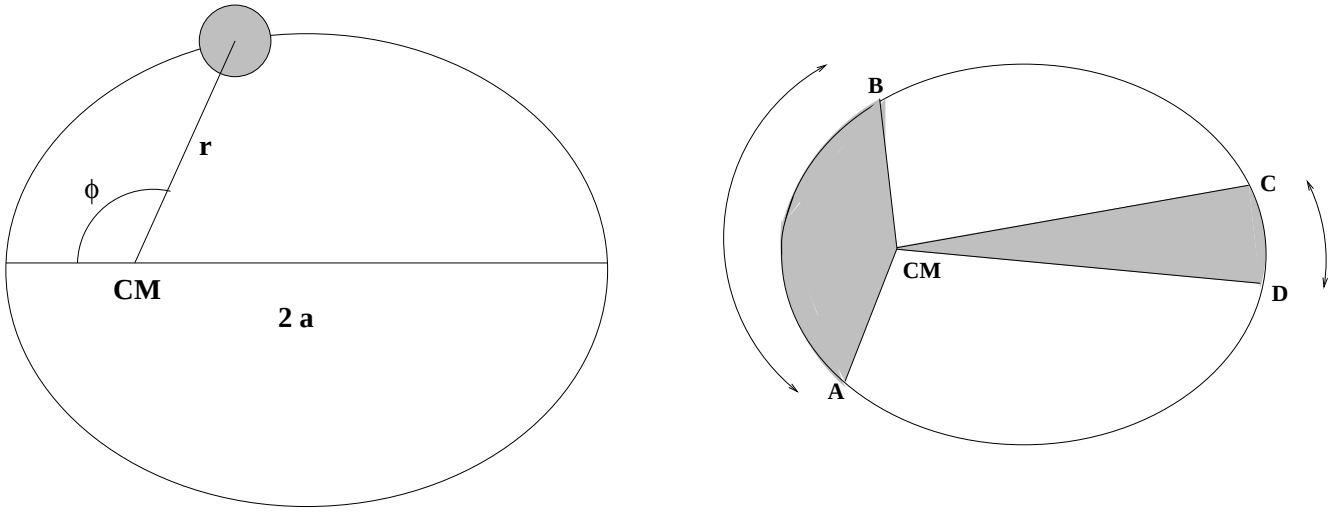


Figure 2.1: Left: schematic view of the orbit of a planet around the centre of mass CM of the Solar System. Right: schematic view of Kepler's second law. The areas A CM B and C CM D are equal and correspond thus to identical time intervals.

By definition, the Astronomical Unit (1 AU = 149 598 000 km) corresponds to the mean distance between the Sun and the Earth. Given the low eccentricity of the Earth's orbit, this is nearly the same as the semi-major axis of the Earth's orbit. Planets and dwarf planets have orbits that are roughly in the same plane as the orbit of the Earth (known as the ecliptic) and orbit the Sun in a counter-clockwise direction as viewed from above the Sun's north pole.

The Sun is not only the source of light in the Solar System, but it also generates a continuous flow of highly ionized particles which is called the solar wind. This wind creates a hot, tenuous atmosphere called the heliosphere that extends out to at least 100 AU. The eleven year solar cycle as well as the changing activity of the Sun (flares, coronal mass ejection,...) disturb the heliosphere, creating what is called space weather.
The interaction between the particles of the solar wind and the Earth's magnetic field creates the aurorae seen near the polar regions of the magnetic fields. The Earth's magnetic field protects its atmosphere from direct interaction with the solar wind. On the contrary, the atmospheres of planets that have no strong magnetic field (such as Venus and Mars) are progressively eroded by the action of the solar wind.

The four inner planets are also called terrestrial planets. These objects are composed of minerals such as silicates that form a solid crust surrounding a semi-liquid mantle and a core consisting of metals such as iron or nickel. These planets display tectonic activities (volcanoes, rift valleys,...) or have done so in the past. Olympus Mons at the surface of Mars is the highest volcano known in our Solar System.
Venus, Earth and Mars have a substantial atmosphere, whilst Mercury's tenous atmosphere consists of atoms eroded from its surface by the solar wind. Venus has an atmosphere much denser than that of the Earth and the surface temperature reaches over 400° C as a result of a strong greenhouse effect. Mars' atmosphere mainly consists of CO_2 and is much more tenuous than the Earth's atmosphere. The Earth's atmosphere is unique in the sense that it's composition is largely affected by biological activities (leading to a rather large concentration of O_2).
Mars has two small moons (Deimos and Phobos)[1]. The Earth's Moon is much larger and is believed to have formed

[1]They were believed to be captured asteroids, but the detection of phyllosilicates that form in the presence of water and a mean density of 1.86 g cm^{-3}, too low to have survived to capture by a planet, both suggest that they rather formed through collision with a large planetoid.

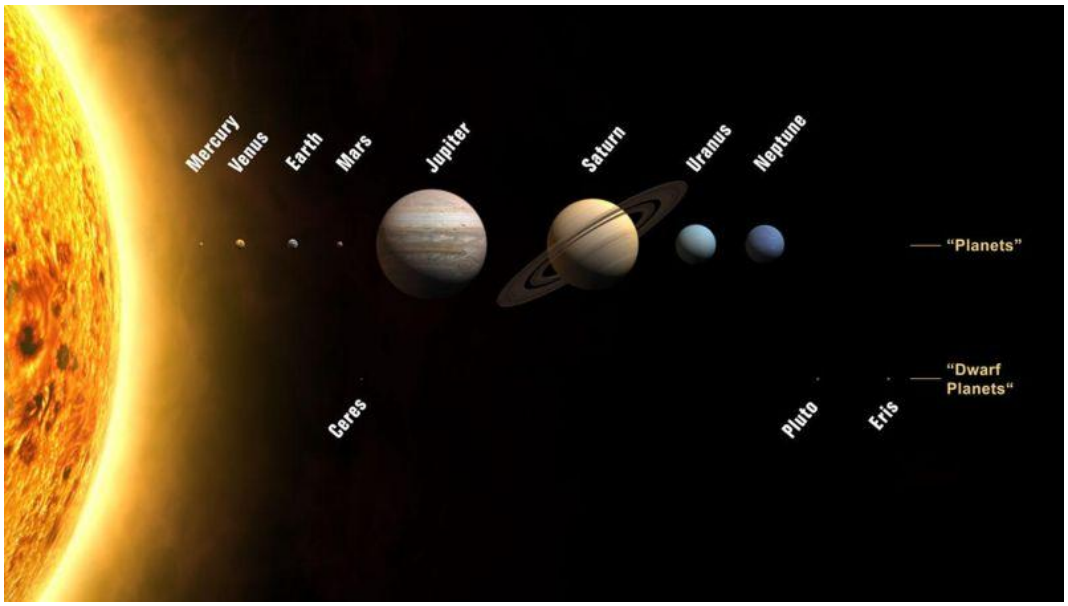


Figure 2.2: Planets and dwarf planets of the Solar System. The sizes are to scale, but orbital separations are not.

from a part of the Earth after collision with an object of the size of Mars.
The asteroid belt occupies the orbit between Mars and Jupiter. Most asteroids range in size from a few millimeters to hundreds of kilometers and are believed to be remnants of the formation of the Solar System that failed to form a planet because of the perturbations by Jupiter's gravitational field. Although there are a huge number of asteroids in the belt, the actual density of the belt is rather low and spacecraft routinely pass through it without damage.
The outer planets differ substantially from the inner ones. They have a large number of moons and are all encircled by rings of dust and particles. The outer planets are mainly gas and ice giants: whilst Jupiter and Saturn consist mostly of hydrogen and helium, Uranus and Neptune contain also water ice as well as ammonia and methane. Jupiter has a large number of moons (79). The four largest Jovian satellites (Ganymede, Callisto, Io and Europa) are similar to the terrestrial planets. Ganymede is even larger than planet Mercury. Saturn also has a very large number of satellites (62). The largest one, Titan, is again bigger than Mercury and is the only moon in the Solar System that has a dense atmosphere. Uranus is unique in the sense that its rotational axis is heavily tilted (almost 90° with respect to the ecliptic). Neptune, finally, is more massive than Uranus.
Pluto and Charon form a binary dwarf planet: they orbit around their common centre of mass which lies outside the surface of Pluto. Pluto and Charon are actually part of the Kuiper belt, a region of asteroids and debris extending between 30 and 50 AU from the Sun.

Comets are small bodies with diameters of a few kilometers that are mainly composed of different types of ices. They travel on highly eccentric orbits. When a comet approaches the Sun, the solar radiation causes its surface to sublimate and ionise, hence creating a coma and a long characteristic tail.

The frontier between the Solar System and the interstellar space is not well defined. Its location is set by the balance between the solar wind and the interstellar medium. It is currently believed that the heliopause, i.e. the boundary of the heliosphere is located at about 140 AU, depending on the direction. In 2003, Voyager 1 sent back data suggesting that it had crossed the termination shock of the heliopause at 94 AU. It must be stressed that the gravitational influence of the Sun extends to about 125 000 AU, well beyond the heliopause.

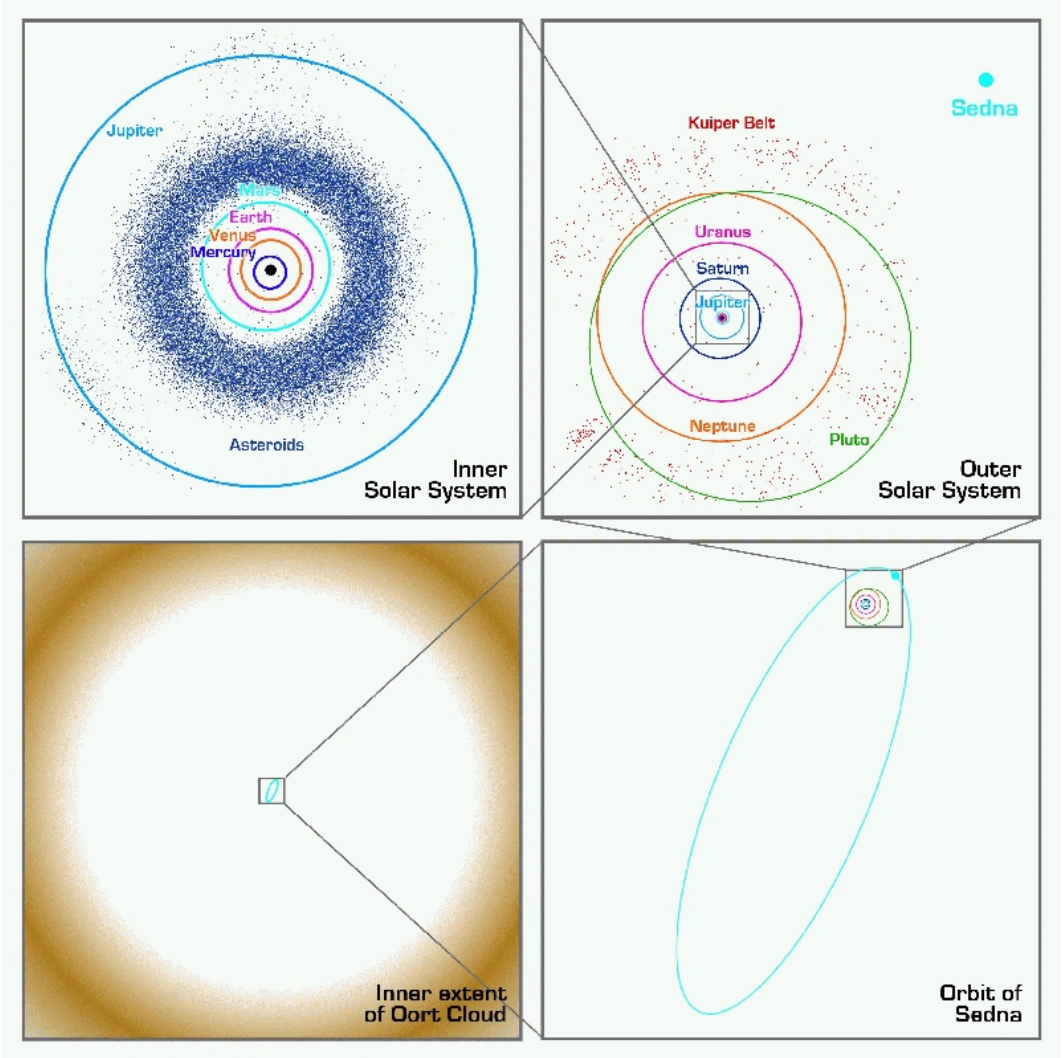


Figure 2.3: Schematic representation of the orbits of the planets and minor bodies of the Solar System. The sizes of the orbits are to scale.

There are relatively few stars within 10 light years of the Sun. The closest neighbour is the triple system α Cen at a distance of 4.4 light years.

2.2 The Roche potential and the Lagrangian points

Beside orbiting the bodies of the Solar System, spacecraft frequently orbit around the so-called Lagrangian points. Consider a system formed by two objects of mass m_1 and m_2 revolving on circular orbits around their centre of gravity. We assume that both objects are small compared to their orbital separation (a), so that they can be considered to be point-like. A test particle of mass $m_3 << \min(m_1, m_2)$ is submitted to the gravitational forces of m_1 and m_2 and since the latter objects are moving, the whole problem occurs in a non-inertial frame of reference. Therefore, in addition to the gravitational forces, one has to account for the fictitious centrifugal and Coriolis

Table 2.1: Main characteristics of the planets of the Solar System and their orbit. The table lists the semi-major axes of the orbits (a in AU), the sidereal period for the revolution of the planet around the Sun, the eccentricity of the orbit as well as its inclination with respect to the ecliptic. The radii and masses of the planets are given in units of the Earth's radius (6378 km) and mass (5.976×10^{27} g). The escape velocity in the last column is the velocity that one has to communicate to a spacecraft to leave the gravity field of the planet. Note the large eccentricity and the high orbital inclination of the dwarf planet Pluto given in the last row. Note that the radius and the mass of the Sun are $R_\odot = 6.9599 \times 10^{10}$ cm and $M_\odot = 1.989 \times 10^{33}$ g, respectively.

Planet	a (AU)	Sidereal period (years)	e	i ($^\circ$)	Radius ($R_\oplus$)	Mass ($M_\oplus$)	Escape velocity (km s^{-1})
Mercury	0.3871	0.2409	0.206	7.00	0.38	0.055	4.2
Venus	0.7233	0.6152	0.007	3.39	0.95	0.815	10.3
Earth	1.0000	1.0000	0.017	–	1.00	1.000	11.2
Mars	1.5237	1.8809	0.093	1.85	0.53	0.108	5.0
Jupiter	5.2028	11.862	0.048	1.30	11.18	317.8	61
Saturn	9.5388	29.458	0.056	2.49	9.42	95.15	37
Uranus	19.182	84.014	0.047	0.77	3.84	14.54	22
Neptune	30.058	164.79	0.009	1.77	3.93	17.23	25
Pluto	39.44	247.7	0.250	17.33			

forces. Hence, the potential "seen" by m_3 can be written:

$$\Phi = -\frac{G\,m_1}{|\underline{r} - \underline{r}_1|} - \frac{G\,m_2}{|\underline{r} - \underline{r}_2|} - \frac{1}{2}\,|\underline{\omega} \wedge \underline{r}|^2 \tag{2.3}$$

where $\underline{r}$ is the positional vector of m_3 with respect to the centre of mass of m_1 and m_2; $\underline{r}_1$ and $\underline{r}_2$ are the positions of the two masses m_1 and m_2 respectively with respect to their common centre of mass. $\omega = [\frac{G\,(m_1+m_2)}{a^3}]^{1/2}$ is the angular velocity.
This so-called Roche potential can be expressed in a dimensionless form as $\Omega = -\frac{\Phi\,a}{G\,m_1}$. This then yields the following expression of the potential:

$$\Omega = \frac{1}{r_1} + \frac{q}{r_2} + \frac{q+1}{2}\,(x^2 + y^2) - q\,x \tag{2.4}$$

$$q = \frac{m_2}{m_1}, \qquad r_1 = \sqrt{x^2 + y^2 + z^2}, \qquad r_2 = \sqrt{(x-1)^2 + y^2 + z^2}$$

The equipotentials of the Roche potential in the orbital plane are shown in Figure 2.4. Near the surface of m_1 (respectively m_2), the contours are essentially spherical and they deform into a drop-like shape as one moves away from m_1 (resp. m_2). The first contour that includes both m_1 and m_2 marks the limit of the *sphere of influence* of m_1. Within this region of space the attraction by m_1 dominates over all other forces. The point between m_1 and m_2 where the two spheres of influence touch is called the first Lagrangian point L_1. Here the attraction forces of m_1 and m_2 and the centrifugal force equilibrate each other and the test particle is hence not submitted to any net force. There are actually five Lagrangian points that are points of relative equilibrium given by $\underline{\nabla}\,\Omega = 0$. These points are all located in the orbital plane ($z = 0$). L_1, L_2 and L_3 are located along the axis of the system ($y = 0$) and their positions are solution of the non-linear equation

$$\frac{-x}{|x|^3} - \frac{q(x-1)}{|x-1|^3} + (q+1)\,x - q = 0 \tag{2.5}$$

The other two points, L_4 and L_5, are located at $x = 1/2$ and $y = \pm\sqrt{3}/2$ (i.e. their positions do not depend on the value of q unlike what happens for L_1, L_2 and L_3).

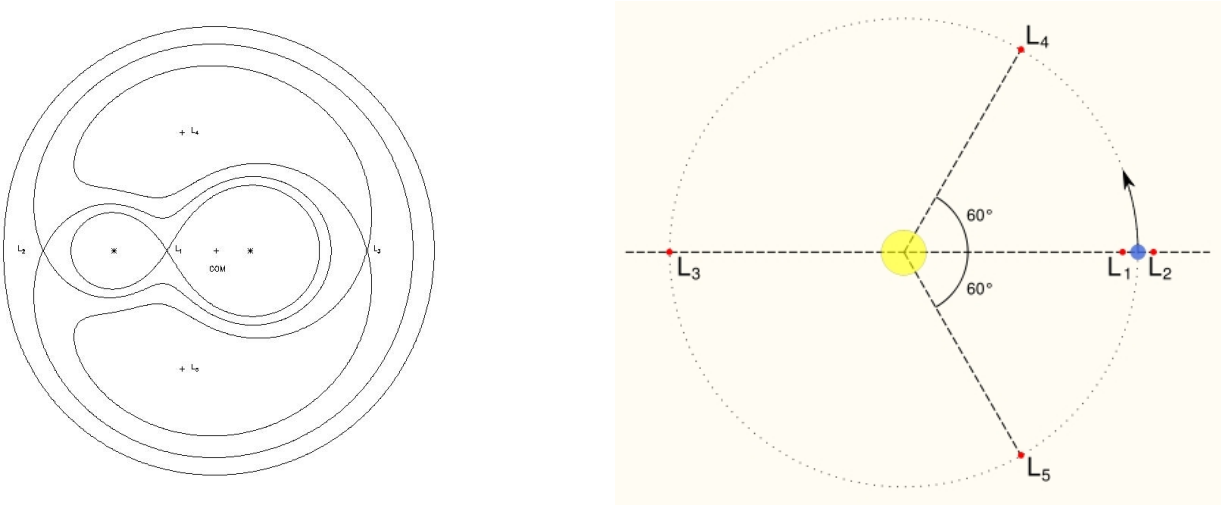


Figure 2.4: Left: the equipotential contours of the Roche potential in the orbital plane. Right: schematic view of the location of the Lagrangian points of the Earth - Sun system.

The L_1 and L_2 Langrangian points of the system Earth - Sun are located at about 1.5 million km from the Earth. These points are of major interest for space science. In fact, these points are positions of relative equilibrium and are hence ideal for formation flight experiments. Moreover, their thermal environment is very stable. For instance, an astronomy spacecraft in L_2 is always on the night side of the Earth, hence avoiding the effects of the changing thermal radiation of the Earth. This advantage has been used by spacecraft such as WMAP, Herschel and Planck, and will be used for ongoing and future missions such as Gaia and JWST. Solar observations (e.g. SOHO) are particularly interesting from L_1 which offers an un-interrupted view of the Sun.

Chapter 3

How to travel in space?

Spacecraft can have a variety of destinations in the Solar System, they can either orbit the Earth (on a number of different types of orbits), orbit one of the Lagrangian points or travel to another planet. All the motions are ruled by the laws of gravity and before we will consider the issue of the choice of an orbit/trajectory, we first recall some fundamental results of classical mechanics. We then establish and discuss the rocket equation (also known as Tsiolkovsky's equation) that expresses the possibilities and limitations of the current generation of spacecraft. In the last section, we review the various solutions that exist for interplanetary missions and how gravity can assist us in reaching the velocities necessary to travel to remote destinations in the Solar System.

3.1 Newton's law of gravity

Consider two masses M and m (e.g. a planet and a spacecraft respectively). In what follows, we neglect the influence of other objects of the Solar System. In other words, we consider that the motion of m occurs well inside the sphere of influence of M (for a definition of the sphere of influence, see the previous chapter). If the two masses act as point-like objects, they attract each other with a force according to Newton's law of gravity (see also lectures on *Celestial Mechanics*):

$$m\,\underline{\ddot{r}} = -\frac{G\,M\,m}{r^2}\,\underline{e_r} \tag{3.1}$$

where $\underline{r} = r\,\underline{e_r}$ is the vector joining the centres of M and m. The first result that is obvious from Eq. 3.1 is the conservation of the angular momentum. In fact,

$$m\,\underline{r} \wedge \underline{\ddot{r}} = \underline{0}$$

hence,

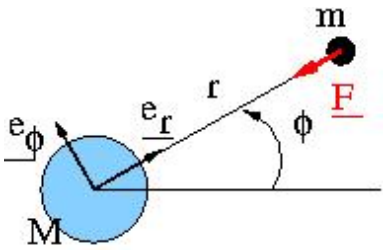
$$m\,\underline{r} \wedge \underline{\dot{r}} = \underline{L} = \underline{Cst}$$

Therefore, the velocity and the position are always perpendicular to a constant vector $\underline{L}$ and the motion thus occurs in a **plane**. In this plane, we then adopt the polar coordinate system (see the cartoon). ϕ is defined with respect to a constant direction of reference. Hence, the velocity of m around M can be expressed as $\underline{v} = \dot{r}\,\underline{e_r} + r\,\dot{\phi}\,\underline{e_\phi}$.
The conservation of the angular momentum can now be written as

$$r^2\,\dot{\phi} = h = Cst \tag{3.2}$$

where h is the norm of the angular momentum per unit mass.
On the other hand, equation 3.1 also yields the evolution of the kinetic energy per unit mass $T = \frac{1}{2}\,v^2$:

$$\frac{dT}{dt} = -\frac{G\,M}{r^2}\,\underline{e_r} \cdot \underline{\dot{r}}$$



Since $-\frac{G\,M}{r^2}\,\underline{e_r}$ can be expressed as $-\underline{\nabla}(-\frac{G\,M}{r})$, we find that the total energy per unit mass E is preserved and thus

$$\frac{1}{2}\,v^2 - \frac{G\,M}{r} = E \tag{3.3}$$

This relation can also be written as

$$\frac{1}{2}\,(\dot{r}^2 + r^2\,\dot{\phi}^2) - \frac{G\,M}{r} = E$$

Given that $r\,\dot{\phi} = \frac{h}{r}$, we finally obtain:

$$\frac{1}{2}\,\dot{r}^2 + \frac{h^2}{2\,r^2} - \frac{G\,M}{r} = E \tag{3.4}$$

It is quite convenient to define an effective potential

$$U_{\text{eff}}(r) = \frac{h^2}{2\,r^2} - \frac{G\,M}{r} \tag{3.5}$$

This relation is shown in the figure below and we will use it to discuss the various types of trajectories.

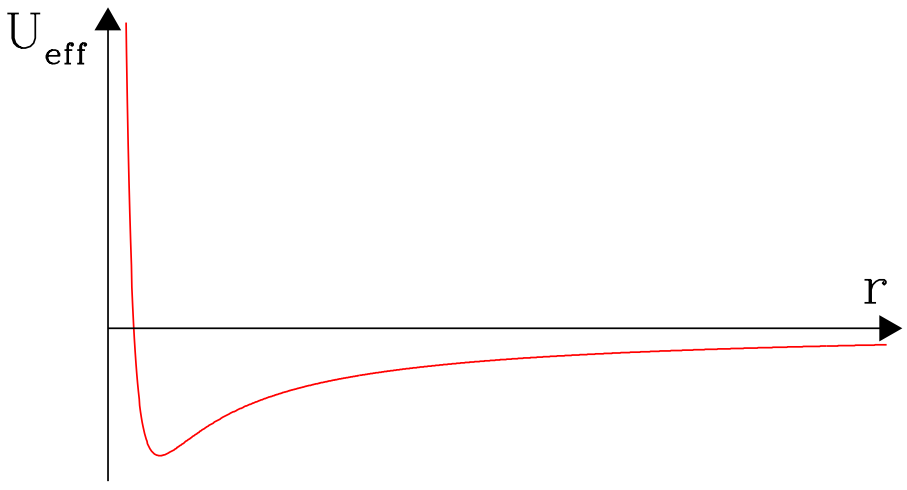


Figure 3.1: Schematic representation of the effective potential.

For any given trajectory, the conservation of total energy implies that $\frac{1}{2}\,\dot{r}^2 = E - U_{\text{eff}}(r)$. Hence, since $\dot{r}^2$ must be a positive quantity, the motion is only possible provided that $E \geq \min\,(U_{\text{eff}}(r))$.

Let us first consider the case where $E = \min\left(U_{\text{eff}}(r)\right)$. In this case, there is a single value of r that is allowed and the orbit is thus circular. The radius of the orbit can be obtained from

$$\frac{d}{dr}U_{\text{eff}}(r) = \frac{d}{dr}\left(\frac{h^2}{2\,r^2} - \frac{G\,M}{r}\right) = -\frac{h^2}{r^3} + \frac{G\,M}{r^2} = 0$$

which leads to $r = \frac{h^2}{G\,M}$. For a circular orbit, $\dot{r} = 0$ and $h = r\,v_c$. As a result, we obtain the expression of the orbital velocity

$$v_c = \sqrt{\frac{G\,M}{r}}$$

This is an important result. In fact, it tells us what is the velocity required to put a satellite into a circular orbit of radius r around a planet of mass M. If we replace r by the equatorial radius of the Earth and M by the Earth's mass, we obtain the so-called **first cosmic velocity**.

$$V_c = \sqrt{\frac{G\,M_\oplus}{R_\oplus}}$$

This value ($V_c = 7.904\,\text{km}\,\text{s}^{-1}$) is the absolute minimum velocity needed to put a satellite into an orbit around the Earth.

Next, we consider the situation where $\min\left(U_{\text{eff}}(r)\right) < E < 0$. In this case, we are dealing with a closed orbit around the planet and the trajectory is an ellipse (see lectures on *Celestial Mechanics*). The best known illustrations of this case are the planetary orbits that we have discussed in the previous chapter. The separation between m and M changes periodically and the point of closest approach is called the pericentre. In the case of a geocentric or heliocentric orbit, we talk about perigee or perihelion, respectively. Conversely, the largest separation corresponds to the apocentre (apogee and aphelion for orbits around the Earth and the Sun, respectively). As a result of Kepler's second law (which is of course equivalent to the conservation of angular momentum), the velocity along an elliptical orbit changes with the separation. The velocity is largest at pericentre and minimum at apocentre. From an energetic point of view, it is most advantageous to change the orbit by accelerating or decelerating at pericentre. In fact, the energy variation ($V\,\Delta V$) is largest at pericentre because of the maximum velocity over this part of the orbit. For a satellite orbit around the Earth, the velocity at pericentre can be expressed as

$$V_{\text{per}} = V_c\sqrt{\frac{2\,R_\oplus}{R_{\text{per}}} - \frac{R_\oplus}{a}}$$

where a is the semi-major axis of the orbit, $R_{\text{per}} = a\,(1-e)$ is the distance from the Earth's centre at perigee and V_c is the first cosmic velocity. This relation immediately tells us that the result of an acceleration at perigee (i.e. an increase of V_{per}) is an increase of a (hence an increase of the eccentricity). Therefore, if a spacecraft accelerates at perigee, the altitude of apogee increases. In many cases, a spacecraft is put in a low-Earth parking orbit immediately after launch. The apogee of the orbit is then progressively raised by applying thrusts on successive perigee passages. Conversely, braking the spacecraft velocity at perigee will reduce the altitude at apogee and if this process is repeated, then the orbit will circularize progressively. Finally, it is possible to modify the pericentre distance by applying an acceleration or deceleration at apocentre (which increases respectively reduces the pericentre altitude).

The first open trajectory (i.e. that allows the spacecraft to go to an infinite distance from M after an infinite time) is a parabola and corresponds to $E = 0$. In this case, the equation of energy becomes

$$\frac{1}{2}\,v^2 - \frac{G\,M}{r} = 0$$

Therefore, at any position around this parabola, one can write $\frac{1}{2}\,v^2 = \frac{G\,M}{r}$. In particular, if we consider a probe launched from Earth, the velocity required to put it on a parabolic trajectory is given by

$$V_{\rm esc} = \sqrt{\frac{2\,G\,M_\oplus}{R_\oplus}}$$

This is the **second cosmic velocity** ($V_{\rm esc} = 11.178\,{\rm km\,s^{-1}}$) which corresponds to the minimum velocity required for a spacecraft to escape from the gravitational field of the Earth.
In a similar way, one can define the **third cosmic velocity** as the minimum velocity one has to provide to a spacecraft to make it leave the Solar System (i.e. the Sun's attraction) starting from the Earth's orbit (i.e. from a heliocentric distance of 1 AU). This velocity amounts to $\sqrt{\frac{2\,G\,M_\odot}{a_\oplus}} = 42.1\,{\rm km\,s^{-1}}$ where $a_\oplus = 1\,{\rm AU}$.

The last type of trajectory that we shall consider here corresponds to $E > 0$. In this case, we are dealing with a hyperbola and the spacecraft will again escape to infinity, but unlike the parabola, this time its velocity at infinity will not be zero. From the energy equation, it follows immediately that $V_\infty^2 = 2\,E$. E being constant over the trajectory, one finds that

$$V_\infty^2 = V_{\rm per}^2 - V_{\rm esc}^2\,\frac{R_\oplus}{R_{\rm per}}$$

where $V_{\rm per}$ and $R_{\rm per}$ are the velocity and distance at perigee. The perigee distance of a typical escape orbit is about 1.05 – 1.1 $\rm R_\oplus$. Assuming $R_{\rm per} = 1.05\,R_\oplus$, the velocities at perigee to reach an escape orbit with $V_\infty = 3$ or $7\,{\rm km\,s^{-1}}$ are 11.31 or $12.96\,{\rm km\,s^{-1}}$, respectively.
For a launch from Earth onto a hyperbolic orbit, one can define

$$C3 = V_\infty^2 = V_{\rm per}^2 - V_{\rm esc}^2$$

As we will see in Sect. 3.2, $C3$ is a measure of the difficulty and hence the cost to send a spacecraft to a large distance from Earth with an asymptotic velocity V_∞.

3.2 The rocket equation

A launcher for space applications has to fulfil three different tasks that impact on its design:

- overcome the gravity field,
- cross a rather dense atmosphere,
- provide the required horizontal velocity to the payload.

These considerations lead to several constraints that can be formulated in a very general way:

- the thrust at take-off must be larger than the weight of the launcher (including the payload),
- the launcher must have an aerodynamic shape,
- the engines of the launcher must be able to operate in the vacuum of empty space.

Currently, all launchers rely (at least partially) on rockets for injecting the payload into orbit. Most rockets use very energetic chemical propellants that react in a combustion chamber to form hot combustion gases that are then ejected at high velocity through a nozzle (see Sects. 3.2.1 and 3.2.2 below). The thrust of a rocket engine is the reaction experienced by the rocket as a result of the ejection of the high velocity gas.

Consider a rocket travelling in gravity-free space. The so-called rocket equation or Tsiolkovsky equation expresses the conservation of the total momentum

$$\frac{d}{dt}(m\,\underline{v}) = \underline{0}$$

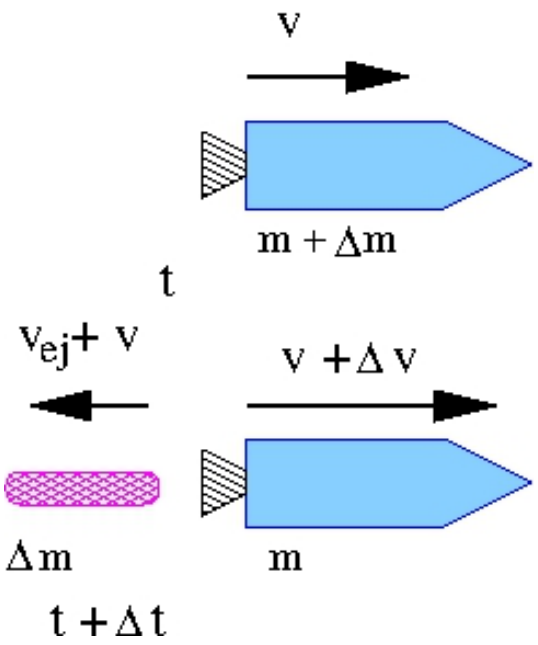


Figure 3.2: Schematic illustration of rocket propulsion. The rocket is shown at time t and $t + \Delta\, t$.

If we assume that the rocket engine is operated for a time $\Delta\, t$, and that during that time interval the gas is ejected with a constant speed relative to the rocket, the above relation can be expressed as:

$$\Delta(m\,\underline{v}) = \underline{0}$$

At time t, the rocket and fuel mass is $m + \Delta\, m$ and the rocket moves at velocity $\underline{v}$ with respect to an inertial frame of reference. At $t + \Delta\, t$, a mass $\Delta\, m$ has been ejected and the rocket mass is m. The rocket now moves at velocity $\underline{v} + \Delta\,\underline{v}$ (still with respect to the same inertial frame of reference) and the gas has been ejected at a velocity $\underline{v}_{\rm ej}$ with respect to the rocket. Typical ejection velocities of conventional rocket engines are of the order of $3\,{\rm km\,s^{-1}}$.

$$(m + \Delta m)\,\underline{v} = m\,(\underline{v} + \Delta\underline{v}) + \Delta m\,(\underline{v} + \underline{v}_{\rm ej})$$

$$\Rightarrow \Delta\underline{v} = \frac{-\Delta m\,\underline{v}_{\rm ej}}{m}$$

Now, if we divide by $\Delta\, t$ and let $\Delta\, t$ approach zero, this can be written

$$\frac{d}{dt}\underline{v} = \frac{-1}{m}\left|\frac{d\,m}{dt}\right|\,\underline{v}_{\rm ej}$$

$$\Rightarrow \int_{t_1}^{t_2}\frac{d\,v}{dt}\,dt = -\int_{t_1}^{t_2}\frac{1}{m}\,\frac{d\,m}{dt}\,v_{\rm ej}\,dt$$

$$\Rightarrow \Delta v = -\ln\left(\frac{m_{t_2}}{m_{t_1}}\right) v_{\rm ej} = v(t_2) - v(t_1) \tag{3.6}$$

Therefore:

$$\frac{m_{final}}{m_{init}} = \exp\left(\frac{-\Delta v_{\rm total}}{v_{\rm ej}}\right) \tag{3.7}$$

This is the rocket equation for a single-stage rocket. It actually expresses the requirements in terms of the initial mass needed to accelerate a final mass to a given velocity. However, the final mass is not equal to the mass of the payload m_u. In fact, one has to account for the additional mass needed for the fuel tanks and the supporting structure. These usually correspond to an additional mass of order 15% of the mass of the propellant (which is equal to $m_{init} - m_{final}$). Therefore, $m_{final} = m_u + 0.15\, m_{prop}$ and $m_{init} = m_u + 1.15\, m_{prop}$.
As a result, we find:

$$m_{prop} = m_u \frac{\exp\left(\frac{\Delta v_{\rm total}}{v_{\rm ej}}\right) - 1}{1.15 - 0.15\, \exp\left(\frac{\Delta v_{\rm total}}{v_{\rm ej}}\right)}$$

$$m_u = m_{init} \frac{1.15 - 0.15\, \exp\left(\frac{\Delta v_{\rm total}}{v_{\rm ej}}\right)}{\exp\left(\frac{\Delta v_{\rm total}}{v_{\rm ej}}\right)}$$

Hence, for a useful mass that becomes zero, the maximum velocity that can be reached by a single-stage rocket is roughly twice the ejection velocity. Thus the maximum velocity increase that can be achieved for an ejection velocity of 3 km s^{-1} is about 6.3 km s^{-1}. This is not sufficient to put a satellite into a stable orbit around the Earth, not to speak about an interplanetary mission.
The solution to this dilemma is a multi-stage rocket which carries its propellant in smaller separate tanks that are discarded (together with their engines) when empty. Therefore, no energy is wasted to accelerate empty tanks and higher velocity increases are possible. For multi-stage rockets, the rocket equation has to be applied to each stage (consisting of the tank and its engines) separately and the effective increase in velocity is the sum of the individual velocity increments. For each stage, the payload mass consists of the mass of all subsequent stages plus the ultimate payload itself. To benefit as much as possible from the Earth's rotational velocity and reduce the mass penalty for a given Δv, a good practice is to launch due East from near the equator.
The performance of a rocket is often expressed as a function of the $C3$ parameter defined in Sect. 3.1. An example is shown below for the Soyuz-Fregat launcher.
When gravity losses are taken into account (for an ascending rocket that moves along the vertical direction), equation 3.6 becomes

$$\Delta\, v = -\ln\left(\frac{m_{t2}}{m_{t1}}\right) v_{\rm ej} - \int_{t1}^{t2} g\, dt \tag{3.8}$$

This situation will be considered in more details in exercise 10.1.
To avoid wasting too much energy and hence fuel, the design of the early phases of a mission needs to be optimized following some elementary rules. First, one needs to remember that a spacecraft can be launched only as the launch site rotates through the plane of the transfer orbit. For instance, for a launch into a Sun-synchronous orbit, the launch window amounts to a few minutes twice each day. Furthermore, if there is a constraint on the launch azimuth at the site, only one of these opportunities remains. After the launch, an orbit transfer is usually required to put the spacecraft into its operational orbit. This is a critical phase and the accuracy of this operation is essential to limit the subsequent use of the secondary propulsion system for orbit and attitude control. One has to keep in mind that each manoeuvre may only transfer a vehicle between intersecting orbits.

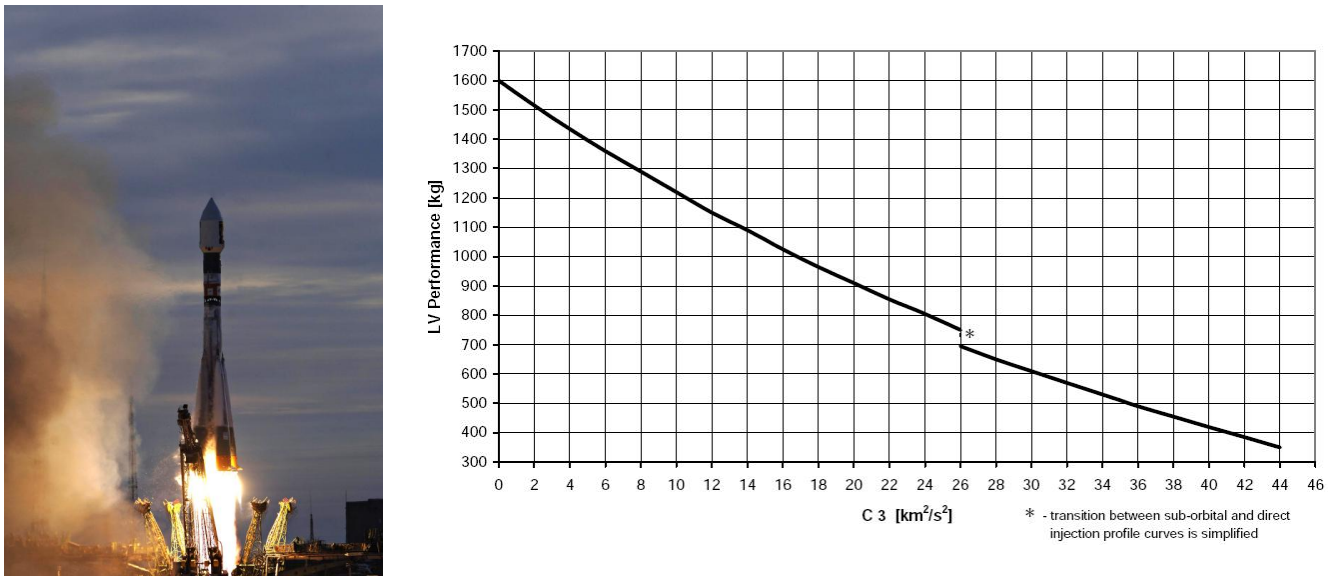


Figure 3.3: The payload mass of the Soyuz-Fregat launcher (left) as a function of the $C3$ parameter (right panel).

3.2.1 Liquid propellant rockets

Liquid propellant rockets produce a hot gas through the chemical reaction (combustion) between a liquid fuel, such as hydrogen (H_2), methane (CH_4) or kerosene ($C_{12}H_{26}$) and an oxidizer, usually liquid oxygen (O_2). The rocket engine consists of an injection system, the combustion chamber and the nozzle. Because the combustion takes place at high pressure and high temperature (typically between 2800 and 3900 K), the chamber and the nozzle have to be cooled. This is usually achieved by flowing a coolant over the back side of the chamber and the nozzle. This coolant is usually one of the propellants.
The nozzle is characterized by a large section ratio between the exit area A_e and the area of the throat. The nozzle converts the thermal energy generated by the combustion into kinetic energy. It does so by reducing the gas pressure from a high value P_c in the combustion chamber to a low value P_e at the exit.
The thrust is given by

$$F = -\frac{d\,m}{dt}\,v_{\rm ej} + (P_e - P_a)\,A_e$$

where P_a is the pressure of the ambient atmosphere ($P_a = 0$ outside the atmosphere). The total thrust hence consists of the impulse thrust, given by the product of the mass loss rate times the ejection speed, plus the pressure thrust which accounts for the unbalanced pressure forces at the nozzle exit. It can be shown that maximum thrust is achieved when $P_e = P_a$. A nozzle is designed for the altitude, and hence the atmospheric pressure, at which it has to operate. In some cases, however, the diameter of the nozzle is limited by the maximum allowable diameter set by the outer diameter of the stage below.
It can be shown that the ejection velocity is given by

$$v_{\rm ej} = \sqrt{\left(\frac{2\,\gamma}{\gamma-1}\right)\left(\frac{\mathcal{R}\,T_c}{\mathcal{M}}\right)\left(1-\left(\frac{P_e}{P_c}\right)^{(\gamma-1)/\gamma)}\right)}$$

where $\gamma = c_p/c_v$ is the specific heat ratio, $\mathcal{R}$ is the universal gas constant and $\mathcal{M}$ is the average molecular weight of the exhaust gases. From this relation, we see that high combustion temperature and pressure and low exhaust gas molecular weight result in high ejection velocity.

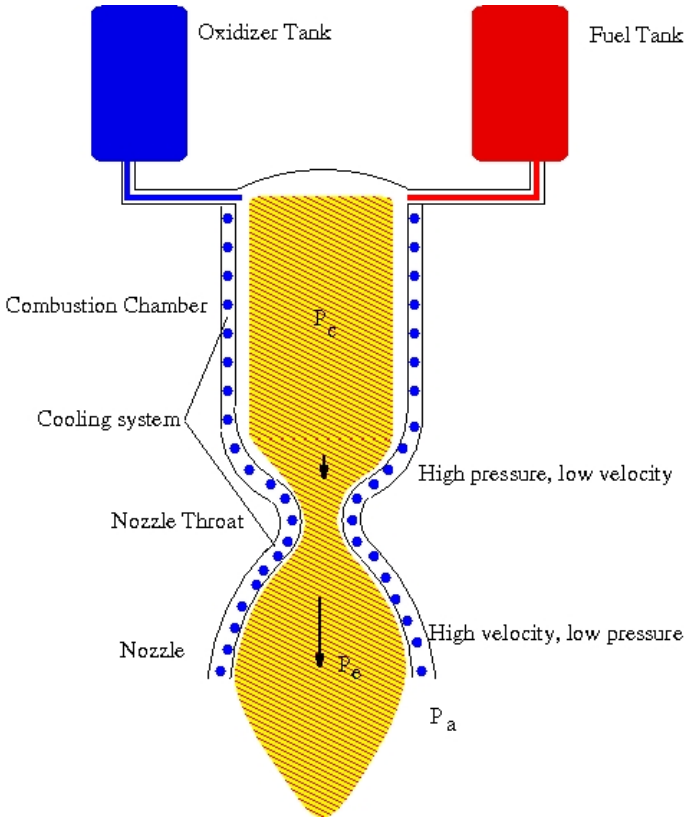


Figure 3.4: Schematic view of a liquid propellant rocket engine. The oxidizer is not only used to react with the fuel, but it also provides the cooling of the engine. Near the nozzle throat, the gas in the combustion has a high pressure, but a low bulk velocity, whilst the gas pressure is significantly reduced at the exit of the nozzle and has much larger exhaust speed instead.

The specific impulse of a rocket is defined as the ratio

$$I_{\text{sp}} = \frac{F}{-\frac{dm}{dt}\, g}$$

This quantity can be seen as the time interval during which the rocket engine provides a thrust equal to the weight of the propellant consumed. As an illustration, we consider the Merlin 1D engine of the Falcon Heavy launcher[1]. At sea level, this engine produces a thrust of 845 kN and a specific impulse of 282 s.

3.2.2 Solid fuel rockets

Solid fuel rockets store the propellants in solid form. The fuel and oxidizer are typically powders held together in the same tube. The solid fuel's geometry determines the surfaces that are involved in the combustion. It is important to note that after ignition, operation of a solid fuel rocket only stops once all the fuel has been used up. There are two broad categories of solid fuel rockets. In the first case, the fuel/oxidizer mix forms a cylindrical block and the combustion front has a constant section (the section of the cylindrical tube), hence producing a roughly constant thrust. The combustion progresses upwards across the propellant until all the propellant has been consumed. In the second case, the combustion surface develops along a cylindrical channel. The geometrical shape of this channel determines the growth rate of the combustion surface and hence the thrust regime (see Fig. 3.5): a fuel block with a cylindrical channel develops its thrust progressively, reaching maximum shortly before exhaustion; a fuel block with a cylindrical channel partly filled by a central cylinder of fuel has a roughly constant thrust; and finally, a cruciform channel produces its maximum thrust shortly after ignition. The shape of the fuel block is chosen for the particular type of mission the rocket has to accomplish.

[1]This launcher features 28 Merlin 1D engines, 27 of them in the first stage and the auxiliary boosters.

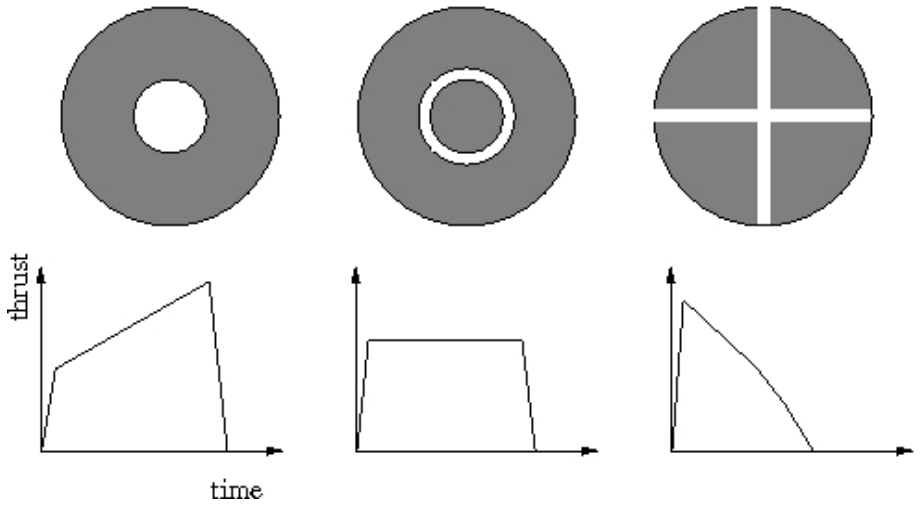


Figure 3.5: Cross sections of three commonly used shapes of the fuel blocks in solid fuel rockets. The lower panels illustrate the evolution of the thrust with time after ignition.

In solid fuel rockets, the combustion products consist of both gas and smoke: whilst the pressure is produced by the sole gas, the smoke nevertheless contributes to the thrust of the engine, due to its mass and ejection velocity.

3.2.3 The current generation of European launchers

Currently, ESA clients have the possibility to use three different launchers.
Ariane V is the most evolved launcher of the Ariane family. It features two cryogenic stages (liquid oxygen and hydrogen) and two solid fuel boosters. Ariane V is currently used in two configurations: Ariane V ECA which can put 10 tons of payload into a geosynchronous transfer orbit and Ariane V ES which is designed for delivering payloads into low- or medium-Earth orbits. Ariane V ES has the capability of putting 20 tons into low-Earth orbit and is used for instance for putting the Automatic Transfer Vehicle into a rendezvous orbit with the ISS.
For medium-sized satellites, ESA has concluded an agreement with the Russian agency Roscosmos to launch Soyuz-ST vehicles from Kourou. Owing to Kourou's location near the equator, Soyuz-ST can carry up to 3 tons into geostationary transfer orbit, compared to 1.7 tons that can be launched with the same vehicle from Baikonour[2].
Soyuz-ST features four liquid-fuel stages. The three main stages are powered by liquid oxygen and kerosene. The upper stage of Soyuz-ST, called Fregat, is designed as an orbital vehicle: it can be restarted up to twenty times, allowing it to deliver payload into a wide variety of orbits and to perform rather complex missions. The maiden flight of a Soyuz-ST launcher from Kourou took place on 21 October 2011.
Finally, the Vega launcher is designed for payload masses between 300 kg and 2.5 tons, depending on the orbit: for instance, launched from Kourou, Vega can put 1.5 tons into a polar orbit at 700 km altitude. Vega features three solid-propellant stages and a liquid propellant upper stage. The maiden flight took place on 13 February 2012.

3.2.4 The future of launchers

Access to space is expensive. For instance, putting 1 kg into LEO costs about 8000 € for unmanned missions, but up to 350 000 € for manned missions. One of the reasons for these costs are the launchers. Several strategies are

[2]The rocket launch site of Baikonour is actually located 370 km from the city of Baikonour. During the cold war, the Soviet government did not want to reveal the actual location of its space centre.

considered to reduce these costs.
Recovering the first stage of a launcher has become a standard procedure for Space-X, and its competitor Blue Origin is also working in that direction. But does this really provide an enormous saving? The cost of developing this technology was substantial (probably around 1 billion USD) and it will take a number of successful launches to recover it. Moreover, this concept relies on retrorockets to slow down the stage and enable a soft landing. This manoeuvre therefore requires up to one third of the fuel of the stage. This induces a penalty as far as the launch capabilities are concerned. For instance, the Falcon Heavy launcher can put 26.7 tons into a geostationary transfer orbit when none of its stages is to be recovered. When only the boosters are to be recovered, this mass goes down to a bit more than 10 tons, and it is further reduced to a bit over 8 tons in case the first main stage is also to be recovered. Alternative techniques are currently being studied in Europe, such as the recovery of the engines of the stage only (e.g. by means of parachutes).

Alternative options to introduce reusability, and to overcome the limitations of conventional rockets are being studied by various actors. Some of them may still sound like science fiction, but they are seriously being investigated.

- The Dream Chaser, developed by Sierra Nevada Corporation, is a Space Shuttle like re-usable spacecraft designed to carry up to seven people to a space station. It should be launched vertically on top of an Atlas 5 rocket and land, upon return, as a plane on a runway.

- The Synthetic Air-Breathing Rocket Engine (SABRE) is a hybrid hydrogen-fueled propulsion engine breathing within Earth atmosphere, up to 28 km altitude, and acting as a conventional rocket engine at higher altitude using its own liquid oxygen reservoir. This engine is designed for the Skylon project of the Reaction Engine Limited company (www.reactionengines.co.uk). First on-ground tests of the SABRE engine are foreseen around 2020. The development is partially funded by ESA.

- To reduce the costs of an access to space, the Spanish company Zero2infinity proposes to implement launches from an altitude of 40 km that would be reached thanks to a balloon filled with helium. The balloon would carry a small three-stage rocket to this altitude that would then be able to put a 75 kg payload into a 400 km low Earth orbit (LEO).

- Another option for an air-launch platform was developed by Stratolaunch (www.stratolaunch.com, see Fig. 3.6). This concept involved a plane with a 117 m wingspan that should carry the rocket to an altitude of 10.7 km. The carrier aircraft features a dual fuselages and carries six Boeing 747 engines. The announced performances were 450 kg for LEO. In June 2019, the Stratolaunch corporation closed operations, only two months after the first flight of the carrier plane and eight months after the death of its founder Paul Allen (co-founder of Microsoft).

- Nuclear thermal propulsion: thrust is provided by heating hydrogen via the contact with nuclear reactors and is ejected through nozzles. NASA has worked on this technology between 1955 and 1972. This led to the NERVA (Nuclear Engin for Rocket Vehicle Application) engine (see Fig. 3.7). Compared to chemical propulsion, an optimized version might provide a factor of two improvement of the performances.

- Mass-drivers are electromagnetic 'catapults' with moving parts in magnetic levitation above a rail. The mass driver magnetically accelerates a magnetizable carrier which contains the payload. Upon reaching the required velocity, the carrier and the payload separate. Whilst the payload continues its journey, the carrier is slowed down to be reused for another launch. A 1 km rail could accelerate a 20 kg mass to 10.5 km s^{-1}. Due to the friction by the air, this type of propulsion would be better suited for air-less bodies (e.g. on the Moon), unless one would put the rail in vacuum tubes extending to altitudes of more than 20 km.

Figure 3.6: Artist view of the Stratolaunch concept.

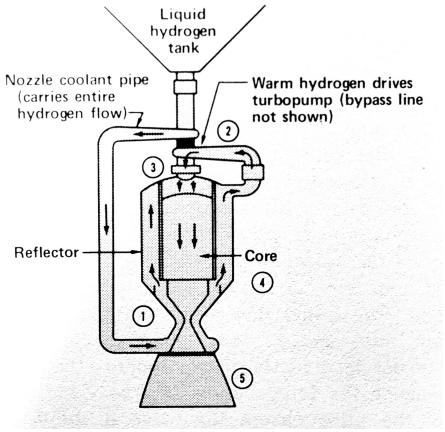


Figure 3.7: Schematic view of the NERVA engine.

- External pulsed plasma propulsion relies upon pulsed nuclear explosions behind the rocket which produce a hot plasma that acts upon a metal pusher plate attached to the spacecraft through damped springs. The successive explosions of small nuclear charges would then lead to a continuous acceleration of the spacecraft. The bottlenecks of this technology are the radioactive fallout and the radiation effects on the material of the pusher and the crew.

Finally, let us stress that regular launches of a launcher are a necessary condition to maintain its reliability and to preserve the know-how of the technical teams. Indeed, such a know-how usually takes a long time and a lot of money to build, but is difficult to maintain since the teams dissolve very quickly when the project gets stuck (see for instance the loss of the know-how of the Apollo programme in the US with the advent of the Space-Shuttle programme, which causes the current dilemma in the US space policy).

3.3 The various categories of orbits

The choice of an orbit, or more generally of a trajectory, is an important step in the design of any space mission. This is at the core of what is called mission analysis and will be discussed in this section and the forthcoming Sect. 3.5. At the start of a space project, the mission requirements are evaluated to obtain an overview of the available trajectory options. For each option, the mission analyst computes the information needed for a proper trade-off between the different options and to define one or more baseline and back-up solutions for further detailed analysis, definition and optimisation. This information usually includes the timeline of major events, the launcher injection orbit and mass, the ΔV budget, the power and thermal aspects related to the orbit (such as eclipses and distance from the Sun), the distance from Earth, the Sun-Earth-spacecraft angle and its influence on telecommunications, the visibility of science targets, as well as a qualitative assessment of complexity and operational risk.

Mission analysis is especially important and complex for interplanetary trajectories (see Sect. 3.5). As an example, let us consider Bepi Colombo, ESA's project for a mission to Mercury. Originally, when the Ariane V launcher was to be used, just two flybys of Venus and two of Mercury were required, in combination with solar-electric propulsion. The whole trip could then be done in less than three years. However, the Ariane rocket became unaffordable and solutions with the less powerful Soyuz launcher had to be found. To compensate for the missing

thrust, one lunar flyby and an Earth flyby were introduced. Furthermore, the solar arrays had to be reduced in size, cutting the available ion engine thrust by half. As a consequence, the transfer duration increased to five years.

3.3.1 Orbits around the Earth or another planet

The most common type of spacecraft orbits are circular or elliptic orbits around the Earth. There are many different possible configurations depending on the orbital inclination (equatorial or polar orbit with all the possible intermediate situations), the eccentricity (circular orbits or highly eccentric ones) and the orbital period (short periods for low-Earth orbits which are usually circular; geostationary orbits; wide highly eccentric orbits...). The choice of an orbit depends mainly on the job the satellite is supposed to accomplish.

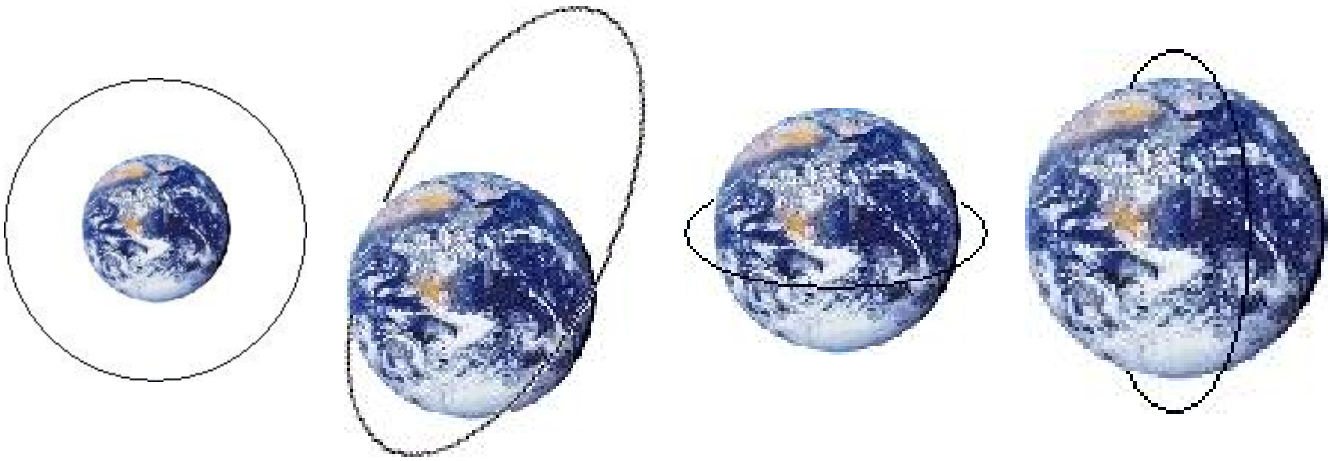


Figure 3.8: Schematic view of various kinds of orbits. From left to right: circular orbit in the equatorial plane, elliptical orbit with relatively high inclination, LEO equatorial and LEO polar orbit.

Let us start with the geostationary orbit. This is the most common choice for telecommunication satellites and for many weather satellites (e.g. Meteosat) used for short-term forecast. These satellites orbit exactly over the Earth's equator (zero inclination) with a period of 24 hours. For an observer on the ground, the satellite seems to be suspended on the same spot in the sky. For such an orbit, the ground stations do not need to track. The velocity on a circular orbit can be expressed as

$$V = \frac{2\,\pi\,r}{P_{\rm orb}} = \sqrt{\frac{G\,M_\oplus}{r}}$$

hence $r = \left(\frac{(G\,M)^{1/2}\,P_{\rm orb}}{2\,\pi}\right)^{2/3}$. For a geostationary orbit $P_{\rm orb} = 24\,\text{hours}$, which leads to $r = 42\,241\,\text{km}$ (i.e. an altitude of 35 860 km). There are several limitations to the usage of a geostationary orbit: since the satellites are on equatorial orbits, their elevation angle above the horizon decreases as the latitude of the ground station increases or the difference in longitude becomes too large. A global coverage in longitude can be achieved with three geostationary satellites separated by 120° in longitude.
When selecting the orbit, several considerations come into play. One concerns the power needed for telecommunications. Indeed, communications with a satellite in a geostationary (GEO) orbit have to comply with a signal 1000 times weaker than for the same satellite in an orbit at 1000 km altitude. This is not a very convenient situation, especially for applications involving small multi-directional antennas on the ground, such as mobile phones. Therefore, mobile phone telecommunications rather use a constellation of lower orbit satellites, such as the Iridium network, to achieve a good coverage of the entire Earth.

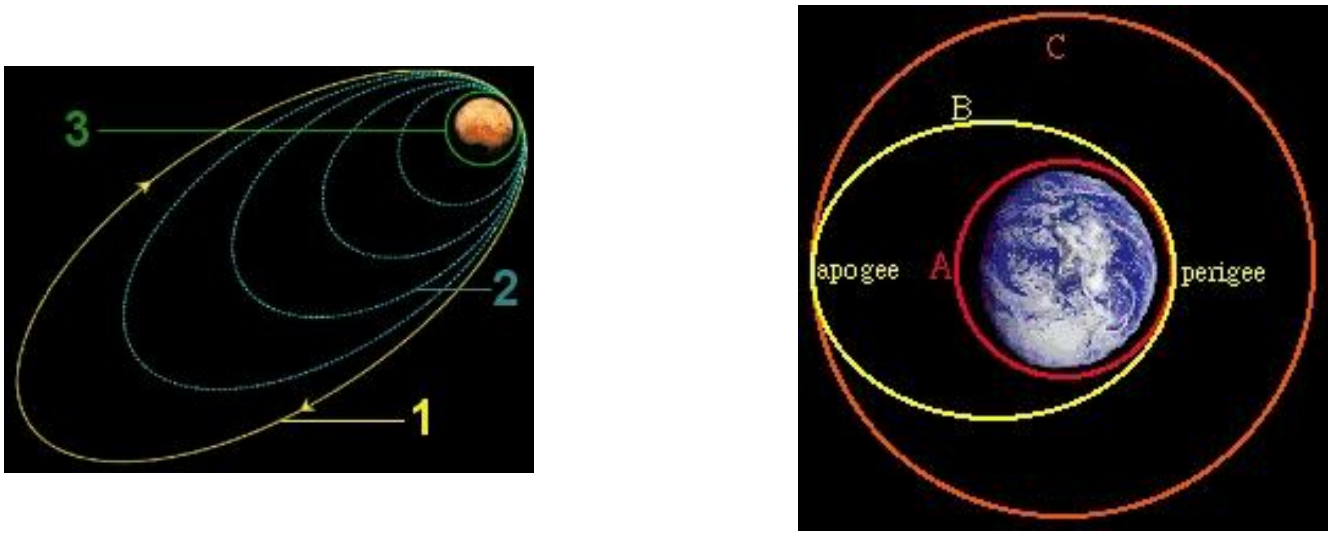


Figure 3.9: Left: schematic illustration of the effect of aerobraking on an initially highly elliptical orbit around planet Mars. Right: Hohmann transfer orbit between a circular LEO and a MEO around the Earth.

Satellites that need to map the Earth (monitoring of the climate...) or another planet are ideally put into an orbit that provides access to all parts of the planet. To get the most detailed view, the satellites should revolve at a rather modest altitude. These requirements can be met by low-Earth orbits (LEOs) which are commonly defined as orbits with altitudes between 300 and about 1000 km. To access as large a part of the planet's surface as possible, the inclination of the orbit[3] has to be rather close to 90°. A satellite in a polar orbit will go over all positions of the planet which is useful for mapping, but also for medium-term weather forecast. However, if a continuous access from a specific ground station is required then a constellation of satellites is needed. LEOs have the advantage that the power necessary for communication with the ground is rather modest. For scientific (astronomical) satellites, LEOs are of interest because they stay below the radiation belts over the entire orbit, although they might cross the South Atlantic Anomaly. The drawback is that the short orbital periods do not allow to observe a given source over extended periods of time.

A major issue with LEOs is the residual atmosphere. Satellites in orbit around the Earth experience friction from the atmosphere up to altitudes of 1000 km. This slows down the satellite and as a result, the altitude decays until the satellite falls back to the Earth. The density of the atmospheres and thus also the atmospheric drag decrease exponentially with altitude. For instance, a spacecraft orbiting at an altitude of 250 km experiences a 1000 times larger atmospheric drag than the same spacecraft orbiting at 800 km altitude. As a result, a satellite at an altitude of 200 km will stay in its orbit only for several months, without altitude correction manoeuvres. This effect is sometimes used to deliberately modify the orbit of a satellite. In fact, aerobraking is a controlled manoeuvre that reduces the altitude of the apocentre by taking advantage of the drag of the atmosphere at pericentre. This technique was first tried on the Magellan probe after the end of its nominal mission around Venus and was used lateron to place Mars Global Surveyor on its final orbit. Aerobraking takes a very long time (several months) and due to sudden changes in the atmospheric density, it becomes sometimes a risky endeavour. The kinetic energy dissipated by aerobraking is converted into heat and this poses a major problem for the thermal control during the manoeuvre.

Another problem is that satellites in low inclination LEOs spend a rather long fraction of their orbit in the Earth's shadow so that rather efficient batteries are needed to operate them while their solar panels receive no light. Be-

[3]The inclination of the orbit is given, with respect to the equatorial plane, between 0 and 180°. Values below 90° correspond to prograde revolution (in the same sense as the planet's rotation), whilst values larger than this limit indicate retrograde orbits.

cause one cannot benefit from the Earth's rotation, it is much more difficult to put a satellite into a polar orbit rather than a low inclination orbit. To change the inclination of an orbit, one has to apply a propulsion during the passage over one of the nodes of the orbit (i.e. when the satellite crosses the equatorial plane). For a circular orbit, the velocity increment to be provided at one of the nodes to produce a change in inclination $\Delta\, i$ (in radians) is to first order $\Delta\, V = \frac{2\,\pi\,a}{P_{\rm orb}}\,\Delta\, i$.

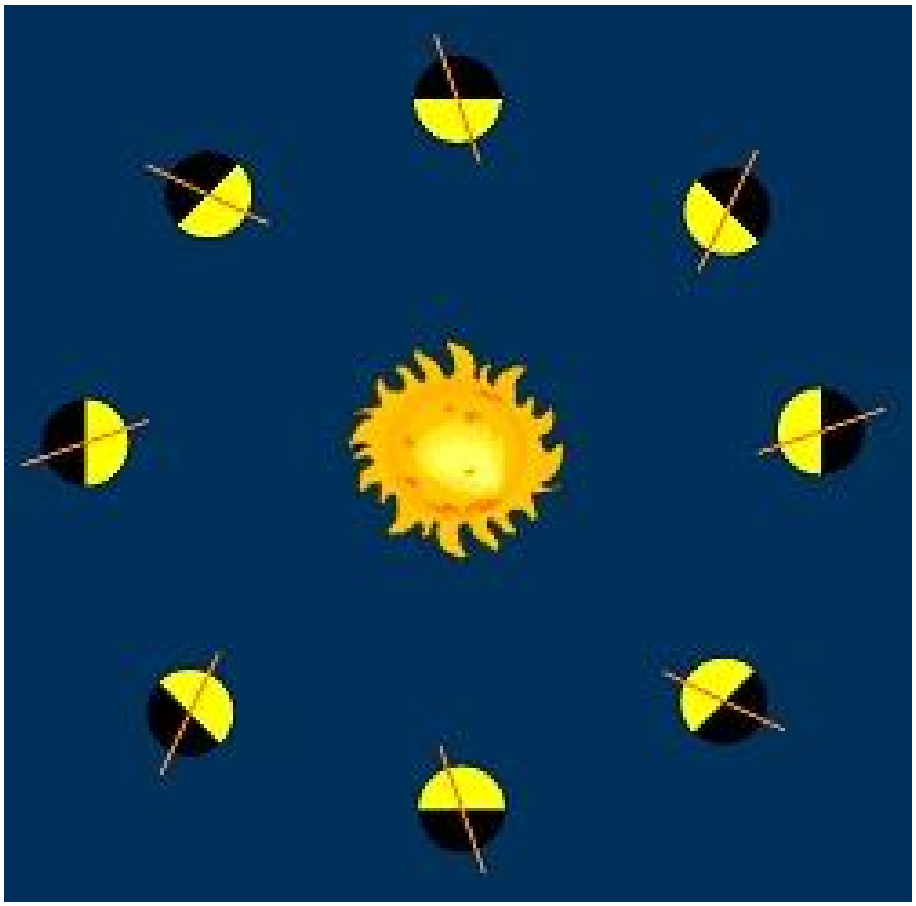


Figure 3.10: Schematic representation of the progressive rotation of the orbital plane of a Sun-synchronous orbit.

Sun-synchronous orbits pass over the same locations on Earth at the same local solar time each day. To do so, the plane of the orbit has to precess slowly. This is achieved by taking advantage of the non-spherical shape (oblateness) of the rotating Earth. For a planet with a shape that is not a perfect sphere, the force of gravity can be written as $\underline{F} = -\underline{\nabla}\, V$ with

$$V = -\frac{G\,M}{r}\,\left[1 - \sum_{n=2}^{\infty}\,J_n\,\left(\frac{R}{r}\right)^n\,P_n(\cos\theta) + \sum_{n=2}^{\infty}\,\sum_{m=1}^{n}\,J_{nm}\,\left(\frac{R}{r}\right)^n\,P_n^m(\cos m\,(\lambda - \lambda_{nm}))\right] \tag{3.9}$$

the geopotential expressed as a function of geographical latitude (θ) and longitude (λ). For details, see the lectures on *Celestial Mechanics*.

For a circular orbit, the resulting rate of node precession (in rad s^{-1}) is given by:

$$\frac{d\Omega}{dt} = -\frac{3\,\pi\,R_\oplus^2}{a^2\,P_{\rm orb}}\,J_2\,\cos i$$

where Ω is the right ascension of the ascending node, a is the radius of the satellite's orbit, i its inclination, and $P_{\rm orb}$ its period. $J_2 = 1.082 \times 10^{-3}$ is the second dynamic factor $J_2 = \frac{2\,\epsilon_\oplus}{3} - \frac{R_\oplus^3\,\omega_\oplus^2}{3\,G\,M_\oplus}$ that expresses the deviation of the Earth's shape from a sphere. For a specific value of the orbital inclination, the perturbation by the equatorial bulge triggers exactly the necessary rotation rate for the orbital plane to accomplish a full rotation in exactly one year. Typical Sun-synchronous orbits have an altitude around 600 – 800 km, a period in the range 96 – 100 minutes and an inclination of 98° (i.e. the orbit is retrograde compared to the Earth's rotation). An advantage of this orbit for Earth observing satellites is that the ground surface illumination angle is nearly the same at each passage of the satellite. This is useful for spy or remote sensing satellites. A special case are the dawn/dusk orbits where the satellite passes around sunrise or sunset, so that the satellite rides the terminator between day and night. In such an orbit, the satellites' solar panels are never shadowed by the Earth. This is useful for scientific satellites that carry instruments that point the night side of the Earth continuously. However, it does not offer an interesting solution for remote sensing since the Sun is always low on the horizon at the sub-satellite point, hence leading to a low illumination. A noon-midnight Sun-synchronous orbit does not offer a good choice for remote sensing either, since the definition of the scenes to image is rather low (due to the short shadows) and there is a potential risk of detector dazzling by specular reflection from sea surface. The optimal solution is therefore in between the two extreme configurations.

Medium Earth orbits (MEOs) with altitudes around 10 000 km are often used for navigation and communication satellites, such as the GPS constellation (see below). For many scientific applications it is advantageous to use a highly eccentric high-Earth orbit (HEO). The motivations are to reduce differential (tidal) forces on spacecraft in formation flying and to allow long monitoring observations of targets that are not possible in LEOs. These satellites spend a long time outside the radiation belts
A special type of HEOs are the Molniya orbits. The Molnyia have an inclination of 63.4° and an orbital period of 12 hours (actually half a sidereal day). A satellite in such a highly eccentric ($e \sim 0.7$) orbit spends most of its time (around apogee) over a specific area of the Earth. This solution is very interesting for telecommunication at high (positive or negative) latitudes, where communication from GEO satellites would require considerable power due to the low elevation above the horizon. This is the case for instance for Russia and the former USSR. Molniya orbits have thus been used for telecommunication in the USSR and the name 'Molniya' actually stems from the names of a series of Soviet communication satellites. Each satellite on such an orbit spends about 8 hours per day over Russia and as a result, only three satellites shifted in Ω by 120° are needed to achieve a full 24 hours coverage.
The inclination of the orbit is actually a key parameter. In fact, for an eccentric orbit, the longitude of perigee and hence the longitude of apogee as well are slowly moving due to orbit perturbations. This effect can be expressed as

$$\frac{d\omega}{dt} = -\frac{3\,\pi\,R_\oplus^2}{2\,a^2\,P_{\rm orb}}\,J_2\,\frac{5\,\cos^2 i - 1}{\sqrt{1-e^2}}$$

where J_2, ω, a, e and i are the second dynamic factor, the longitude of the perigee, the semi-major axis, the eccentricity and the inclination of the orbit respectively (see lectures on *Celestial Mechanics*).
For an inclination of $i = 63.43^\circ$ ($\cos i = \frac{1}{\sqrt{5}}$), this perturbation is zero and hence the longitudes of the perigee and apogee remain constant over a long period of time.
A drawback of this orbit is that it crosses the van Allen radiation belts twice per revolution.

3.3.2 Orbits around the Lagrangian points

The Lagrangian points L_1 and L_2 have become important destinations over the last decades. However, they are only stable against a perturbation in the plane perpendicular to the line joining the two main masses m_1 and m_2. On the contrary, any perturbation along this axis will result in the spacecraft being either accelerated towards one

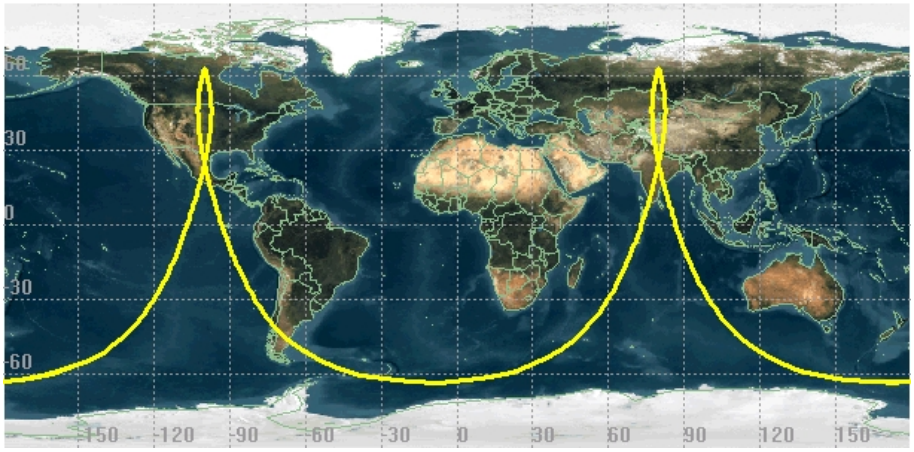


Figure 3.11: Projection of the Molniya orbit on the Earth. Since the orbital period is exactly half the rotation period of the Earth, only every second apogee occurs over Russia, the other one happens over North America.

of the main masses or being driven away from the system[4]. Although, the L_1, L_2 and L_3 points are unstable, there exist families of so-called halo and Lissajous orbits around these points that require a relatively modest effort to maintain the spacecraft in its orbit. The halo orbit designation is used for periodic three-dimensional orbits, whilst the term Lissajous orbit is used to refer to quasi-periodic, three-dimensional orbits. A spacecraft in a halo orbit travels in a closed, repeating path about the Lagrangian point as a result of the interaction between the gravitational pull of the two planetary bodies, the Coriolis and centrifugal accelerations and frequent modest stationkeeping manoeuvres. Spacecraft in Lissajous orbits follow a Lissajous curve (see lectures on *Celestial Mechanics*). They require a modest stationkeeping manoeuvre about every month to keep the satellite on its orbit.

The Lagrangian points are of major interest for uninterrupted observations of the Sun or the deep space. They also provide a roughly constant illumination of the solar panels and the thermal control is easier because the thermal load is constant in time. Many of the missions that have been proposed to use formation flight are ideally placed in orbit around L_2 because of the benign gravitational fields there. The main drawbacks of these orbits are the relatively long transfer durations, the large distances for telemetry and the need of regular stationkeeping manoeuvres.

In 1968, Robert Farquhar proposed the use of spacecraft in halo orbits around the L_2 point of the Earth-Moon system as a communications relay station for an Apollo mission to the far side of the Moon. Indeed, such an orbit would ensure continuous communication with the Earth and the far side of the Moon. However this proposal was turned down. The first mission to use a halo orbit was the International Sun-Earth Explorer (ISEE-3), launched in 1978. ISEE-3 traveled to the Sun-Earth L_1 point and remained there for several years until its initial mission was completed in June 1982[5].

ESA is currently studying a space situational awareness mission that would combine observations of the Sun from the L_1 and L_5 points. The advantage of the L_5 point in this context would be to offer a stereoscopic view on solar mass ejection events directed towards Earth. However, the requirements on communication for both probes are

[4]Note that the same comment holds for the L_3 point, whilst the L_4 and L_5 points are stable under certain circumstances depending on the value of the m_2/m_1 mass ratio (see lectures on *Celestial Mechanics*). In the case of the system Sun - Jupiter, there are several thousand asteroids, the so-called Trojans, that orbit around the L_4 and L_5 points.

[5]The spacecraft was then put into an transfer orbit that included 5 gravity assisted lunar flybys before it reached a heliocentric orbit ahead of Earth (355 days period, aphelion = 1.03 AU, perihelion = 0.93 AU, $i = 0.1°$) to encounter comet P/Giacobini-Zinner. On 11 September 1985, the veteran NASA spacecraft flew through the tail of the comet. ISEE-3 returned to the Earth-Moon system in August 2014. Attempts to reactivate the mission failed due to a failure of the thrusters.

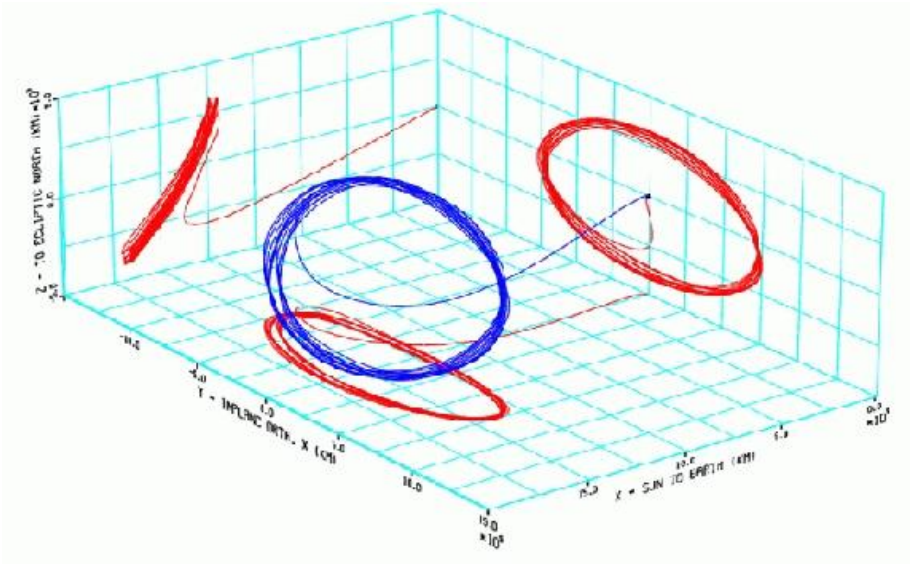


Figure 3.12: Halo orbit around the L_2 point. The Earth is located at the origin of the axes and the projection of the orbit on the various planes is shown.

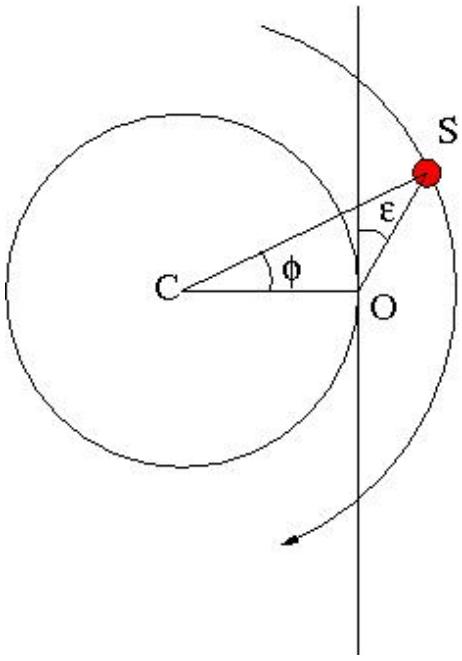


Figure 3.13: Illustration of the limited range over which a satellite (S) is visible from the ground (above an elevation angle ε as seen from an observer located at position O).

very different. L_1 being only 1.5 million km away from Earth, communication can be done with a relatively small on-board antenna and a ground station antenna of less than 15 m. On the other hand, the much larger distance to L_5 (150 million km) implies the need of an on-board antenna of 1.5 m and of a ground antenna of at least 30 m.

3.4 Satellite constellations

As stressed above, satellites in LEO have rather short orbital periods and are thus only visible over a limited duration from a given ground station. In addition, when the satellite is near the horizon, the signal received by the ground station is considerably attenuated by the atmosphere. Therefore, for the Earth, the surface coverage is restricted to a region in which the satellite elevation above the horizon is greater than $\varepsilon \sim 5 - 10°$.
This can be converted into a maximum fraction of the satellite's orbit during which it will remain visible. From Fig. 3.13, we note that

$$\phi = -\varepsilon + \arccos\left(\frac{R_\oplus}{R_\oplus + h}\,\cos\varepsilon\right)$$

The maximum duration[6] of an overhead pass is $2\,\phi/\sqrt{\omega_\oplus^2 + \omega^2 - 2\,\omega_\oplus\,\omega\cos i}$, where $\omega_\oplus$ and ω are respectively the Earth's and the satellite's angular velocities and i is the inclination of the spacecraft's orbit with respect to the equator.
For some specific applications, a permanent coverage of the ground station is required. If this is to be done with non-GEO satellites, the only option is a constellation of satellites. The Global Positioning System (GPS) offers a nice example of such a constellation. Indeed, to satisfy the global coverage requirements, the GPS constellation comprises 24 satellites deployed in 6 orbital planes, inclined by $55°$ to the equator. Each (nearly-circular) orbit plane contains 4 spacecraft, equally spaced in true anomaly and the nodes of the planes are equally spaced in right ascension around the equator. The orbital period is half a sidereal day.

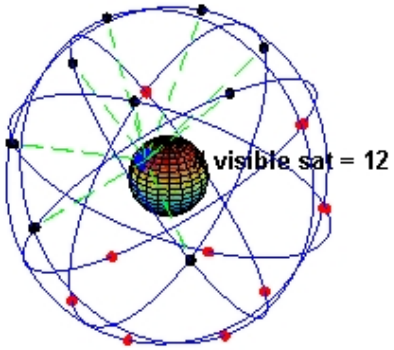


Figure 3.14: Schematic view of the constellation of satellites that build up the global positioning system.

The design of a constellation is a complex problem and a clever in-plane phasing is critical to avoid collisions between the satellites of the constellation. The complexity of the problem obviously increases the larger the number of satellites. For instance, for the 5G Starlink internet network, Space-X intends to deploy 4425 mini-satellites on 83 orbital planes at 1200 km altitude! Moreover, the constellation could be complemented by 7518 other satellites on lower orbits.

3.5 Interplanetary trajectories

For interplanetary probes, there are a number of possible mission profiles: flyby, orbiter or landers pose different constraints on the spacecraft's trajectory. To send a massive probe to a distant planet, one has to find the optimum solution in terms of the velocity variations ΔV that must be provided. In what follows, we adopt the so-called

[6] This maximum occurs when the ground station lies inside the orbital plane of the spacecraft at the time of the overhead pass.

patched conics method to describe the interplanetary trajectories. In this approximation, one assumes that at any time only one central body acts on the spacecraft, which is then said to be inside the sphere of influence of this body.

3.5.1 The Hohmann transfer orbit

For strictly co-planar and circular planetary orbits, the optimal solution is a Hohmann transfer orbit. The Hohmann orbit is half of an elliptic orbit tangent to the orbit of the planet of departure (Earth) and to the orbit of the planet of destination. The semi-major axis of the Hohmann transfer orbit is $a = \frac{R_1+R_2}{2}$.

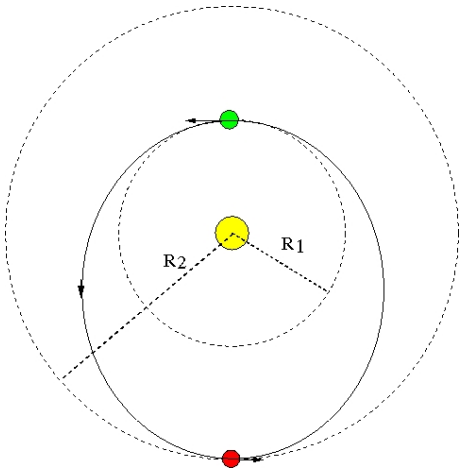


Figure 3.15: Schematic illustration of a Hohmann transfer orbit.

Table 3.1: ΔV over escape required for orbit insertion in the coplanar approximation and assuming $R_{\rm per}/R_c = 1.05$. For non-coplanar orbits, the amount of energy required increases substantially.

Planet	a (AU)	Period (years)	Synodic period (years)	i (°)	e	ΔV (km s^{-1})
Mercury	0.387	0.241	0.32	7.0	0.205	8.70
Venus	0.723	0.615	1.60	3.4	0.007	0.62
Mars	1.52	1.88	2.14	1.9	0.094	1.07
Jupiter	5.20	11.9	1.09	1.3	0.049	3.55
Saturn	9.58	29.4	1.04	2.5	0.057	4.24

The relative velocities at departure and at arrival with respect to the orbital velocity of the Earth and the planet of destination are respectively:

$$V_{\rm dep}^{\rm rel} = \sqrt{\frac{G\,M_\odot}{R_1}}\left(\sqrt{\frac{2\,R_2}{R_1+R_2}} - 1\right)$$

$$V_{\rm arr}^{\rm rel} = \sqrt{\frac{G\,M_\odot}{R_2}}\left(1 - \sqrt{\frac{2\,R_1}{R_1+R_2}}\right)$$

where $\sqrt{\frac{G\,M_\odot}{R_1}} = 29.8\,\mathrm{km\,s^{-1}}$ for the Earth.
The difficulty and cost of an interplanetary trip are measured by the velocity changes needed to reach the destination. The necessary effort is usually expressed as ΔV over escape. This corresponds to the difference between the velocity at pericentre (for a hyperbolic orbit) and the escape velocity. This is the velocity increment one has to provide to leave (respectively get into) a weakly bound orbit around the planet of departure (respectively arrival).

$$\Delta V = \left[\frac{V_{\rm esc}^2\,R_c}{R_{\rm per}} + V_\infty^2\right]^{1/2} - V_{\rm esc}\sqrt{\frac{R_c}{R_{\rm per}}}$$

where $R_{\rm per}$ is the distance at pericentre and R_c the radius of the planet of departure or arrival. The total ΔV (the sum of the various $\Delta\,V$ required in the manoeuvre) to get into an orbit around a specific target is listed in Table 3.1.

In a Hohmann transfer, one starts by firing the spacecraft engines to provide the acceleration needed to put the spacecraft on the elliptical orbit around the Sun. Upon arrival, the velocity must be modified at closest approach to the target planet so as to allow capture of the spacecraft into a weakly bound orbit around the planet. This is a critical component since the corresponding $\Delta\,V$ has to be provided entirely by the spacecraft propulsion system. As an alternative, when there exists an atmosphere, one can try to use aerocapture where the atmospheric drag is used to slow down the probe. The disadvantage is that this procedure has to be performed in a single pass, hence one needs to dissipate a power of several Megawatt per square meter. This technique is therefore very risky and has not been implemented to date.

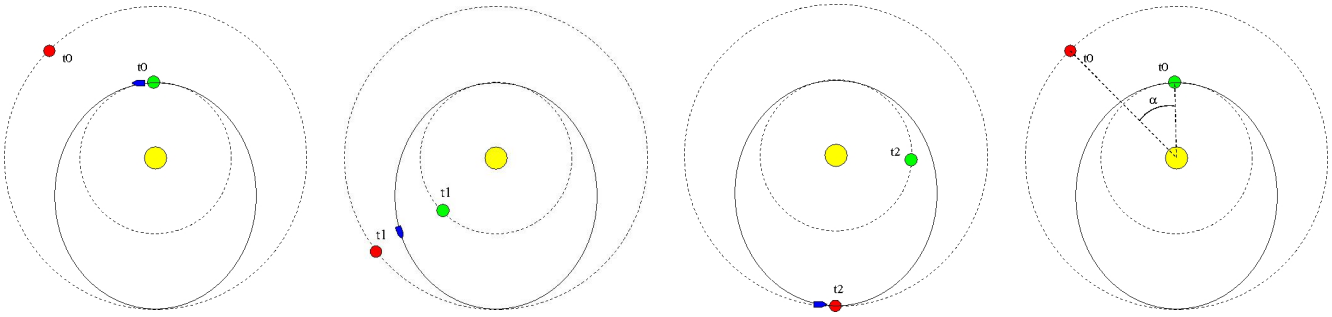


Figure 3.16: First three panels, from left to right: positions of the Earth, the target planet and the spacecraft at different moments of the cruise on a Hohmann transfer orbit. The last panel illustrates the constraint on the position of the target planet at launch.

When the spacecraft is launched on a Hohmann orbit, it is actually launched towards a direction that is empty. The target planet itself is moving while the spacecraft is cruising and the date of the launch has to be calculated in such a way as to ensure an encouter between the spacecraft and the planet once the spacecraft reaches the aphelion of the Hohmann orbit. The period of the Hohmann orbit (in years) is $(\frac{R_1+R_2}{2})^{3/2}$. To get from Earth to the destination takes half this time. This leads to the concept of **launch windows**. In fact, the Earth and the planet of destination have to be in a specific configuration (angle α, see Fig. 3.16) at launch.
Assuming circular orbits, a favourable configuration repeats with the synodic period: $P_{\rm syn} = \frac{P_1\,P_2}{P_1-P_2}$ for inner planets and $P_{\rm syn} = \frac{P_1\,P_2}{P_2-P_1}$ for outer planets (where all periods are expressed in years).
The Hohmann transfer orbit relies on the assumption of coplanar circular planetary orbits. Whilst non-zero eccentricities are usually not a major problem (although they impact the duration of the transfer and the choice of the optimal launch window), the main complications arise from situations where the inclination is different from zero. In fact, the planes of the orbits of the Earth and the target of destination intersect only along the line of nodes. Therefore, a Hohmann-type orbit is only possible if the target is at one of these nodes at the time of the rendez-

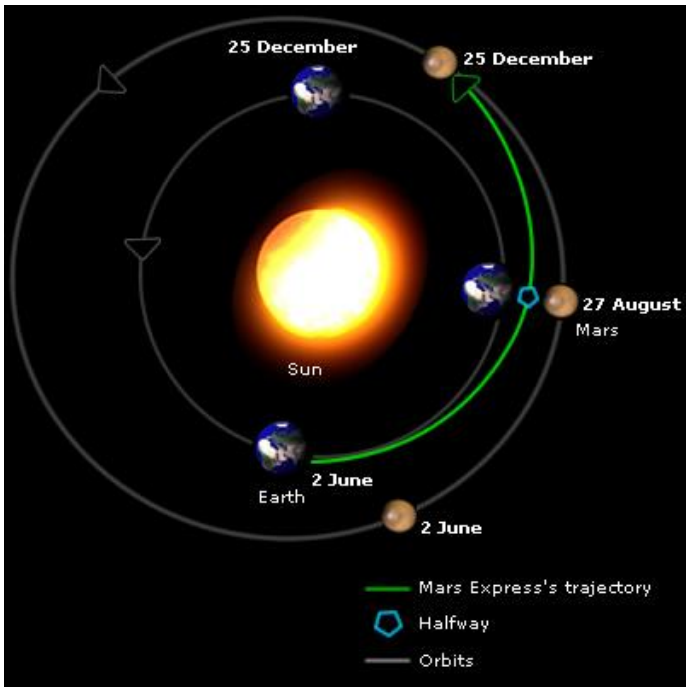


Figure 3.17: Illustration of the Hohmann-like trajectory of the Mars Express probe to planet Mars.

vous. Alternatively, one can change the orbital inclination of the spacecraft orbit, but this is a very expensive manoeuvre.

3.5.2 Gravity assisted manoeuvres

An elegant solution to procure the ΔV increments is by gravity assisted manoeuvres. The idea is to pass close to a planet to change the heliocentric velocity vector of the spacecraft. A spacecraft travelling around the Sun encounters a planet, or more exactly enters the sphere of influence of this planet. In the frame of reference of the planet, the trajectory of the probe is essentially a branch of a hyperbola. The energy is conserved in this frame of reference, but the velocity vector is rotated. When the orbital velocity of the planet is added back, to consider the result in the heliocentric frame of reference, there is a net effect on the heliocentric velocity.
In the gravity assist manoeuvre, the spacecraft's velocity vector in the frame of reference of the planet is rotated by an angle

$$2 \arcsin \left[\frac{1}{1 + \frac{R_{\mathrm{per}} V_r^2}{R_c V_c^2}} \right]$$

and the velocity in the heliocentric frame of reference increases at most by

$$\frac{2 V_r}{1 + \frac{R_{\mathrm{per}} V_r^2}{R_c V_c^2}}$$

The largest velocity increase is achieved by flying over the rear of the planet (i.e. when the pericentre occurs at 180° from the planet's velocity vector, see Fig. 3.18). Here R_c, R_{per}, V_r and V_c represent respectively the radius of the planet, the distance of the probe from the planet's centre at closest approach, the relative velocity of the probe (with respect to the planet) at infinity and the first cosmic velocity for the planet.

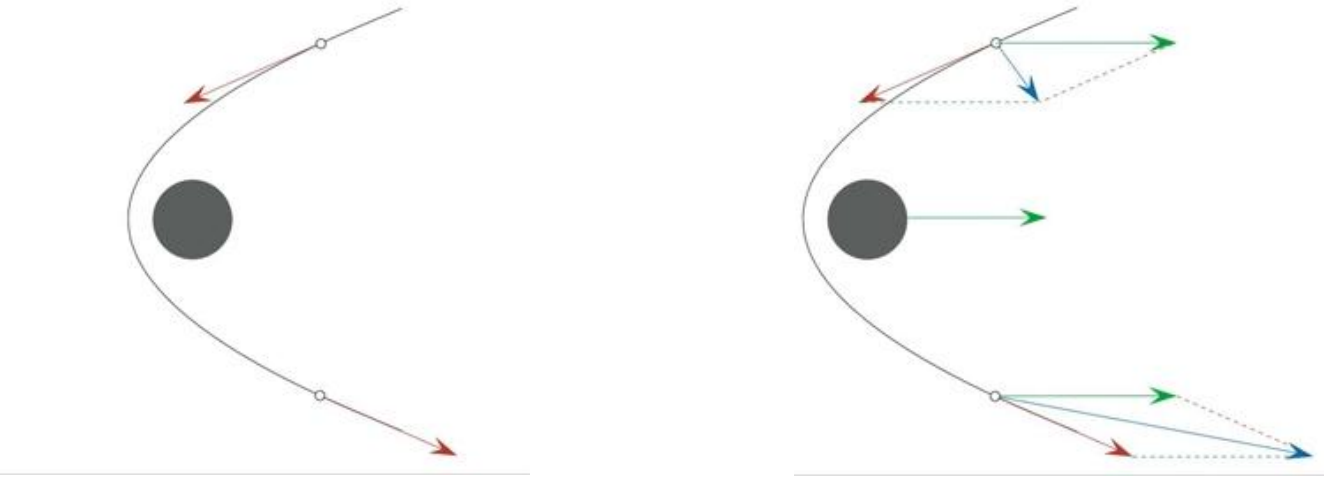


Figure 3.18: Left: the hyperbolic trajectory of the spacecraft in the frame of reference of the planet. Right: the planet's velocity is added back and the net effect on the heliocentric velocity of the probe becomes visible.

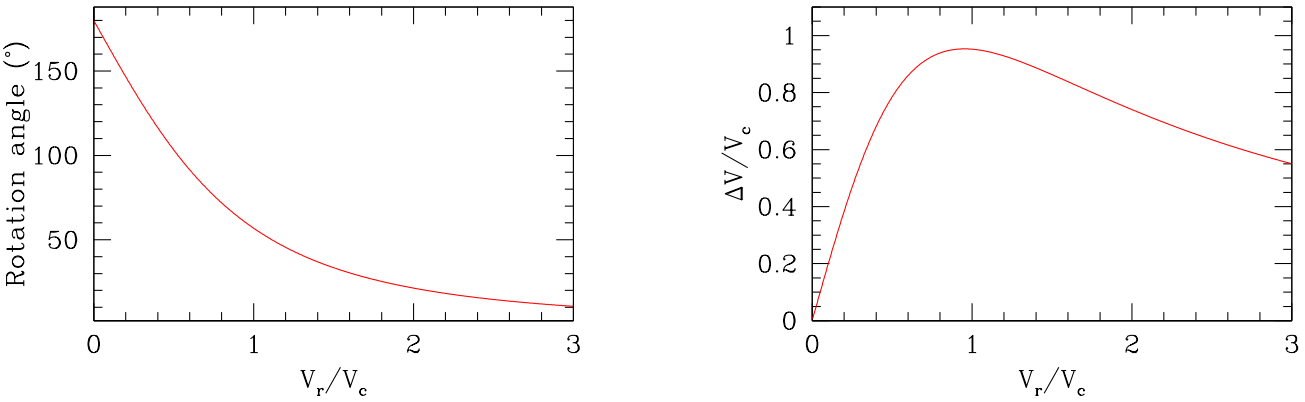


Figure 3.19: Left: the variation of the angle of the velocity vector during the fly-by manoeuvre as a function of V_r/V_c. Right: the velocity increase in terms of $V_c = \sqrt{\frac{G\,M_c}{R_c}}$ during the fly-by as a function of V_r/V_c and assuming a spacecraft flying over the rear side of the planet.

In a typical fly-by manoeuvre the distance of minimum approach to the planetary surface is of the order of 300 km or 1.05 – 1.1 R_c. Exceptions are Jupiter (minimum altitude of 30 000 km to prevent the spacecraft instruments from being exposed to the particle flux in the vicinity of the planet) and Saturn (minimum altitude of 80 000 km corresponding to the outer edge of the A-ring).

As an illustration, we consider the NEAR-Shoemaker mission launched on 17 February 1996. The spacecraft followed a Delta-V Gravity Assist trajectory. Shortly after a fly-by of asteroid Mathilde on 27 June 1997, a deep space manoeuvre on 3 July 1997 positioned the spacecraft in such a way to allow an Earth gravity assist swingby on 22 January 1998 at an altitude of 540 km. This swingby allowed to change the orbital inclination from 0.5 to 10.2°. Simultaneously the aphelion distance was reduced from 2.17 to 1.77 AU to match the orbit of asteroid Eros. A first attempt for orbit insertion about Eros failed because of an aborted ignition of the engine on 20 December 1998. The probe flew-by the asteroid on 23 December 1998 and a new manoeuvre was performed on 3 January 1999 to place NEAR on course for a rendez-vous in February 2000.

Table 3.2: Maximum ΔV that can be obtained in a gravity-assist manoeuvre involving different objects of the Solar System.

Object	V_c km s^{-1}	Minimum altitude (km)	Mass ($M_\oplus$)	Radius ($R_\oplus$)	ΔV (km s^{-1})
Mercury	3.1	200	0.055	0.383	2.9
Venus	7.6	300	0.815	0.949	7.4
Earth	7.9	250	1.000	1.000	7.7
Mars	3.6	200	0.107	0.533	3.4
Jupiter	45	30 000	318	11.2	37
Saturn	27	80 000	95.2	9.45	17
Moon	1.7	200	0.012	0.273	1.6

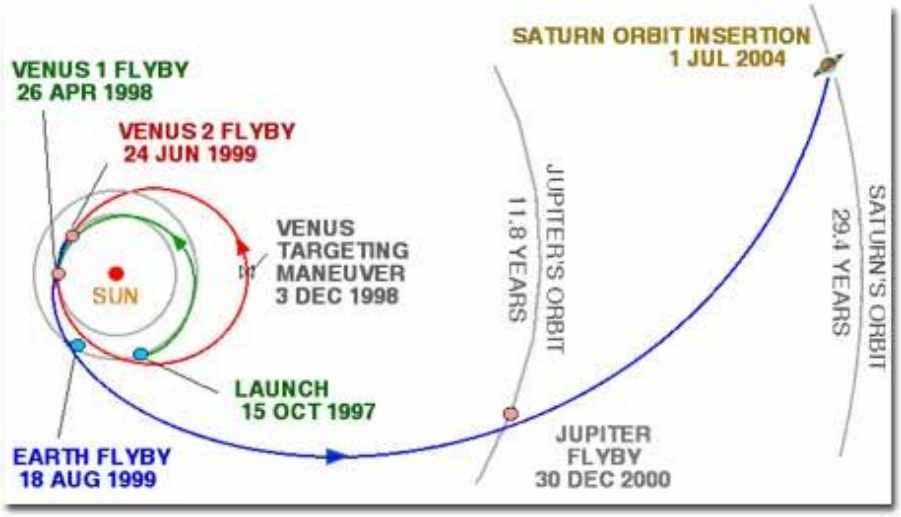


Figure 3.20: The trajectory of the Cassini-Huygens probe on its way to Saturn: the spacecraft flew by Venus (twice), the Earth and Jupiter, taking advantage of their gravity to gain the speed necessary to reach Saturn.

3.6 Low-energy transfers

For some destinations, as an alternative to the classical Hohmann transfer orbits and gravity assist manoeuvres, one can envisage a low-energy transfer that requires considerably less ΔV, but at the expense of a significantly longer transfer time.

Let us consider for instance the transfer of a probe from the Earth to the Moon. This is actually a 4-body problem, involving the Earth, the Moon, the Sun and the spacecraft. The basic idea is to take advantage of the complex (essentially chaotic) nature of the trajectories in such a 4-body problem. To first approximation we can treat this 4-body problem as the superposition of two restricted 3-body problems. Indeed, let us assume that we provide a sufficient ΔV to the probe to send it to the vicinity of the Sun-Earth L_2 point. If the spacecraft has an energy just above that of the L_2 point, it can move in a region around that point and depending on the initial conditions, it will either (1) orbit around the Lagrangian point (Lyapunov orbit), (2) wind on or off this Lyapunov orbit, (3) pass from one region of space to the other or (4) bounce back towards the original region. One thus needs to find a trajectory

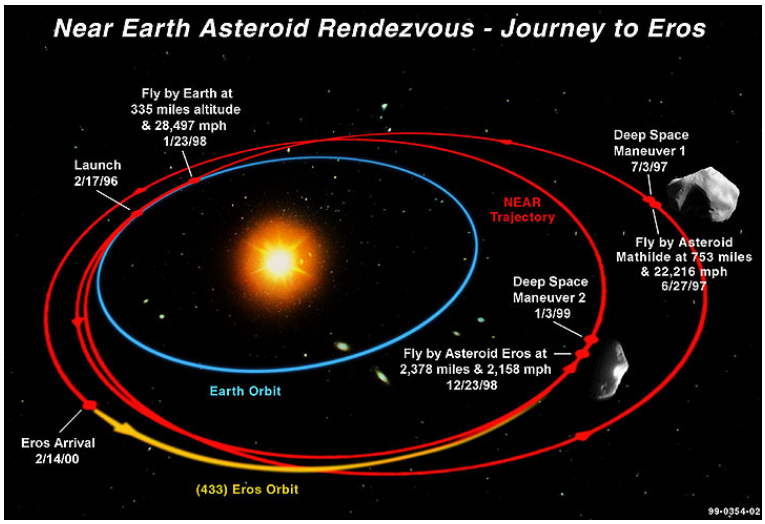


Figure 3.21: The trajectory of the NEAR-Shoemaker probe. Note that the orbit of asteroid Eros has an inclination of roughly 10° with respect to the ecliptic.

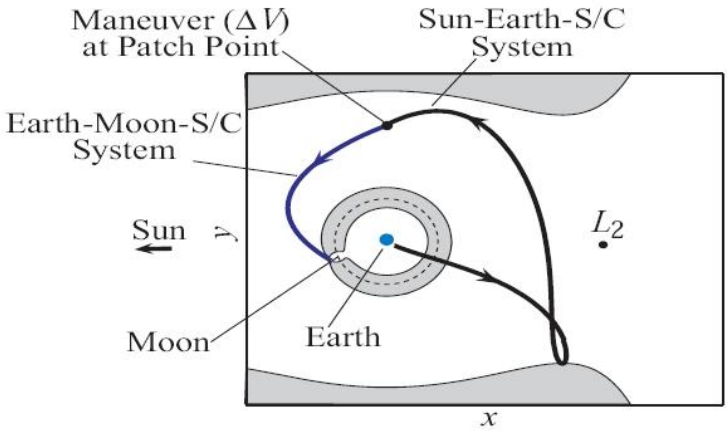


Figure 3.22: Schematic view of a low-energy transfer trajectory shown here in the rotating frame of reference of the Sun-Earth system. The spacecraft first moves towards the Sun-Earth L_2 point before bouncing back towards the Earth. After a modest ΔV manoeuvre, it settles on a trajectory that leads it to the Earth-Moon L_2 point where it transits to the sphere of gravitational influence of the Moon and gets captured.

of the latter type such that the spacecraft first moves towards the Sun-Earth L_2 point and bounces back towards Earth where it then enters the region near the Earth-Moon L_2 point and transits on a type (3) trajectory into the sphere of gravitational influence of the Moon to get eventually captured.
The first use of such a low-energy transfer trajectory was the rescue of the Japanese lunar mission Hiten (Muses-A) in 1991. Hiten was initially designed as an Earth orbiting spacecraft on a circumlunar orbit that should serve as a relay for a small lunar orbiter (Hagoromo) released by the Hiten spacecraft during one of its revolution. However the transmitter of Hagoromo failed and the mission could only be rescued by sending the Hiten spacecraft itself to the Moon. However the remaining propellant on board represented only 10% of the required fuel to reach a lunar orbit. Unlike the standard three-day Hohmann transfer to the Moon, it actually took three months to reach it.

Chapter 4

The environment of the spacecraft

Spacecraft are built on the ground in a clean and stable environment with temperatures near 20° C, moderate humidity around 50%,... The cleanrooms where the spacecraft are assembled and tested are characterized by a low level of environmental pollutants such as dust, airborne microbes, aerosol particles and chemical vapors. Cleanrooms are classified into categories according to the maximum number of particles above a given size that are allowed per unit volume of the air. For instance, in a "class 100" cleanroom there should be at most 100 particles of size $0.5\,\mu$m or more per cubic foot of air. This is equivalent to an "ISO 5" cleanroom, defined by the number of particles of size $\geq 0.1\,\mu$m that should not exceed $100\,000\,\mathrm{m}^{-3}$.

During transportation and integration into the shroud of the launcher, the spacecraft is submitted to a changing environment; the temperature and pressure vary a lot during transportation to the launch site aboard of an airplane, vibrations are an issue even before the launch, and humidity is especially critical for the launch sites nearest to the equator (Kourou, Cape Canaveral,... that are the most interesting in terms of ΔV). The launch itself is a very critical phase in terms of mechanical and acoustic vibrations. All these aspects need to be considered in the design of the spacecraft and require an extensive testing on the ground (especially as far as the vibrations are concerned).

Figure 4.1: Integration of ESA's *Herschel* infrared space observatory in the cleanroom of ESTEC (Noordwijk).

However, these are not the issues that we consider in this chapter. Here, we rather focus on the constraints that arise from the environment of the spacecraft either in orbit around the Earth or another planet or on its way across the Solar System. Indeed, once in space, the spacecraft is exposed to the effects of energetic particles (either coming from the Sun or the outer space or trapped by the magnetic field of the Earth or the planet it orbits), the rarefied atmosphere (either of the Earth or the planet it orbits), micro-gravity, the solar radiation, the solar wind, micrometeoroids and space debris,...

4.1 The particle environment

Solar cells, integrated circuits and sensors can be damaged by energetic particles. One usually refers to this as **radiation damage**, here the term *radiation* is used for energetic particles and ionizing electromagnetic radiation (such as γ-rays). The main problem is that these highly energetic charged particles can create thousands of free electrons when they interact with the material of the spacecraft and this high number of free charges can lead to malfunctions. For instance, the Telstar 401 communication satellite was lost in January 1997 as the result of a large coronal mass ejection event. The electronics on satellites must therefore be **hardened against radiation** to operate reliably.
Shielding does not systematically offer a good protection against radiation damage. In fact, the high-energy particles react with the material of the shielding leading to the formation of secondary particles that lose their energy through ionisation and bremsstrahlung. However, in some situations, it can be the only viable solution. This is the case for instance of the Juno mission that orbits Jupiter and is exposed to the harsh particle environment of Jupiter's radiation belts. In addition to radiation-hardened electrical wirings, Juno carries a 172 kg titanium vault (1 cm thick walls of Ti) that shields the flight computer and the main electronics of the instruments, reducing the level of radiation by a factor 800.

The fact that the particles in the plasma around the spacecraft are not neutral can produce electric discharges (leading to undesired switch on or off of the electronics, deterioration of the external thermal shielding, deterioration of the solar panels and/or the detectors). Single Event Upsets (SEUs) correspond to a change of state (e.g. ionization of a silicon component) caused by the impact of a high-energy particle on a sensitive element of an electronic component. These events are generally non-destructive, but can lead to a short circuit that can eventually destroy the circuit. For LEOs, SEUs occur most frequently over or near the South Atlantic Anomaly. The best way to circumvent this problem is to foresee **redundant electronics**. Electric charges are created by the impact of charged particles, but also by photo-ionization by the solar radiation. The part of the spacecraft that is exposed to the Sun will be charged positively compared to the part that is in the shadow. Therefore, to avoid short circuits, the outer surface of the satellite should be made as conducting as possible.
Radiation damage is not limited to malfunctions of the electronics. Indeed, the microscopic defects produced by high-energy particles and photons also result in a progressive alteration of the macroscopic properties (elasticity, optical opacity,...) of the material that makes up the spacecraft.

The term **space weather** is generally used to describe the changing particle environmental conditions in space. In the near-Earth space, there are essentially three sources of charged particles contributing to the space weather: the solar wind, the Van Allen belts and cosmic rays.

4.1.1 The solar wind and the Sun-Earth interaction

The solar wind is a stream of charged particles (a plasma mainly consisting of protons and electrons) ejected from the outer atmosphere (the corona) of the Sun. The driving mechanism of the solar wind is still not fully understood, but it is established that magnetic phenomena play a key role. Near the plane of the ecliptic, the so-called slow

solar wind usually has a speed in the range 200 – 600 km s^{-1}. Outside this plane, the fast solar wind can reach velocities of more than 800 km s^{-1}.
The Earth's magnetic field protects the atmosphere against the direct impact of the solar wind. The interaction between the solar wind and the Earth's magnetic field leads to the formation of aurora. Generally speaking, the solar wind is responsible for the shape of the Earth's magnetosphere. The magnetopause, i.e. the frontier between the solar wind and the magnetosphere results from the balance between the solar wind and the geo-magnetic field. The front end of the magnetopause is located at about 10 Earth radii towards the Sun, whilst the back end is located at more than 60 Earth radii in the opposite direction. The location of the magnetopause can change by several Earth radii as a result of the fluctuations of the solar wind and this can lead to situations where the geostationary orbit is directly exposed to the solar wind.

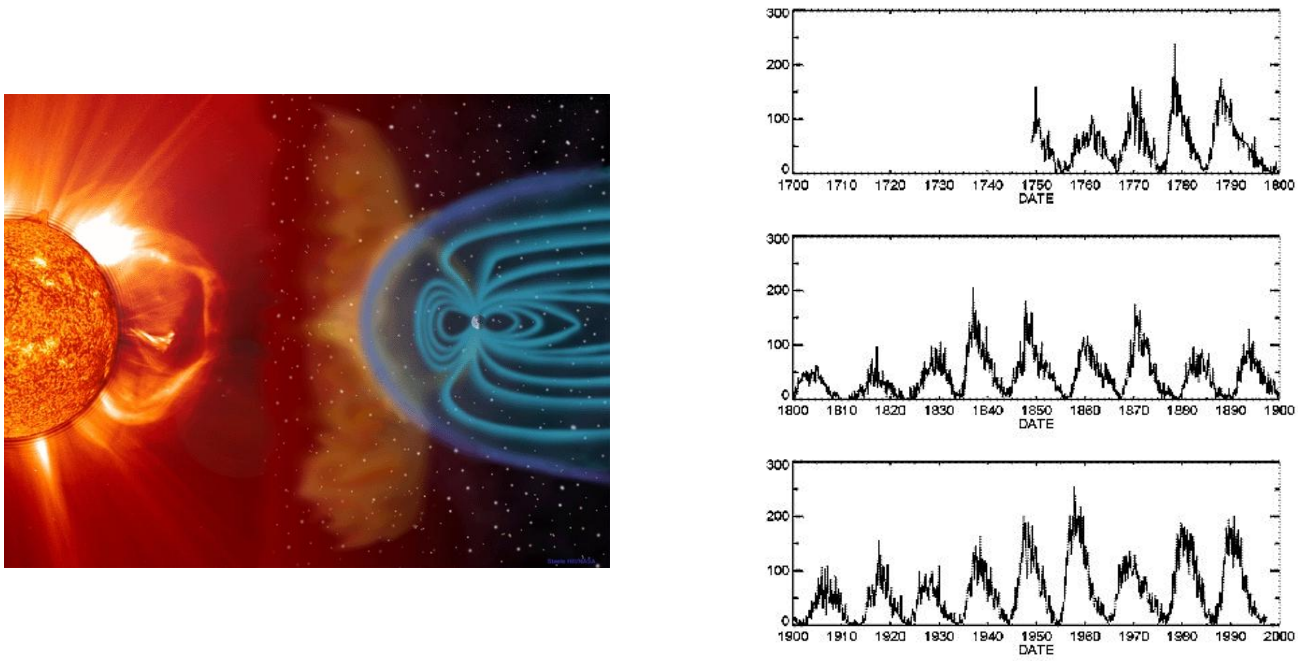


Figure 4.2: Left: Schematic view of the solar wind interacting with the Earth's magnetosphere. Right: the solar activity cycle illustrated by the number of sunspots as a function of time.

The solar wind is sometimes disturbed by coronal mass ejection (CME) events. These events are associated with sunspots and manifest themselves by sudden increases of the brightness of a small part of the chromosphere or the lower corona near a sunspot. The X-ray and chromospheric Hα emission increases significantly and particles (mainly protons and α particles) accelerated to energies of about 50 GeV are emitted. After the initial rise, both the radiative and the particle flux decay roughly exponentially. When the material ejected during a CME reaches the Earth, it distorts the magnetosphere and can lead to geomagnetic storms (strong auroral activity). The X-rays travel at the speed of light and reach the Earth after 8 minutes, whilst the particles take much longer (between a few hours and a few days). This allows to take some protective measures (move the crew of a manned spacecraft to the section that has the best shielding, switch off sensitive instruments on board of unmanned satellites,...).
Since these eruptions are associated with sunspots, their frequency is modulated by the 11 year sunspot cycle, also called the solar magnetic activity cycle. The solar magnetic activity cycle is the main engine behind all solar phenomena driving space weather. It modulates the number of sunspots, the flux at high energies from the far UV to X-rays, the frequency of flares, CME events and other solar eruptive phenomena. Solar minimum and maximum

refer respectively to the epochs of minimum and maximum sunspot numbers. For instance, solar flares are about 50 times more frequent at solar maximum than at minimum. The solar activity cycle also indirectly modulates the flux of galactic cosmic rays entering the Solar System. The average duration of the cycle is 11.1 years[1], although this duration can differ quite substantially from one cycle to the other. Important variations in the amplitude of the cycle have been observed. In the period between 1645 and 1715, virtually no sunspots were observed. This period is known as the Maunder minimum and coincides with an episode of very cold weather in Western Europe.

Solar flares are divided into categories (B, C, M and X) according to their X-ray flux I (in the wavelength domain between 1 and 8 Å). If $I < 10^{-6}\,\mathrm{W\,m^{-2}}$, the flare is of category B, for $10^{-6} \leq I < 10^{-5}\,\mathrm{W\,m^{-2}}$, the flare is of category C, for $10^{-5} \leq I < 10^{-4}\,\mathrm{W\,m^{-2}}$, the flare is of category M, and finally the strongest flares with $I \geq 10^{-4}\,\mathrm{W\,m^{-2}}$ are of category X. The latter events can represent a serious threat for manned space missions outside the Earth's magnetosphere (such as a future mission to Mars).

4.1.2 The Van Allen radiation belts

The Van Allen radiation belts are tori of plasma around the Earth, held in place by the geomagnetic field. The larger outer belt consists mainly of high-energy (0.1 – 10 MeV) electrons that are trapped by the magnetic field lines. It extends from an altitude of about 13 000 to 65 000 km above the Earth's surface and has its maximum between 14 500 and 19 000 km. The inner belt extends over a range of altitudes between 700 and 10 000 km above the surface of the Earth and contains a high concentration of energetic (> 100 MeV) protons. The inner belt is believed to be populated by the decay of neutrons that result from the collision of cosmic rays with the upper atmosphere, whilst the outer belt is most likely populated by particles from the solar wind captured by the Earth's magnetosphere.
The Earth's magnetic field is roughly dipolar with an equatorial field strength of 0.3 G and a polar field strength of 0.6 G. However, the dipole is off-centre by about 430 km towards the South-East of Asia and this leads to the depression in the geomagnetic field known as the South Atlantic Anomaly (SAA) where the inner Van Allen belt reaches deeper towards the Earth. For a given altitude, the radiation intensity is higher in the SAA than elsewhere. Therefore, LEO satellites, especially those with $35^\circ \leq i \leq 60^\circ$ are exposed to higher radiation doses each time they cross the SAA. For the International Space Station ($i = 51.6^\circ$), an extra shielding has been foreseen. Actually, the SAA slowly drifts towards the west at a rate of about $0.3^\circ\ \mathrm{yr}^{-1}$.

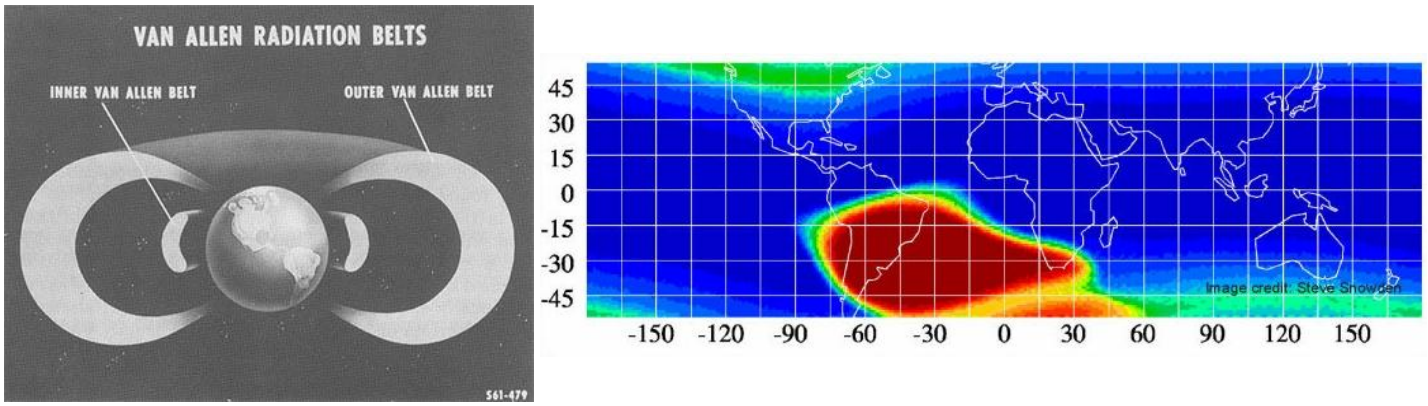


Figure 4.3: Left: Schematic view of the Van Allen radiation belts. Right: the South Atlantic Anomaly.

The Earth is not the only planet of the Solar System that has a magnetic field. Jupiter has a very powerful magnetic field that collects a large number of particles in the planet's radiation belts. The magnetic field of Jupiter is quite

[1] Actually, the solar magnetic cycle corresponds to a reversal of the polarity of the Sun's magnetic field on a period of about 22 years.

strong, and the fluxes of energetic electrons and ions are among the highest found in the Solar System. Part of the particles are in a plasma torus created by Io. In fact, the latter moon has a very strong volcanic activity and the gas particles are captured by Jupiter's magnetosphere. An important issue concerns the moon Europa which is likely to be the target of future space missions. Europa is located in the inner magnetosphere of Jupiter (at a radial distance of ~ 9.5 planetary radii), a region populated mainly by plasma derived from the Io plasma torus. The plasma there consists of protons, as well as oxygen and sulfur ions, along with their electrons. The plasma is dragged around Jupiter by the magnetic field as the planet rotates. Since the magnetic field rotates at a faster rate than Europa revolves around Jupiter, the plasma hits the moon with a relative velocity of about $120\,\mathrm{km\,s^{-1}}$ on its orbital trailing hemisphere.

4.1.3 Cosmic rays

Cosmic rays are energetic particles originating from outer space. Cosmic rays have energies up to 10^{20} eV and consist mainly of protons (87%) and α particles (12%). Their energies span about 14 orders of magnitude, reflecting the variety of sources from which they originate. Particles with energies up to 10^{14} eV are likely to be accelerated by shocks in the remnants of supernova explosions. However, the origin of particles with even higher energies (up to 10^{20} eV) remains mysterious. This is because these particles are deflected by the Galactic magnetic field. As a result, particles with energies above 10^{10} eV have an isotropic distribution of their directions when they reach Earth. The trajectories of lower energy particles are further deflected by the Earth's magnetic field.
The solar activity reduces the flux of cosmic rays. During a solar eruption, the plasma ejected by the eruption carries its own magnetic field and when this plasma surrounds the Earth, it acts as a magnetic screen that deviates the low-energy cosmic rays. The solar wind has thus the effect of decelerating the incoming particles and repelling particles with energies up to 10^9 eV. The cosmic ray density in the inner Solar System is therefore anti-correlated with the overall level of solar activity.
In space, the yearly radiation dose is a factor 200 – 500 higher than on the ground. This is a serious issue for future manned interplanetary exploration missions. Indeed, high-energy cosmic rays can damage DNA, hence increasing the risk of cancer, neurological disorder,... Actually, radiation from coronal mass ejection is easier to shield against than cosmic rays and a possible solution is therefore to launch such a manned interplanetary mission during an epoch of maximum solar activity.

4.2 The thermal environment

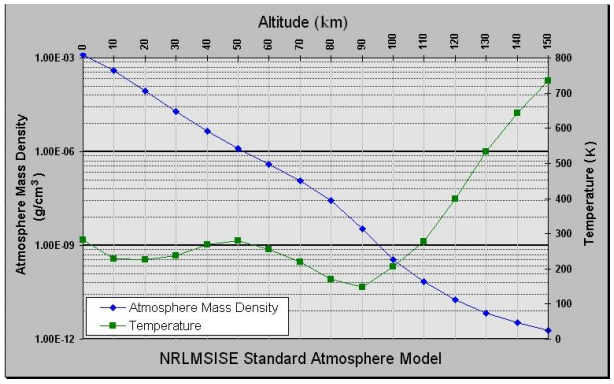
Whilst the kinetic temperature of the gas in the upper layers of the Earth's atmosphere increases strongly, the atmosphere's density decreases strongly with altitude. As a result, the mean free path of the molecules becomes larger than the satellite's dimensions. For instance, beyond altitudes of 160 km, the mean free path exceeds 50 m. Therefore neither conduction nor convection play a role in the heat exchange between the satellite and its environment, which is solely based on radiative processes.
The electromagnetic spectrum of the Sun can be approximated by a black body of temperature around 5800 K. The radiation produced by the Sun hence displays its maximum near 5000 Å. Whilst the far-UV (wavelengths below 1300 Å) radiation represents only a tiny fraction (about 2%) of the total light emitted by the Sun, it can increase quite significantly during solar maximum. Due to its ionizing effect, this radiation can degrade the coating of a satellite.
Integrated over all wavelengths, the solar radiation corresponds to a flux of $1370\,\mathrm{W\,m^{-2}}$ at the orbit of the Earth. This flux decreases as r^{-2}, where r is the distance from the Sun. For instance, at the orbit of Mercury, the solar flux amounts to $9145\,\mathrm{W\,m^{-2}}$, whilst the flux decreases to $50\,\mathrm{W\,m^{-2}}$ near the orbit of Jupiter. Actually, the distance of minimum approach to the Sun for a spacecraft is essentially set by thermal considerations. At the radius of the

orbit foreseen for Solar Orbiter (0.28 AU = 42 $R_\odot$), the power received by the spacecraft amounts to 17.5 kW m^{-2}.

Since radiation is the only process that plays a role in the heat exchanges with the environment, one can easily compute the equilibrium temperature of a spherical black body (of radius R) at the distance of $r = 1$ AU from the Sun. The incident flux amounts to

$$(1 - c_A)\,\pi\,R^2\,\frac{L_\odot}{4\,\pi\,r^2}$$

where c_A is the albedo coefficient (see below) of the sphere, which is equal to zero for a black body. The amount



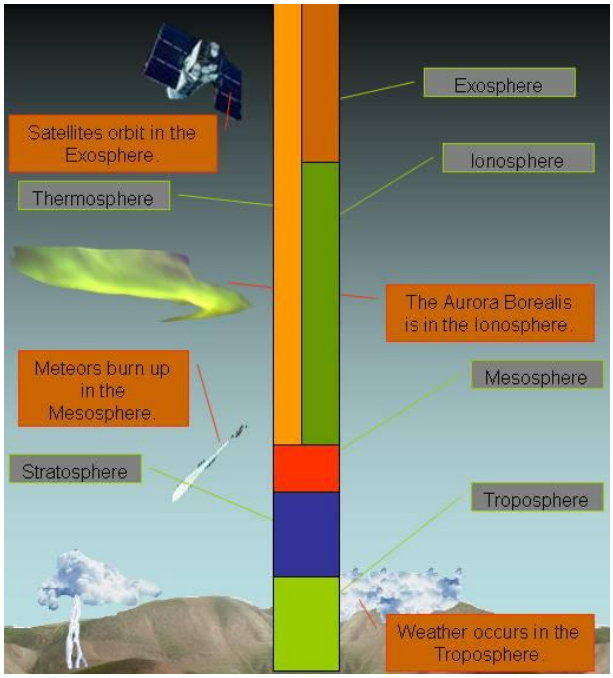
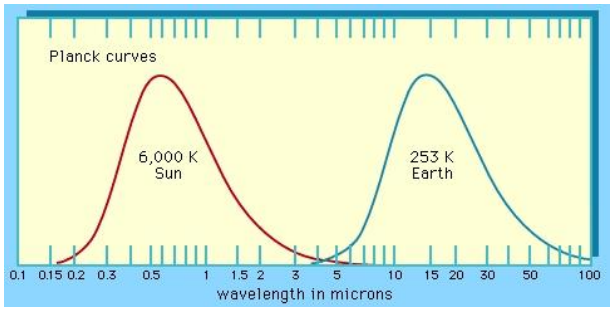


Figure 4.4: Left: the density and temperature profile of the Earth's atmosphere as a function of altitude. Right: schematic view of the various layers of the Earth's atmosphere.



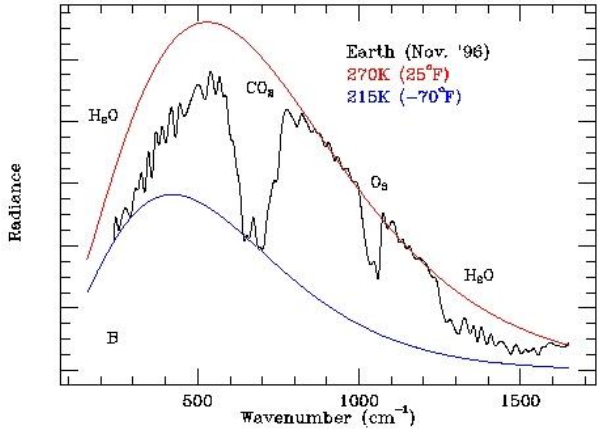


Figure 4.5: Left: schematic comparison between the spectral distributions of the Sun and the Earth. The fluxes are not to scale. Right: the Earth's spectrum as seen from space.

of flux radiated by the sphere is then

$$4\,\pi\,R^2\,\alpha\,\sigma\,T^4$$

where $\sigma = 5.67 \times 10^{-5}$ erg cm^{-2} K^{-4} is the Stefan-Boltzmann constant and α is the emissivity relative to a black body (hence $\alpha = 1$ here). In this way, one finds that the temperature of equilibrium of a spherical black body amounts to 288 K (15° C) at 1 AU from the Sun, whilst it would be 465 K near the orbit of Mercury and 123 K at the distance of Jupiter.

In addition to the solar illumination, the light reflected by the Earth (or another planet around which the spacecraft orbits) is also relevant. This is quantified by the albedo coefficient c_A defined as the ratio between the reflected light and the incident electromagnetic radiation. One usually makes the first order approximation that the albedo is wavelength independent or refers to the light integrated over the entire range of wavelengths. Typical albedo values are 0.8 – 0.9 for a layer of fresh snow, 0.4 – 0.8 for clouds, about 0.25 – 0.30 for a desert, 0.05 – 0.10 for a forest and around 0.05 – 0.10 for the ocean (depending on the angle of incidence of the Sun light). On average, the Earth's albedo is about 0.30. For Venus, the average albedo amounts to 0.75, whilst its value is only 0.12 for the Moon.

The reflected light from the Earth is not to be confused with the Earth's own thermal emission which results from the fact that the surface of the Earth is heated up by the Sun's light. Whilst the incident and reflected Sun light have their maximum in the optical wavelength domain, the radiation emitted by the Earth itself peaks in the infrared domain.

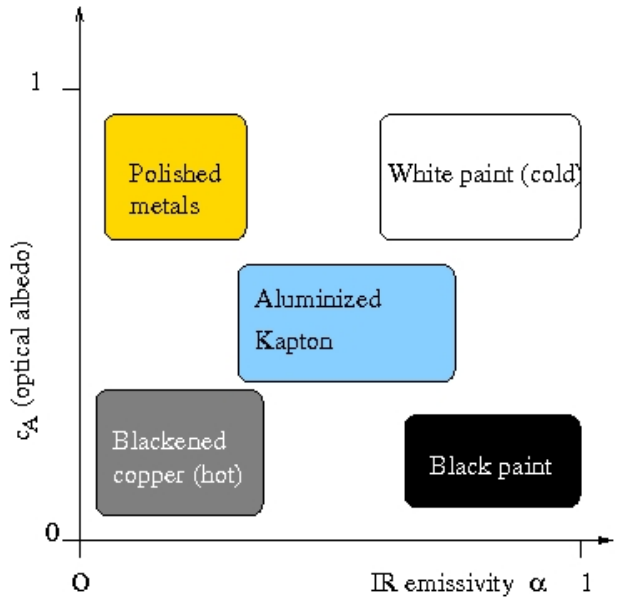


Figure 4.6: Schematic view of the role of coatings in passive thermal control of a spacecraft.

Those parts of the spacecraft that look permanently at the deep space (such as mirrors of telescopes) can become very cold (the thermal background of the Universe has a black body temperature of 3 K) and it is therefore mandatory to control their temperature. Thermal control of a satellite is an important topic to make sure that the detectors and electronics are operated at an appropriate temperature, to avoid differential thermal expansion of optical components, etc. To achieve this, one can use either passive or active thermal control. Passive cooling relies on various sorts of coatings and offers usually a rather cheap and low-weight solution (see Fig. 4.6). A white paint is used to produce a cold coating (low absorption efficiency, but high emissivity); black paints are used for situations where both the absorption and emission efficiencies should be near unity; finally a hot coating is obtained when

the absorption exceeds the emissivity (aluminium, polished gold, black chrome). A major problem with passive thermal control is the progressive degradation of the coating. This can result from radiation effects (see above), micrometeorites (see below) or highly reactive atomic oxygen that is the main component of the ionosphere[2] where the LEO satellites evolve. Active thermal control is required when passive techniques are not sufficient. This can include (electric) heaters as well as cooling systems (e.g. cryogenics for IR space observatories).
Thermal considerations are also extremely important for landers. Indeed, as a spacecraft approaches a body surrounded by an atmosphere, it encounters the effect of a roughly exponentially increasing atmospheric density. While the nature of the interaction between the vehicle structure and the atmosphere changes with altitude and hence atmospheric density, the net effect is a substantial heat transfer to the surface of the vehicle. The heat exchange equation for a probe of radius R can be written

$$m\,c_p\frac{dT}{dt} = f_\odot\,(1-c_A)\,\pi\,R^2\,F_\odot + f_{\rm Planet}\,(1-c_A)\,\pi\,R^2\,F_{\rm Planet} - 4\,\pi\,R^2\,\alpha\,\sigma\,T^4 + 4\,\pi\,R^2\,\kappa\,(T_a - T)$$

here κ is the convective heat transfer coefficient and T_a is the ambient temperature, $f_\odot$ and $f_{\rm Planet}$ are the geometric viewing factors between the spacecraft and the Sun or the planet respectively, $F_\odot$ is the direct solar flux (attenuated by absorption from the planet's atmosphere) and $F_{\rm Planet}$ the total flux (thermal emission plus reflected solar light) from the planet. In a thick atmosphere, convection dominates and T tends towards T_a.

4.3 Micrometeorites and space debris

The natural debris environment of the Solar System contains meteoroids with dimensions ranging from a few microns up to a few meters. The vast majority of these objects are tiny (less than 1 mm), but their velocities with respect to a spacecraft can be substantial, typically of order $20\,\mathrm{km\,s^{-1}}$. A major component of this population comes from comets that leave a tail of meteoroids behind them when they approach the Sun. During its revolution around the Sun, the Earth regularly crosses several clouds of micrometeoroids left behind by comets. It is estimated that every year about 10 000 – 20 000 tons of natural meteorites fall on the Earth[3]. On the Soviet space station Salyout, about one impact of a micro-meteorite was registered every six days. However, these natural micrometeoroids are usually not considered as a high risk for spaceflight, although the long term exposure of a spacecraft to micrometeorites can degrade its coating.
The situation is quite different for artificial debris that are usually more massive and denser (thus more dangerous) and remain in orbit around the Earth. About 94% of the artificial objects in orbit around the Earth are space junk[4]. The launch of a payload into space generates considerable quantities of space debris. As an example, we consider the Sputnik 1 satellite. Its mass was 84 kg, but the launch also put 6500 kg of the Semyorka main stage and a 100 kg protective fairing onto the same orbit. The payload thus amounted only to 1.3% of the injected mass. Moreover, the satellite worked only for 21 days and spent 3/4 of its in-orbit life as a debris.
For sizes above 1 mm, the artificial debris dominate over natural meteoroids. Since October 1957, more than 4600 launches placed some 6000 satellites into orbit. About 7% of these are either on HEOs or interplanetary trajectories, but the majority of the space junk is in LEO or GEO. The mass of man-made debris in LEO is estimated to about 6000 tons, for HEO and GEO, these numbers are about 1800 and 1600 tons, respectively. Most of the mass is concentrated in larger objects (more than 10 cm diameter) that can be tracked from the ground.

[2] The ionosphere is formed by ions produced by the solar radiation upon the atmosphere and extends from about 50 – 70 km up to about 2000 km.

[3] This is the estimated mass that reaches the ground, most of the mass of the meteorites is actually consumed during their descend across the Earth's atmosphere.

[4] About 60% are fragments from in-orbit explosions or collisions, 16% are dysfunctional satellites, 11% are orbital rocket bodies, 7% other mission-related objects, only 7% are active satellites.

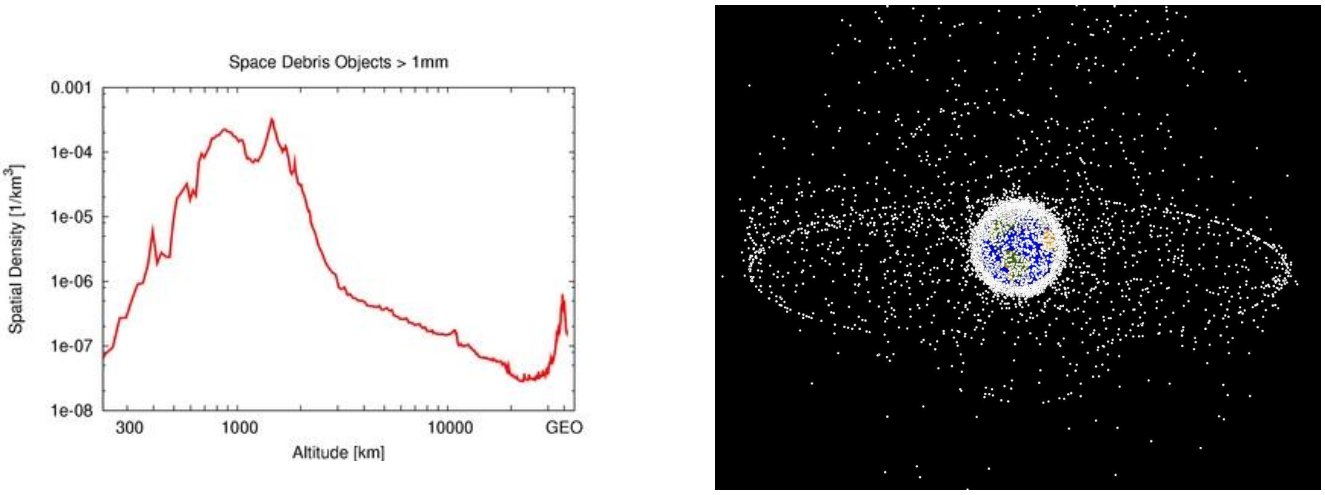


Figure 4.7: Left: Spatial density of space debris by altitude following the ESA MASTER-2001 model. Right: schematic view of the distribution of debris around the Earth.

Space debris include many different categories of objects, spanning a wide range of sizes and velocities, from spent rocket stages and obsolete satellites to explosion fragments, paint flakes... Objects with a large surface/mass ratio are most efficiently affected by the atmospheric drag, but many space debris are rather compact. A particular danger comes from the so-called yo-yo wheels that were used on previous generations of spin-stabilized satellites. Yo-yos are small massive wheels that are spun up to make a satellite rotate in the opposite sense by conservation of angular momentum, and are then ejected while the satellite continues spinning.
Given the current density of satellites in orbit around the Earth, atmospheric drag and re-entry alone are no longer sufficient to clean the orbit. The typical lifetime of a debris strongly depends upon its orbiting altitude. Indeed, at altitudes where atmospheric drag is negligible, lunar perturbations and solar wind drag are the only perturbing forces that can lower the altitude of the orbit. As these are very small forces, this process takes a very long time. For circular orbits at an altitude of 400 km, the typical lifetime of a debris is about 1 year, whilst it reaches 200 years at 800 km altitude and millions of years for the GEO! For the highly eccentric geostationary transfer orbit (perigee altitude of 200 km), the lifetime still amounts to about 10 years.
The estimated number of close encounters (less than 5 km distance) between spacecraft and debris is nowadays about 1600 per week. There thus exists a serious risk for collisions between satellites and space debris which translates into about 1.5 avoidance manoeuvre per year and per spacecraft. A collision with a debris of size less than 0.1 mm will usually only erode the surface of the spacecraft, whilst a debris of size up to 1 cm can lead to perforation and hence significant damage depending on the equipment that is hit. Very significant damage up to the destruction of the spacecraft result from collisions with debris of larger size. The first known collision between an active spacecraft and a debris occurred in 1996 when the French military satellite Cerise was hit by a small catalogued debris produced in the explosion of an Ariane 1 rocket stage ten years earlier. A boom was torn off and the Cerise satellite was severely damaged. The most severe accident so far occurred in February 2009 when an operational telecommunication satellite (Iridium 33) collided with a deactivated satellite (Kosmos 2251) at an altitude of 790 km above sea level. Both satellites were destroyed in this collision and two large clouds of at least 500 debris were created. Such collisions can generate thousands of debris thus leading to a runaway process that becomes a threat to space activities in the long term. Indeed, six years after the event, 75% of the fragments were still in the catalogue. Space debris also represent a serious threat for astronauts. Indeed, the suits for extra-vehicular activities resist against impacts of debris with a size below 1 mm, but larger debris can have fatal consequences. To reduce the consequences of debris impacts, the ISS features a so-called Whipple Bumper shield. This is a thin sheet of aluminium separated from the spacecraft by an open space. The sheet causes small pieces of debris to

explode when they strike the surface thereby reducing their energy and consequences when they hit the actual skin of the spacecraft.

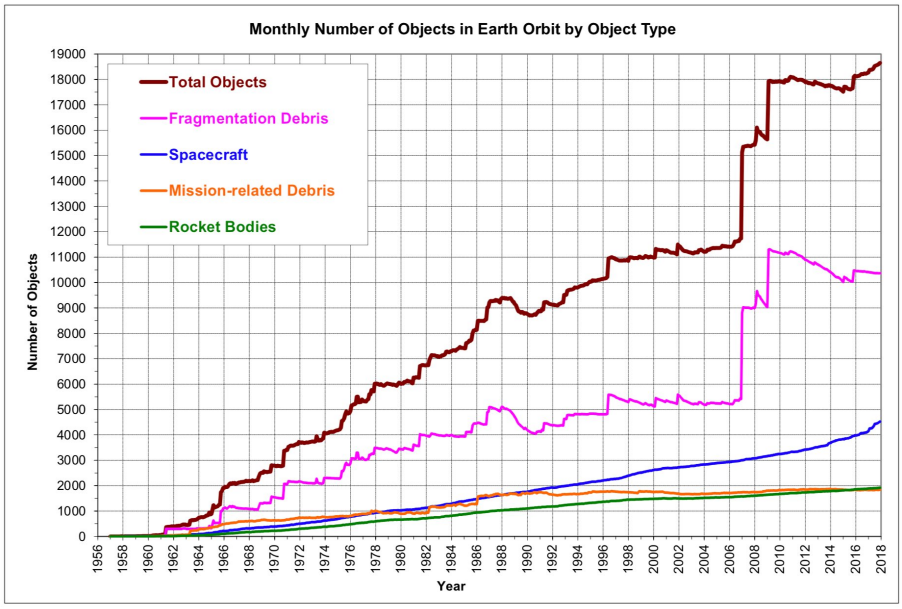


Figure 4.8: Evolution of the number and type of catalogued debris of size larger than 10 cm.

In 1978, the NASA scientist Donald Kessler described a scenario, known as the Kessler syndrome, where the collisions between space debris and spacecraft lead to an exponential increase of the number of debris, hence increasing the probability of further collisions and eventually rendering space exploration impossible (especially at altitudes between 800 and 2000 km) for hundreds of years.

Several measures are taken to reduce the risks associated with space debris: one of these is the passivation of spent rocket stages by releasing the residual fuel to decrease the risk of an in-orbit explosion that could generate a huge number of new debris[5]. At the end of a mission, all latent fuel reservoirs of the spacecraft need to be vented and all batteries have to be discharged to avoid a risk of accidental post-mission explosion. For LEOs below 2000 km it is foreseen to de-orbit satellites that are no longer operational by bringing them to an orbit where the atmospheric drag will cause them to re-enter the atmosphere within 25 years of mission completion. Concerning the geostationary orbit, it is foreseen to move satellites that are no longer operational to a graveyard orbit 300 km above the GEO. We note however that even satellites which have been moved to a graveyard orbit may again become dangerous due to the action of solar radiation pressure which might render their orbit more eccentric and ultimately lower the altitude of the perigee to make it cross the GEO, thus threatening active satellites.

The above measures mostly concern debris mitigation. They might not be sufficient to prevent an exponential increase of the number of debris though. Indeed, long-term debris environment projections indicate that even a complete stop of launch activities could not prevent the onset of the Kessler syndrom due to runaway collisions at some altitudes. Several concepts for active debris removal are thus currently under development (Mark & Kamath

[5]In 1973, the (at the time of the accident) un-manned Soviet space station Salyut 2 was damaged by the explosion of the upper stage of the Proton-K rocket, a few days after launch. When the debris hit the station, the solar panels were torn off and the hull was perforated, leading to depressurization.

2019). They include

- collector spacecraft that approach the debris and use a robotic arm, a net, or other devices to capture it,
- high-power lasers, either on the ground or in space, that sublimate the (small) debris,
- ion-beam devices that emit a beam of plasma towards a debris with the goal to produce a thrust on the debris,
- long (several km) conductive theters attached to a debris to let the Earth magnetic field and the Lorentz force pull the debris towards the atmosphere,
- solar sails, notably to remove debris from the GEO,
- microsatellites docking onto a dysfunctional satellite and using their own propulsion system to alter its orbit,
- efficient exploitation of perturbations of the debris' orbit by the Moon, the shape of the Earth, etc., to alter the orbital parameters, notably via a chaotic growth of the eccentricity,
- more exotic solutions, like a platform that ejects a foam onto the debris that subsequently grows in volume, leading to an increase of the drag, etc.

All these solutions currently have a low maturity level and further developments are needed before they can be used in practice.

About 20 000 orbiting objects (including about 800 operational satellites) of size larger than 10 cm are currently tracked and catalogued, 38% of them are dysfunctional satellites or upper stages. The total number of debris larger than 10 cm is estimated at 29 000. But, the number of debris larger than 1 cm is estimated to be much higher, around 600 000. Actually, some space debris were deliberately created! In January 2007, a Chinese anti-satellite weapon test (against a Chinese weather satellite) created more than 2000 pieces of about 10 cm diameter at an altitude of 865 km. The number of pieces larger than 1 cm resulting from this 'test' is estimated to be more than 35 000. The in-orbit lifetime of the debris created during this event is expected to be of order 35 years. In April 2011 and January 2012, debris from this test threatened the ISS.
Currently, the large and massive, inactive Envisat satellite, drifting at 785 km altitude, is a potential trigger of a runaway collision scenario. If no measure is taken, Envisat will remain in orbit for about a century! Another possible trigger is the Japanese X-ray observatory Hitomi. Hitomi was launched in February 2016 on a 575 km altitude LEO. About six weeks later, the satellite experienced an attitude anomaly. Due to an inadequate programming, the on-board attitude control system considered that the satellite was spinning rapidly though it was not. The activation of the reaction wheels and the thrusters coupled with erroneous values of the spacecraft's moments of inertia following the deployment of an optical bench structure, led to the satellite spinning out of control (one rotation every 5.2 s) and to the break-up of the satellite into several pieces. The main body of the Hitomi satellite will remain in space until solar activity will reach its next maximum in the mid 2020's. Space debris can also be generated as a result of aging of the spacecraft. This happened in August 2017 to the Indonesian telecommunication satellite Telkom-1 which broke apart on the geostationary orbit.

The space agencies use mathematical models to assess the risk of a collision with a micrometeoroid or a space debris. These models are calibrated using data from solar panels that have flown in space and were brought back to Earth to be examined in a laboratory (such as the solar panels of the Hubble Space Telescope). These data allow to predict the expected flux of debris according to their size. The chemical analysis of the residues in the impact features allows discriminating between natural micrometeorites and space debris. It must be stressed though that

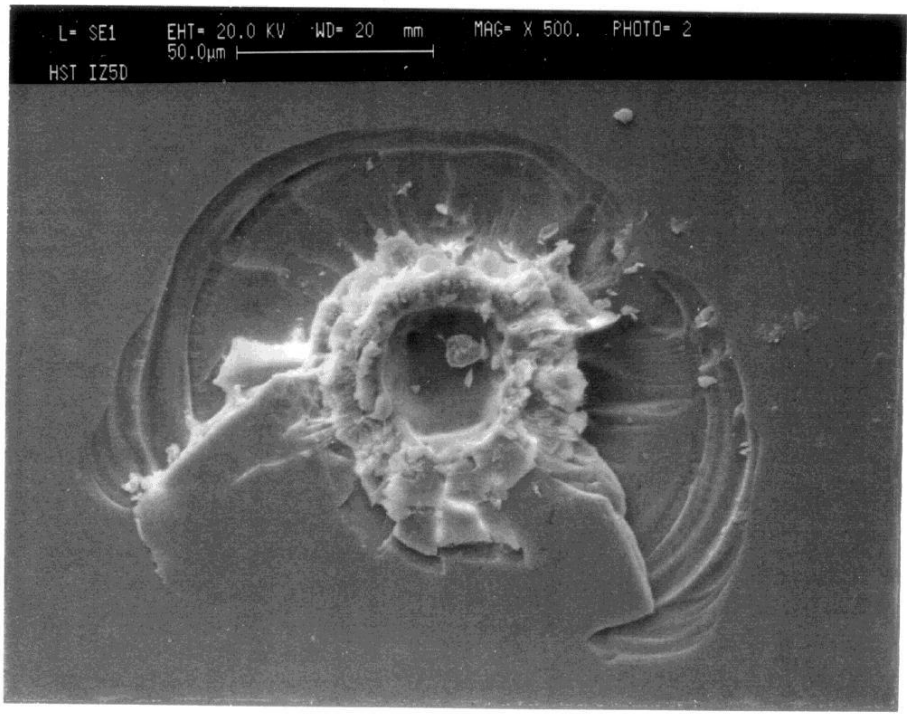


Figure 4.9: The impact of a micrometeorite on a solar panel of the HST as seen under the microscope.

the predictive power of such models is very sensitive to the uncertainties on the orbits of the debris. Indeed, calculations had predicted that the Iridium 33 and Kosmos 2251 satellites should have missed each other by 584 m.

The probability of accidental collisions between spacecraft that could lead to a runaway process obviously increases with the number of objects in orbit and we are currently far away from a complete stop of launch activities. In 2018, there were 2100 operating satellites catalogued by the Satellite Industry Association. This is 300 more than in 2017. Part of this increase is due to the *New Space* phenomenon and could lead to an inflation of the number of satellites orbiting the Earth. Indeed, several private companies have ambitious projects to launch huge constellations of satellites for world-wide high performance internet connections. The most ambitious one, the Starlink constellation designed by Space-X should start with 1 600 LEO satellites at 550 km altitude, and eventually contain a total of 12 000 satellites at altitudes between 1 100 and 1 300 km. The first 60 satellites were deployed in May 2019. Beside Starlink, there are other competing projects, such as Kuiper (Amazon) with 3 236 satellites, Hongyan (China) with 320 satellites, Oneweb with at least 600 satellites, etc. The risk of in-orbit collisions is going to increase tremendeously with such constellations, especially for those who adopt a polar or near-polar orbit.

Whilst the heavily populated LEO and GEO orbits have nowadays standardised, but non-constraining, procedures defined for spacecraft at end-of-life, there currently exist no such international standards for Lagrangian point missions. Indeed, at first sight, collision risks around L_1 or L_2 are relatively low: there are few spacecraft there, and the Lissajous or halo orbits are much bigger than a GEO orbit. Moreover, because the L_1 and L_2 Langrangian points are unstable, the spacecraft will eventually be ejected from its orbit if no stationkeeping manoeuvres are performed anymore. This ejection can happen either towards the Earth, bringing the spacecraft into a wide geocentric orbit (orbital period of several weeks), or away from Earth, putting it into a heliocentric orbit. A geocentric orbit would be heavily perturbed by the Moon and there could be a risk for collision with the Earth.

To prevent any accident, ESA has adopted a safety procedure for the Herschel and Planck spacecraft (which were orbiting the Sun-Earth L_2 point) at the end of their lifes. In the summer and automn 2013, both spacecraft were put

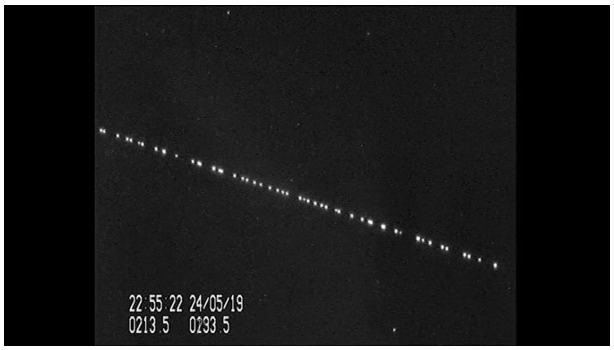


Figure 4.10: The first 60 satellites of the Starlink constellation deployed in May 2019 as seen shortly after their deployment from the ground. Beside causing a major risk in the context of space debris, these satellites also contribute to light pollution of deep-space astrophysical observations.

into a heliocentric disposal trajectory and were passivated by emptying their reservoirs. The heliocentric parking orbit was chosen in such a way to avoid subsequent close approach to the Earth over centuries.

4.4 Microgravity and outgassing

Spaceflight does not imply the absence of gravity ($\underline{\nabla}\,\Omega \neq \underline{0}$, except at the Lagrangian points). Actually, the Earth's attraction is only compensated by the spacecraft's velocity. The resulting microgravity is of course of interest for many scientific experiments, but can also be a problem in practice. In fact, it leads to problems with the storage of the propulsion liquids in the spacecraft tank.
For missions to minor bodies of the Solar System, one needs to account for the non-spherical shape of the body that the spacecraft is orbiting (see also lectures on *Celestial Mechanics*). Indeed, the motion of a spacecraft around a small, low-mass, rotating, irregularly shaped body such as a comet or an asteroid can only be approximated by a Keplerian orbit as long as the probe is rather far away from the body. Near the body, the gravitational potential must be expressed as a series of spherical harmonics. Subsequent tracking of the motion of the orbiter will then allow to determine the coefficients in the expression of the gravitational potential (see Eq.3.9). The operating environment of a lander or orbiter near such a minor body is a low-energy one. Orbiters move at walking pace (typically from $5\,\mathrm{cm\,s^{-1}}$ to $1.5\,\mathrm{m\,s^{-1}}$; orbital periods being of order 24 to more than 600 hours) and during trajectory manoeuvres one must take care not to reach the escape velocity (typically from $10\,\mathrm{cm\,s^{-1}}$ to a few $\mathrm{m\,s^{-1}}$).

Outgassing corresponds to the progressive release into space of gas that was trapped, frozen, absorbed or adsorbed in some material constituting the spacecraft. This process starts when the local pressure drops below the vapor pressure of the material. Outgassing leads to the formation of a gaseous halo around the spacecraft. Therefore, one preferentially uses materials that have a low rate of outgassing to build a spacecraft. The most common sources of outgassing are moistures, sealants, lubricants and adhesives, but even metals and glasses can release gases trapped in cracks or impurities. When this gas encounters a cold surface, the molecules deposit themselves on this surface. This is particularly critical for mirrors that look at the deep space permanently as well as for detectors that are cooled down to reduce the noise level. This contamination can become especially problematic if the molecules are subsequently modified by the effect of solar UV radiation which can lead to chemical reactions resulting in opaque products. Heating up the contaminated parts can help (e.g. the Narrow Angle Camera aboard the Cassini-Huygens probe) except if the chemical reactions have modified the properties of the contamination layer.

Chapter 5

The basic components of a scientific spacecraft

The difficulties to put an instrument into space and the space environment itself result in a number of constraints that apply to the design of any space mission. The most obvious ones are:

- the restrictions on the in-orbit mass which is essentially limited by the capabilities of the launch vehicle and the spacecraft's orbit (see Table 5.1);
- the size of the spacecraft is limited by its mass and the volume of the shroud of the launcher (see Table 5.1);
- the power of the spacecraft is limited by the size of the solar panels or the output of the generator in general (the latter being also limited by the mass);
- the data rate is limited by the on-board storage and processing capabilities as well as by the antenna and the communication capabilities (limited by ground station availability and coverage);
- the budget limits the number and complexity of spacecraft and the manpower available for operation.

Table 5.1: Performances of several launchers currently proposed for ESA science missions. The second column lists the maximum diameter of the fairing. The subsequent columns yield the mass that can be put into a low Earth orbit (LEO with a typical altitude of 300 km), a Sun-synchronous orbit, a geostationary transfer orbit (GTO), a halo orbit around a Lagrangian point or an interplanetary escape trajectory.

Launcher	Diameter (m)	Mass LEO (kg)	Mass SSO (kg)	Mass GTO (kg)	Mass L_1 or L_2 (kg)	Mass Escape (kg)
Ariane V ECA	4.57	> 10000	> 10000	9600	6600	4300 ($V_\infty = 3.5\,\mathrm{km\,s^{-1}}$)
Soyuz Fregat 2B	3.80	5300	4900	3060	2000	1600 ($V_\infty = 0\,\mathrm{km\,s^{-1}}$)
Vega	2.38	2300	1500		500	–
Rockot-KM	2.10	1850	1000	–	500	–

As a result, in addition to the **mass budget**, there are a number of issues that need to be addressed for every spacecraft, regardless of its specific mission: the instruments need **electrical power** to operate, the **attitude and velocity** of the spacecraft need to be controlled and **communication** with the ground is paramount to up- and downlink information and data. Another aspect concerns the **ground segment** that is mandatory to successfully operate the mission. All these aspects will be discussed in the present chapter. Before, we do so, let us recall that, in order to be selected for a space mission, a given technology must be *space qualified*. This implies a huge number of tests under specific conditions. When proposing a new space experiment, the proponents must provide an evaluation of the maturity of both the spacecraft and the payload unit. This is usually done in terms of the

so-called Technology Readiness Level (TRL, see Table 5.2). The higher the TRL of a given technology, the better its chances for being selected.

Table 5.2: Technology Readiness Level definitions as used by ESA.

TRL	Description
1	Basic principles observed and reported
2	Technology concept and/or application formulated
3	Analytical and experimental critical functions and/or characteristic proof of concept
4	Component and/or breadboard validation in laboratory environment
5	Component and/or breadboard validation in relevant environment
6	System model or prototype demonstration in a relevant environment (ground or space)
7	System prototype demonstration in a space environment
8	Actual system completed and flight qualified through test and demonstration (ground or space)
9	Actual system flight proven through successful mission operations

Each spacecraft can be considered as consisting of two major parts: (1) the payload that comprises the application-specific scientific instruments and experiments and (2) the service module that is commonly referred to as the bus or the platform. In this chapter we focus on the functionalities of the latter. In fact, the bus includes all subsystems that support the payload. These are typically

- the mechanical structure of the spacecraft on which all components are mounted,
- the equipment used to generate and distribute the electrical power,
- the units needed for telemetry and communication with the ground station,
- the attitude and orbit control system (AOCS) that allows monitoring, maintaining and/or modifying the attitude and velocity of the spacecraft,
- and the thermal control subsystem.

5.1 Electrical power supply

The procurement of electrical power is often a major issue and leads usually to restrictions on the power budget of the various instruments. Depending on the mission requirements (mainly the maximum power needed and the duration of the mission), various solutions can be considered for the generation of the electrical power (see Fig. 5.1).
In most situations, the electrical power available at a specific time not only depends upon the capabilities of the generator, but also on the storage capacities and the quality of the on-board batteries (so-called secondary batteries) that convert electrical power into chemical energy and restitute electrical power when solar cells are not illuminated or at times of peak power loads. Currently, most spacecraft batteries use NiCd cells. For atmospheric probes, one often uses so-called primary batteries, which are non-rechargeable batteries, designed to provide power over a short duration of operations (generally hours or days). The main constraint on these devices comes from the long durations of the spacecraft cruise (often many years) before the batteries are actually used.
The service module must provide the necessary equipment that includes among other things the solar arrays or the RTG, the batteries, the solar array control units, power converters,...

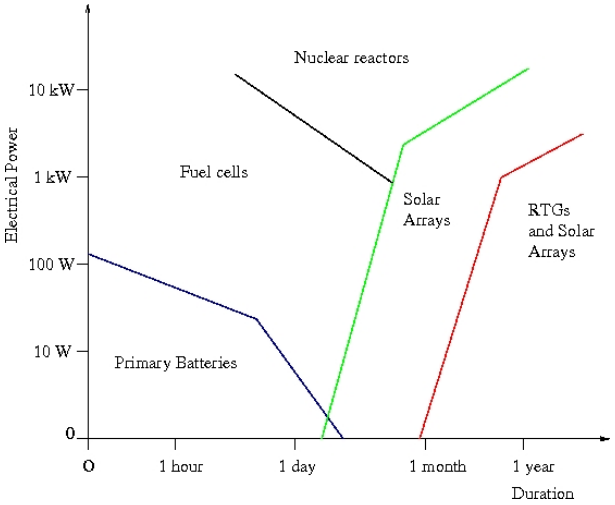


Figure 5.1: Schematic view of the various options to generate electrical power on board of a spacecraft and the regimes over which they offer an optimum solution. Note that fuel cells offer a high specific power (275 W kg^{-1} for the fuel cells on board the Space Shuttle), but their usage is mainly restricted to manned missions because of their cost and complexity (especially with respect to the storage of hydrogen).

5.1.1 Solar panels

Most spacecraft actually use photovoltaic solar panels to derive electricity from sunlight. Solar panels must have a rather large surface that can be oriented towards the Sun to produce enough energy. The solar panels are usually built in such a way that they can be pivoted to some extent to remain pointed at the Sun whilst the spacecraft is moving. The most efficient solar cells currently available on the market have an efficiency of 29% under ideal conditions, i.e. they can convert up to 29% of the incident solar light into electrical power. Since the solar radiation at 1 AU corresponds to a flux of 1370 W m^{-2}, the maximum electrical power that can be produced is of order 350 W for a solar panel of 1 m^2 collecting area[1]. In addition, since the flux decreases as r^{-2}, where r is the distance from the Sun, solar panels are rarely used beyond the orbit of Mars. The Rosetta space probe was the first space mission that travelled beyond the asteroid belt, while fully relying on solar cells. Rosetta used two 14 m long solar panels with a total collecting area of 64 m^2 that produce 8.7 kW in the inner Solar System and allowed the mission to operate as far as the orbit of Jupiter (5.25 AU) with a remaining power production of 400 W.

Figure 5.2: Left: artist view of the Rosetta spacecraft with its huge solar panels. Right: since the service mission in August 2007, the International Space Station (ISS) has a total solar panel collecting area of 4500 m^2 that provides 110 kW of electrical power.

[1]For most current generations of solar arrays, the typical performance is actually around 80 W m^{-2} and the electric specific power is about 40 W kg^{-1}.

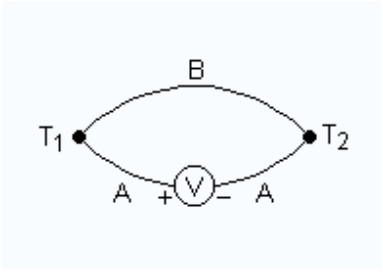
The first spacecraft in orbit around Jupiter that relies on solar panels only is Juno. The mission is designed in such a way that the maximum power will only be required for 6 hours near perijove, out of the 11 days orbit. The rest of the time, the power consumption is modest and the batteries recharge. ESA is currently designing the JUICE mission to Jupiter's icy moon Ganymede. Since ESA does not have access to radio-isotope thermoelectric generators, this mission will also have to rely entirely on solar panels.
In the future, it is planned to improve the performances of solar panels through the use of new photovoltaic cell material and solar concentrators that intensify the incident sunlight.

5.1.2 Radio-isotope thermoelectric generators

In the outer parts of the Solar System, sunlight is too weak to supply sufficient power. Radio-isotope thermoelectric generators (RTG) are used as power sources instead. The principle of these devices is to convert the heat released by the decay of a radioactive material into electricity.
The conversion of heat into electrical power is based on the Seebeck effect. The latter stems from the fact that a voltage appears between two ends of a metal bar when a temperature difference exists in the bar. This voltage is very small, of order several μV per degree of temperature difference. The thermoelectric current is created in the presence of two different metals (A) and (B) that are subject to the same temperature gradient. The voltage that can be generated is given by

$$\Delta V = \int_{T_1}^{T_2} (S_B(T) - S_A(T))\, dT$$

where S_A and S_B are the Seebeck coefficients of material A and B respectively. This assembly constitutes a thermocouple and several thermocouples connected in series with each other form a thermopile.



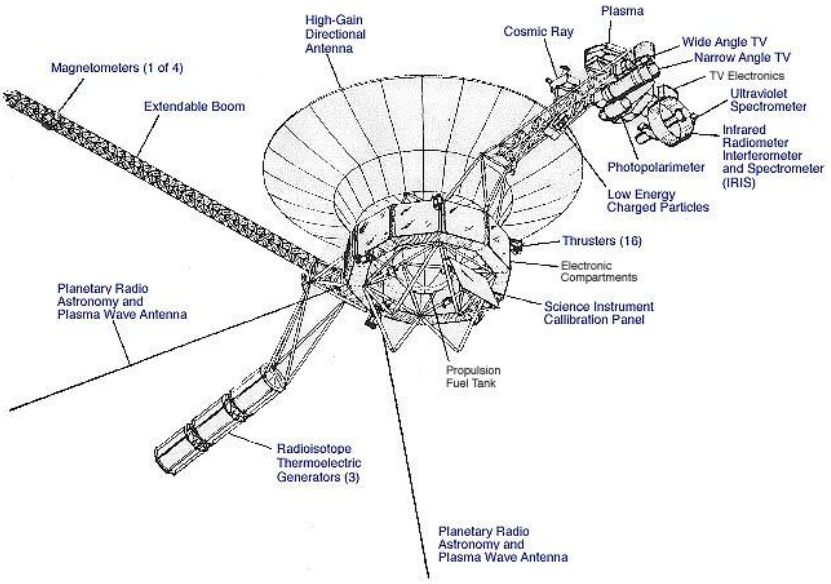


Figure 5.3: Left: schematic view of a thermocouple. Right: diagram of the Voyager 1 spacecraft showing the location of the RTG (in the lower left of the figure).

The radioactive material used in RTGs must have a half-life long enough to produce energy at a relatively continous rate over the lifetime of the mission. At the same time, the half-life needs to be short enough so that the material decays sufficiently quickly to generate enough heat. Also, the high-energy particles and radiation produced by the decay must have a rather low penetration power (such as α particles) to avoid the need of too heavy a shielding of

the rest of the spacecraft. ^{238}Pu which decays with a half-life of 87.7 years into ^{234}U and an α particle, is nowadays the most widely used fuel for RTGs in space applications (e.g. on board Cassini-Huygens)[2].

Another prominent example of the usage of such an RTG are the Voyager probes where the device provided approximately 470 W of 30 V current at launch. Since the radioactive half-life is about 87.7 years, the RTGs lose a factor $1-(0.5)^{1/87.7}$ of their power output per year. In 2007, 29 years after launch, such an RTG should still produce $470 \times 0.5^{29/87.7} = 371$ W. However, the bi-metallic thermocouples that convert the heat into electric power degrade as well, so that the available power was actually down to 60% of its initial value. In the case of the Cassini probe, the power production by the RTGs was 880 W at launch in 1997, whilst it was down to 640 W in 2011. Note that the electrical specific power of RTGs is rather low. For instance, it was 2 W kg^{-1} for the RTGs on board the Viking probes, 5 W kg^{-1} for that of Galileo and 27 W kg^{-1} for Cassini-Huygens (at launch). This is mainly because of the low efficiency (typically 5%) of the thermopiles. One option to increase the efficiency of RTGs would be to implement in addition thermophotovoltaic cells that use the IR light emitted by the hot decaying material to produce electricity. The overall efficiency could thereby reach 20%. A practical issue with this solution is the degradation of the photoelectric cells in the presence of ionizing radiation.

5.2 The attitude and orbit control

Attitude control is the control of the orientation of a spacecraft. Typical components of the attitude and orbit control system (AOCS) are the Sun and Earth sensors, star trackers[3], reaction and momentum wheels, inertial measurement and reference units, liquid or solid rockets, ion engines, thrusters, solar sails...

One distinguishes between three-axes stabilized and spin-stabilized spacecraft. The former are kept inertially stable whilst the latter achieve stability through constant rotation about one axis. In this configuration, the communication payload is usually mounted on a shelf which is despun relative to the rest of the spacecraft allowing it to point the Earth continuously.

In a low-Earth orbit, a spacecraft with one axis much larger than the other will spontaneously orient itself so that its long axis points towards the centre of mass of the planet. This effect is due to the tidal forces that act on the satellite and can help to passively control the attitude.

5.2.1 Gyroscopes

Attitude control of spacecraft relies to a large extent on the usage of gyroscopic reaction and momentum wheels. Gyroscopes are devices allowing to measure rotation or maintain orientation in three-dimensional space. They are based on the principle of the conservation of angular momentum. The device consists of a spinning wheel on an axle. Once the wheel is spinning, it tends to resist changes to its orientation. These gyroscopes are used either as reaction wheels or momentum wheels. Reaction wheels allow to rotate the spacecraft around its centre of mass without using fuel. The wheels are spun up in the sense opposite to the desired re-orientation of the spacecraft. Both spin-up and spin-down of the wheel are controlled by electric motors. Since their mass is small compared to the mass of the spacecraft, they allow small and very accurate rotations of the spacecraft and are thus ideal for maintaining the spacecraft pointed at a specific target. While reaction wheels have a zero average rotational velocity, momentum wheels have high rotation speeds (about 5 000 cycles/minute) and are used for gyroscopic stabilisation of the spacecraft.

[2] In the USA, the launch of an RTG requires a launch order signed by the President.

[3] Star trackers are optical devices that measure the direction to one or several pre-selected bright stars.

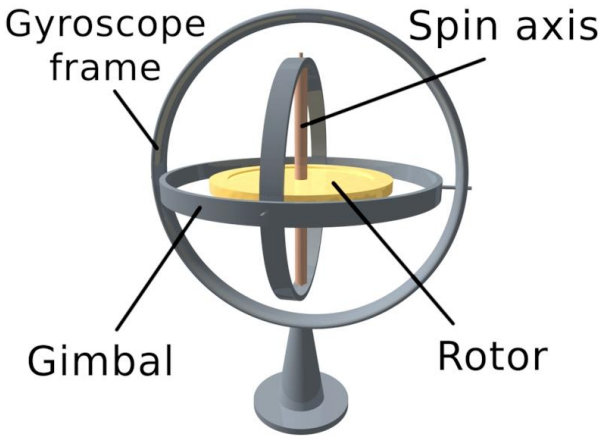


Figure 5.4: Left: a satellite of the Intelsat series during integration. These satellites were spin-stabilized. To generate the electrical power, the outer surface of the satellite is hence covered with solar cells. Right: schematic representation of a gyroscope.

5.2.2 Electric propulsion

Most spacecraft use conventional chemical thrusters (monopropellant rockets) as devices for attitude control. In these devices, the liquid propellant (usually hydrazine) is injected as a spray into the thrust chamber where it interacts with a catalyst. The incoming propellant then heats up and through contact with the catalyst, it decomposes. Chemical propulsion then creates thrust by thermodynamically expanding the heated gas through a nozzle. The main limitations of these devices stem from the capacity of the fuel tank and through the rocket equation from the limited ejection velocity that can be achieved.

An attractive alternative comes from the so-called ion engines. The basic idea is to create a plasma of charged particles and accelerate it through the action of an electrostatic field. The ion engine thus consists of several components: (1) an ion source, (2) a focusing and accelerating device, (3) a system allowing electrons and ions to recombine outside the spacecraft. The ions are created in a magnetically confined ionisation chamber using either the contact with a metal at high temperature or an electric discharge. A series of electrodes focalises the ions into a beam and accelerates them. The electrons are extracted from the chamber by an anode and pumped to an external cathode. Once the ions are ejected out of the spacecraft, the electrons emitted by the cathode allow them to neutralize. The most commonly used propellants are cesium, sodium, lithium and xenon.

Whilst the force produced by these engines is very low (of order 0.2 N), its big advantage is that it can operate over extended durations and hence the cumulated acceleration can be quite large. For instance, the ionic propulsion system of the Deep Space One probe operated for a cumulated operation time of two years. The specific impulse $v_{\rm ej}/g_0$ of an ion engine is also considerably higher than that of a conventional chemical propulsion thruster, because the exhaust speeds $v_{\rm ej}$ can be as large as $30\,{\rm km\,s^{-1}}$.

Table 5.3: Comparison between chemical and electric spacecraft propulsion. The classical hydrazine thruster produces less than a tenth of the total impulse of the SMART 1 engine, whilst using 65% of the propellant mass. The Soyuz Fregat main engine produces 14 times the total impulse of the SMART 1 electric engine, but uses almost 70 times more propellant mass to achieve this result.

	Monopropellant thruster	Fregat Main Engine	SMART 1 Thruster
Propellant	Hydrazine	Nitrogen tetroxide, dimethyl hydrazine	Xenon
Specific impulse $v_{\rm ej}/g_0$ (s)	200	320	1640
Thrust $\lvert\frac{dm}{dt}\rvert\, v_{\rm ej}$ (N)	1	1.96×10^4	6.80×10^{-2}
Thrust time (h)	46	0.24	5000
Propellant mass (kg)	52	5350	80
Total impulse (Ns)	1.1×10^5	1.72×10^7	1.2×10^6

Electric propulsion systems have been used for orbit adjustments for GEO and LEO satellites. NASA's Deep Space One and ESA's SMART 1 spacecraft were the first vehicles to use xenon ion engines as the primary (in-space) propulsion modes. The electric energy required for ion thrusters is usually quite considerable and has to be generated on board either by means of solar panels or RTGs. In the SMART 1 spacecraft, 1.2 kW of power were required to produce a nominal thrust of 0.068 N. For Deep Space One, 2.3 kW were needed to generate a thrust of 0.09 N.

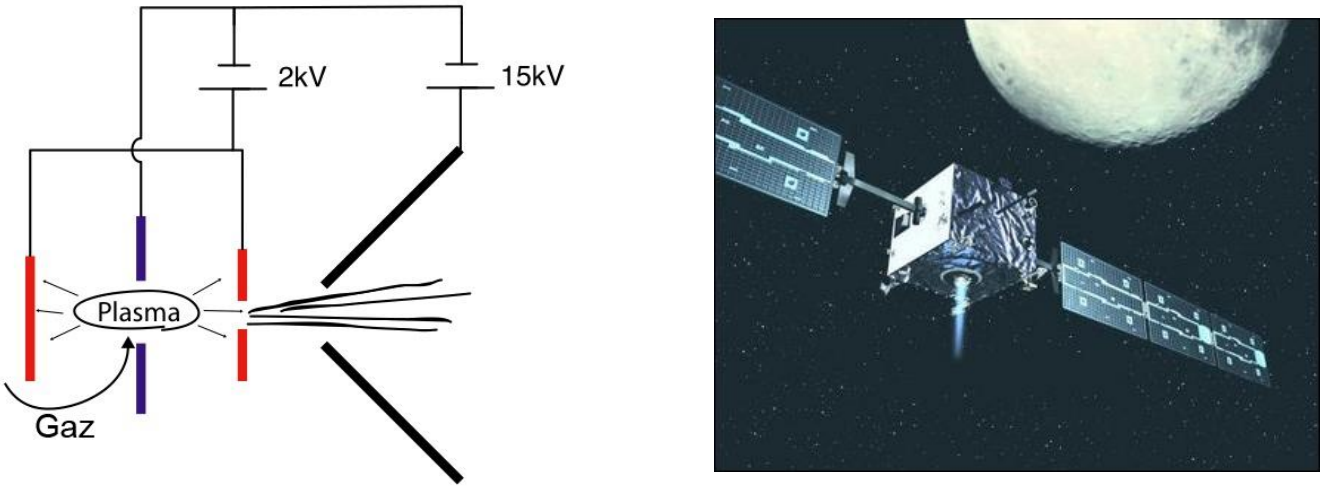


Figure 5.5: Left: schematic view of an ion engine. Right: the SMART 1 spacecraft.

A new generation of electric propulsion systems is currently under development by Ad Astra Rocket Company (www.adastrarocket.com). This so-called VASIMR (VAriable Specific Impulse Magnetoplasma Rocket) device uses radio waves to ionize the propellant gas and magnetic fields to heat the resulting plasma to generate the thrust. The system relies upon three linked magnetic cells. In the first stage, neutral gas (Ar, Xe or H) is first ionized using a helicon radio wave generated by a superconducting magnet. The resulting cold plasma is then energized in an ion cyclotron heating stage. The cyclotron motion of the ions is converted into linear motion by means of a magetic nozzle (see Fig. 5.6).
This system can work in two regimes, combining either a high thrust with a low specific impulse or a low thrust with a high specific impulse, and is therefore able to better meet the instantaneous mission requirements, although

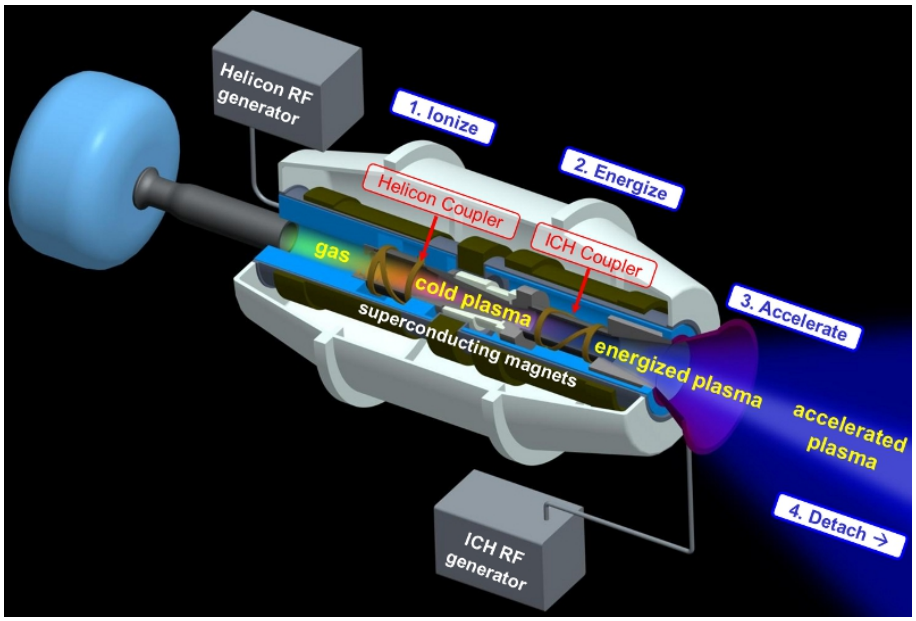


Figure 5.6: Schematic illustration of the principles of the VASIMR engine. ©Ad Astra Rocket Company.

it cannot be used to launch a payload from the Earth's surface.
In the VASIMR concept, the plasma is driven entirely by electromagnetic waves and the engine has no electrodes in contact with the hot plasma. This design reduces the heat flux from the plasma to the structure, allowing to reach higher plasma temperatures and eliminating problems with the erosion of the electrodes, therefore increasing the reliability and the lifetime of the engine. In 2009, a series of tests have been performed on the VF-200 prototype, producing a thrust of 5 N with an electrical power of 200 kW. In 2015, NASA decided not to test the VF-200 prototype on the ISS, and in 2019 VASIMR was no longer supported by NASA funding. The future of this engine is thus unclear.
A trip to Mars with a VASIMR engine requires a power of 200 MW (one tenth of the power production of a nuclear power plant). Hence, nuclear reactors are required to generate the necessary power. However, for this energy production mode the mass-to-power ratio is currently not optimized (about 50 kg kW^{-1}).

5.2.3 Solar sails

In the future, an alternative mode for spacecraft propulsion could rely on the usage of solar sails. The concept is rather simple: one uses the radiation pressure $\frac{h\nu}{c}$ communicated by the photons of the solar radiation field when they reflect off a large surface (the solar sail) to accelerate a probe that has already reached either the first or the second cosmic velocity. The larger the sail and the higher its reflectivity, the stronger will be the force acting on the sail. The big advantage of this concept is that it does not rely on any use of consumables (no propellant is needed). Let us consider a solar sail with a 100% reflectivity at 1 AU from the Sun. Since the average wavelength of solar photons is about 5000 Å, the solar flux corresponds to about 3×10^{21} photons m^{-2} s^{-1}. Multiplying this number by the momentum of the photons then yields a force of 5×10^{-6} N per square metre of the sail. Hence a sail surface of about 200 000 m^2 is needed to accelerate a mass of 1 kg by 1 m s^{-2}. Note that the contribution of the solar wind to the acceleration is actually negligible. The major difficulty with this technique is the need to deploy a huge, ultra-thin, light-weight sail in space. In addition, the sail has to be sufficiently rigid and highly reflective. This technology was used for the Japanese technology-demonstrator mission IKAROS, launched in May 2010.

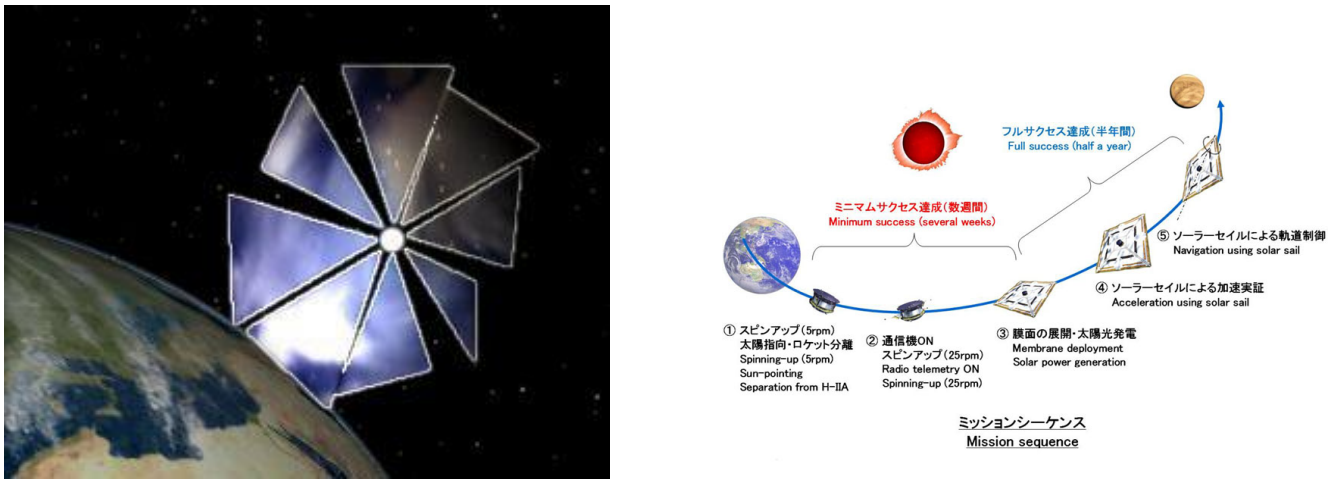


Figure 5.7: Left: Artist representation of the solar sail Cosmos 1. The orbit insertion of this experiment failed in June 2005. Right: Schematic view of the IKAROS mission, the first spacecraft to successfully use a solar sail (from the JAXA website).

The IKAROS solar sail had a size of $200\,\mathrm{m}^2$, 10% of which were covered with solar cells to generate electricity. The mission was designed for a fly-by of Venus which was successfully accomplished on 8 December 2010 (see Fig. 5.7). Attitude control of the sail was achieved using variable reflectance liquid crystal panels. Deployment of a much smaller solar sail was successfully tested on the NanoSail-D spacecraft, a nano-satellite ($30 \times 10 \times 10\,\mathrm{cm}^3$) in low-Earth orbit that deployed its $10\,\mathrm{m}^2$ solar sail in January 2011. The sail was only $7.5\,\mu$m thick and was made of a special polymer called CP-1. In the latter experiment, drag from the residual atmosphere was the main force beside gravity acting on the spacecraft. The effect of radiation pressure was actually very small.
Solar sails can be used either to expand or to shrink the orbit of a spacecraft moving around the Sun (Fig. 5.8) In the future, engineers and scientists hope to use solar sails for de-orbiting deactivated satellites and space debris as well as for interstellar travelling.

5.3 Communication and the ground segment

The on board central computer is an essential component of any spacecraft. It not only controls the vital functions of the spacecraft, but also takes care of the data handling. Airborne or spaceborne software has to meet several safety requirements to ensure that it will perform reliably in the relevant environment. This is summarized by the Design Assurance Level (DAL) which specifies the requirements depending on the potential consequences associated with a failure scenario.

Communication with the spacecraft is required for telemetry, command up-load, position and velocity determination. The distance and the velocity of the spacecraft are measured by ranging (round-trip light time) and the Doppler effect respectively. To make use of the Doppler effect for navigation, it is necessary to know the frequency of the emitted signal very accurately. Most interplanetary spacecraft therefore carry an ultra stable oscillator that maintains a highly stable frequency. However, the most accurate atomic clocks that allow reaching accuracies down to $0.1\,\mathrm{mm\,s}^{-1}$ for velocities and $\sim 1\,\mathrm{m}$ for distances are too large (size comparable to a refrigerator) to be put on the spacecraft. Measuring the velocity and distance of the spaceship hence requires a two-way transponded link between the ground station and the spacecraft. In this coherent mode, the downlink frequency is automatically set to the frequency of the uplink signal. An accurate determination of the angular position can be achieved

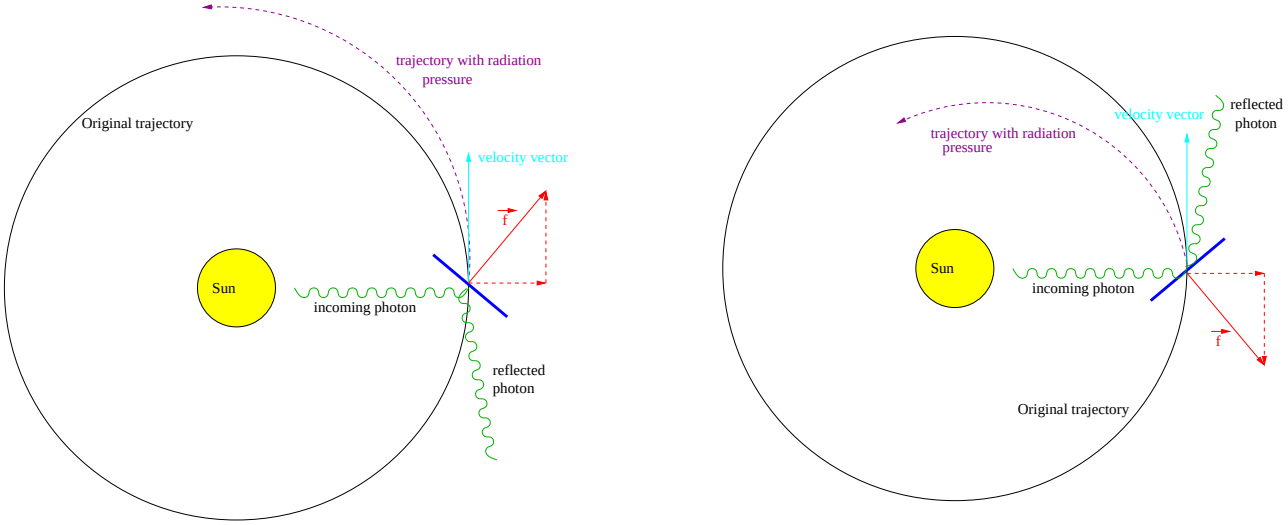


Figure 5.8: Schematic illustration of the operations of a solar sail. As any spacecraft, the solar sail (shown in blue) moves around the Sun. Here we assume that the trajectory without the effect of radiation pressure would be circular. Depending on the orientation of the sail, radiation pressure will produce a net force that will either act in the same sense as the motion of the spacecraft (left) or in the opposite sense (right). In the former case, the spacecraft will accelerate and the orbit expands (left panel), whereas in the latter case, it will decelerate and the orbit shrinks, bringing the spacecraft closer to the Sun (right panel). The component of the radiative force that is opposite to the direction of the Sun, will result in a small reduction of the Sun's gravitational attraction.

through the delta differential one-way range method. The latter consists of the differentiation of signals received simultaneously from the spacecraft at two different ground stations.

On Earth, large amounts of power are available for radio communications. The situation is however much different in space. The power of the on-board transmitter is limited to a few hundred Watt for telecommunication satellites, whilst for scientific satellites, the power of the transmitter is restricted to about 20 W. The communications use different bands over the microwave part of the electromagnetic spectrum. The S, C, X and K_a-bands range in frequencies from 2 to 4 GHz, 4 to 8 GHz, 7 to 12.5 GHz and from 18 to 40 GHz respectively. In the X (resp. C) band, satellites usually use 7.25 – 7.75 GHz (resp. 3.7 – 4.2 GHz) for downlink and 7.9 – 8.4 GHz (resp. 5.9 – 6.4 GHz) for uplink. These bands are regulated by the International Telecommunication Union. In practice, they are also limited by the transmission of the Earth's ionosphere, which absorbs frequencies below about 100 MHz, whilst this cut-off is around 50 MHz for the Martian ionosphere. Communications with probes in the Jovian system, must avoid frequencies affected by the strong synchrotron emission produced by Jupiter's radiation belts.

Most spacecraft have two antennas, a low-gain omni-directional antenna and a high-gain antenna that must remain accurately pointed at the Earth during high-rate telemetry. The gain of an antenna is the ratio between the power transmitted by the antenna in its working direction to the power that would be transmitted over the same beam-width by an isotropic antenna. A low-gain antenna with a wide beam is useful for a moving target and/or for emergency telemetry and is also used during descent and landing for atmospheric probes and landers. The larger the dish of an antenna, the higher its gain and the narrower the beam. When the diameter of the antenna doubles, the gain increases by a factor four.

A schematic view of the on-board communication chain is shown in Fig. 5.9. Depending on the data rate and the situation, the spacecraft switches between the high-gain and the low-gain antennas. The transponder demodulates the received signal into spacecraft commands, generates a downlink carrier and modulates the downlink carrier with telemetry. The power amplifier amplifies the downlink signal for transmission through the antenna. For a

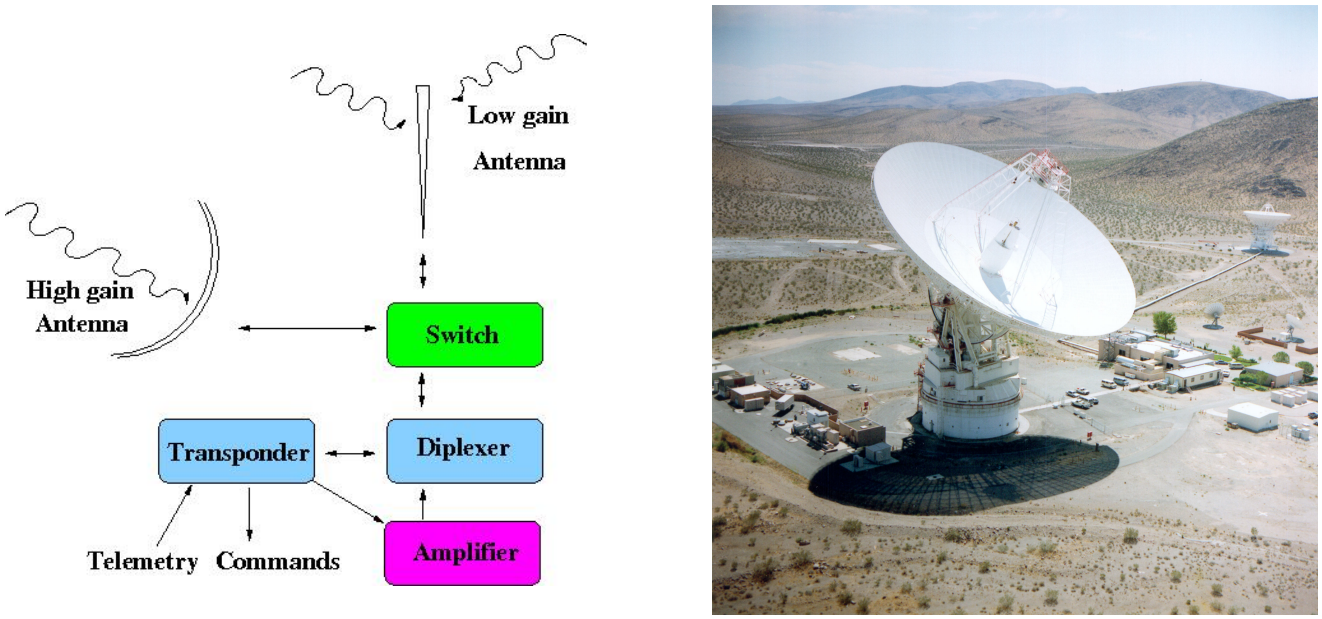


Figure 5.9: Left: schematic illustration of the communication system on board of a satellite. Right: the 70 m diameter DSN radio dish at Goldstone in California.

high-gain antenna, the data rate is proportional to $P_t\, A_t\, A_r\, \nu^2/R^2$ where P_t is the power of the transmitter, A_t is the area of the transmitter, A_r the area of the receiver, ν the frequency and R is the distance between the receiver and the antenna. The ν^2 dependence shows the benefit to go from the S to the X band. The K_a band and, to a lesser extent, also the X band are characterized by severe losses due to rain and a lower power conversion efficiency. Hence there is no big (ν^2) benefit to go from the X band to the K_a band.
The NASA Deep Space Network (DSN) is an international network of facilities to support interplanetary spacecraft missions. It provides the largest and most sensitive scientific telecommunication system in the world. The main antennas of the Deep Space Network are located in Goldstone (California), Canberra (Australia) and Madrid (Spain), approximately 120° apart around the planet, thus allowing a constant communication as the Earth rotates. The ground stations consist of several high-gain, parabolic-reflector steerable antenna dishes and their front-end equipment (amplifiers,...). For very distant spacecraft (e.g. Cassini-Huygens), the uplink signal forms such a narrow beam that it must be pointed slightly ahead of where the spacecraft is in the sky. Then, by the time the signal has travelled, the spacecraft has moved into the beam and can receive the signal.

A major component of any space mission is the ground segment. The latter not only includes the antennas on the ground that are required to communicate with the spacecraft, but also a number of people that take care of various tasks that are essential for a successful operation of the mission. For ESA missions, these people are gathered in two distinct groups, the SOC (Science Operation Centre) and the MOC (Mission Operation Centre). The SOC is mainly composed of scientists and some engineers. The SOC directly interacts with the general scientific user. It coordinates calls for proposals for new observations, it prepares and optimizes the schedule of the scientific observations, takes care of the calibration of the instruments on board the satellite and provides a first rough data processing before the data are delivered to the user. The MOC consists of engineers and spacecraft controllers who actually monitor the health of the spacecraft and its instruments and issue the commands that are needed to operate the mission. Regular contact with the ground segment is crucial. For instance, ESA maintains a permanent contact with the XMM-Newton spacecraft which allows to note and diagnose potential problems much faster than in case of a single or a few contacts per day.

Chapter 6

Atmospheric probes and landers

The exploration of the Solar System is currently the only discipline in Space Science where in-situ experiments are possible. The most spectacular examples are provided by atmospheric descent probes and landers. Whilst these are challenging experiments, as demonstrated by their relatively low rate of success of about 50%, they offer the possibility to gain considerable insight about the chemistry, physics and geology of other planets and eventually to search for extant or extinct life. The way a probe can land on a Solar System body depends on the presence or absence of an atmosphere and the gravity of the body.

6.1 Atmospheric entry

6.1.1 The initial phases of atmospheric (re)entry

The entry and descent through an atmosphere constitute either the final part of a mission, or the final stage of its journey. This is a critical step which requires finding a balance between deceleration, heat load and the accuracy of landing.
Let us consider a thin, isothermal, chemically homogeneous planetary atmosphere in hydrostatic equilibrium. Its density profile can be expressed as

$$\rho = \rho_0 \, \exp \frac{-z}{H}$$

where the scale height $H = \frac{k\,T}{\overline{m}\,g}$ with $\overline{m}$ being the mean molecular mass of the atmosphere. Aerodynamic drag acts parallel to the (relative) velocity vector of the spacecraft, in such a way that

$$\frac{d\,v}{dt} = -\frac{\rho\,S\,C_D\,v^2}{2\,m} = -\frac{\rho\,v^2}{2\,BC}$$

Here C_D is the aerodynamic drag coefficient and S is the cross section of the spacecraft perpendicular to its relative velocity vector (the so-called *wetted surface*). m is the mass of the spacecraft and BC the ballistic coefficient. The kinetic energy of the probe is used to heat the layers of air around the probe. The lower BC, the quicker the deceleration. Since we have assumed that the atmosphere is thin (i.e. its height is very small compared to the planetary radius), we can further assume that g and thus H are constant over the height of the atmosphere. Hence, to first approximation, the atmospheric entry will happen under a constant angle γ and the variation of the altitude of the spacecraft is thus given by $\frac{d\,z}{dt} = -v\,\sin\gamma$. Introducing the parameter $\eta(z) = \frac{\rho(z)\,S\,H\,C_D}{2\,m}$, we then obtain that

$$\frac{d\,v}{dt} = -\eta(z)\,\frac{v_0^2}{H}\,\exp\left(-\frac{2\,\eta(z)}{\sin\gamma}\right)$$

Therefore, the peak deceleration happens for $\eta(z) = \sin\gamma/2$ and the value of the peak deceleration amounts to $-\sin\gamma\,\frac{v_0^2}{2\,e\,H}$ (see Fig. 6.1). The steeper the entry angle, the higher the peak deceleration and thus the heat load experienced by the spacecraft. A shallow entry allows to reduce the maximum heat load, but because of the lower deceleration, the re-entry takes longer and thus the heat load affects the vehicle for a longer time. Note that the peak deceleration value is independent of η and hence of the drag coefficient C_D, although the altitude at which the peak deceleration occurs and the maximum heat load (see below) both depend upon C_D. The above results apply to a so-called ballistic entry (i.e. to a spacecraft that produces no aerodynamic lift, see below).

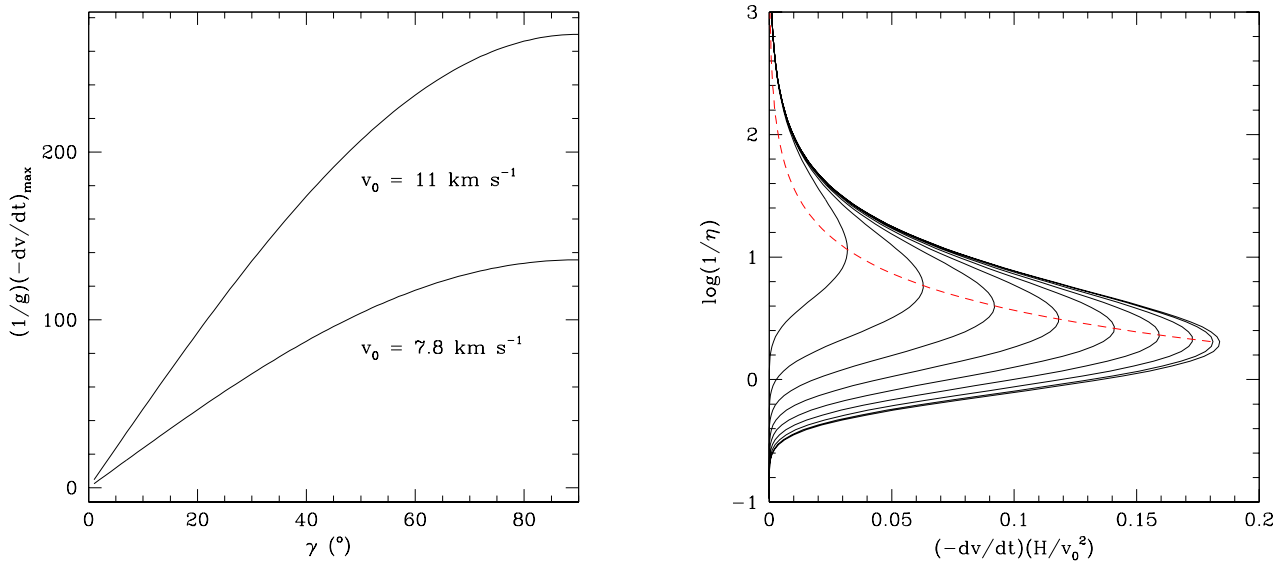


Figure 6.1: Left: maximum deceleration for a spacecraft entering the Earth's atmosphere ($H = 8.4\,\mathrm{km}$) along a ballistic trajectory as a function of the angle γ for two different values of the spacecraft velocity (the first and second cosmic velocities). The deceleration is expressed in units of $g = 9.81\,\mathrm{m\,s^{-2}}$. Right: relation between η and the deceleration of a spacecraft along a ballistic trajectory for different values of γ. From left to right the various solid lines correspond to $\gamma = 10^\circ$ to 90° in steps of 10°. The dashed line yields the maximum deceleration experienced by the spacecraft.

During atmospheric entry, the heat loads are proportional to the third power of the entry speed (relative to the atmosphere). Therefore, for a probe coming from the day-side of the planet, it is most advantageous to approach the planet in such a way that the atmospheric entry takes place on the receding (evening) side of the planet (see Fig. 6.2). Note that the peak heat flow will generally occur at a different altitude than the peak deceleration.

As pointed out above, the kinetic energy of the spacecraft is dissipated as heat. For a probe entering the atmosphere directly from a parabolic trajectory, one has $v_0 = V_{\mathrm{esc}}$ and the ensuing heat load (see Table 6.1) is more than sufficient to destroy the spacecraft if all the heat is to be absorbed by the latter. In the upper atmosphere, where the density is very low, the gas flow around the spacecraft is ballistic and only the forward facing part of the probe undergoes collisions with molecules. In the deeper layers, a hydrodynamic shock forms around the supersonically moving probe and the gas ahead of the probe is compressed into a dense layer and heated. The energy that is dissipated to heat this layer of air is taken away from the heat load on the spacecraft itself. To first order one finds that, the larger the mass of the layer of shocked air, the lower the heat load on the spacecraft.

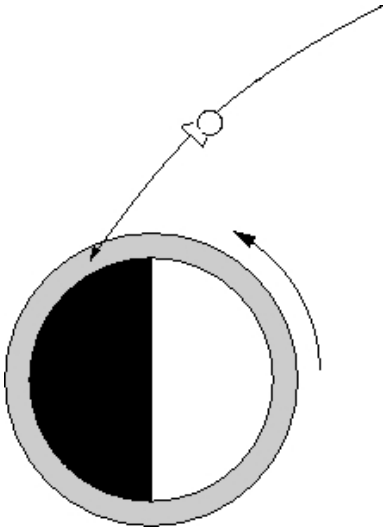


Figure 6.2: Schematic view from above the planet's pole of an entry probe approaching a planet from the day-side and aiming at an atmospheric entry near the evening terminator.

Table 6.1: Kinetic energy per unit mass for a probe entering the atmosphere of a planet at the planet's escape velocity. This kinetic energy must be dissipated as heat.

	$\mathrm{MJ\,kg^{-1}}$
Venus	54
Earth	63
Mars	13
Jupiter	1770
Saturn	630
Titan	3.5
Uranus	226
Neptune	276

Actually, the larger the drag coefficient C_D, the lower the heat load on the spacecraft. This, slightly counter-intuitive result stems from the fact that a streamlined vehicle will have the shock attached to its tip, leading to a concentration of the heat. On the contrary, a blunt shape (large C_D) of the entry vehicle forces the air to move along with the vehicle and to form a cushion that detaches the shock wave from the vehicle and hence pushes the heated layers away from the vehicle's surface. A very simple geometry to achieve this objective is a sphere or a spherical section with a converging conical afterbody. This geometry yields a low aerodynamic lift[1]. For a spacecraft with a spherical section heat-shield, a modest aerodynamic lift can be generated by entering the atmosphere under a certain angle of attack (e.g. the Apollo command module). An alternative design is the sphere-cone consisting of a spherical section attached to a conical surface (see Fig. 6.3). For space-exploration probes, the half angle of the cone is usually rather large (typically 45° or larger) whilst for inter-continental ballistic missiles, the cone half angle is typically 10°.

[1] A pure sphere generates zero lift.

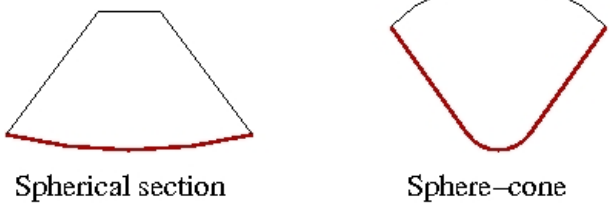


Figure 6.3: Top: schematic view of a spherical section heat-shield and a sphere-cone.

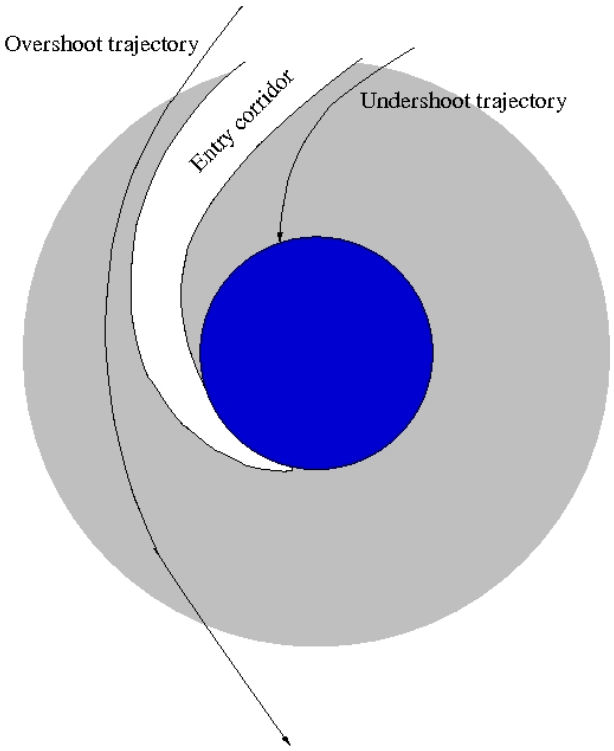


Figure 6.4: Right: schematic view of the entry corridor for atmospheric entry.

A successful entry into the atmosphere of a planet must be done with a rather narrow range of entry angles (see Fig. 6.4). This range is set by the spacecraft's trajectory (which conditions its speed), the maximum acceptable rate of deceleration, the maximum acceptable level of aerodynamic heating and the duration over which these loads exist. Humans can endure a level of 3 g for about 10 min and up to 10 g for a shorter time of about 2 min. Prolonged exposure to high g levels damages the human body and can even be fatal. During his ballistic flight, astronaut Alan Shepard experienced maximum loads of 6.5 g during launch and 11.6 g during atmospheric re-entry. For manned spacecraft, the maximum acceptable deceleration is about 10 g, whilst an unmanned probe can endure more than 100 g depending on its structure. Too steep an entry angle leads to too strong a deceleration and to overheating of the spacecraft. This situation is called undershooting and can lead to the disintegration (by break-up or burn-up) of the spacecraft. On the other hand, if the spacecraft enters the atmosphere with too shallow an entry angle, the drag will not be sufficient to slow down the vehicle and it will instead re-emerge of the atmosphere. This situation is called overshooting.

The trajectory across the atmosphere strongly depends upon the lift-over-drag (L/D) ratio of the vehicle. Lift is a force perpendicular to the direction of motion that results from the air moving over the surface of the vehicle. Most spacecraft have a shape that generates little aerodynamic lift and their motion is thus mainly controlled by gravity and atmospheric drag only. The latter slows the vehicle to the point where parachutes can be deployed for a soft landing. This situation ($L/D = 0$) is called a ballistic entry. Examples are the Mercury command modules, the entry capsules of the Huygens landers or the Mars Exploration Rovers. The equations above and the plots in Fig. 6.1 refer to such a ballistic entry.
Alternatively one can design the spacecraft in such a way as to ensure that it generates a significant amount of aerodynamic lift. This approach was implemented on most manned spacecraft (except the Mercury capsules[2]) to

[2]The latter entered the atmosphere with $\gamma \simeq 1.5^\circ$ and the peak deceleration was over 8 g.

keep the peak deceleration sufficiently low. For instance, the Soyuz capsule re-enters the atmosphere at a velocity of $7.6\,\mathrm{km\,s^{-1}}$ and under an angle $\gamma = 1.35°$. The combination of this small angle with a moderate lift allows to keep the maximum deceleration below 4 g. In the case of the Apollo command module, the angle of re-entry was in the range 5.3 to 7.7° and the peak deceleration typically reached 7.2 g. The largest lift was implemented on the Space Shuttle ($L/D \simeq 1$), reducing the peak deceleration below 1 g and allowing the pilot to have some control over the spacecraft's trajectory and to land it in much the same way as a normal plane. The price to pay was the longer duration of the re-entry and thus the longer duration over which the spacecraft had to resist the heat load.

As seen above, a heat shield is required to avoid the destruction of the spacecraft. If contamination is not an issue (i.e. no sensitive measurements are to be performed after the main phase of deceleration), an ablative heat shield can be used, that vaporizes when heated sufficiently. For instance, for vehicles that enter the Earth's atmosphere, ablative heat shields are appropriate. Because of its high melting point (3775 K) and its capability to absorb large quantities of heat to reach the melting point ($6.7\,\mathrm{MJ\,kg^{-1}}$, when starting at a temperature of 300 K) and to vaporize ($29.7\,\mathrm{MJ\,kg^{-1}}$), carbon is a very interesting material for a thermal protection system. The ablative materials are mostly composites with phenolic resins providing a matrix in which refracting particles or fibres are embedded. This configuration facilitates the escape of the gas released by pyrolysis during ablation. For instance the Galileo probe entered into the Jovian atmosphere at an approach speed of $47.5\,\mathrm{km\,s^{-1}}$. With a mass of the probe of 340 kg, a total kinetic energy of $3.8\,10^5$ MJ had to be dissipated. The maximum temperature reached during entry was about 15 000 K and it is estimated that 90 kg of ablative material were lost. As an alternative to ablation, one can use radiative cooling of the interface between the spacecraft and the air. This solution was implemented for the Space Shuttle. The ceramic tiles of the Shuttle's heat shield were made of a composite material combining a high emissivity ($\epsilon \sim 0.8$) with an efficient insulation (using highly refined silicates). In this way, most of the heat was radiated away rather than being absorbed by the vehicle's structure.
To ensure stability of the vehicle during the atmospheric entry and avoid exposing unshielded parts to the airflow, the centre of mass of the spacecraft must be located below the aeroshell's centre of pressure. This can be achieved for instance by bringing the more massive components of the spacecraft, such as the batteries, closer to the heat-shield.
Finally, we note that during the supersonic part of the entry, the plasma that forms around the probe blocks the radio signals and communications are interrupted.

6.1.2 The final stages of the descent

When the spacecraft has reached the lower regions of the atmosphere and has been slowed down significantly, the heat-shield is ejected and the parachutes are deployed. The terminal descent velocity can be computed from

$$\frac{1}{2}\,S\,C_D\,\rho\,v_t^2 = m\,g$$

Introducing the ballistic coefficient $BC = \frac{2\,m}{S\,C_D}$, one finds that $v_t^2 = BC\,\frac{g}{\rho}$. Typical parachutes have C_D of about 0.5.
For terrestrial applications, parachutes often use silk. For probes to other planets, outgassing must be limited and the most commonly used materials are polyester and nylon. The parachute material has to meet several conditions. For instance, it needs to be radio-transparent to allow communications of the probe with an orbiter during the descent (e.g. the Huygens probe landing on Titan had to keep contact with the Cassini spacecraft). Parachutes need to be decontaminated and sterilized with particular care, because they usually represent the largest surface area of the probe.

The terminal descent velocities reached with parachutes are often too large for a soft landing[3] and retro-rockets can be used to null the descent velocity or bring it into the range $1 - 10\,\mathrm{m\,s^{-1}}$. Particular attention must be paid to avoid contamination of the landing site by exhaust gases from the descent engine. For this purpose, the Viking landers used a hydrazine landing rocket featuring 18 nozzles to spread the hydrogen and nitrogen exhaust over a wide area. This technique limited surface heating and avoided stripping away too large a layer of surface material.

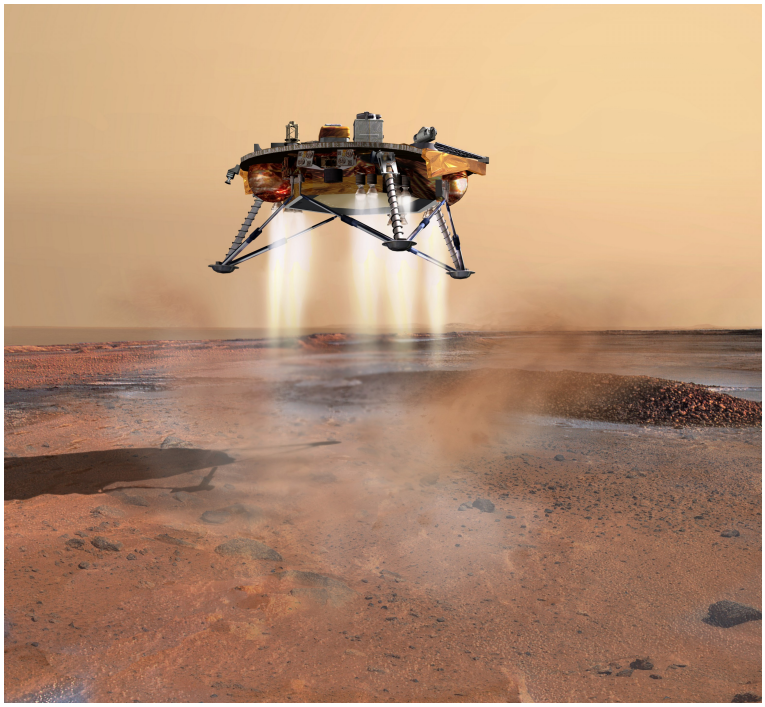


Figure 6.5: Artist impression of the Phoenix probe landing on Mars.

Alternatively, the residual kinetic energy before touchdown has to be dissipated either by a damped elastic structure (piston-like landing legs), fluid damping (airbags) or plastic deformation of a crushable material (foam, honeycomb material). Airbags are inflated shortly before landing using gas tanks or chemical generators.

Landers for Venus do not require a parachute for the final stages of the descent. Indeed, the atmosphere of Venus is so dense that an aerobrake ring attached to the probe is sufficient to slow it down. For instance, the Soviet Venera missions used a parachute only for the initial phases after entering the atmosphere. From an altitude of 50 km on, the final descent was done without parachutes.
Finally, in some cases, the descent velocity is deliberately kept at a large value of $60 - 300\,\mathrm{m\,s^{-1}}$. This is the case of penetrators, i.e. vehicles using their kinetic energy of arrival to place a payload beneath the planetary surface.

6.2 Descent to an airless body

Let us assume that the probe is on a closed orbit about an airless body. One first applies a thrust opposite to the velocity vector. For small bodies, such as asteroids or comets, the gravity is quite low and the descent can

[3]This is especially true for missions to planet Mars which has an atmosphere that is not dense enough to reduce the descent velocity via the sole use of parachutes.

take several hours, i.e. a significant fraction of the body's rotation period. Moreover, the gravity field is highly non-spherical, making navigation inside this gravity field a complex operation. During the descent, the altitude is constantly monitored using laser distance ranging. Thrusters are used to reduce the impact velocity. However, if in-situ measurements are foreseen, the last meters are usually done in free fall to avoid contamination by exhaust gases and throwing up dust from the surface. In principle, any vehicle that can make a soft landing can also make a take-off provided it has enough fuel left. For sample return, storable, non-cryogenic, propellants are needed for the ascent stage.

Upon landing on an asteroid or a comet, rebound must be avoided. Furthermore, drilling into the surface and outgassing of the comet's nucleus could actually lift off the lander. To avoid this, landers on small bodies feature an anchoring system. As an example, let us consider the Philae probe, which was carried by the Rosetta spacecraft to the comet 67P/Churyumov-Gerasimenko. Philae was designed to land on the comet and was therefore equipped with three devices to help it attach itself to the surface: hold-down thrusters, two anchoring harpoons and foot screws on each of the three feet. The harpoons were located beneath the lander and should have been driven by a gas generator that accelerates a projectile that pulls a cable out of a magazine. When the projectile leaves the harpoon, a rewind motor gets started that pulls up the cable until it reaches a predefined tension. However not everything went as expected. On 12 November 2014, Philae was released from Rosetta. The descent to the comet took 7 hours. Upon touchdown on the comet's surface (within the targeted area), the hold-down thrusters failed and the harpoons did not fire either. As a result the lander bounced off the surface at a velocity of 38 $\mathrm{cm\,s^{-1}}$ (about 75% of the comet's escape velocity). This first bounce lasted 111 min and Philae travelled about 1 km. The lander then touched the surface a second time and bounced off at 3 $\mathrm{cm\,s^{-1}}$ for another 7 min before it finally came to rest. The failure of the anchoring system left the probe in an unsecured, heavily shaded position on the comet. The initially targeted landing side benefitted from solar illumination more than 55% of the time allowing the probe's solar panels to recharge the batteries. At Philae's actual location, the solar illumination was much lower and therefore Rosetta lost its contact with Philae after 57 hrs, once the primary batteries of the lander were discharged[4].

6.3 Balloons

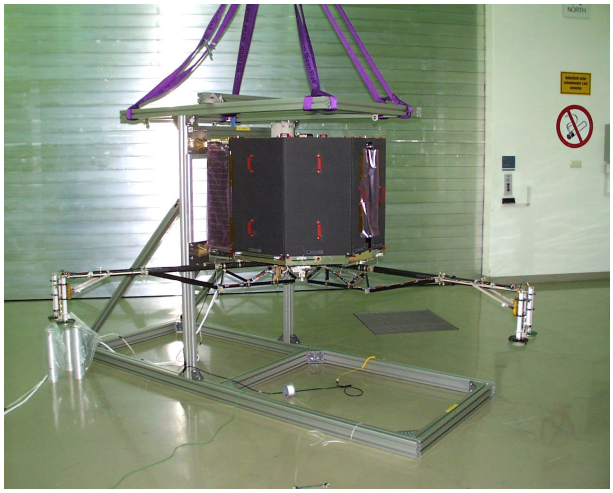
For some situations, balloons are more advantageous to study a planet than landers. A balloon floats in the air when the buoyancy force $(\rho_{\mathrm{air}} - \rho_{\mathrm{gas}})\, V_{\mathrm{balloon}}\, g$ is equal to the weight of the probe (i.e. the balloon plus its payload), $m_{\mathrm{probe}}\, g$. Here V_{balloon} stands for the balloon's volume, ρ_{air} is the local density of the air and g the acceleration of gravity. Therefore,

$$m_{\mathrm{probe}} = \rho_{\mathrm{air}}\, V_{\mathrm{balloon}} \left(1 - \frac{P_{\mathrm{gas}}\, \mu_{\mathrm{gas}}\, T_{\mathrm{air}}}{P_{\mathrm{air}}\, \mu_{\mathrm{air}}\, T_{\mathrm{gas}}}\right)$$

with P, μ and T standing for the pressure, mean molecular weight and temperatures respectively. A light-gas ballon has $\mu_{\mathrm{gas}} < \mu_{\mathrm{air}}$.
Generally speaking there are five categories of balloons:

- extensible balloons that are inflated with a given mass of gas, sealed and then allowed to rise until they burst.
- hot air balloons (Montgolfière) are open structures where the atmospheric gas is used as buoyant gas, but is heated in such a way that the temperature difference allows the balloon to float in the air.
- equal- or zero pressure balloons have $P_{\mathrm{gas}} = P_{\mathrm{air}}$. They are designed to operate at a specific altitude. If the balloon moves upwards, gas must be vented to avoid stress in the balloon's envelope and bring it back

[4]The probe briefly woke up on 13 June 2015 and established intermittent contacts with the Rosetta orbiter until 9 July 2015.



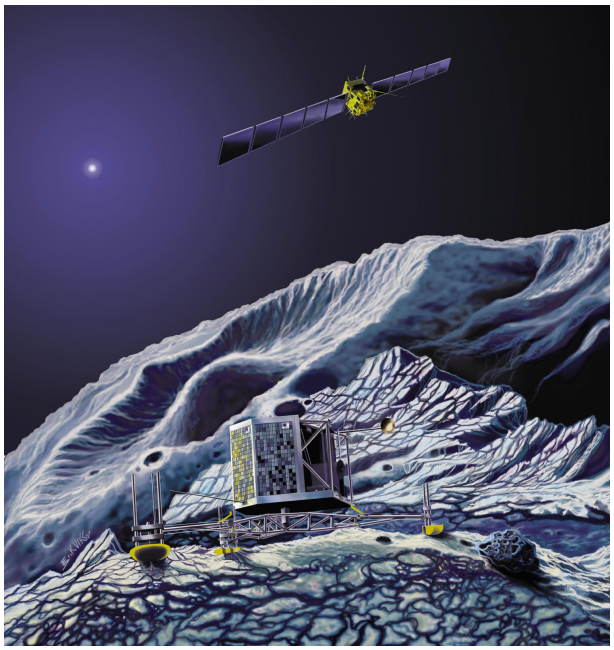


Figure 6.6: Left: the Philae lander during ground tests before the launch of the Rosetta spacecraft. Right: artist's view of Philae after (a soft) landing on the surface of comet 67P/Churyumov-Gerasimenko.

to its nominal altitude. If, on the contrary, it moves downwards, ballast has to be dropped. Both corrections actually limit the balloon's lifetime.

- super-pressure balloons that feature a non-extensible fabric. These balloons have an internal pressure that exceeds the ambient atmospheric pressure, $P_{\rm gas} > P_{\rm air}$. They operate at an iso-density altitude and move roughly horizontally with the wind. The temperature of the buoyant gas varies with solar radiation (day/night cycle). This configuration allows for longer lifetimes than zero pressure balloons, but at the same time a stronger (hence heavier) material must be used for manufacturing the fabric (mostly made of mylar-polyethylene sandwiches).

- balloons with phase changes. In this case, the fluid inside the envelope can exist either as a gas or liquid at the operational temperature. The altitude of the balloon thus oscillates around the altitude of the condensation point.

The last three options have been considered for the implementation of planetary missions, but, so far, only super-pressure balloons have actually been used.
A typical sequence for the use of a planetary balloon is shown in Fig. 6.7. It is more mass-efficient and hence more interesting to deploy and inflate a balloon in the air rather than from the planetary surface. In this way, one does avoid the need of a landing system.
The advantages of balloons are that they require neither propulsive engines nor fuel. The disadvantages are the difficulties to actively control the balloon's motion and its limited lifetime. The latter is generally of a few days at most and is controlled by the batteries (needed for communications), the ballast (needed for altitude control) and the leak rate through the balloon's skin.

The first, and so far the only, balloons that have been used to explore another planet are the Soviet VeGa 1 and 2 balloons that have flown through the atmosphere of Venus in June 1985. Indeed, balloons are especially interesting

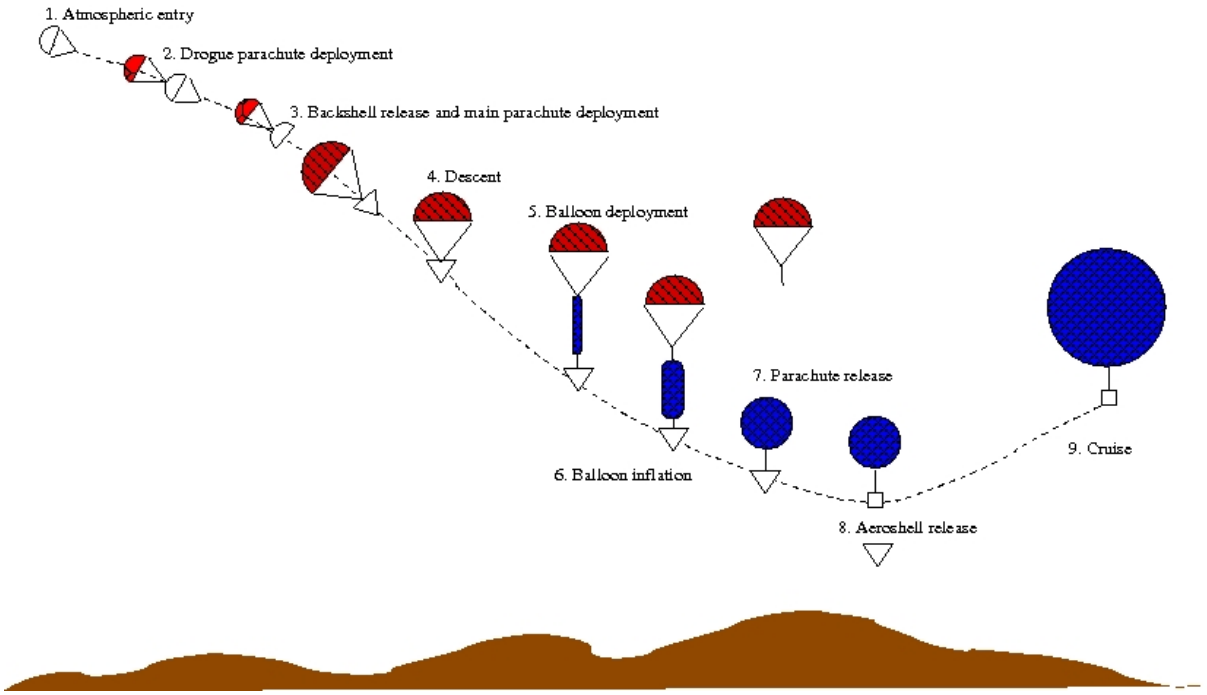


Figure 6.7: Schematic view of the operational sequence of an atmospheric balloon.

for Venus where landers have very short lifetimes, whilst balloons can evolve in the more benign upper layers of the atmosphere[5]. The VeGa aerostats floated at an altitude of 54 km where the temperature ranges between 30 and 40° C, and the atmospheric pressure is 535 mbar (instead of 465° C and $\sim$ 90 atm at the surface). The balloons had a diameter of 3.4 m and were made of teflon, transparent to the radio frequency used for telecommunications. Each balloon was filled with 2.1 kg of helium at a slight over-pressure of 0.03 bar. A gondola of 6.9 kg was suspended 13 m below the balloon and featured instrumentation to measure the local atmospheric parameters, winds, lightning and cloud properties. The total mass of the probe was 21.5 kg. The duration of the operations was about 46 hrs. This was at least partially limited by the lifetime of the lithium batteries.
Mars is a more difficult place for balloon operations because of the low atmospheric pressure and large diurnal temperature variations. Because of the latter, a Mars balloon would have to land during the night and rise during day-time when heated by solar irradiation. Such a project was initially foreseen for the Mars 94 probe.

6.4 In-situ measurements

Some scientific measurements require the acquisition of samples of material either to be returned to Earth or to be delivered to instrumentation inside the lander for analysis. For this purpose the landers feature sampler arms that can drill or scrape the ground and collect samples to be analysed. Beside a wealth of pictures and meteorological measurements, the Viking landers performed many analyses of the Martian soil thanks to a soil scoop mounted on a robotic arm.

More sophisticated measurements can be performed with rovers. The first planetary rovers were Lunokhod 1 and 2 deployed by the Luna 17 and 21 probes in November 1970 and January 1973 respectively. These rovers were

[5]The main problem at these altitudes comes from the sulphuric acid clouds.

remote-controlled from Earth. They drove during the lunar day and hibernated during the night. They featured a large convex lid with an array of solar cells on its inside. The lid was opened during day time to recharge the batteries. During the night, the lid was closed and the internal instruments were kept warm by the heat generated by ^{210}Po radioactive heater units (RHUs). The rovers had eight wheels, each with an independent suspension, electrical motor and brake. Because of the vacuum, the electrical motors were enlosed in pressurized containers and a special lubricant was needed for the mechanical parts. Both rovers featured a number of scientific instruments that returned a huge quantity of images and measurements. Lunokhod 1 operated for almost 11 months, travelling 10.5 km. Lunokhod 2 was equipped with several television cameras that returned images to Earth, allowing a team of controllers to drive the rover in near-real time. Lunokhod 2 operated for about four months and covered a distance of 37 km.

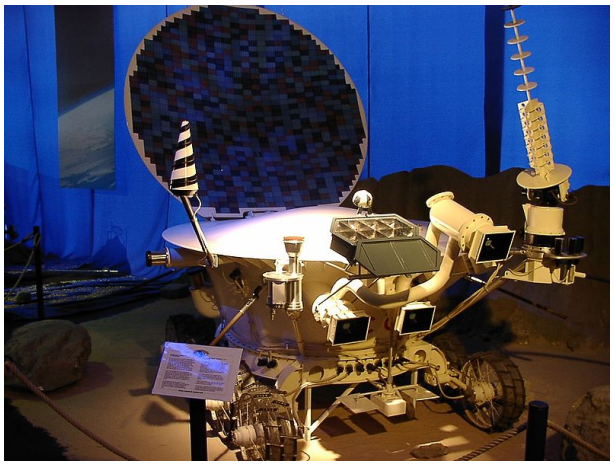


Figure 6.8: Model of the Lunokhod Moon exploration rover.

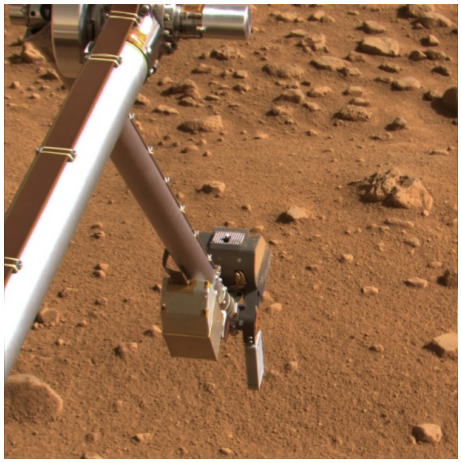


Figure 6.9: The robotic arm of the Phoenix lander. The arm features a rasp and a scoop that allow collecting material and a thermal and conductivity probe (pointing downwards on this image).

It is quite interesting that the Lunokhod design largely inspired the Sojourner rover that landed with the Mars Pathfinder probe on Mars in 1997. Sojourner served as a prototype technology demonstrator mission for the Mars Exploration Rover missions (Spirit and Opportunity).
In December 2013, the Chinese mission Chang'e 3 landed in the Mare Imbrium ('Sea of Rains') on the Moon. Beside the lander itself, Chang'e 3 carried a small (1.5 m heigh) 6-wheel rover called Yutu. The design of the rover was largely inspired by Mars Pathfinder. Its mass (120 kg), included 20 kg of scientific instruments. The rover featured several ^{238}Pu radioisotopic heating units for preserving its interior from getting too cold during the lunar night. The electrical power was produced by two solar arrays, implying that the rover could only operate during lunar days. Unfortunately, during its second lunar day, the rover suffered a 'mechanical abnormality' bringing its mission to a premature end. In January 2019, Chang'e 4 performed the first soft landing on the far side of the Moon. The lander carried the rover Yutu 2 and communication with the ground was done via a relay satellite Queqiao on a halo orbit around the Earth-Moon L_2 point.

6.5 Protection of the planetary environments

A major issue with atmospheric probes and landers is to avoid interplanetary biological and organic contamination. Actually, the problem of planetary protection covers two aspects

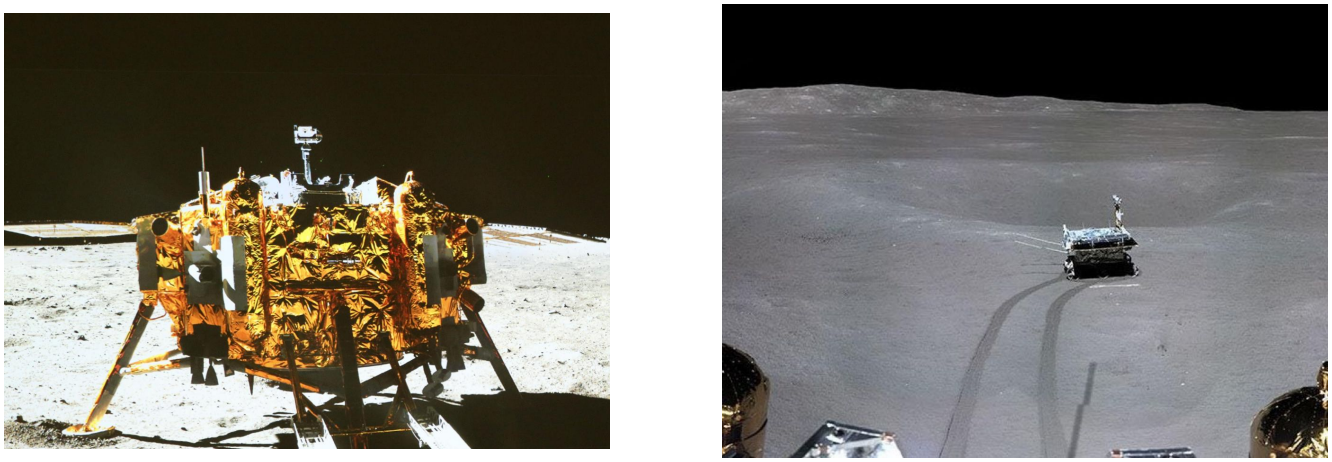


Figure 6.10: Left: the Chang'e 3 lander on the Moon as pictured by Yutu. The lander itself carries several instruments, including a near-UV telescope. Right: the Yutu 2 rover as pictured by the Chang'e 4 platform in the Von Kármán crater on the far side of the Moon.

- avoid contamination of planetary bodies by terrestrial organic material, either alive or death;
- avoid importing material from an extraterrestrial biosphere to Earth.

Article 9 of the UN Outer Space Treaty (adopted in 1967) prohibits harmful contamination of the outer space and celestial bodies. Yet, contamination, deliberate or accidental, has been perpetrated by a number of actors. This has been the case of the Apollo missions where the astronauts left bags containing human faeces on the lunar surface, but also, more recently when the Chang'e 4 lander featured a biosphere containing cotton seeds. Another example is the privately funded Beresheet probe that crashed onto the Moon and carried a (non-declared) container with tardigrade organisms.
For robotic probes, accidental contamination of a spacecraft mainly comes from the people manipulating it prior to launch: even with a sterile suit, a human being sheds several thousand skin flakes per minute. At first sight, one would expect that the harsh conditions in outer space (absence of an atmosphere, heavy bombardment by energetic particles,unattenuated UV radiation from the Sun, large temperature excursions) would help solve these issues by killing any form of life carried by a man-made probe. However, this is far from certain as illustrated by the Surveyor 3 lunar probe. The un-manned probe Surveyor 3 landed near Oceanus Procellarum on 20 April 1967. Thirty-one months later, the crew of Apollo 12 recovered the camera of Surveyor 3 and brought it back to Earth under sterile conditions. When the camera was opened on the ground, it was found that the polyurethane foam insulation covering the electronic circuit had been colonized by about 50 – 100 viable specimen of the streptococus mitis bacteria[6] that had survived in the absence of air and water and despite being subject to large temperature variations and harsh radiation environments.

The main problem arises from the fact that many bacteria can take a spore form under stressful conditions. Spores are water deficient forms that can be revived and resist to a wide range of adverse conditions thanks to a protective coat. Currently it is estimated that about 25% of all microbes can survive within a spacecraft for one year and that 1% can survive for six years. Hence, very specific techniques are required to sterilize a spacecraft. These techniques are quantified by the quantity of surviving microbes after a given time of treatment: the D10 parameter corresponds to the time needed to kill all but 10% of the initial population.

[6]This is a harmless bacteria commonly present in the human nose, mouth or throat.

Most known microbes cannot withstand temperatures beyond 110° C, although dormant spores can resist to wide temperature excursions. Nevertheless, in most cases, heating damages the DNA of the spore. Radiation can also be used to damage the DNA by destroying molecules inside the DNA chain through ionization and production of free radicals. However, a treatment with ionizing radiation to sterilize the spacecraft must be done with caution as it can also damage the on-board electronics. One can also expose the spacecraft to a lethal vacuum that leads to water evaporation. However, this is a rather slow process for bacteria.

Not all destinations in space require the same level of planetary protection. For instance, the Moon is considered part of the terrestrial biosphere, hence does not require special planetary protection measures for missions from Earth to the Moon[7]. On the contrary, missions to Mars, Europa or Enceladus have to meet very strict regulations. For instance, the Martian wind will spread a local contamination over wide areas of the planetary surface. Therefore, to prevent contamination of Mars by terrestrial organisms, the Viking landers had to undergo a very special treatment. The hydrazine used for the thrusters was purified and each lander, upon assembly and enclosure within the aeroshell, was baked at a temperature of 121° C for seven days. The landers inside their aeroshell were then encased in a bioshield. The bioshield remained in place until the spacecraft left its parking orbit around the Earth. The planetary protection protocols developed for the Viking landers are still used as a reference for many missions to Mars.

Over recent years, the issue of planetary protection has come to the forefront of space policy. On the one hand, it turned out that even the Viking procedure was not 100% efficient in sterilizing the spacecraft. On the other hand, private space ventures, planning to exploit space in an *economically viable way*, require a clear definition of the notion of harmful contamination. As a result, the Committee on Space Research (COSPAR) has proposed to divide space missions into several categories which define the requirements in terms of planetary protection.

- Category I includes missions to a target body that is not of interest for exobiology. No protection requirements are formulated for such missions.
- Category II comprises all missions to targets that are of interest for exobiology, but where there is only a slim chance that contamination by the spacecraft might compromise future studies. This is the case for fly-by, orbiter or landers aiming at exploring the planets Venus, Jupiter, Saturn, Uranus or Neptune, the moons Ganymede, Callisto, Titan and Triton, the Kuiper-Belt objects (including Pluto and Charon), as well as the dwarf planet Ceres, carbonaceous chondrite asteroids and comets. In these cases, a document providing relevant information on the mission is required. One of the main issues to address in this documentation is an analysis of intended or accidental impacts on bodies that might harbour a liquid water reservoir below their surface.
- Category III concerns fly-by or orbiter missions to a target of interest for exobiology (Mars, Europa, Enceladus) and where there is significant chance that contamination might compromise future studies. For such missions, in addition to a more extensive documentation, procedures for trajectory biasing (i.e. avoiding impact by any part of the space vehicle on the target body for a given period of time) need to be implemented and the spacecraft must be assembled in a cleanroom.
- Category IV refers to probes and landers on the same targets as for category III missions. In addition to the requirements already put on category III missions, a full inventory of all organic material carried by the spacecraft, a bioshield and partial sterilization of the direct contact hardware are required. Planned landing ellipses must also be evaluated against the risk to contaminate regions of special interest (i.e. regions with sufficient water activity and sufficiently warm temperatures to allow replication of terrestrial organisms).

[7]One must however avoid contamination of the Moon itself by material from an extraterrestrial biosphere as this could lead to a secondary contamination of the Earth.

- Category V concerns Earth-return missions. Here two situations are to be considered. For missions to bodies that are not expected to carry indigenous life forms (e.g. the Moon, Venus), unrestricted Earth return is allowed. Restricted Earth return applies to all other category V missions: the samples must be stored in a hermetic container that must then be sealed and sterilized from the outside before bringing it back to Earth. Destructive impact upon return must be avoided and all returned hardware is to be handled under strict containment in specialized laboratories. In particular, the spacecraft must provide a means to break the chain of contact with the target body: no uncontained hardware that has been in contact with material from the target body shall return to the Earth's biosphere.

Future manned missions conducting surface operations on planet Mars are critical as it will not be possible to perform the entire operations within closed systems and the crew will inevitably be exposed to Martian material. Such missions must feature a quarantine capability for individuals as well as for the entire crew.
For robotic or manned sample return missions to Mars, Europa or Enceladus, approval for continuation of the mission is required at three critical times: (1) prior to launch from Earth, (2) prior to leaving the target body for return to Earth, and (3) prior to commitment to Earth atmospheric re-entry.

Chapter 7

Instrumentation in space-borne astrophysics

Modern astrophysics would be much more limited without an access to space. Whilst ground-based facilities have allowed tremendous progress in our understanding of the Universe, especially over the second half of the 20th century, many phenomena can only be studied outside of the Earth's atmosphere and highly accurate measurements require stable conditions that cannot be created on the ground.
In this section, we will discuss the application-specific payload of several astronomy missions. In fact, the design and the ingredients of a spacecraft depend on the scientific objectives, on the wavelength range to be explored,... For each of the disciplines reviewed here, we will first introduce the basic principles and motivations of the space experiment. We then present a past satellite as well as an existing mission and/or future mission concept and we finally highlight some of the science results that have been obtained or are expected.

7.1 Astrometry

Let us start with astrometric missions. Astrometry is one of the oldest disciplines of astronomy. Around 130 BC, the Greek astronomer Hipparchus (ca. 190 – 120 BC) compiled a catalogue of about 1000 fixed stars. Beside their positions in the sky (determined with an accuracy of about 1°, he also listed their brightness (introducing for the first time the concept of magnitudes). This work was improved lateron by various Greek astronomers and in the 16th century, the Danish astronomer Tycho Brahe (1546 – 1601) determined stellar positions with an unprecedented accuracy of about one arcminute (although he had no optical instruments at his disposal). With the advent of telescopes, astrometric precision improved tremendously, especially through the use of meridian circles in the first half of the 20th century. Before, the first astrometric space mission was launched, measurements from the ground achieved accuracies of about 0.1 arcsec.

Nowadays, astrometry is back at the forefront of research. In broad terms, modern astrometry has two major goals:

- provide a high-precision, non-rotating reference frame to which the motions of objects in our Solar System and in the Galaxy can be referred.
- derive, through the use of this framework, a number of key stellar parameters, such as the distances, luminosities, masses, 3-D distribution and motion of stars within our Galaxy.

The best (i.e. least model-dependent) way to measure the distance of a star is through the determination of the parallax angle Π (in arcsec). The principle is rather simple: the parallax is measured between two opposite positions of the Earth's motion around the Sun. The apparent change in the angular position of a nearby star with respect to the reference frame provided by very distant stars actually reflects the orbital motion of the Earth. Since the distances of stars are much larger than the dimensions of the Solar System, the parallax angles are always

very small. The distance is then given by $d = \frac{1}{\Pi}$ where d is expressed in parsec[1]. Another important astrometric parameter is the proper motion, i.e. the annual angular displacement of a star resulting from its motion in the Galaxy. The kinematics of the various stellar populations in the Galaxy are key ingredients for the study of the dynamics of the Galaxy and of stellar clusters.
Whilst the nearest star, Proxima Centauri, has a parallax of 0.772 arcsec, both, the parallax and the proper motion, are usually extremely small quantities and are therefore commonly expressed in milliarcsec[2] or fractions of milliarcsec. It has to be stressed that the amplitude of these motions is so small that the Sun's motion through space, the finite (but unknown) distance of the background stars and their motion, the multiplicity of stars and the deformation of space-time due to the gravitational field of Solar System bodies must be taken into account. As a result, the measurements required to accurately determine the distances and proper motions of distant stars must be extremely precise, in fact far more accurate than what can be done from the ground (where the Earth's atmosphere as well as the gravitational and thermal flexure of the telescope affect the measurement). To solve many of these issues, one can attempt to do astrometry from space. This was first done by ESA's **Hipparcos** mission.

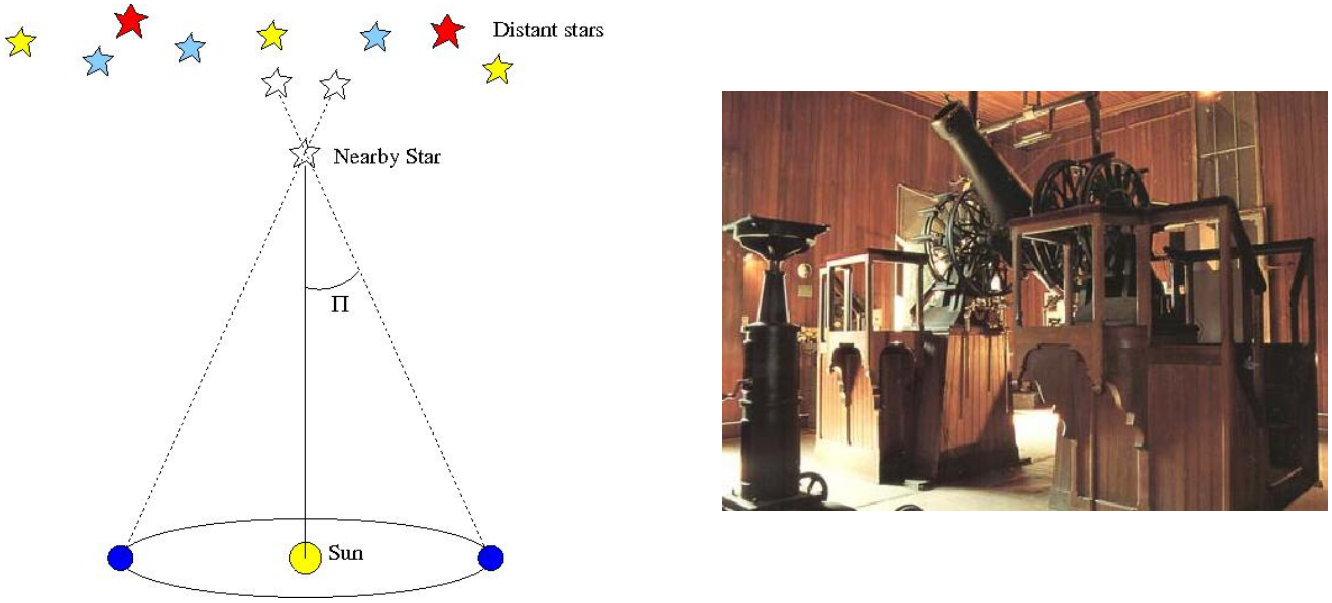


Figure 7.1: Left: schematic illustration of the principle of the parallax measurement of a relatively nearby star. Right: the meridian circle of the Toulouse observatory installed in 1891.

7.1.1 The Hipparcos mission

The Hipparcos mission was approved by ESA within its scientific programme in 1980 and the satellite was launched by an Ariane 4 in August 1989. Due to a failure of the apogee booster, the satellite did not reach the proposed geostationary orbit and rather remained on a highly eccentric orbit that took it through the Van Allen radiation belts twice per orbit. Following a revision of the mission operation, the satellite could nevertheless be operated successfully until August 1993 despite being on the 'wrong' orbit. A full analysis of the data collected during the mission led to the publication of the catalogue in 1997. The total cost of the mission was about 400 M€.

[1]The parsec (3.0857×10^{18} cm = 206265 AU) is the distance under which an object of 1 AU extension is seen with an angular size of 1 arcsec.

[2]1 arcsec corresponds to the size of an object of 0.5 mm diameter seen at a distance of 100 m. 1 milliarcsec corresponds to an object of 2 m height on the Moon.

The principle of the Hipparcos mission was to perform simultaneous observations of two viewing directions separated by $58°$. The payload was centered around an optical Schmidt telescope (diameter of the main mirror = 29 cm) featuring a beam combining mirror that brought the light from the two fields of view (each one of dimensions $0.9° \times 0.9°$ to a common focal surface. The satellite was spinning around its axis and the direction of the spin axis was precessing slowly. In this way, every direction of the sky was observed about 100 times over the lifetime of the mission. As the satellite was spinning around its axis, thereby scanning the sky, the images of each star inside the two fields of view were modulated by a regular grid of 2688 transparent parallel slits located at the focal surface, and the resulting periodic signal was recorded by an image tube detector at a frequency of 1200 Hz. The data of this instrument were sent to the ground along with all the relevant information on the spacecraft's attitude.

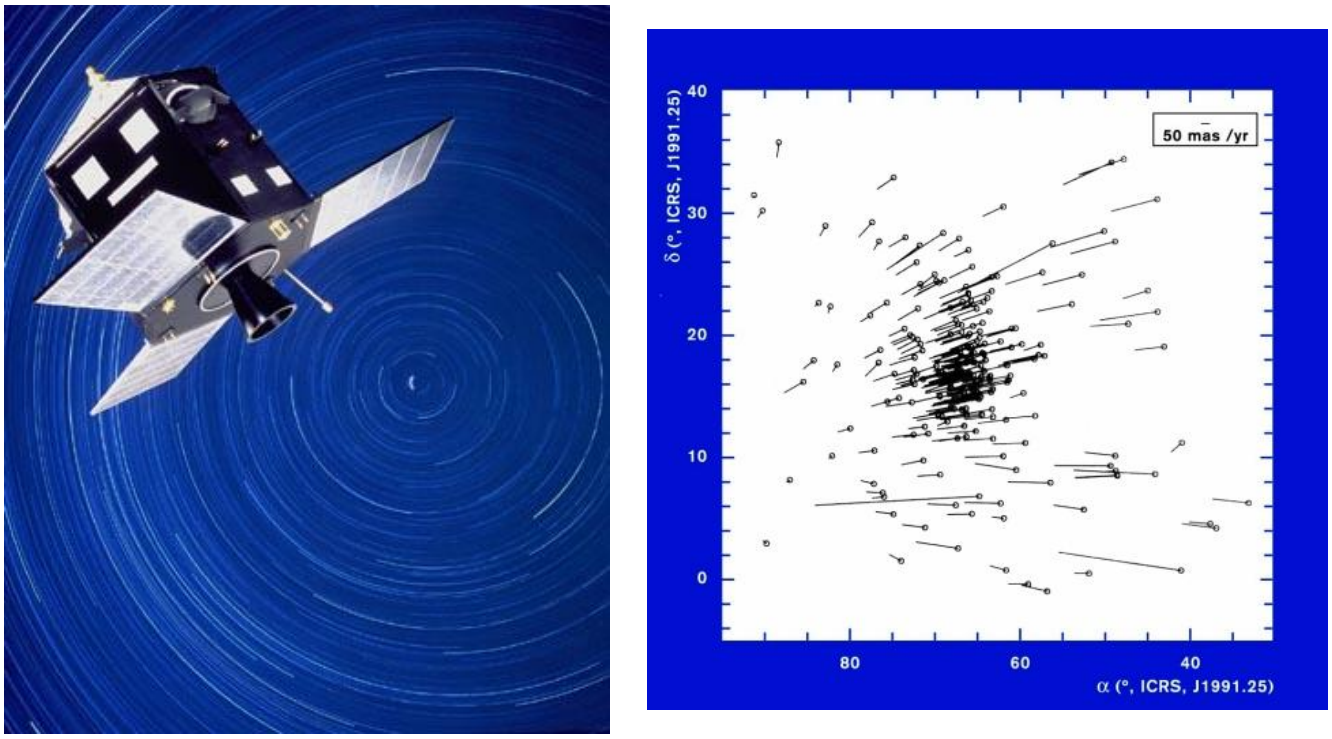


Figure 7.2: Left: artist view of the Hipparcos spacecraft. Right: the proper motion of the stars in the field of the Hyades cluster as measured with Hipparcos.

Actually, Hipparcos significantly exceeded its original mission goals. 120 000 stars with a completeness to magnitude 7.3 – 9, depending on the galactic latitude, were measured with an accuracy of 1 milliarcsec in position (originally foreseen to be 2 – 4 milliarcsec), parallax and proper motion. The Hertzsprung-Russell diagram[3] of the stars within a radius of 100 pc could be established with distances accurate to the 10% level. Furthermore, an auxiliary instrument, the Tycho star mapper produced less accurate, but still unprecedented, measurements for 2.5 million stars (99% of all stars to magnitude 11). Originally, the Tycho experiment was expected to provide measurements of at least 400 000 stars.
The quality of the astrometric Hipparcos catalogue can be appreciated by the following numbers: the final median errors for stars brighter than 9th magnitude are 0.77 and 0.64 milliarcsec for the mean angular positions (two

[3]The Hertzsprung-Russell diagram (HRD) is a plot of the stellar luminosity versus the colour (or surface temperature) of the star. The HRD is an extremely important tool for understanding stellar evolution and stellar populations. The stars do not fall on random positions in this diagram, but occupy specific locations, such as the main sequence where a star spends most of its lifetime. Theoretical models indicate that the main-sequence corresponds to the phase of nuclear burning of hydrogen in the core of the star. The location of a star in the HRD depends on its mass, evolutionary stage (hence age) and chemical composition.

coordinates), 0.97 milliarcsec for the parallaxes, and 0.88 and 0.74 milliarcsec yr^{-1} for the proper motions (two coordinates). As a 'by-product', photometric measurements for each of the about 110 epochs of observation were published, leading to the discovery of a number (8200) of previously unknown variable stars and eclipsing binary systems (917 such systems were observed, 343 of them were not known before). The median accuracy of the photometric data (for constant stars) was 0.0015 mag. In addition, detailed information on 2900 astrometric binaries or multiple systems were also obtained and astrometric orbital solutions were derived (see e.g. Fig. 7.4). Several aspects combined to make Hipparcos such a success. The satellite observed the entire sky and was of course not affected by atmospheric perturbations. Instrumental stability (due to the lack of gravitational instrumental flexure as well as an active thermal control) further reduced the systematic error budget. Differential angular measurements were made over large angles, at many different orientations and at many different epochs. The parallaxes are expressed in a quasi-inertial frame of reference by relating the final catalogue to observations of extragalactic objects such as quasars with accurate positions measured in the radio domain.

Among the most important results of the Hipparcos mission were the detection of the warp of the Milky Way. Most of the stars of our Galaxy orbit around the centre of the Milky Way in a flattened disk. However, as Hipparcos has shown, this disk is not really flat. There are several possible explanations for this situation: it could either result from the gravitational pull of the Magellanic Clouds (which are two satellite galaxies of the Milky Way) or might be due to the distribution of dark matter in the halo of the Galaxy.

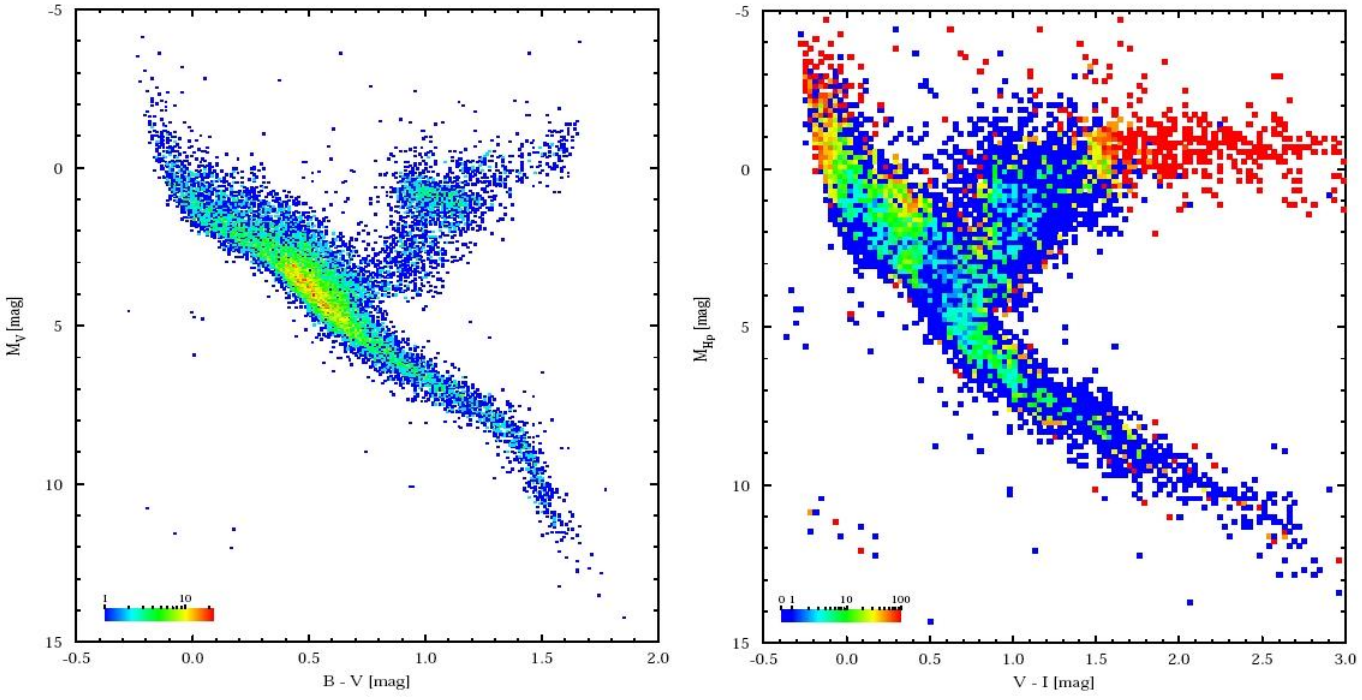


Figure 7.3: Left: the Hertzsprung-Russell diagram (HRD) of 16 631 single stars from the Hipparcos catalogue that have relative distance precisions of better than 10% and colour errors $\sigma(B-V)$ of less than 0.025 mag. The colours indicate the number of stars in a cell of 0.01 mag in $B-V$ and 0.05 mag in V. Right: HRD in the M_{Hp} vs. $V-I$ space. The colour code indicates the percentage of variable stars in each cell of 0.03 mag in $V-I$ and 0.15 mag in M_{Hp}.

The Hipparcos data have given new insight into the distance, structure, dynamics and age of several open clusters, such as the Pleiades and the Hyades. For instance, the Hyades are located at around 150 light years from Earth.

The arrows in Fig. 7.2 indicate what will be the position of the stars in the Hyades field of view in 100 000 years time. Those stars that actually belong to the cluster have similar proper motions and will thus remain grouped. Those stars that are field objects unrelated to the cluster instead have proper motions that are independent of that of the cluster. This is an important criterion for cluster membership.
Hipparcos has measured the distance of a number of cepheids (named after the prototype object δ Cep). The objects of this class are pulsating stars that display periodic changes of their luminosity and of their surface temperature (see lectures on *Variable Stars*). These pulsations follow a period-luminosity relation that can be used to infer the absolute magnitude of a cepheid star from its pulsational period: the higher the luminosity of the star, the longer its pulsational period. The Hipparcos parallaxes allowed to calibrate this period-luminosity relation which is used as one of the main standard rulers to determine distances in modern astrophysics and plays thus a key role in the evaluation of the distances of extragalactic objects making it an important tool for observational cosmology.
Last, but not least, the Hipparcos data have improved our understanding of stellar evolution, allowing the elaboration of an accurate, observational Hertzsprung-Russell diagram (see Fig. 7.3).

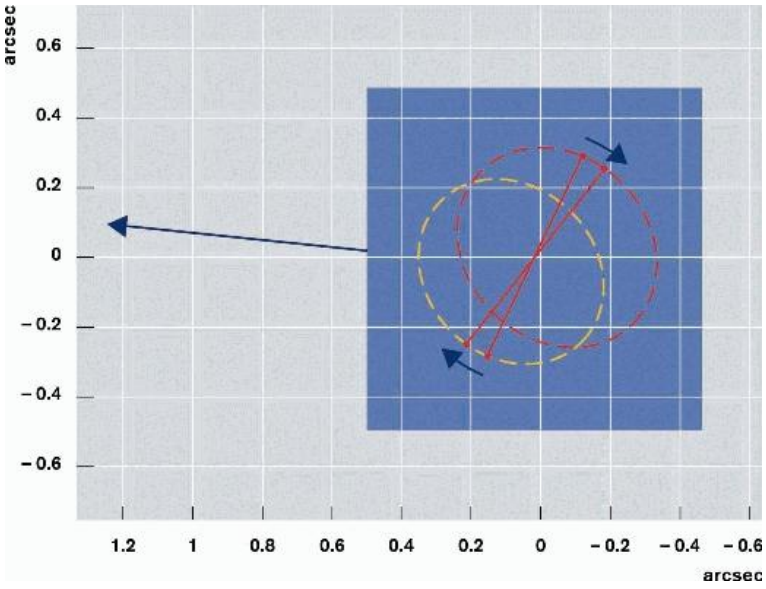


Figure 7.4: The astrometric orbit of the binary system BD-12° 2918 (= HIP 46706). The stars revolve around their centre of mass every 18.3 years. The system is 34 light-years away from Earth ($\Pi = (94.95 \pm 4.31)$ milliarcsec and its proper motion is substantial (723, 53) milliarcsec yr^{-1}. The astrometric orbital solution yields masses of 0.42 and 0.41 $M_\odot$ for the components.

7.1.2 The Gaia mission

Gaia (Global Astrometric Interferometer for Astrophysics) is an ambitious ESA mission aiming at establishing a 3-D chart of our Galaxy. Although the name of the mission refers to interferometry, the design that was adopted does not rely on genuine interferometry, but rather uses the same basic principles as for the Hipparcos satellite. The Gaia satellite carries a single instrument that comprises three major functions: astrometry, spectro-photometry and spectroscopy. The instrument uses two telescopes that have viewing directions separated by a large basic angle (106.5°) and their images are combined into a single focal plane. Each function (astrometry, photometry, spectroscopy) uses a dedicated section of the large (0.5 m $\times$ 1 m) mosaic of 106 CCDs of the focal plane. The thermal stability and the low-mass requirements of the mirrors are met by the use of silicon carbide ultra-stable material. This allows to meet the stability requirement of the basic angle between the two telescopes.
As Hipparcos, Gaia also scans the sky through rotation about a spin axis and precession around the Sun direction. The accuracy is achieved through a relative positional measurement in the combined focal plane and a global data analysis on the ground. Over its five years nominal lifetime, Gaia should obtain some 70 measurements of each portion of the sky. Measuring the instantaneous image centroids from each data set sent to the ground allows to determine the relative separation of the thousands of stars simultaneously present in the two fields of view. Since

the spacecraft operates in a continuously scanning motion, a constant stream of relative angular measurements is built up as the fields of view sweep across the sky. The astrometric measurements are done by 62 CCDs of the focal plane, read out in time-delayed integration (TDI) mode, synchronised to the scanning speed of the satellite. The stars entering the combined field of view are first detected by a column of sky mapper CCDs. The detection of a star triggers the creation of a TDI read-out window around the object on the subsequent CCDs.

Table 7.1: Expected astrometric performances of the Gaia mission before (third column) and after (fourth column) commissioning.

Spectral type	V (mag)	$\sigma(\Pi)$ (μarcsec)	$\sigma(\Pi)$ (μarcsec)
B1 V	< 10	< 7	5 – 14
	15	< 25	26
	20	< 300	600
G2 V	< 10	< 7	5 – 14
	15	< 24	24
	20	< 300	540
M6 V	< 10	< 7	5 – 14
	15	< 12	9
	20	< 100	130

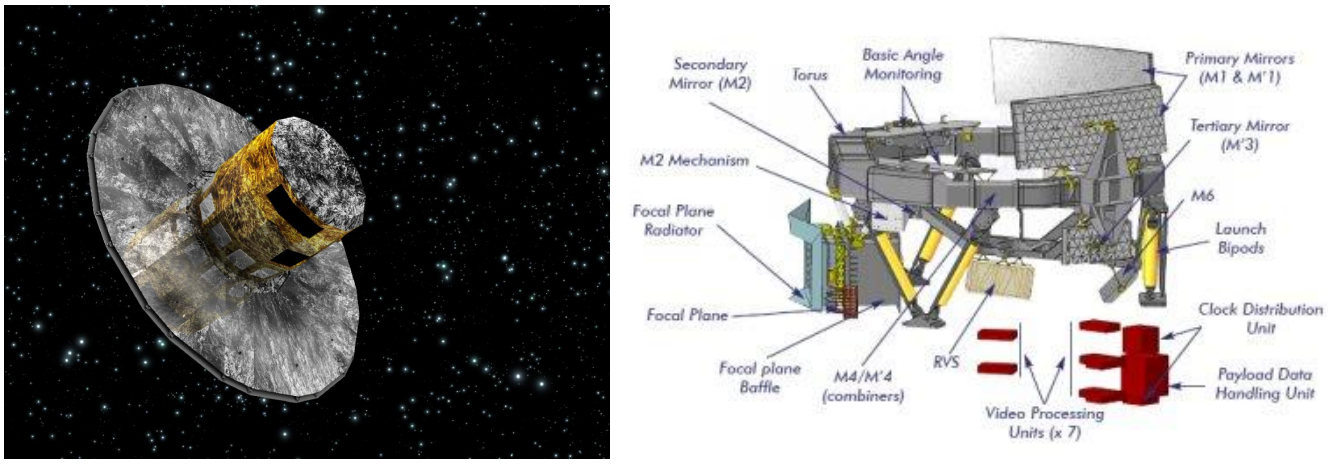


Figure 7.5: Left: artist's view of the Gaia spacecraft. Right: schematic view of the payload module of the Gaia spacecraft.

The main mission goal is the study of the composition, formation and evolution of our Galaxy. To this end, Gaia measures the positions of about one billion stars down to a magnitude of about 20 (this corresponds to 1% of the total stellar population of the Milky Way) with an unprecedented accuracy. For about 20 million stars, out to 2.5 kpc, the distance should be determined with an accuracy of better than 1%. This will result in a comprehensive luminosity calibration over all spectral types. The final Gaia catalogue will also provide insight into the distance and velocity distribution of all stellar populations of our Galaxy, including the spatial and dynamical structure of the disk and the halo. Indeed, our Galaxy consists of different locations that harbor different stellar populations. The formation history of these populations imprints on their orbit around the Galactic centre and their chemical composition. Gaia observes stars out to distances of 8.5 kpc, e.g. near the centre of the Galaxy as well as in

globular clusters in the Galactic halo. The 3-D velocity will be determined by the combination of the astrometric information with the radial velocity inferred from the spectroscopic data (see below). Accurate knowledge of the stellar velocities and positions will allow to identify structures in the Galaxy and give insight into the Galactic dynamics driven by gravitational interactions such as mergers with smaller galaxies.

Gaia should also discover a large number (several thousand) of Jupiter-like exoplanets with orbital periods between 1.5 and 9 years through the astrometric wobble that these planets induce upon their parent star.

Last, but not least, Gaia should allow to test Einstein's theory of General Relativity. In fact, the accuracy of the positions that will be measured is such that usually negligible gravitational effects (such as the bending of light by massive objects in the Solar System) show up in the data.

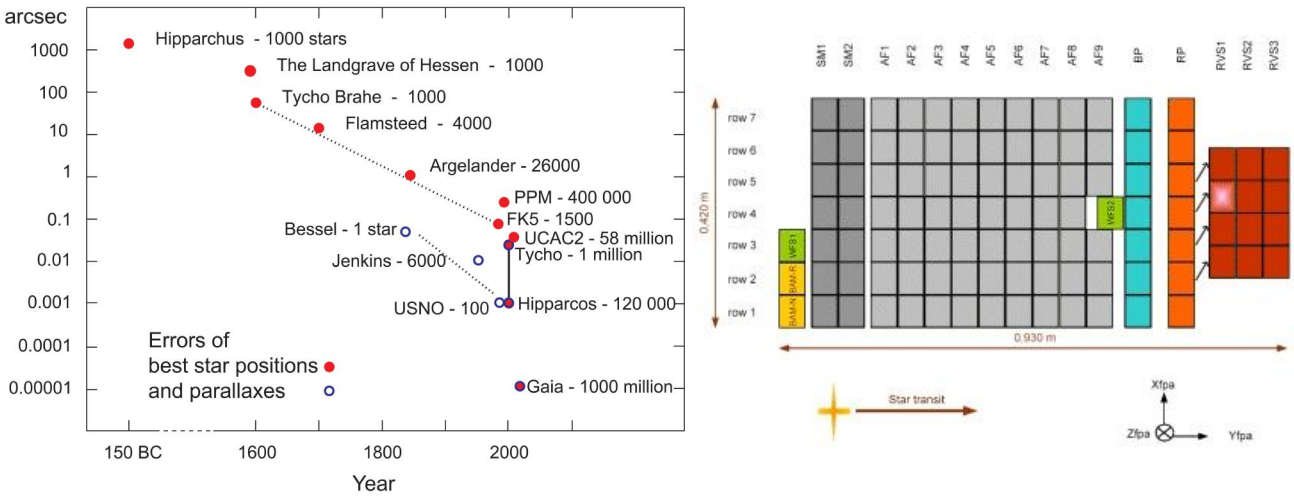


Figure 7.6: Left: the evolution of the accuracy of astrometric measurements from ancient times up to Gaia. Right: schematic view of the focal plane of the Gaia payload. The first 3 CCDs as well as one CCD of the 12th column are used to monitor the basic angle between the two directions of observation. The next 14 CCDs are the sky mappers (SM1 and SM2) that perform the detection of the star and trigger the definition of a read-out window around the target. Thanks to the use of a mask in the intermediate images from each telescope, SM1 only detects objects from the first telescope, whilst SM2 only sees objects from the second telescope. This is followed by 62 astrometry CCDs and two columns each of 7 CCDs for the spectro-photometry (one blue and one red column). Finally, the RVS instrument consists of 12 CCDs and covers a smaller field of view than the astrometric and photometric instruments. As a result, each object will be measured about 40 times with the RVS over the mission lifetime (instead of 70 times for the astrometric and photometric channels).

But, Gaia not only obtains astrometric measurements. It also performs broad-band spectro-photometry and higher resolution spectroscopy (for stars brighter than magnitude 16.5) centered on the wavelength domain of the Ca triplet at $\lambda\lambda$ 8498, 8542 and 8552 Å with a resolving power of $\lambda/\Delta\,\lambda \sim 11\,500$. The latter information is extremely important to measure the radial velocities of the stars and hence evaluate the 3-D velocity vector (together with the proper motion information). The wavelength domain (8470 – 8740 Å) of the spectroscopic instrument (RVS) has been selected to coincide with the energy distribution peaks of G and K-type stars that make up the majority of the stars to be studied.

The RVS is an integral field spectrograph. The spectral dispersion is oriented along the scanning direction. The radial velocities of the stars will be determined during the data processing on ground through the cross-correlation with a mask or a template spectrum. For stars brighter than 15th magnitude, the RVS provides radial velocities for each transit, allowing to study variations of the radial velocity due to binarity for instance.

The Gaia payload further provides spectro-photometry over the wavelength range 3200 to 10 000 Å. For this purpose, there are two low dispersion optics (prisms) located in the light path of the combined beam from the two telescopes. These prisms are located in front of the detector arrays and disperse the light along the scan direction.
The spectro-photometry allows to quantify the spectral energy distribution of the detected objects, yielding in the end astrophysical quantities such as the luminosity, effective temperature, mass, age and chemical composition. It will also help quantify and correct the chromatic displacement in the astrometric measurements.
The Gaia focal plane is the largest ever developed for a scientific mission and contains a very large number (106) of CCDs. Hence radiation damage is a major issue and intensive research was conducted to make the CCDs more robust against radiation.
The data processing is one of the most challenging ever in astronomy. Scientists from all over Europe have joined their efforts to set up a data processing centre (DPAC) that has developed and runs the pipeline for the data reduction.

Gaia was successfully launched by a Soyuz Fregat launcher from Kourou on 19 December 2013. The satellite was put into a Lissajous orbit around L_2. The first seven months in space were devoted to the commissioning phase: the telescope alignment was finalized, the best focus was determined, the spin rate was optimized and the on-board software adjusted. During this phase, a series of problems were discovered and solutions were investigated.

- The optical brightness of the spacecraft as seen from Earth was found to be 2.5 magnitudes fainter than expected, requiring a change in the strategy for the ground-based optical tracking needed for orbit reconstruction.

- The mass flow sensor of one of the thrusters of the micro propulsion system was found to produce an erroneous feedback to the attitude and orbit control system (AOCS). The AOCS software was adjusted to account for this error.

- The basic angle was found to vary at the milliarcsec level with a period of 6 hours (the spacecraft's spin period). The variations of the basic angle are monitored and accounted for in the data processing.

- A strong and variable level of stray light was detected. The variable stray light pattern was found to repeat with the spin period, suggesting diffraction of Sun light by protruding fibres at the edge of the Sun shield[4]. An additional background component arises from nominal stray light from the collective effect of bright objects such as groups of stars in the Milky Way outside the field of view. For stars up to magnitude 15, the impact of the stray light is usually moderate, the largest impact being on the RVS spectrometer (accuracy of $15\,\mathrm{km\,s^{-1}}$ instead of $3\,\mathrm{km\,s^{-1}}$). For fainter stars, the astrometric and photometric accuracies are also somewhat impacted.

- One of the telescopes was affected by a loss of transmission equivalent to 1 magnitude. This was attributed to contamination of the optics by water outgassing of the service module. Decontamination by heating of the payload was implemented.

Operations of Gaia are currently extended by 17 months until December 2020. The second Gaia data release, containing astrometry for 1.3 billion sources up to a magnitude of 21 took place in April 2018. Parallax uncertainties for stars with $G < 15$ are typically 0.04 milliarcsec. Further improvements of the accuracy are expected in the forthcoming releases that will use data from more years of measurements.

[4]Cutting these fibres in the clean rooms on the ground was considered too risky as it could have led to contamination of the spacecraft.

7.2 X-ray astrophysics

X-ray astrophysics is one of the most obvious activities in space-borne astrophysics. In fact, the X-ray radiation from cosmic sources cannot be observed from the ground because of the huge absorption by the Earth's atmosphere. This is why X-ray astrophysics is a young discipline. In fact, the very first detections of cosmic X-ray radiation date back to the early sixties of the 20th century. The first cosmic X-ray source to be discovered was Scorpius X-1 in June 1962. The source was detected by an American team led by Riccardo Giacconi during a rocket flight organised with the goal to detect the X-ray emission from the Moon[5]. The nature of Sco X-1 remained mysterious for several years, although this object was subsequently re-observed and was found to be the brightest X-ray source in the sky. The difficulties to identify the optical counterpart of the X-ray emission arose from the combination of two effects. First, there was no X-ray satellite available at that time and the angular resolution of the instruments on board of the rockets was rather poor. Second, X-ray emission arises from physical processes that do not necessarily produce a bright optical emission and a bright X-ray source can indeed have a faint optical counterpart. The solution came in 1966, when it was found that a faint (12th magnitude) star in the direction towards Sco X-1 actually displayed a modulation of its radial velocity with the same period as the variations of the X-ray emission of Sco X-1. The source was the first object of a new category, the X-ray binaries (see below).
Actually, the first dedicated X-ray satellite, Uhuru, was launched in 1970 (see below). The very first X-ray observatory, capable of taking pictures in the X-ray domain was the Einstein satellite launched in 1978. Nowadays, X-ray astrophysics is in a sort of golden age, with two large observatories being available: the NASA satellite Chandra, and the ESA observatory XMM-Newton (see below). The number of known X-ray sources has increased tremendously and beyond the discovery of new sources, X-ray satellites nowadays do real astrophysics. However, to some extent the discipline is still in its infancy. In fact, the combined collecting area of the three X-ray telescopes aboard XMM-Newton, the most sensitive X-ray satellite to date, is only about 4000 cm^2. This is equivalent to the collecting area of an optical telescope with a diameter of 70 cm! Therefore the collecting areas of professional X-ray instruments are only slightly larger than those of optical telescopes of amateur astronomers. Another point concerns the resolving power of X-ray spectrographs in space. In most cases, they hardly exceed about $\lambda/\Delta\lambda \sim 1000$. However, the most sensitive and most powerful plasma diagnostics require high-resolution X-ray spectroscopy.
In the early years of X-ray astronomy, the number of sources was rather limited, but nowadays, a wealth of objects have been detected. Indeed, the targets observed with modern X-ray observatories include Solar System objects such as planets and comets, stars (hot stars as well as rather cool objects), X-ray binaries (harbouring a compact object such as a neutron star or a black hole), clusters of galaxies, active galactic nuclei,...

7.2.1 The Uhuru satellite

After a series of rocket flight experiments, the very first satellite dedicated to observations in the X-ray domain was Uhuru, launched in December 1970 from a modified oil drilling platform off the coast of Kenya. The launch site was chosen because of its proximity to the equator, allowing to put a larger mass into a roughly equatorial orbit. The satellite was put into a 96 minutes period LEO with a low (3°) orbital inclination to avoid the SAA. Uhuru was operated until March 1973.
To achieve its main science goal, i.e. the very first survey of the entire sky in X-rays (energy domain from 2 to 20 keV), the satellite was spinning around its axis every 12 minutes.
Uhuru did not carry an X-ray telescope. Instead, the payload consisted of two proportional counters placed behind a set of X-ray collimators with full widths at half maximum of $0.52^\circ \times 5.2^\circ$ and $5.2^\circ \times 5.2^\circ$. The two collimator assemblies pointed into opposite directions and simply consisted of a mechanical structure looking like a honeycomb. X-ray collimators are non-imaging and non focusing instruments and in this case, the effective area of the

[5]This was done in the context of the preparation of the Apollo programme.

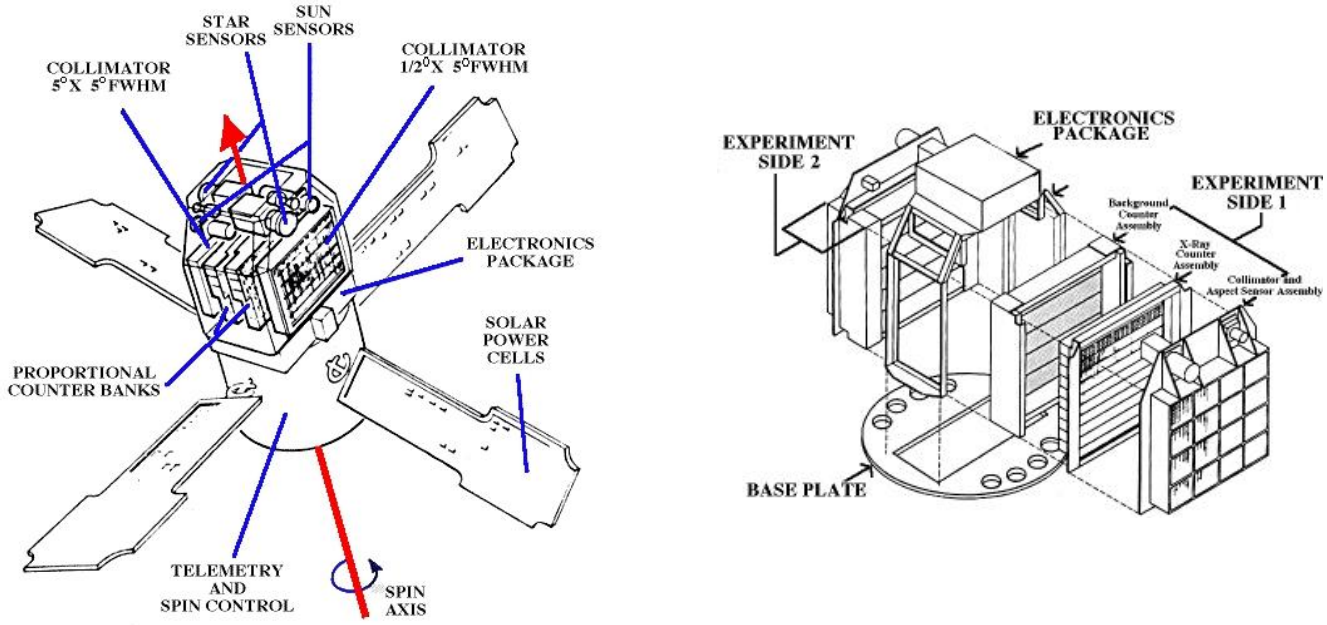


Figure 7.7: Left: the main components of the Uhuru spacecraft. The height of the satellite was about 1.1 m. Right: the X-ray collimator and proportional counter assembly making up the main instrument aboard Uhuru.

instrument is set by the geometrical area of the detector itself multiplied by the efficiency of the detector.

Proportional counter devices, such as those aboard Uhuru, are detectors that consist of a windowed cell holding a gas that can be ionized by an incoming high-energy photon. Once a photon penetrates the chamber, it creates a cloud of free electrons, the number of free electrons being proportional to the energy of the incident photon. The cloud of free electrons then drifts towards a readout electrode and the on-board electronics records the amount of charges created inside the chamber. A proportional counter hence provides a measure of the time of arrival of the photon, as well as a broad evaluation of its energy. The proportional counters aboard Uhuru were sensitive to X-rays with energies in the range 2 – 20 keV. The entrance window of the detector was made of beryllium which set the lower energy limit of the detector, whilst the upper energy limit was determined by the ioniziation cross-section of the gas in the detection chamber. As stated above, a proportional counter device allows to obtain some crude information on the energy of the photon and in the case of the Uhuru satellite, the detector allowed to classify the incoming photons into either of eight energy channels.

The effective collecting area of the Uhuru satellite was about 700 cm^2 comparable to that of the Chandra telescope. Yet, the sensitivity achieved by Uhuru was about $1.5 \times 10^{-11}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$, i.e. five orders of magnitude worse than that of Chandra. The reason for this is simply the lack of focusing capabilities in the case of Uhuru. The location of bright discrete sources could be determined with an accuracy of a few arcmin^2 at best.

The science output of the Uhuru mission was a catalogue of 339 X-ray sources for which positions, brightnesses, and in some cases, rough spectral distributions and light curves were obtained. The satellite made a major contribution to the understanding of X-ray binaries. For instance, in the case of Cen X-3, Uhuru discovered pulsations with a period of 4.8 s as well as X-ray eclipses with a period of 2.09 days showing that Cen X-3 consists of a spinning neutron star orbiting a massive star and accreting material from the latter.

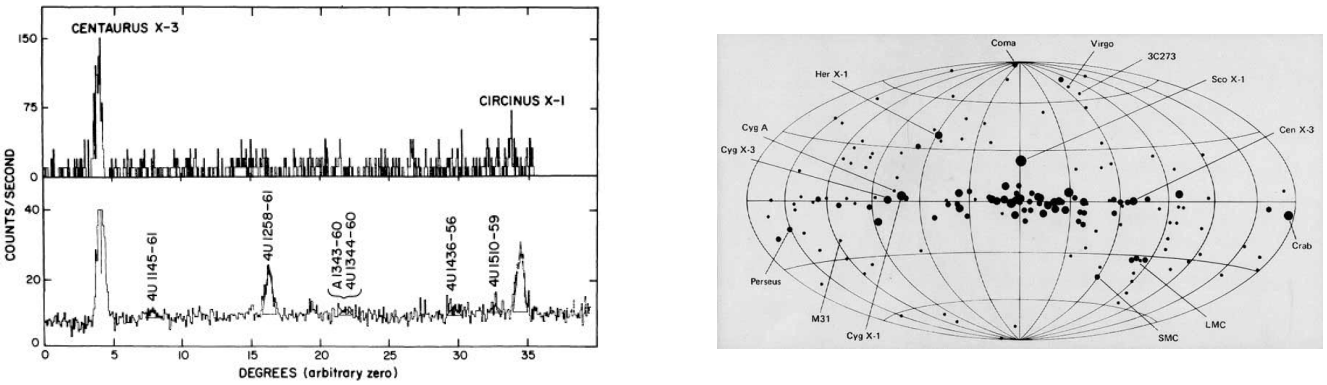


Figure 7.8: Left: a scan of the Uhuru satellite along the Galactic plane. The upper panel illustrates the signal produced during a single scan over a given region. The lower panel illustrates the same region but now all available scans have been combined to yield a much improved sensitivity. Right: the location of the 339 X-ray sources of the Uhuru catalogue shown in Galactic coordinates.

7.2.2 Optics in X-ray astronomy

One of the major problems with high-energy radiation such as X-rays is its high penetrating power. These radiations are not easily reflected and building an efficient focusing optics requires the use of specific techniques such as grazing incidence. In fact, one can define a critical angle between the incident ray of light and the surface of the mirror below which the light is totally reflected. For angles larger than the critical angle, some fraction of the incident light goes through the material. In the X-ray domain, the critical angles are rather small (a few degrees). In other words, the X-ray photons will only reflect off the surface of the mirror if they hit it under grazing incidence and this tendency increases with increasing photon energy: the higher the photon energy, the lower the value of the critical angle for a specific material.

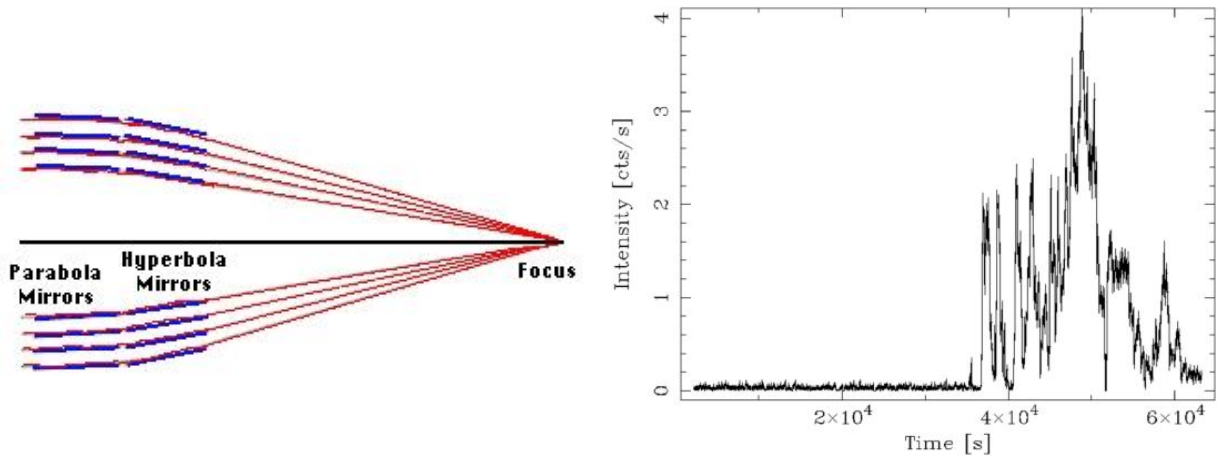


Figure 7.9: Left: schematic view of the principle of a Wolter I mirror assembly. Right: the evolution of the instrumental background of an EPIC camera aboard XMM-Newton during a soft proton flare.

Modern X-ray observatories such as ESA's XMM-Newton observatory or NASA's Chandra satellite, therefore rely on somewhat special mirrors where the reflection occurs under grazing incidence. There are two drawbacks with these mirrors. First of all, the effective area of reflection is much smaller than the geometrical area: $A_{\text{eff}} \sim$

$A_{\rm geom} \sin\theta$ where θ is the grazing reflection angle. This is the main reason why the collecting area of modern X-ray telescopes remains rather modest. The second limitation concerns the focal length of such X-ray mirrors. In fact, because of the grazing incidence reflection, these focal lengths are very high. To reduce them, one can combine grazing incidence reflections on two surfaces. This is the principle of the so-called Wolter-I mirrors[6]. In such a configuration, the X-rays undergo a first grazing incidence reflection off a paraboloid mirror, before being reflected a second time by a hyperboloid.
Vignetting is an intrinsic property of grazing incidence optics. Indeed, the reflectivity is higher for low grazing incidence angles (i.e. for targets close to the optical axis) than for off-axis sources. In all cases, there is a trade-off to be found between a large effective area, high angular resolution and a large field of view. The best angular resolution requires stiff monolithic mirrors. This is achieved for large masses (heavy material needed to reach this stiffness), leading to mass constraints. Currently, the highest angular resolution is achieved by the mirrors of NASA's Chandra telescope. Chandra features four Wolter I type mirrors made from Zerodur glass.
One quite un-expected feature of the Wolter I mirrors is their behaviour with respect to 'soft protons'. In fact, after the launch of Chandra (and lateron XMM), it was realized that soft protons with an energy of about 100 keV from the Earth radiation belts are focused by the grazing incidence optics pretty much in the same way as X-rays and as a result they reach the focal plane leading to high background episodes. These soft protons usually appear in the form of flares and can substantially affect observations that require a low background (such as studies of diffuse X-ray emission or detection experiments).

7.2.3 The XMM-Newton observatory

In December 1999, ESA's X-ray Multi-Mirror observatory XMM-Newton was launched by an Ariane V rocket from Kourou. The satellite was initially placed into a temporary orbit, with a perigee height of 850 km and an apogee of 114 000 km, and then utilised its own propulsion system to raise the perigee: at each of the first five apogees, the thrusters of the satellite were fired to raise the perigee to an altitude of 7000 km. The spacecraft reached its operational orbit about eight days after launch. In the meantime, the telescope tube was emptied of residual gases, the sunshield deployed, and finally the doors of the mirror modules were opened.
XMM-Newton's highly eccentric operational orbit has been chosen mainly for two reasons. First of all, the scientific instruments need to work outside the high-energy particle environment of the Van Allen radiation belts. Second, a highly eccentric orbit offers the possibility to carry out long duration observations around apogee passage that are not interrupted by the frequent passages in the Earth's shadow that affect LEO satellites. To allow an optimal communication with the ground stations, the orbital period of the satellite was chosen to be 48 hours, exactly twice the Earth's rotation period. Currently, two ground stations situated at Kourou (French Guiana) and Perth (Australia) are sufficient to maintain contact between the spacecraft and the Mission Control Centre (MOC) in Darmstadt (Germany). The scientific data received on the ground are then forwarded to the Science Operations Centre (SOC) in Villafranca (Spain) from where they are dispatched to the scientific community.

XMM-Newton's orbit at the beginning of the operations was tilted at 40° to the Earth's equator, with its apogee in the Southern Hemisphere. The orbital eccentricity was 0.79 and the perigee altitude was 7000 km. The orbital parameters evolve as the mission progresses. For instance, the perigee altitude varies between 7000 and 22 000 km, while the apogee altitude varies between 114 000 and 100 000 km over the mission lifetime. The observations by XMM-Newton's instruments are conducted outside the main radiation belts which extend to about 40 000 km. The spacecraft stays below this altitude for about 8 hours per orbit and the instruments are switched off during the passage across the radiation belts. As a result, roughly 40 hours of each orbit are available for scientific observations.

[6]Named after the German physicist Hans Wolter (1911 - 1978) who designed several types of X-ray mirrors combining paraboloid surfaces with hyperboloids or ellipsoids.

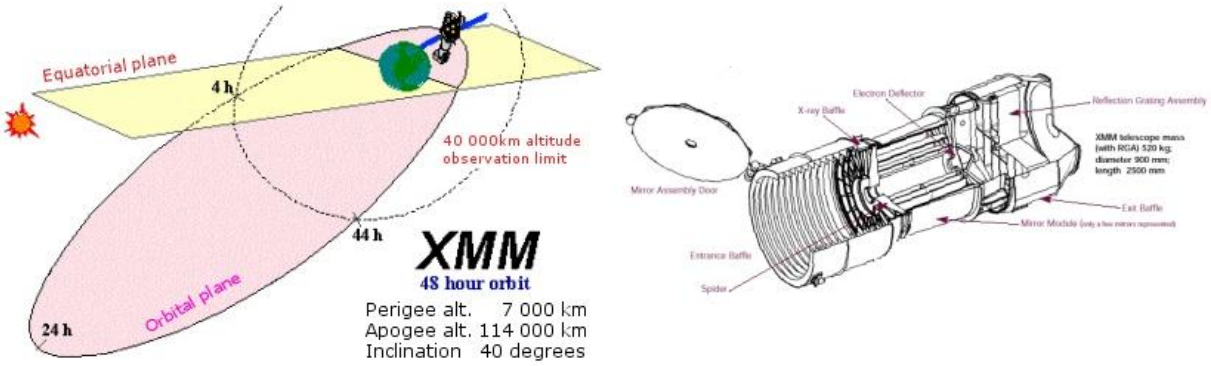


Figure 7.10: Left: the orbit of the XMM-Newton spacecraft. Right: scheme of the mirror assembly of the XMM-Newton telescope.

The main goal of XMM-Newton is to perform high-quality spectroscopic observations to establish the physical parameters (temperature, density,...) and chemical composition of high temperature plasmas in the Universe. To meet these objectives, the satellite hosts three mirror assemblies, each one consisting of 58 gold-coated Wolter I mirrors, nested inside each other in a coaxial and cofocal configuration. To allow focusing photons with energies as high as 7 keV, the grazing incidence used for the XMM-Newton mirrors is as low as 30'. The focal length of the telescope is 7.5 m and the diameter of the largest mirrors is 70 cm. The latter parameter was mainly limited by the dimensions of the shroud of the Ariane V launcher.
Each of the three telescopes includes, apart from the mirror modules, entrance baffles for visible and X-ray stray light suppression and an electron deflector which produces a circumferential magnetic field to prevent soft electrons from reaching the focal plane detector. The X-ray baffles are located in front of the mirror systems and act as collimators to prevent single reflection rays, reducing considerably the amount of stray light from off-axis sources.

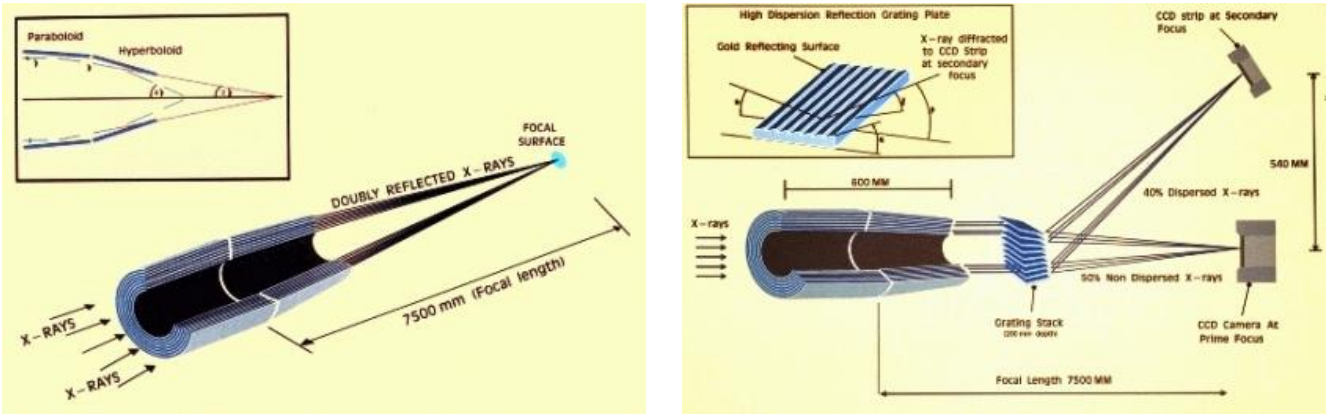


Figure 7.11: Left: the light path in the XMM-Newton telescope for the EPIC cameras. Right: same but for the RGS instruments.

The XMM-Newton spacecraft hosts six scientific instruments that can be operated simultaneously.
At the prime focus of each of the X-ray telescopes, there are three European Photon Imaging Cameras (EPIC, see Fig. 7.11). The three EPIC cameras offer the possibility to perform extremely sensitive imaging observations in the energy range from 0.2 to 12 keV over the telescope's field of view of 30 arcmin diameter. Their angular resolution is about 6 arcsec (FWHM) and the charge-coupled device (CCD) detectors at the heart of the EPIC cameras allow

to obtain medium-resolution spectroscopy ($E/\Delta E \sim 20$ – 50) of the X-ray sources. CCD detectors are very efficient to detect X-rays but are also highly vulnerable to radiation damage. Each CCD is read out continuously and the image data are processed and compressed by the on-board computer, to make them compatible with the spacecraft's telemetry.

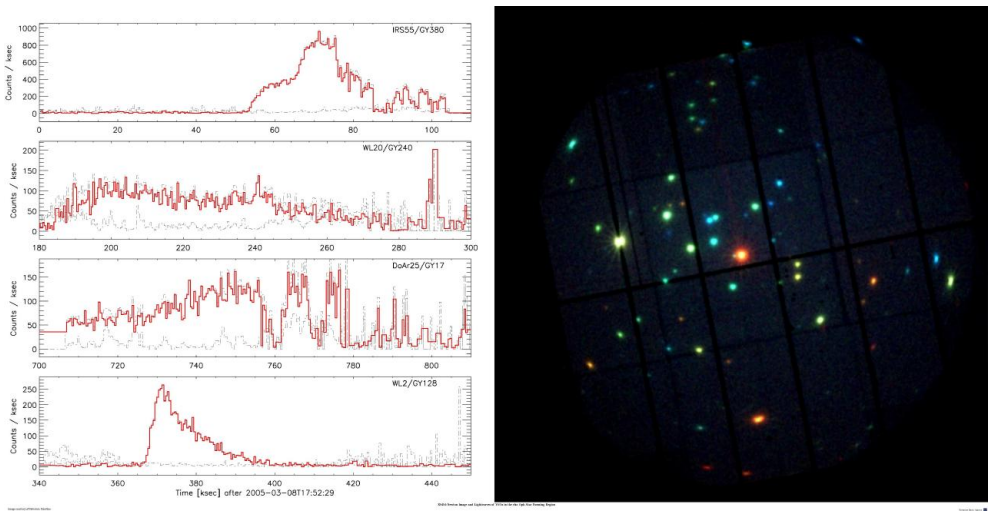


Figure 7.12: EPIC image of the ρ Oph star forming region (right panel). The various colours indicate different photon energies. The light curves of some variable pre-main sequence stars are shown in the left panel.

In two of the telescopes, a reflection grating array is mounted behind the mirror module (Fig. 7.11). This grating structure reflects about half of the incident X-rays to a secondary focus, with its own CCD camera. These instruments, called the Reflection Grating Spectrometers (RGS), allow to obtain high-resolution spectra of bright X-ray sources. Actually, XMM-Newton is the first mission to be equipped with reflection gratings operating in the X-ray band. Early X-ray missions carried Bragg-crystal spectrometers and subsequent missions like EXOSAT and Chandra used transmission gratings. The two grating arrays on XMM-Newton are each composed of 182 grating plates. Each plate of the reflection grating array consists of a silicon carbide substrate coated with a thin (2000 Å) layer of gold. The reflection gratings are mirrors with 600 grooves per mm. X-rays reflected off the top and the valley of the grooves interfere with each other and cause a 'spectral image' whereby X-radiation of different wavelengths (or energy) are reflected under slightly different angles.

The analysis of these high-resolution spectra allows determining the physical characteristics (density, temperature, ionisation state, chemical abundances, mass motions and redshift) of the emitting plasma and its surrounding environment.

Finally, the satellite features an optical Ritchey-Chrétien telescope[7] with a primary mirror of 30 cm. This so-called Optical Monitor (OM) is a very sensitive optical and ultraviolet telescope, which is mounted on the mirror support platform of XMM-Newton alongside the X-ray mirror modules. The Optical Monitor telescope can observe simultaneously the central 17' $\times$ 17' region of the X-ray field of view. The instrument is equipped with a filter wheel that provides coverage between 1700 and 6500 Å. In orbit, this 30 cm telescope is as sensitive as a 4 m telescope on the ground. XMM-Newton is the first ever X-ray mission with such a multi-wavelength component.

Each day, XMM-Newton sees more sources in a small part of the sky than UHURU found across the whole sky during its three years in operation. The pointed observations with XMM-Newton thus detect significant numbers of previously unknown serendipitous X-ray sources. Combining the data from many observations therefore yields a wealth of serendipitous sources. For instance, the 3XMM catalogue, created from source detections drawn from 7427 pointed observations made with the EPIC cameras between February 2000 and December 2012, lists about 370 000 unique X-ray sources with a median flux of 2.4×10^{-14} erg cm^{-2} s^{-1}. Therefore, the 3XMM catalogue

[7] A Ritchey-Chrétien telescope design gives a high-quality image over a relatively wide field of view. See next section.

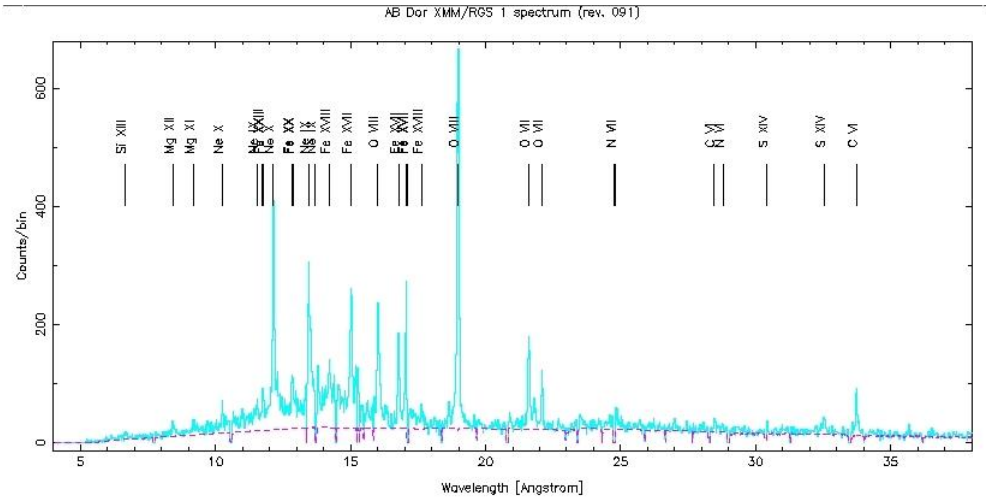


Figure 7.13: RGS spectrum of the K1 IV-V active star AB Dor.

is the largest collection of X-ray objects compiled to date. The total area covered on the sky by the combined observation fields is about 794 square degrees, i.e. only about 2 % of the entire sky.
Although XMM-Newton was initially built with a goal lifetime of ten year, it is currently in its nineteenth year of operations and there is enough fuel aboard the satellite for at least another four years[8]. In November 2018, ESA extended the mission operations of XMM-Newton until December 2020 with an indicative extension up to 2022. This longevity tremendeously increases the scientific return of the mission, but is also a challenge as far as knowledge management is concerned. Indeed, key people who designed the spacecraft and its instruments are nearing retirement or have already retired. Moreover, the way XMM-Newton is currently operated is very different from what was anticipated in 1988 when the design started. This example underlines the importance of an extensive documentation of every single part of a spacecraft during the conception and building phases, to ensure efficient operations over a long period of time.

7.2.4 Examples of scientific questions addressed by X-ray observatories

X-ray astronomy deals with high-temperature plasma often associated with extreme phenomena in the Universe. Nowadays, almost every type of cosmic objects (from comets in our Solar System to distant quasars) has been detected in the X-ray domain. A full discussion of these aspects is well beyond the scope of these lectures, and is rather the topic of the course on *High Energy Astrophysics*. Here, we will only discuss one specific example of science applications of X-ray astrophysics. Indeed, one of the most important phenomena in this context is accretion.
Accretion occurs in a variety of astrophysical situations:

- protostars accrete material from their surrounding to reach the mass necessary to ignite nuclear reactions in their core,
- compact objects (i.e. white dwarfs, neutron stars or black holes) in binary systems accrete the material from their companion star,
- supermassive black holes at the nuclei of galaxies accrete huge amounts of material from their host galaxy.

Sco X-1, the very first cosmic X-ray source that was discovered (see above) is an example of the category of **X-ray binaries**. These systems contain either a neutron star or a black hole that orbits around a normal star. If

[8]There is however not enough fuel to actively de-orbit the spacecraft.

the normal star is more massive than $\sim 8\,\mathrm{M}_\odot$, one speaks about a high-mass X-ray binary. On the other hand, for systems where the normal star is of solar-type or less massive, one speaks about low-mass X-ray binaries. During its evolution, a star expands when it exhausts its nuclear fuel in the stellar core. When the star is in a binary system, it cannot expand to arbitrarily large radii. In fact, the attraction of the companion will lead to a mass transfer if the expanding star reaches its critical volume which is set by the so-called Roche lobe, i.e. the equipotential of the Roche potential that goes through the L_1 Lagrangian point of the binary system.

If the binary system contains a compact object, a flow of material from the donor star towards the compact star sets in. However, due to its angular momentum, the material will not simply fall on the compact object but rather spirals around it in an accretion disk. As the matter approaches the surface of the neutron star or the event horizon of the black hole, it gives away its kinetic energy in the form of heat leading to copious emission of X-ray radiation.

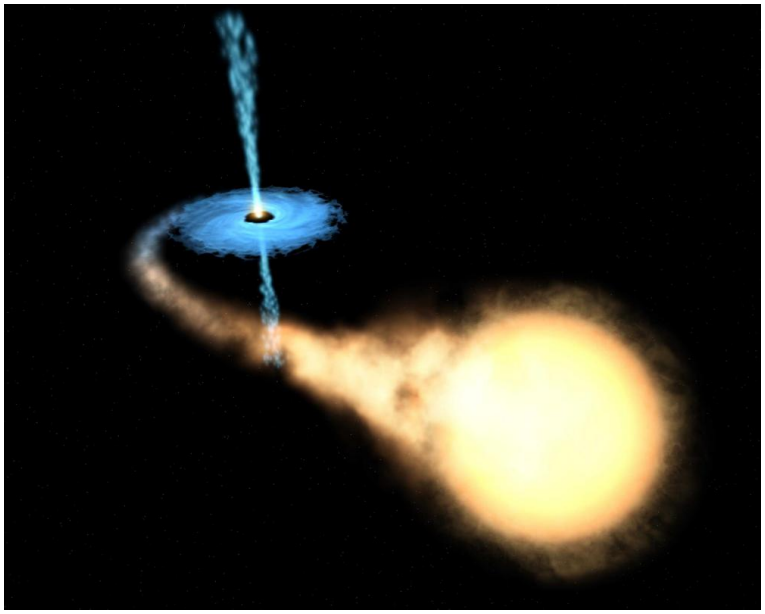


Figure 7.14: Artist view of an X-ray binary system where a black hole accretes material from a normal star through an accretion disk. Part of the material is ejected in a jet perpendicular to the accretion disk. These systems are nowadays called micro-quasars.

The vast majority of the first X-ray sources were indeed such X-ray binaries. If the compact star is a young neutron star that has a rather strong magnetic field, the material that is being accreted is funnelled by the magnetic field towards the magnetic poles and a hot spot appears at the point where the material impacts the surface of the neutron star. Due to the (rapid) rotation of the neutron star, the X-rays emitted by this hot spot are modulated by the rotation period, leading to the appearance of X-ray pulsars as in the case of Cen X-3.

X-ray observations also play a major role in our understanding of supermassive black holes (SBHs) in the centres of active galactic nuclei (AGN). In a way similar to what happens in X-ray binaries, these SBHs accrete huge amounts of material from their host galaxy thereby generating a copious X-ray emission. Since the X-ray emission forms quite close to the innermost stable orbit around the black hole, X-ray spectroscopy of these objects can be used to infer the properties of these exotic objects and to study physical phenomena that occur in strong gravitation fields. An important feature in this context is the iron line near 6.4 – 6.7 keV. Observations of this line in AGN frequently reveal a broad and heavily skewed line. The broadening is attributed to the disk rotation, whilst the extended red wing of the line stems from the gravitational redshift of the line photons near the event horizon of the SBH.

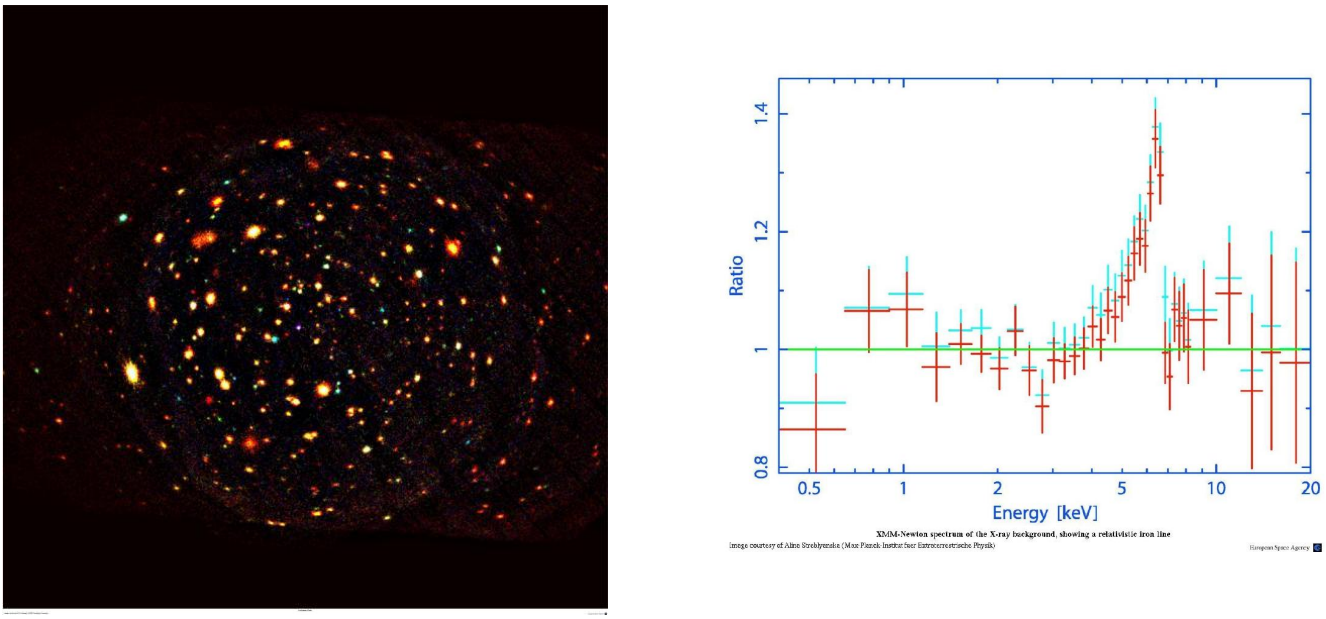


Figure 7.15: Left: three colour XMM-Newton image of the Lockman Hole, a region of the sky where the amount of intervening foreground material is rather low, allowing a relatively unobstructed view on the extragalactic Universe. Red, green and blue colours correspond to photons in the energy range 0.5 – 2.0, 2.0 – 4.5 and 4.5 – 10 keV. Almost all of the sources in the field are either clusters of galaxies or distant AGN. Right: average Fe line profile in the AGN sources of the Lockman Hole. The shape of the line clearly reveals the signature of relativistic effects.

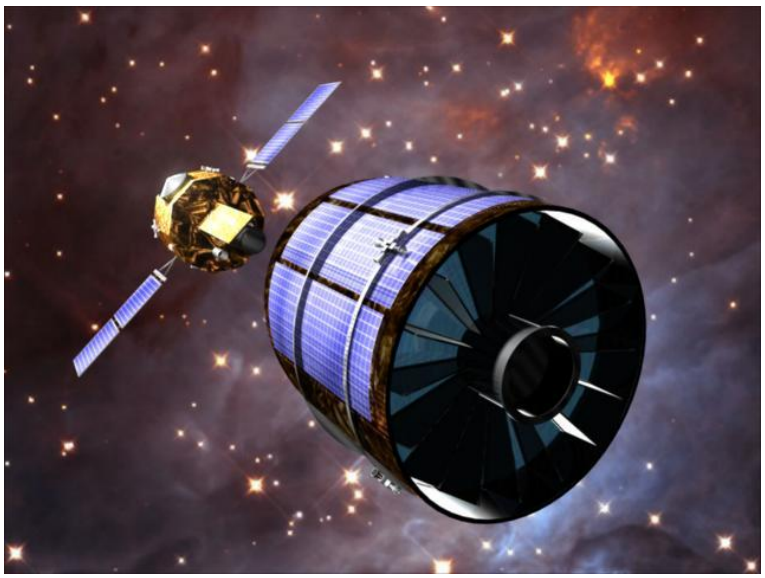


Figure 7.16: Artist view of the formation flight concept, initially foreseen for the XEUS mission.

7.2.5 The future: Athena

In 2007, the advisory structures of the European Space Agency recommended the technological developments of a next generation X-ray telescope for a launch in the timeframe of the first Cosmic Vision 2015-2025 large mission launch window (i.e. launch originally expected around 2018, now delayed to 2022 at the earliest). This mission concept, called XEUS (the X-ray Evolving Universe Spectrometer mission) was designed to consist of two spacecraft in formation flight on a halo orbit around L_2. In this concept, the two spacecraft were expected to

maintain a mutual distance of 35 m with an accuracy of better than 1 mm^3. XEUS's sensitivity was expected to be 4×10^{-18} erg cm^{-2} s^{-1}, about a hundred times superior to that of XMM-Newton. The mirror spacecraft was initially foreseen to have a collecting area of 5 m^2 at 1 keV with an imaging resolution of 5 or even 2 arcsec. To achieve this large collecting area with a relatively low mass, a novel technology needs to be developed. Currently, the most promising option is high-precision silicon pore optics. These are manufactured from flat and polished silicon wafers. The wafers are then chemically treated to produce a rib-structure. The ribbed plates are subsequently stacked by a robotic process to form an approximate Wolter I assembly. The stacked pairs are eventually mounted in a petal segment of the mirror.

In the course of the summer 2008, an assessment of the cost of the project led the ESA management to reconsider a collaboration with NASA and JAXA. The discussions between ESA, NASA and JAXA led to the merger of XEUS with the NASA project Constellation-X. The new mission concept IXO (International X-ray Observatory) was proposed to the NASA decadal survey. The IXO design gave up the concept of formation flying at the benefit of an extendable optical bench. The collecting area was reduced to 3 m^2 at 1 keV with 2020 as the earliest possible launch date. In view of the budgetary cuts at NASA, ESA decided in the spring of 2011 that collaboration with NASA was no longer an option and the European X-ray community was requested to work on an ESA-only mission concept that was called Athena. The Athena design featured two X-ray telescopes with a focal length of 11 m, an angular resolution of 5" and a total collecting area of 0.6 m^2 at 1.25 keV (for each telescope). In the spring 2012, the downselection of the first L-class mission of the Cosmic Vision plan took place. Unfortunately, Athena was not selected. In the winter 2012-2013, ESA released a call for ideas for the second and third L-class missions. The European X-ray astrophysics community replied to this call, proposing an upgraded and more mature version of Athena featuring a single X-ray telescope with two instruments (a wide field imager and an X-ray integral field calorimetric spectrograph) on a movable focal plane platform. In June 2014, Athena was selected for the second slot for an L-class mission (launch expected in 2030) and the European X-ray astrophysics community is now actively preparing the next generation large X-ray telescope.

7.3 UV astronomy

The photons emitted by the astrophysical sources are the only messengers that astrophysicists can use to infer the physical processes in the emitting plasmas. To obtain as complete a description of the Universe as possible, it is therefore mandatory to perform observations over the wavelength domains that are the most relevant to characterize a given phenomenon. This statement also holds for the ultraviolet (UV) domain. However, below about 3000 Å, no UV radiation is detectable from the ground. One can at least partially overcome this problem by making high-altitude observations from mountains, aircraft or balloons. However, important resonance spectral transitions (i.e. involving the ground state) occur at wavelengths $\lambda < 2000$ Å which can only be observed from space.

Table 7.2: Commonly used divisions in UV astrophysics

	Wavelength (Å)
Near-UV	3000 – 4000
Middle-UV	2000 – 3000
Far-UV	1000 – 2000
Extreme-UV	100 – 1000

Ultraviolet line spectrum measurements are used to discern the chemical composition, densities, and temperatures of the interstellar medium, and the temperature and composition of hot young stars. UV observations can also

provide essential information about the evolution of galaxies. The ultraviolet Universe looks quite different from the familiar stars and galaxies seen in visible light. Most stars are actually relatively cool objects emitting much of their electromagnetic radiation in the visible part of the spectrum. Ultraviolet radiation is the signature of hotter objects, typically in the early and late stages of their evolution.

The first UV observation outside the Earth's atmosphere was the detection of UV radiation from the Sun by a V2 rocket flight in October 1946. In 1962, the Orbiting Solar Observatory (OSO-1) was launched. This satellite observed coronal lines in the solar spectrum at wavelengths below 400 Å. OAO-2 launched in 1968 was the first satellite to perform a sky survey in the 1000 – 3000 Å domain. The Copernicus satellite (OAO-3) obtained spectroscopic observations in the 710 – 3280 Å domain from 1972 until 1981.

Observations in the near- to far-UV domain require somewhat special optics. For this energy range, normal incidence reflection is still possible and most telescopes are based on the Ritchey-Chrétien design. Ritchey-Chrétien telescopes are Cassegrain telescopes with both the primary and the secondary mirror of hyperbolic shape to eliminate the coma and spherical aberration. High ultraviolet reflectivity mirror coatings are realised with gold and silicon carbide. As far as refractive optics are concerned, magnesium fluoride lenses are transparent down to 1100 Å. For the short wavelength photons of the extreme-UV domain, one has to use grazing incidence Wolter-type mirrors. In the UV domain, an efficient control of the spacecraft's outgassing is fundamental as the UV reflectivity is strongly affected by molecular contamination.
Detectors used in the UV domain are photoelectric devices or CCDs. CCDs are usually not sensitive to UV radiation below 4000 Å, but can be used in conjunction with a multi-channel plate. Such devices consist of a wafer of hexagonally packed lead glass tubes within which arriving photons produce electron showers (photoelectric effects). The electron showers are incident on a phosphor and are hence converted to optical light that is subsequently detected by the CCD (see also Sect. 7.3.2 for a slightly different design of a microchannel plate detector).

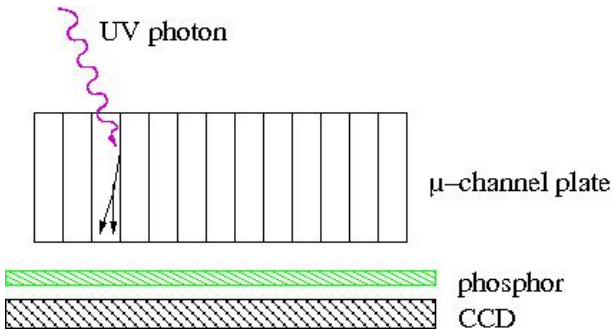


Figure 7.17: Schematic view of the microchannel plate detector for UV observations. Upon interaction with the lead glass tube, the UV photon creates a shower of electrons that impacts on a phosphor where the charges are converted to visible light that is detected by the CCD.

7.3.1 IUE

The International Ultraviolet Explorer (IUE) was a collaborative project between NASA, the UK Science Research Council and ESA. IUE was designed to obtain spectroscopic observations in the UV domain. NASA provided the spacecraft, telescope, spectrographs and one ground station, ESA the solar panels and the second ground station, and the UK the four spectrograph detectors. The satellite was launched from Cape Canaveral in January 1978 and the initial lifetime was foreseen to be three years, but the satellite exceeded the expectations and was operated until 1996. Even then the satellite was still functioning, but was actually switched off for budgetary reasons.
IUE was the first space observatory to be operated in real time by astronomers who visited the groundstations in Greenbelt (USA) and Villafranca (Spain). To allow communications with these ground observatories, the satellite was put into a slightly eccentric ($e = 0.24$) and inclined ($28.6°$) geo-synchronous orbit around $300°$ East longitude.

Over the 19 years of the mission's lifetime, the orbit evolved through the perturbative effects of the non-spherical Earth and the gravitational influence of the Sun and the Moon. The combined action of these forces reduced the orbital eccentricity, increased the inclination to $35.9°$ and pushed the satellite west-ward. The latter drifts had to be corrected with the on-board thrusters.
The spacecraft was 3-axis stabilized with six gyroscopes. The electrical power of 424 W was provided by two panels of solar arrays and two Ni-Cd batteries were used to supply power during the eclipses. The communications with the ground used the S band with a power of 6 W.

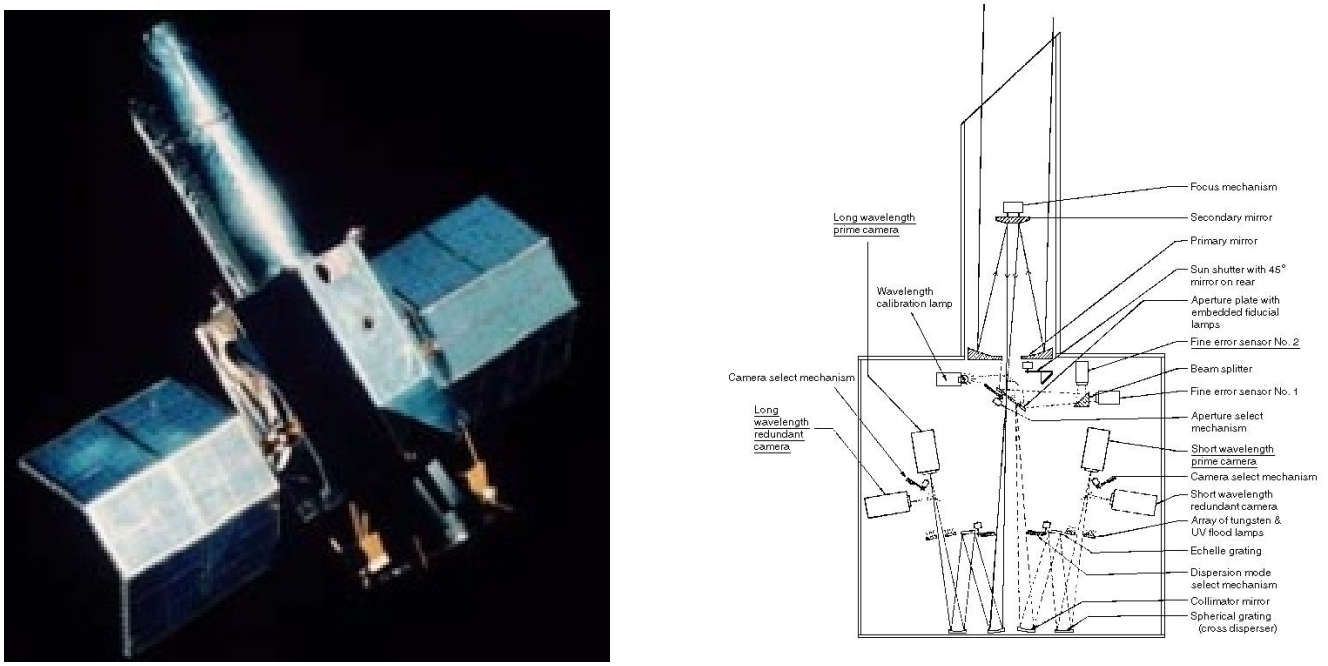


Figure 7.18: Left: artist view of the IUE spacecraft. Right: optical scheme of the telescope and spectrograph aboard IUE.

The satellite used a Ritchey Chrétien telescope featuring a beryllium paraboloid primary mirror with a diameter of 45 cm. The focal plane contained an echelle spectrograph with four UV cameras: one primary long wavelength camera, one primary short wavelength camera, and a backup for each of them. With those cameras, the echelle spectrograph covered the wavelength ranges from 1150 to 1980 Å and 1800 to 3200 Å with a resolving power up to 18 000.

Astronomers made over 104 000 observations of almost 10 000 different objects using IUE. The targets included planets, comets, stars, interstellar gas, supernovae, planetary aurorae, galaxies, and quasars. Among the science highlights of the mission, one can cite the detection of an aurora on Jupiter, the proof that massive stars undergo strong and variable mass-loss through stellar winds, the study of the progenitor of supernova SN 1987A, the discovery of high velocity winds in cataclysmic variables, the determination of the size of the active regions in Seyfert galaxies, etc.

7.3.2 GALEX

After the termination of the IUE mission, ultraviolet spectroscopy relied mainly on

- the NASA/ESA Hubble Space Telescope which carried several instruments with UV capabilities: FOS,

GHRS, STIS,...

- the NASA Extreme Ultraviolet Explorer (EUVE) mission that used grazing incidence optics to observe in the 70 – 760 Å domain over the years between 1992 and 2001. The number of objects observable in this wavelength range is quite limited because most radiation below the hydrogen Lyman limit (912 Å) is absorbed by the hydrogen in the interstellar medium. As a result, only relatively nearby stars could be observed with EUVE.

- the NASA/Canada/France mission Far Ultraviolet Spectroscopic Explorer (FUSE) that operated in the 905 – 1187 Å wavelength range from 1999 until 2007.

The Hubble Space Telescope and FUSE have been the most recent major space telescopes to view the near and far UV spectrum of the sky, though other UV instruments have flown on sounding rockets and on board the Space Shuttle. Currently, there is no new UV spectroscopy mission foreseen.

The Galaxy Evolution Explorer (GALEX) was launched in April 2003 by a three-stage Pegasus XL rocket. The rocket was carried underneath an L-1011 Stargazer aircraft to an altitude of 11.9 km. The Pegasus rocket was then released by the Stargazer plane and fell free in a horizontal position for five seconds before the first stage of the rocket was ignited. The Pegasus rocket then placed GALEX into a circular LEO at an altitude of 697 km with an inclination of 29°.

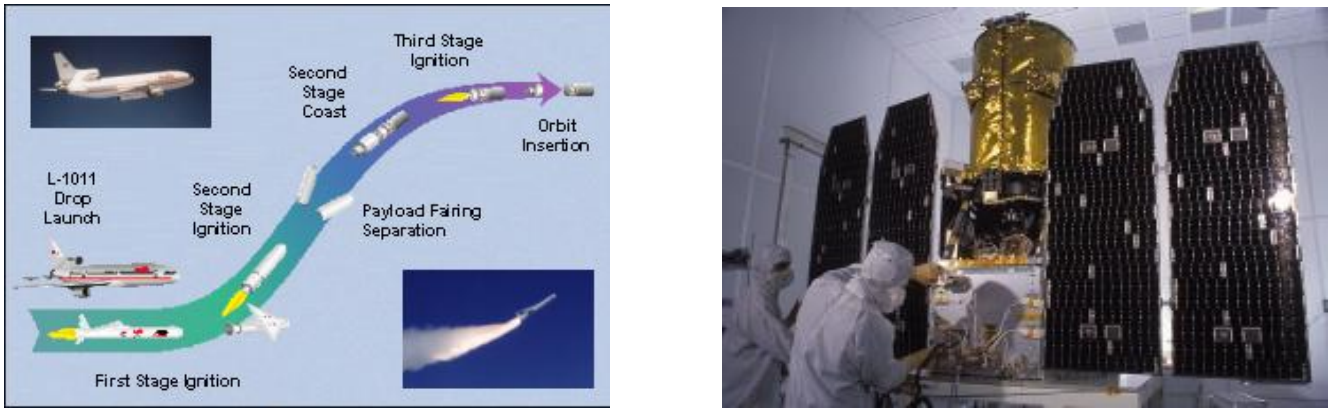


Figure 7.19: Left: the launch sequence of the GALEX satellite with the Pegasus rocket carried by the L-1011 aircraft. Right: the GALEX satellite during integration.

GALEX was a collaboration between NASA as well as several US, Korean and French universities. GALEX featured a Ritchey-Chrétien telescope with a primary mirror of 50 cm in diameter and a secondary of 22 cm. Thanks to its aspheric design, the telescope allowed to obtain good quality images (PSF with 4 – 6 arcsec FWHM) over a very wide field of view (1.2° diameter). A special optical coating reduced the glow from the Earth's atmosphere as well as from zodiacal dust. To ensure that the telescope staid focused, i.e. to maintain the 70 cm distance between the two mirrors despite the changing thermal environment of a LEO, the mirrors were held in place by a support in Invar (an alloy of Fe, Ni and some traces of C and Cr that has a very low coefficient of thermal expansion). To ensure optimal performances of the telescope, an operation temperature between 0 and 27° C was needed. This was achieved through an active thermal control: the temperature was monitored with thermistors that triggered the operation of heaters as required. When needed, these heaters could also be used to heat up the optics of the telescope to get rid of a molecular contamination.

GALEX performed both imaging and spectroscopic observations. An optical wheel assembly with two circular holes, one holding a CaF_2 imaging window and the other featuring a CaF_2 transmission grism[9] (75 grooves mm^{-1}), controled whether GALEX took images or spectra. The light then passed through the dichroic beam splitter[10], a specially manufactured crystal which separated the light into far-UV (1350 – 1740 Å) and middle-UV (1750 – 2800 Å). The grism provided spectroscopy with a resolving power of 80 – 150 in the middle UV and 250 – 300 in the far-UV.
The detectors used by GALEX were large (6.5 cm diameter) microchannel plates. These microchannel plates are photon counting devices. When the light reaches the detector, it first passes through a window lens (fused silica for the middle-UV, MgF_2 for the far-UV) that helps to keep the image in focus across the large surface of the detector. A thin layer of material (the so-called photo-cathode, Cs_2Te for the middle-UV and CsI for the far-UV) of the detector absorbs the UV photons and releases electrons. The electrons then strike a thin metal plate covered with tiny holes, breaking loose many more electrons from the plate. These free electrons are attracted to a grid of anodes and generate an electric pulse that is registered by the detector electronics. The detectors aboard GALEX had 2 million pixels and were so sensitive that they were only operated when the spacecraft passed in the shadow of the Earth. When the satellite was on the sun-lit side of its orbit, it did not observe, but rather recharged its batteries. Over that part of the orbit, the telescope was pointed away from the Sun and from the bright Earth to avoid damage to the detectors.
Two ground stations (one in Hawaï, the other in Australia) ensured the communication with the satellite for about 10 minutes of each 96 minutes orbital revolution. The rest of the time, the satellite had to rely on its own and the on-board computers had the possibility to store instructions for two weeks of operation. The satellite had the possibility to place itself in a safe mode in case of an emergency.

The goal of the mission was to study the UV radiation from a huge number (several tens of millions) of distant galaxies to trace star formation activity as a function of redshift and hence of age. These data allow to study the evolution of galaxies over a time span of 10 billion years, i.e. about 80% of the current age of the Universe. GALEX perfomed several imaging and spectroscopic surveys: an all-sky imaging survey, a deep imaging survey of some selected extragalactic fields as well as a survey of the 200 galaxies nearest to our Milky Way. Three spectroscopic surveys were designed to provide information on the star formation rate, the distance, the gas and dust content of about 100 000 galaxies.
The combination of the data obtained with GALEX in the UV domain with those from the IR telescope Spitzer provides a powerful tool to investigate the relationship between gas and dust in star formation processes. The legacy of the GALEX mission is an unprecedented archive, both for cosmological research as well as for the study of individual objects.
The GALEX spacecraft was turned off in June 2013 after a decade of operations and three mission extensions. During the last year of operations, GALEX was loaned by NASA to the California Institute of Technology (Pasadena) which used private funds to run the spacecraft. During this last year, GALEX also observed regions in the plane of the Milky Way that were avoided before. GALEX will remain in space for another seven decades before it will re-enter the Earth's atmosphere.

7.3.3 Applications of UV astronomy: hot massive stars and star formation activities

Among the most interesting objects to be studied in the UV domain are hot (i.e. $T_{\rm eff} \geq 30\,000$ K) massive (i.e. $M \geq 10\,M_\odot$) stars of spectral type OB or Wolf-Rayet. Because of their high surface temperatures, these hot stars emit the bulk of their radiation at short wavelengths (i.e. in the UV domain). Although these stars are rather rare

[9] A grism is a right-angled glass prism with a transmission diffractive grating on the hypothenuse surface.
[10] A dichroic material causes light to be split up into two distinct beams of different wavelengths: light is reflected over a certain range of wavelengths and transmitted outside this range.

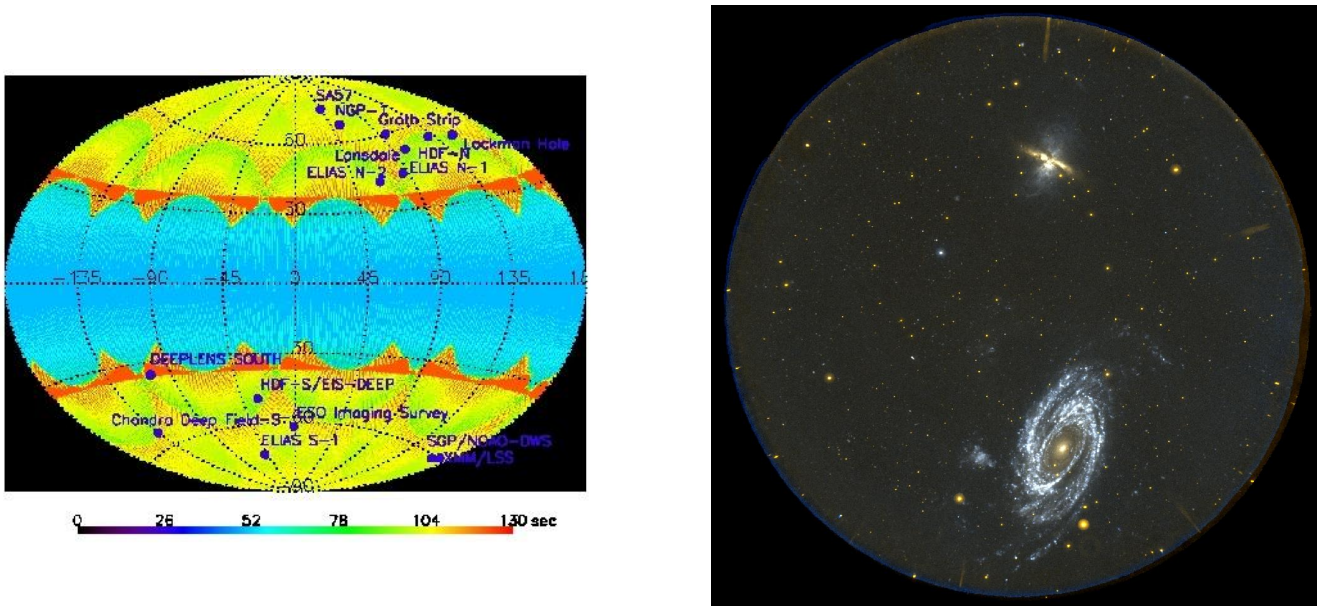


Figure 7.20: Left: exposure map of the GALEX all-sky survey. The deeper extragalactic fields are highlighted. Right: GALEX image of the M 81 and M 82 galaxies. M 81 is the great spiral galaxy in the lower half of the image. The stars in its spiral arms are younger than 100 Myrs. The stellar populations in the centre of this galaxy are much older (10 billion years) and hence cooler (and redder). In the upper half of the image, the starburst galaxy M 82 is seen on its side. The strong supernova activity in this galaxy creates a halo of hot gas that is expelled perpendicular to the disk of the galaxy which is very rich in dust.

objects and have short lifetimes (of order 10 million years to be compared with 10 billion years for solar-mass objects), they play a key role in many processes in the Galaxy and the Universe as a whole. In fact, these stars have powerful stellar winds driven by the radiation pressure due to their high luminosities. These winds associate high mass loss rates (of order $10^{-7} - 10^{-5}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$) and large wind velocities (of order several $1000\,\mathrm{km\,s}^{-1}$). Through these winds and through the gigantic supernova explosion at the end of their life, these massive stars restitute chemically enriched material (synthesized in the core of the star) into the interstellar medium. In fact, most of the chemical elements (heavier than H and He) that make up the present-day Universe did not exist in the early Universe, but were actually formed in the nucleo-synthesis processes either in the cores of previous generations of massive stars or during their deaths as supernova explosions.

UV spectroscopy of massive stars revealed the existence of P-Cygni line profiles that are a clear signature of mass loss via a stellar wind (see Fig. 7.21). To correctly understand the mass loss of massive stars, it is important to determine the geometry of their winds. The long lifetime of IUE enabled astrophysicists to revisit several objects many times and to monitor their variability. An example of such an intensive monitoring is HD 64760, a rapidly rotating B0.5 Ib star ($v \sin i = 238\,\mathrm{km\,s}^{-1}$) that was continously monitored with IUE for almost 16 days in January 1995. This intensive campaign revealed a pattern of rapidly evolving discrete absorption components which migrate across the absorption components of P-Cygni line profiles starting at radial velocities near $-200\,\mathrm{km\,s}^{-1}$ and moving out to $-1500\,\mathrm{km\,s}^{-1}$ in about 12 hours. In addition, there are slowly evolving structures that modulate the profile with a periodicity of 1.2 days (a quarter of the rotation period of the star). These observations are interpreted in terms of corotating wind interaction regions, where outflows of different velocities collide and form a higher density spiral structure in the wind (see Fig. 7.22).

UV observations are also of fundamental importance to understand the star formation activity across the cosmic

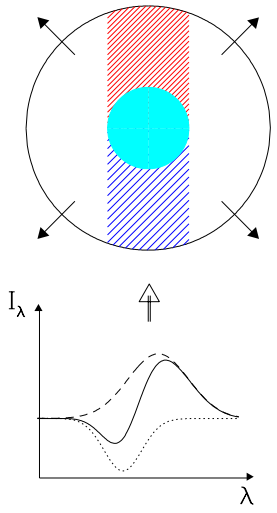


Figure 7.21: Schematic view illustrating the formation of P-Cygni line profiles in the outwards expanding winds of massive stars. The material of the wind along the line of sight moves towards the observer, hence creating a blue-shifted absorption that is superimposed on the emission formed in the side lobes of the expanding wind.

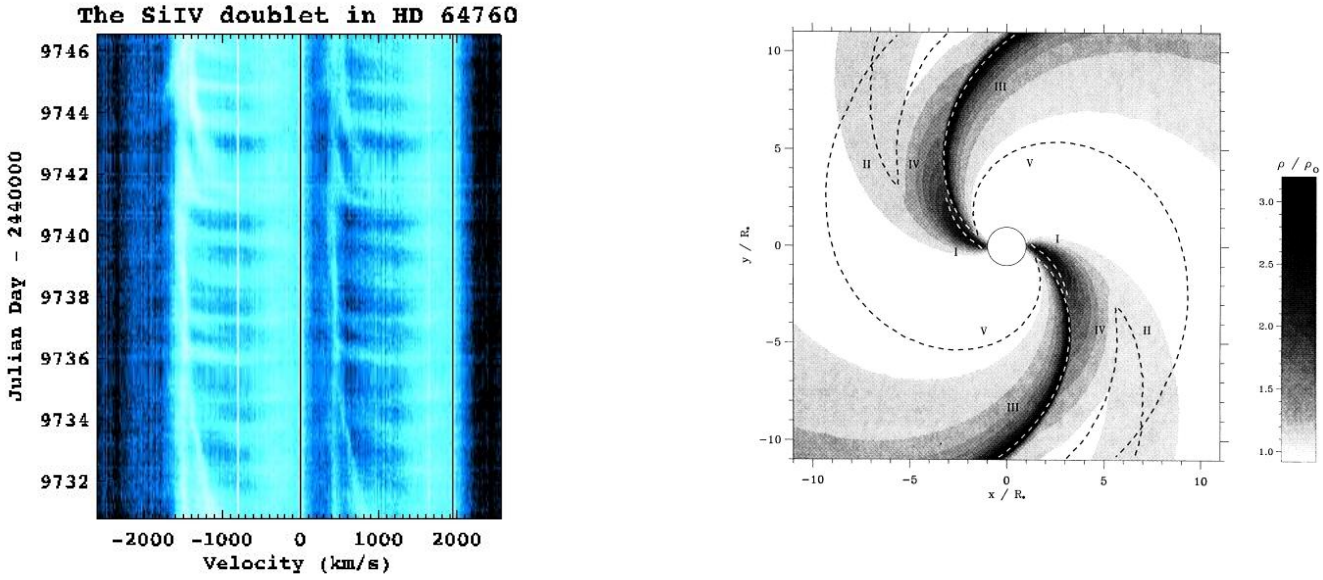


Figure 7.22: Left: grey-scale montage illustrating the variability of the Si IV doublet of HD 64760 as seen during the two weeks monitoring with IUE. Right: the corotating wind interaction region model proposed to explain the variations seen in the P Cygni profiles of HD 64760: a denser and slower wind emitted by two opposite spots on the surface of the star collides with the faster wind emitted over the remaining surface of the star. In combination with the stellar rotation, this leads to the formation of a spiral-like corotating enhanced density region in the wind.

ages. This is one of the major objectives of the GALEX mission.

Galaxies are huge ensembles of stars (between 10^7 and 10^{12}) and clouds of dust and gas held together by their mutual gravitational attraction. They are usually classified into spiral, elliptical or irregular galaxies. Spiral galaxies have a large concentration of stars and matter in the centre, the so-called bulge. In addition they have spiral arms

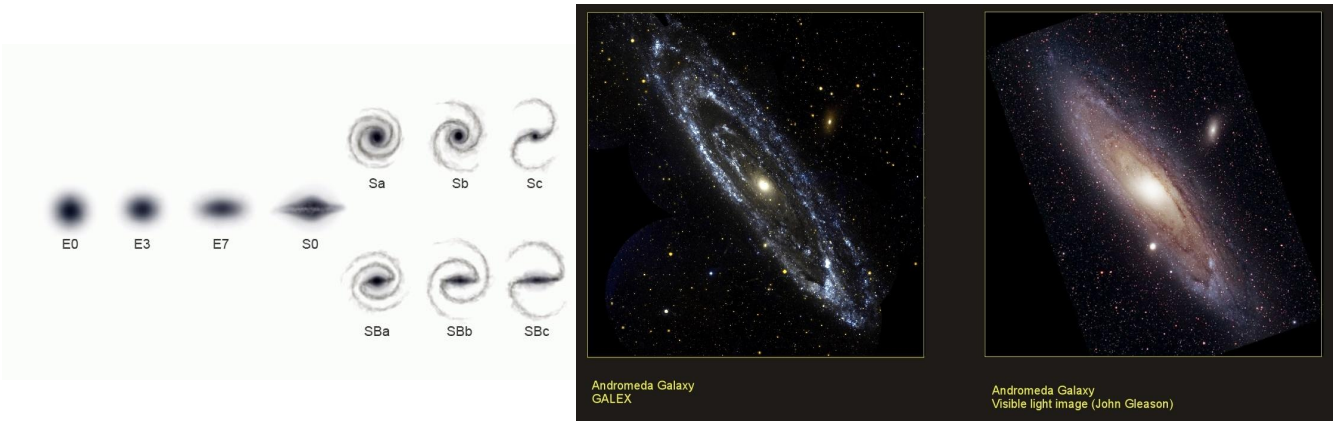


Figure 7.23: Left: types of galaxies according to the Hubble classification. E stands for elliptical, S for spiral and SB for barred spiral galaxies. Right: GALEX and optical observations of the Andromeda galaxy, M 31. In the UV image, the blue and red colours stand for far and middle-UV photons respectively.

that extend outwards. These spiral arms are rich in gas and dust and are often the theater of a strong star formation activity. Active sites of star formation are mainly identified by the presence of hot (hence blue) massive stars of spectral type OB. Since O stars have rather short lifetimes (see above), their presence is a clear indication of a 'recent' star formation event. Spiral galaxies that are bright in the UV are actively forming stars. For instance, in the starburst galaxy M 82, the star formation rate reaches about $10\,M_\odot\ yr^{-1}$, about a factor ten higher than in our own Galaxy. On the contrary, elliptical galaxies do not contain much gas and are hence no longer forming new stars. Their red colour indicates that they mainly host rather old (cool) stars. It should be stressed that, although star formation activity is always associated with strong emission in the UV domain, UV observations will efficiently trace this activity only in galaxies with a relatively low redshift. Indeed, for high redshift objects, on the contrary, the bulk of the light emitted by the starburst events is shifted into the visible or IR domain (the most distant objects will be ideal targets of the future James Webb Space Telescope, JWST).

Chapter 8

The exploration of the Solar System

The study of the Solar System is certainly the field that has witnessed the most spectacular developments thanks to the advent of space exploration. We now have the possibility to perform in-situ measurements both around the planets as well as on their surface. Orbiting space probes have provided us with spectacular images of distant places all over the Solar System. Sample return missions are being considered (or have been implemented for the Moon and for some minor bodies) and even manned missions to other planets are envisaged.

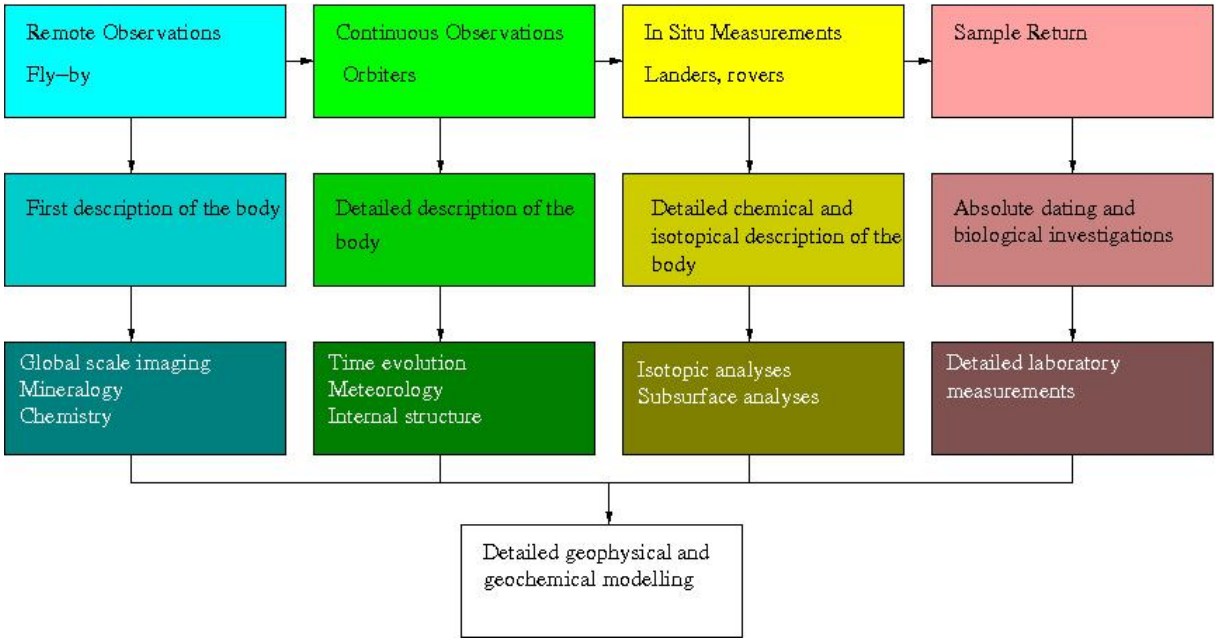


Figure 8.1: Increasingly complex space missions provide increasingly rich scientific information. For Solar System missions, sample return missions are the most sophisticated programmes that can be conducted with robotic spacecraft.

In this chapter, we illustrate some of the progresses made in our understanding of the Solar System by considering three different topics: the study of the Sun and the solar wind, the exploration of planet Mars and finally the investigation of Saturn and its moons.

8.1 The Sun, the solar wind and its interactions with the magnetospheres of the planets

Multiwavelength observations of the Sun are of fundamental importance for the understanding of many phenomena occuring in the atmosphere of our daily star. For instance, observations in the UV and X-ray domain are most relevant in the context of flares, eruptions and coronal heating. They are also fundamental in the context of space weather.
Several planets of the Solar System have a sizeable magnetic field. These are Mercury, the Earth, Jupiter, Saturn, Uranus and Neptune. These planets and their magnetic fields hence provide an obstacle to the flow of the solar wind. The solar wind flows around the magnetosphere, i.e. the volume of space occupied by the magnetic field and plasma population generated in the planet's vicinity. These interactions lead to new phenomena that are ideally studied from space.

8.1.1 Instrumentation in space physics

In-situ measurements of the plasma properties are especially important to understand the interactions between the solar wind and the planets of the Solar System. These mainly concern measurements of the local magnetic and electric fields, the plasma velocities and compositions.

The current generations of magnetometers have a dynamical range of about one million (i.e. the largest field they can actually measure has an intensity $10^6\times$ their resolution). Near the Earth, the largest field values are about 45 000 nT (nano Tesla), whereas they are about 30 nT in the inner heliosphere.

Fluxgate magnetometers rely upon the generation of a voltage that is proportional to the magnetic field strength. The principle is to use a small magnetically susceptible core wrapped by two coils of wire. An alternating electric current flows through one of the coils, hence generating a modulated magnetic field in the core. This constantly changing field induces an electrical current in the second coil that can be measured. In the absence of an external magnetic field, the two currents should match. However, an external magnetic field will generate a phase-shift between the input and output currents.
Magnetometers often have a non-zero offset (i.e. $\vec{B} = \vec{0}$ will not yield a zero measurement) that needs to be calibrated regularly in flight. This is achieved by spacecraft spinning. Indeed, in a constant or slowly varying magnetic field, the average measured value of the $\vec{B}$ component perpendicular to the spin axis over a rotation cycle should be zero. An average different from zero provides a direct measure of the offset. The magnetometer sensors are usually placed on a deployable boom (often 5 to 10 m long) to remove them from the immediate vicinity of the spacecraft and hence minimize perturbations from the fields generated by the electric circuits of the satellite itself. In most cases this is not sufficient to get rid of the magnetic field of the satellite and one has to place compensating magnets on the spacecraft to eliminate its effect on the magnetometers.

Magnetic fields can also be measured by electron drift instruments. Electrons in a magnetic field move in a helicoidal motion around the field lines. The period of the circular motion (the so-called gyroperiod) is given by $T_g = \frac{2\,\pi\,m_e}{e\,B}$. In electron drift devices, electrons are emitted from the spacecraft into two orthogonal directions in the plane perpendicular to $\vec{B}$. Beam 1 returns to the electron detector mounted on the spacecraft after a flight time $\Delta\,t_1 = t_1 - t_0 = T_g\,(1 - \frac{v_{sc}}{v_e})$ where v_e and v_{sc} are the electron and spacecraft velocities respectively. For beam 2, this time is $\Delta\,t_2 = t_2 - t_0 = T_g\,(1 + \frac{v_{sc}}{v_e})$. Hence $T_g = \frac{\Delta\,t_1 + \Delta\,t_2}{2}$ and $B = \frac{4\,\pi\,m_e}{e}\,(\Delta\,t_1 + \Delta\,t_2)^{-1}$.

Electric fields are measured by long (60 to 100 m) wire booms on a spinning spacecraft: a sphere (typically with a diameter of 8 cm) with a conducting surface is attached at the end of each boom, whilst the rest of the boom has an

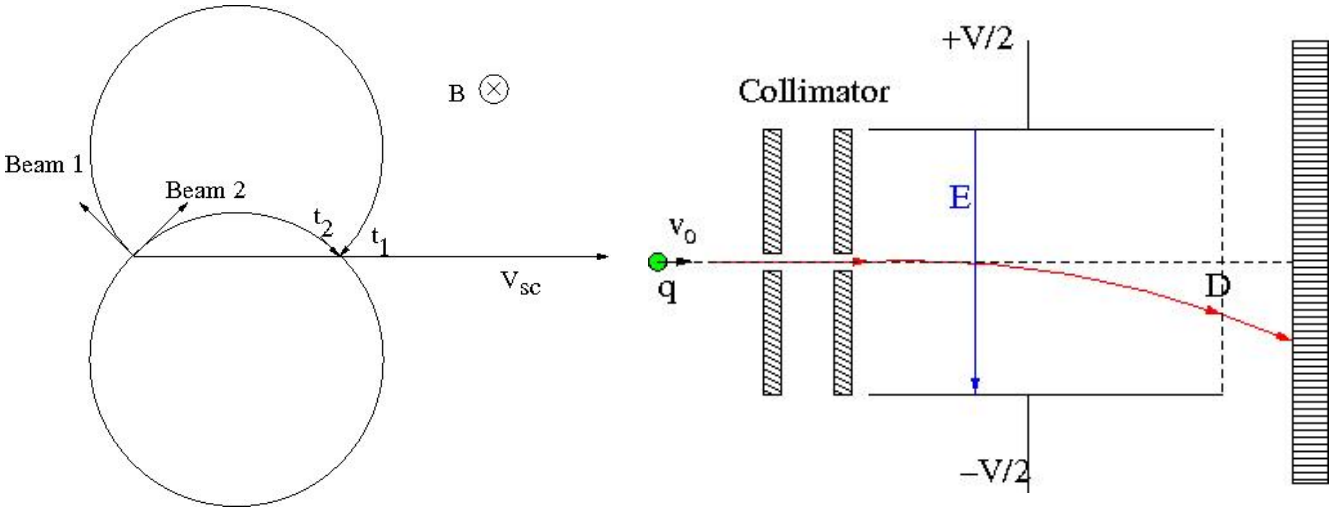


Figure 8.2: Left: The basic principle of an electron drift instrument. Electrons are ejected at velocity v_e in two orthogonal beams. Whilst the electrons of beam 1 return to the detector (which is moving along with the spacecraft at velocity v_{sc}) after a time $\Delta\, t_1$ ($< T_g$), those from beam 2 return only after more than a full gyroperiod (see text). Right: simplified scheme of an energy per charge plasma analyser. The charged particles cross the carbon foil at a deflection D with respect to the axis of the device. Afterwards, they move along a straight line and finally hit the solid state detector.

insulating cover. These booms are usually used by pairs. If the electric field does not vary on the time scale of the spacecraft rotation, then to first order the potential of each sphere is equal to the local potential so that the voltage between the two spheres can be written $\Delta\, V = V_2 - V_1 = -2\, E_{\rm spin}\, l\, \cos\left(\omega_{\rm spin}\, t + \phi\right)$ where $E_{\rm spin}$ is the electric field in the plane perpendicular to the spin axis and l is the length of the boom.

Finally, energy per charge plasma analysers allow to measure the kinetic energy per unit charge E_k/q of the plasma particles. The principle is to measure the deflection of a charged particle of velocity $\vec{v_0}$ under the influence of an electric field perpendicular to the incident velocity field. This allows to determine the local velocity vector of the solar wind which is the sum of the bulk motion of the wind plus the thermal speed. Hence, these devices determine both the bulk velocity of the solar wind as well as its temperature (from the 1 σ dispersion of the velocity measurements). A schematic view of such a device is shown in Fig. 8.2. The charged particles first pass through a collimator that selects only particles with the right velocity vector direction (along the axis of the instrument chamber). Once inside the instrument, the charges are submitted to an electric field $\vec{E}$ perpendicular to $\vec{v_0}$ over a distance L along the axis of the device. The charges then cross a thin carbon foil with a deflection $D = \frac{q\,E}{2\,m\,v_0^2}\, L^2$ from the axis of the device. D thus provides a quantity inversely proportional to the ratio between the kinetic energy and the charge E_k/q. When the charge crosses the carbon foil it strips off some electrons, hence allowing to measure D as well as the time when the particle crosses the foil. Unfortunately, the value of E_k/q alone is not sufficient to determine the nature of the ions. For instance, fully ionized helium (He^{2+}), carbon (C^{6+}), nitrogen (N^{7+}) and oxygen (O^{8+}) all have the same E_k/q ratio if they are moving at the same velocity! Therefore one adds another stage to the instrument where the charges move freely (i.e. with no external electric field) towards a solid state detector. Finally, after a time of flight τ (straight line motion), the particle hits the detector and generates a 'stop' signal. The solid state detector allows to directly measure the energy of the incoming particles. The time of flight can be used to determine the residual energy $E_r = \frac{1}{2}\, m\, (\frac{d}{\tau})^2$ where d is the distance between the carbon foil and the final detector. This then allows to determine m. The residual energy is related to the kinetic energy E_k and the loss of energy due to the crossing of the carbon foil (the latter is calibrated during ground-based tests). Finally,

the knowledge of E_k and m allows to compute q (from the D measurement) and v_0. Actual instruments use this principle, but have a more complex design (allowing for instance to collect charges from many different directions at the same time).

8.1.2 SOHO

The Solar and Heliospheric Observatory (SOHO) was selected as a cornerstone mission of ESA's Horizon 2000 programme in 1984. SOHO is a collaboration between ESA and NASA: the spacecraft was built under the control of ESA and the launch and mission operation (including communication via the DSN) were provided by NASA. SOHO was launched in December 1995 by an Atlas II Centaur rocket from Cape Canaveral. The mass of the spacecraft at launch was 1850 kg and its dimensions are (height $\times$ breadth $\times$ width): 4.3 m $\times$ 2.7 m $\times$ 3.7 m.
After a cruise of four months, SOHO reached its operational halo orbit around L_1. SOHO is indeed the first solar observatory that revolves around L_1 thus benefiting from an uninterrupted view on the Sun.
The mission operations are currently approved until the end of 2020. SOHO is a three-axes stabilised spacecraft, although it lost its gyroscopes in February 1999. In June 1998, the contact with the mission was lost for several weeks due to an erroneous command sent from the ground, but SOHO was brought back to nominal operation four months later.

Figure 8.3: Left: the SOHO spacecraft. Right: the orbit of SOHO around L_1. Note that the distance between L_1 and the Sun is not to scale.

There are twelve instruments aboard SOHO. These are CDS (Coronal Diagnostics Spectrometer), CELIAS (Charge Element and Isotope Analysis System), COSTEP (Comprehensive SupraThermal and Energetic Particle Analyser), EIT (Extreme Ultraviolet Imaging Telescope), ERNE (Energetic and Relativistic Nuclei and Electron experiment), GOLF (Global Oscillations at Low Frequencies), LASCO (Large Angle and Spectrometric Coronagraph), MDI/SOI (Michelson Doppler Imager/Solar Oscillations Investigation), SUMER (Solar Ultraviolet Measurements of Emitted Radiation), SWAN (Solar Wind Anisotropies), UVCS (Ultraviolet Coronagraph Spectrometer) and VIRGO (Variability of Solar Irradiance and Gravity Oscillations). It is not possible to describe all of these instruments here and we will rather concentrate on two of them (CELIAS and VIRGO).

The CELIAS instrument is designed to study the composition of the solar wind as well as of solar and interplanetary energetic particles. CELIAS consists of three different sensors with associated electronics, which are optimized each for a particular aspect of ion composition. These aspects are the elemental, isotopic, and ionic charge composition of the plasma. The main scientific goals of the instrument are particle abundance observations of the solar wind and of solar energetic particles to address questions in solar physics such as: the steady heating process in the corona, the solar wind acceleration processes, the dynamic heating phenomena driven by magnetic fields or

waves (e.g. heating in coronal mass ejections, filament eruptions, flares,...), solar energetic particle acceleration processes, the relation between composition and events/regions in the solar atmosphere...

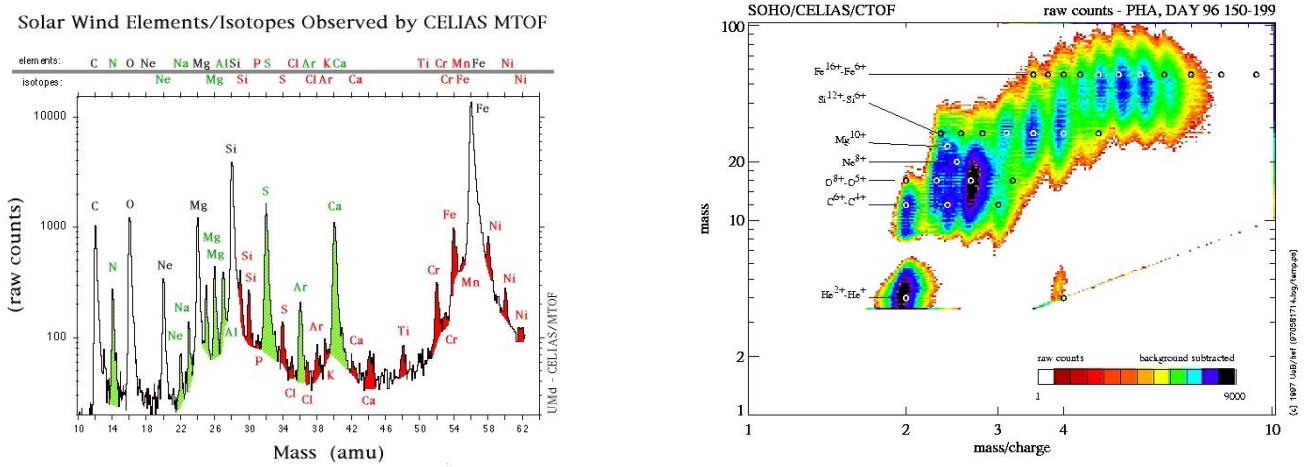


Figure 8.4: Left: CELIAS/MTOF spectrum of the solar wind uncorrected for the efficiencies of the instrument. The MTOF sensor was operated in a mode optimized to detect chemical elements heavier than sulphur. This is why the Ca and Fe peaks are so prominent compared to that of oxygen. The isotopes that were measured for the first time in situ by SOHO are shown in red. Right: The CELIAS/CTOF sensor determines for every ion that enters the detector its mass, charge and energy. The plot shows a histogram of the minor ions (hence excluding hydrogen) in the mass versus mass/charge plane (the data are hence integrated over ion energy).

The CELIAS instrument consists of three different sensor units (CTOF, MTOF, and STOF) coupled to a Digital Processing Unit (DPU). All three CELIAS sensors employ electrostatic deflection systems in combination with time-of-flight (TOF) measurements of the impinging particles. The CTOF and STOF sensors employ also silicon solid state detectors to determine the residual energy of particles. Thus for these two sensors the quantities energy per charge, E_k/q, time-of-flight (TOF), and energy, E_r, are determined for the incoming particles. CTOF and STOF are aimed at different energy ranges, CTOF being devoted to solar wind and low-energy suprathermal ions, while STOF is devoted to higher energy suprathermal and low-energy solar energetic particles. The MTOF sensor is able to measure the composition of the less abundant elements in the solar wind as well as the isotopic composition of the more abundant heavy ions (up to Ni). Finally, a solar EUV monitor, observing in the light of the He II λ 304 Å transition, is structurally connected with STOF.
The solar wind Charge Time of Flight (CTOF) sensor covers the energy per charge (E_k/q) range between 0.1 and 55 kV, which fully covers the solar wind range. The upper limit of 55 kV corresponds to a bulk speed of about 1000 km s^{-1} for Fe^{8+}. The sensor determines the composition, charge state distribution, kinetic temperature, and speed of the more abundant solar wind ions (e.g. He, C, N, O, Ne, Mg, Si, and Fe).

The VIRGO experiment provides continuous high-precision measurements of the solar total and spectral irradiance and spectral radiance variation, continuous measurements of the solar polar and equatorial diameters, as well as frequencies, amplitudes and phases of oscillation modes in the frequency range of 1 μHz to 8 mHz.
The total irradiance is measured with active cavity radiometers: the spectral irradiance by three-channel Sunphotometers (operating between 4020 and 8620 Å) and the radiance with 12 resolution elements on the solar disk using the Luminosity Oscillations Imager (LOI). The LOI is fed by a Ritchey-Chrétien telescope with a primary mirror of 60 mm diameter. The LOI bandpass is centered at 5000 Å and has a width of 50 Å. The detector is a deep diffused silicon photodiode with 12 scientific and 4 guiding pixels. The shape of the 12 pixels was optimised in such a way as to allow detection of low degree ($l < 7$) non-radial pulsation modes.

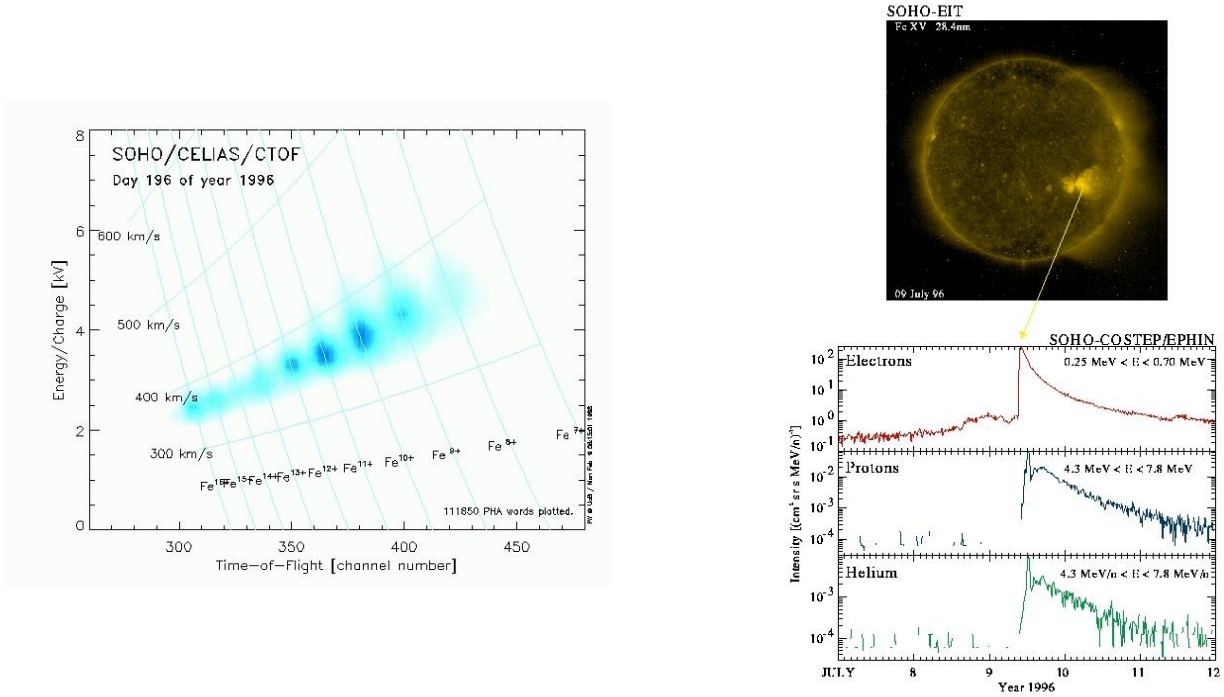


Figure 8.5: Left: The distribution of iron charges as a function of velocity (measured by the E_k/q ratio). The nearly horizontal lines correspond to velocities of 300, 400, 500 and 600 km s^{-1}. Right: A solar energetic particle event (flare and coronal mass ejection event) observed on 9 July 1996 with the COSTEP instrument aboard SOHO. The upper panel shows the Sun during the flare as seen by EIT in the light of the Fe XV λ 284 Å line. The bottom panels show the intensities of the 0.25 – 0.7 MeV electrons, the 4.3 – 7.8 MeV protons and the 4.3 – 7.8 MeV α particles (from top to bottom).

The main scientific objectives of VIRGO can be summarized as the detection and classification of low-degree g modes of solar oscillations[1], the determination of the sound speed, density stratification and rotation in the solar interior, the study of the solar atmosphere, the search for the long periodicities or quasi-periodicities that have been found in other solar parameters, determination of the solar asphericity and its variation with time, the study of the influence of solar active regions and other large-scale structures on total and spectral irradiance (providing data also for terrestrial climate modelling efforts).

8.1.3 Cluster II

The Cluster mission was approved as a counterpart to SOHO in the ESA Horizon 2000 programme. In 1996, the launch of the first Ariane 501 with Cluster as payload failed and the spacecraft were destroyed. In July and August 2000, a new set of four spacecraft, the Cluster II satellites, was launched with two Soyuz rockets from Baikonour into an eccentric polar orbit with a period of 57 hours. The mission operations are still going on and are currently approved until December 2020.
The main science objectives of the Cluster mission are the study of the interaction between the solar wind and the

[1] Although there have been claims of the detection of such gravity modes in the Sun, there is no consensus about these results in the helioseismic community.

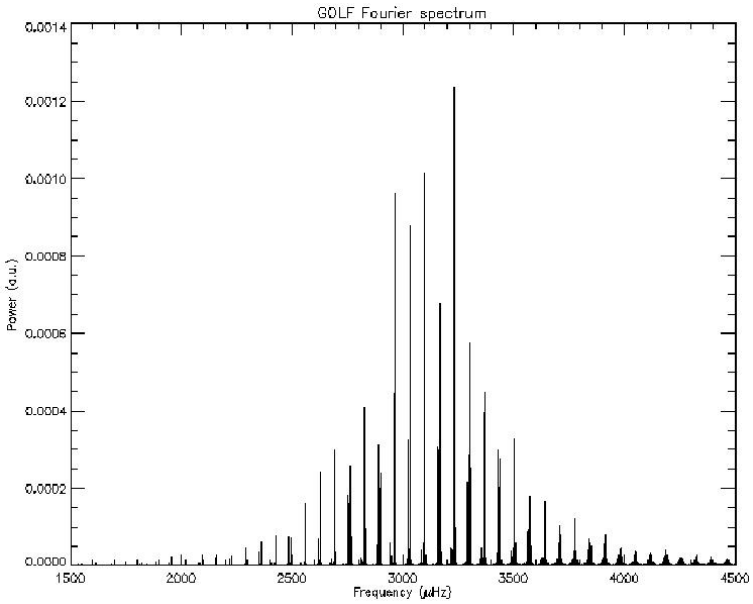
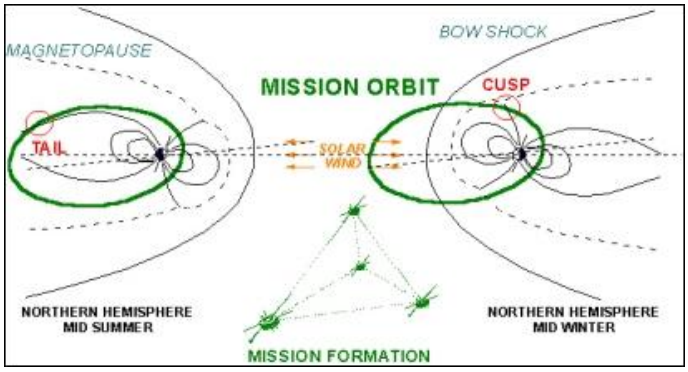


Figure 8.6: The p-mode Fourier spectrum of the Sun as derived from 690 days of GOLF radial velocity observations.

Earth's magnetosphere through an investigation of the small-scale 3-D structure of the Earth's plasma environment. For instance, the polar orbit of the mission allows to study the direct entry of solar wind particles through the polar cusps. To achieve these science goals, Cluster actually features four identical spacecraft flying in a tetrahedral configuration. The separation between the spacecraft varies between 600 and 20 000 km. Each Cluster spacecraft consists of a cylinder of height 1.3 m and diameter 2.9 m. The solar array provides a power of 224 W and the satellites rotate at a rate of 15 rounds per minute. Each satellite carries twelve instruments, including fluxgate magnetometers, electron drift instruments and ion spectrometers. The orbit of each spin-stabilised spacecraft is selected so that each one is located at a vertex of a predetermined tetrahedron when crossing the regions of scientific interest. The separation between the spacecraft is adjusted depending on the spatial scales to be investigated.



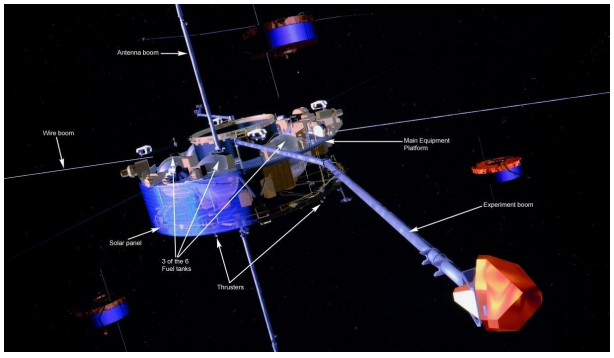


Figure 8.7: Left: Cluster II seasonal orbit, allowing the spacecraft to explore both the bow-shock and the magnetotail. Right: anatomy of one of the Cluster spacecraft.

Having several spacecraft such as in the Cluster mission has the advantage that one can distinguish genuine variability of the plasma at a specific location in the magnetosphere from variations due to the motion of the spacecraft. The orbits of the Cluster spacecraft are eccentric ($e = 0.66$, $a = 11.7\,\mathrm{R}_\oplus$). Whilst the orbits of the spacecraft maintain a constant orientation with respect to an inertial frame of reference, they rotate with respect to the Earth's magnetosphere as the Earth revolves around the Sun. This then allows to explore both the bow-shock between the

solar wind and the magnetosphere, as well as the magnetotail (see Fig. 8.7)[2].

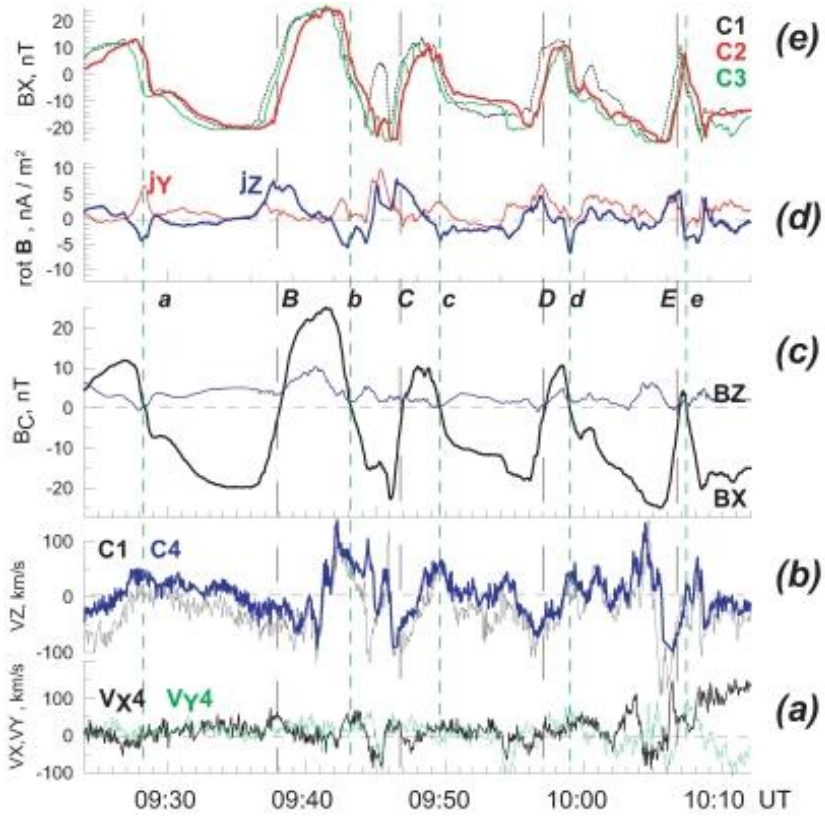


Figure 8.8: Observations of rapid large-amplitude magnetic variations in the centre of the magnetotail by the Cluster II spacecraft. Panels (a) and (b) show the ion plasma flows as measured by spacecraft C1 and C4. Panel (e) illustrates the B_x component as measured by three spacecraft (C1, C2, C3), this then yields the B_x component at the barycentre of the Cluster configuration (panel (c)).

8.1.4 The future: Solar Orbiter

Solar Orbiter, launched on 10 February 2020 is the first M-class mission of ESA's Cosmic Vision plan. After a three years cruise, Solar Orbiter will orbit the Sun on a highly eccentric orbit with a period of 168 days with a perihelion distance of 0.28 AU and an aphelion distance around 0.8 AU. During the mission, regular encounters with Venus will crank the orbit out of the plane of the ecliptic up to a solar latitude of more than $30°$. The thermal load over the orbit foreseen for Solar Orbiter varies between 0.8 and 17.4 kW m^{-2} at farthest aphelion and closest perihelion respectively. The spacecraft will be three-axes stabilized. Because of the thermal constraints, the instruments have to be protected by a thermal shield and the high-gain antenna may have to be folded several times during the mission to place it in the shade of the heat shield. During this time, no data can be downlinked. This sets constraints on the telemetry rate, the on-board storage capabilities and the instrument autonomy.
The payload features a series of instruments, both for in-situ measurements (solar wind analyser, energetic particle detector, magnetometer, radio and plasma wave instrument) and remote sensing (polarimetric imager, extreme UV imager, spectral imager of the corona, X-ray imaging spectrometer,...).

Solar Orbiter is not the first mission that will observe the Sun and the heliosphere from outside the plane of the ecliptic. In fact, the Ulysses mission orbited the Sun on a 6 years period at an inclination of $80°$ with respect to the solar equator (perihelion distance of 1.4 AU and aphelion at ~ 5.4 AU). However, Solar Orbiter will allow a close

[2]Note that a different approach was adopted for NASAs WIND mission that used lunar swingbys to rotate the orbit in such a way that the WIND spacecraft continuously explored the sunward side of the magnetosphere.

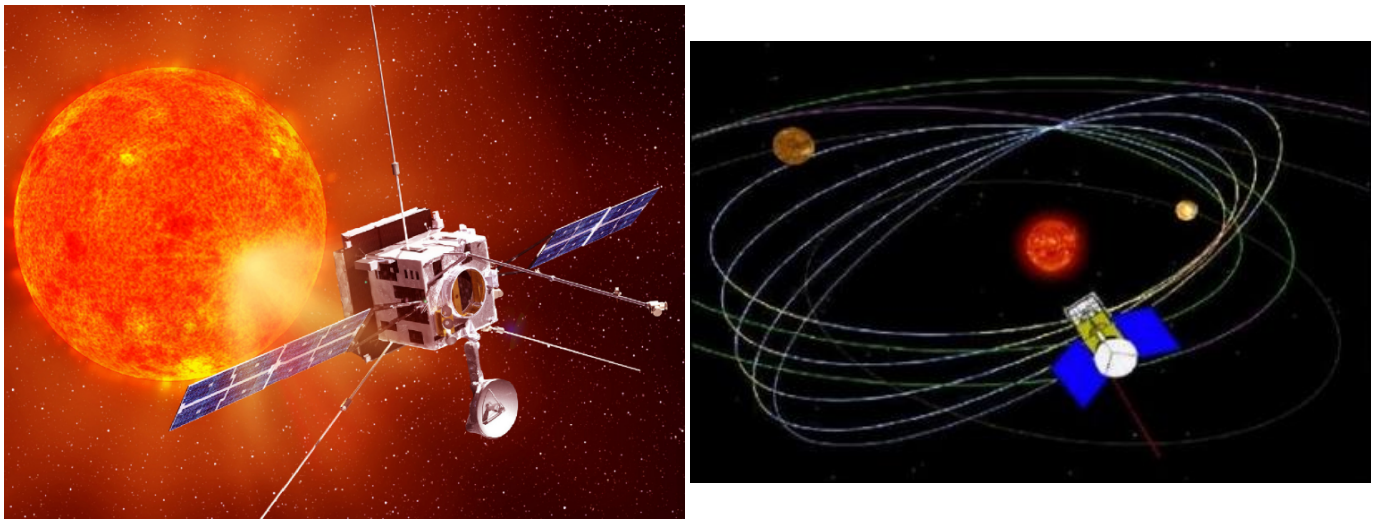


Figure 8.9: Left: artist view of the Solar Orbiter spacecraft flying over the Sun. Right: the evolution of the orbit of Solar Orbiter. The orbit of the spacecraft is progressively cranked out of the plane of the ecliptic through regular encounters with Venus.

up study of the Sun, the first out of the ecliptic plane. Furthermore, the orbit has been designed in such a way that the spacecraft is nearly in corotation with the Sun during the perihelion part of its orbit.

8.2 Missions to planet Mars

Remote sensing as well as in-situ exloration of planetary surfaces have boosted our knowledge on the properties of the planets of the Solar System. Mars is probably one of the best studied objects in this respect. In this section, we will start by discussing some general aspects of planetary orbiters and then present three particular missions: the Mars Exploration Rovers, Curiosity, and Mars Express. Finally, we briefly address some aspects of a future manned mission to Mars.

8.2.1 Remote sensing planetary missions

An important ingredient of any planetary mission is the camera. Geomorphological information can be obtained from broadband visible orbital imaging in the case of planets with thin or absent atmospheres, as in the case of Mars or Mercury, respectively. For planets with thick atmospheres, the opacity as a function of wavelength must be taken into account. For Venus for instance, the only solution is microwave (radar) imaging.

Older missions, such as the Mariner, Viking and Voyager probes, used vidicon based imaging systems. A vidicon is a storage camera tube in which a charge density pattern is formed by the imaged scene on a phosphorous surface (antimony trisulfide Sb_2S_3). This surface is then scanned by a beam of low-velocity electrons emitted by a small cathode-ray tube. The electrical potential varies in proportion to the levels of light. Modern cameras use CCD detectors. These cameras record black and white images through 3 different filters hence allowing the reconstruction of coloured images afterwards during data processing on the ground. Actually, not all CCD cameras have 2D arrays of sensors. For instance, the Mars Orbiter Camera aboard MGS (Mars Global Surveyor) has a detector consisting of a single line of CCD pixels. The 2D images are built up as the image of the Martian surface drifts across the detector whilst the spacecraft orbits the planet. In the near or mid-infrared, the most-frequently used detectors are arrays constructed from InSb, SiAs, Ge or InGaAs substrates. Most of these detectors need to

be cooled down (to less than 77 K for the InSb arrays). The advantage of the Ge and InGaAs detectors is that they can operate at higher temperatures (250 to 270 K).

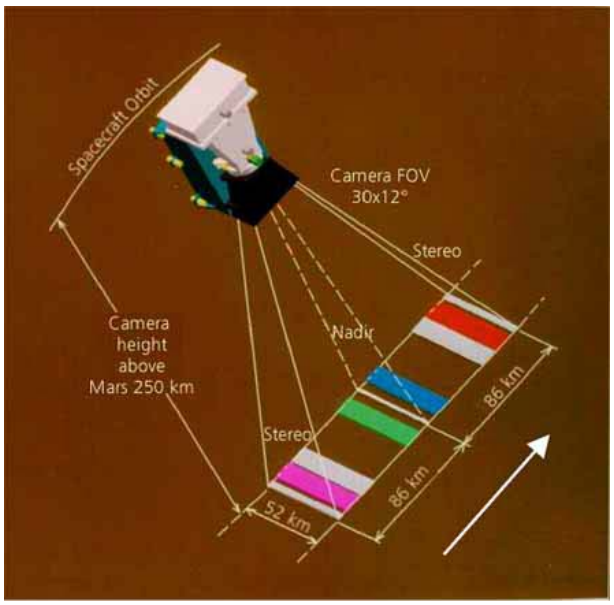


Figure 8.10: The operation principle of the HRSC camera aboard Mars Express. The 9 CCDs are aligned with the flight direction and the stereo channels allow to build 3D images. This stereo channel operates over a 1800 Å wide band centred at 6750 Å. Colour images are obtained by combining these data with additional images obtained through several filters.

Currently, the most sophisticated cameras allow to obtain 3D views of a planetary surface. For instance, the HRSC camera aboard Mars Express offers the possibility to obtain such 3D images through the principle of stereo observations, where views from three directions are combined. For the HRSC, these directions are downward (to the nadir), backward ($-18.9°$) and forward ($+18.9°$) with respect to the motion of the satellite over the ground.

On board planetary orbiters, two types of spectroscopic diagnostics can be taken:

- reflectance spectra: when solar photons hit the surface they are absorbed in the minerals by various processes and the wavelength-dependence of these processes leaves its signature in the reflected or emitted light. This provides information on the chemistry of the minerals in the uppermost microns of the surface (the surface layer that can be probed has a thickness of a few times the photon wavelength).

- thermal emission spectra: the thermal emission of a planet occurs at wavelengths that are larger than those of the reflected Sun-light. The thermal emission spectrum contains molecular absorption bands due to vibrational transitions. Indeed, each molecule consisting of N atoms has $3\,N - 6$ modes of vibration. For water, there are 3 modes that produce bands at 2.663, 2.738 and 6.270 μm. It has to be stressed though that if there exists an atmosphere, it can be difficult to separate the surface signature from that of the atmosphere and the aerosols.

The most sophisticated spectrographs are imaging spectrometers. Actually, they consist of a slit-grating spectrograph. A CCD-type detector records the spectra as a function of spatial position along the slit height. A scanning mirror then allows to collect the spectral information over a 2D area. This procedure results in a so-called data cube consisting in 2D images as a function of wavelength.
Modern planetary missions involving an orbiter frequently use a so-called Ground Penetrating Radar (GPR) for the sub-surface planetary exploration. These radars rely upon microwaves because they are easily transmitted through clouds and the atmosphere and penetrate to a certain depth into the ground. The reflection properties of these microwaves strongly depend on the properties of the sub-surface layers.

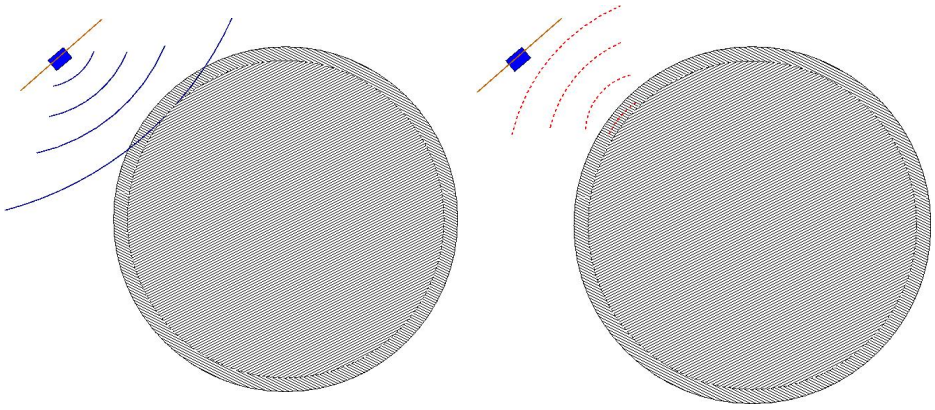


Figure 8.11: Schematic illustration of the principle of a GPR.

The radar equation can be written as $P_r = \frac{P_t G_t}{4\pi R^2} \sigma_{rt} \frac{A_r}{4\pi R^2}$. Here P_t and G_t are the power transmitted and the gain of the emitter, σ_{rt} is the radar cross-section, A_r is the effective area of the scatterer as a receiving antenna and R is the distance between the emitter and the scatterer.
The electrical and magnetic properties of rocks, soils and fluids control the speed of propagation of radar waves. This is usually expressed by the dielectric constant ε. Indeed, whilst most material is largely transparent to microwaves, these waves are strongly absorbed by conductors (the microwaves cause an electric current that heats the material). For instance, the dielectric constant of water is significantly different from that of silicates and the polarisation of the wave strongly depends on the presence of water. The principle of a sounding radar is therefore based on the reflection of radio waves by boundaries between different materials. By analysing the echoes, it is possible to deduce information (such as the composition and physical state) on the kind of material causing the reflection. In principle, a GPR thus provides a powerful tool to search for water below the surface of telluric planets.

The orbiting remote sensing probes allow to characterize sites for landers. The choice of the landing sites depends on scientific and technical criteria. The scientifically most interesting sites depend on the science topics: in the case of Mars, the search for traces of biological activity would ideally focus on former lakes and seas, geochemical research would rather concentrate on volcanic material and climatology research would be highly interested in the polar caps. Practical considerations however limit the choice of the landing sites. For instance, the landers will preferentially land near the equator where more solar radiation is available for the solar arrays and where the temperatures are less extreme than in higher latitude regions. For sample return missions, the equator is also quite interesting since it offers the highest benefit in terms of ΔV provided by the rotation of the planet (for the return journey). The altitude is also an important parameter, since the parachutes need to have the time to brake the motion of the lander. For instance, the residual atmosphere above Olympus Mons (26 km altitude) is not sufficient to allow landing with parachutes there. An ideal landing site should also be flat and smooth. The selection of a landing site must also account for the uncertainties that affect the precise location where the lander will touch the ground. For a planet with an atmosphere such as Mars, this will depend on the accuracy of the entrance trajectory, but also on the actual density of the atmosphere, the wind speed and direction,... For missions to Mars, the total uncertainty on the location of the landing site can typically amount to 200 km.
As an illustration, let us consider the choice of the landing site for the ExoMars 2022 mission[3]. The search for a suitable landing site began in December 2013, when the scientific community was invited to make proposals to the Landing Site Selection Working Group. Four sites were retained for further investigation: Aram Dorsum, Hypanis

[3]ExoMars is a robotic exploration programme in collaboration between ESA and Roscosmos (Russia). The goal is to establish whether life has existed on Mars. ExoMars consists of two missions. ExoMars 2016 features the Trace Gas Orbiter and the Schiaparelli entry, descent and landing demonstrator module. ExoMars 2016 was launched in March 2016 and arrived at Mars in October 2016. The orbiter was successfully inserted into Martian orbit, whilst the lander failed to safely land. ExoMars 2022, comprising a rover and a surface platform, is currently foreseen for launch in 2022.

Vallis, Mawrth Vallis and Oxia Planum. For one year, these candidate sites were evaluated in terms of engineering constraints of descent and landing, and the best possible scientific return of the mission. Indeed, ExoMars 2022 is designed to search for evidence of past or present Martian life. For this purpose, it needs to explore an area with ancient rocks where liquid water was once abundant, and it must be able to drill into the ground and extract samples from 2 m below the surface. To ensure a safe landing, the entry probe needs enough atmosphere to effectively slow down its descent via the parachutes. Therefore, the touchdown location must be in a low-lying area of the planet. Furthermore, it should be rather flat, i.e. free of geological features such as craters, steep slopes, and large rocks. Because of the uncertainties on the density of the Martian atmospheres and on the wind directions, the uncertainty ellipse for ExoMars 2022 is 104 km long by 19 km wide (see Fig. 8.12). The four pre-selected landing sites show evidence for the presence of water in the past (more than $3.6\,10^9$ years ago) and are thus well suited to fulfil the scientific objectives of the mission. Furthermore, they offer good conditions for a safe landing at a scientifically interesting site or within a 1 km drive from such a location. In October 2015, Oxia Planum has been recommended as the primary candidate for the landing site of the ExoMars 2022 mission (see Fig. 8.12). Yet the final decision about the landing site will only be taken one year before launch.

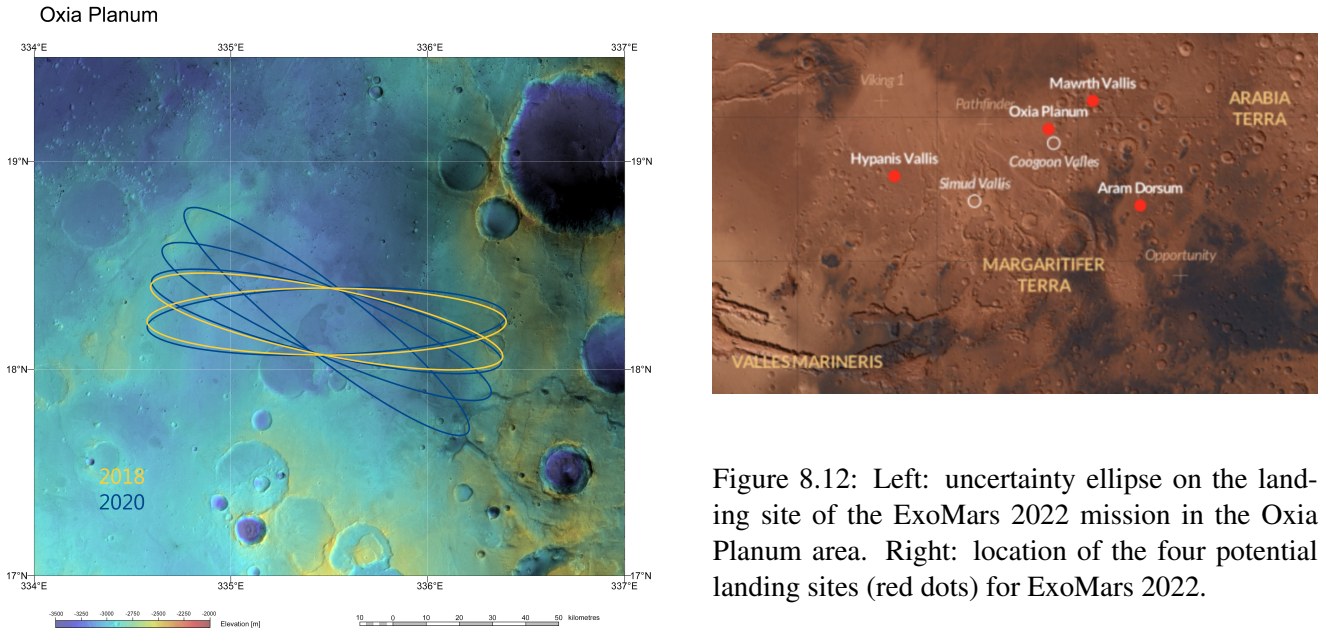


Figure 8.12: Left: uncertainty ellipse on the landing site of the ExoMars 2022 mission in the Oxia Planum area. Right: location of the four potential landing sites (red dots) for ExoMars 2022.

8.2.2 The Mars Exploration Rover Mission

The Mars Exploration Rover mission features two rovers (Spirit and Opportunity) that were launched on 10 June and 7 July 2003 respectively by a Delta II 7925 and a Delta II 7925H (H for heavy) rocket. Whilst the payloads of the two spacecraft were identical, the launch date of the second rover was slightly less favourable in terms of the ΔV required to reach Mars and the use of the more powerful Delta II 7925H launcher was thus mandatory. Actually, the two rovers could not be launched simultaneously due to restrictions at the Cape Canaveral site and the availability of a ground support team.

For each rover, the mass of the spacecraft that travelled to Mars was 1063 kg, consisting of the rover itself (185 kg), the lander platform (348 kg), the aeroshell, the parachutes and the heatshield (287 kg), the cruise stage and the propellant (243 kg).

During their 7-months journey to Mars, the spacecraft were spin-stabilized with a spin rate of 2 rotations per minute. The cruise stage featured several devices to keep the spacecraft on the right trajectory towards Mars. These were a star scanner and a Sun sensor, allowing to determine the position of the spacecraft, as well as two propellant tanks carrying hydrazin propellant, allowing to change the spacecraft velocity, move it laterally or rotate it. In addition, the cruise stage also carried a low-gain and a medium-gain antenna for communications with the ground. Through periodic corrections of the orientation of the spin axis, the antennas remained pointed towards the Earth.
Upon their arrival at Mars, the entry vehicle of the spacecraft (i.e. the aeroshell together with the lander and the rover) penetrated the Martian atmosphere. First, an aeroshell and a parachute decelerated the motion of the lander. Prior to the surface impact, the retro rockets were fired to slow the lander's speed of descent and airbags were inflated to cushion the lander at surface impact. After an initial impact, the lander surrounded by its airbags bounced on the Martian surface until it came to rest. As a next step, the airbags were deflated and retracted and the lander petals were deployed. Finally, the rover deployed its own solar arrays, ready to start the exploration of the Martian surface.

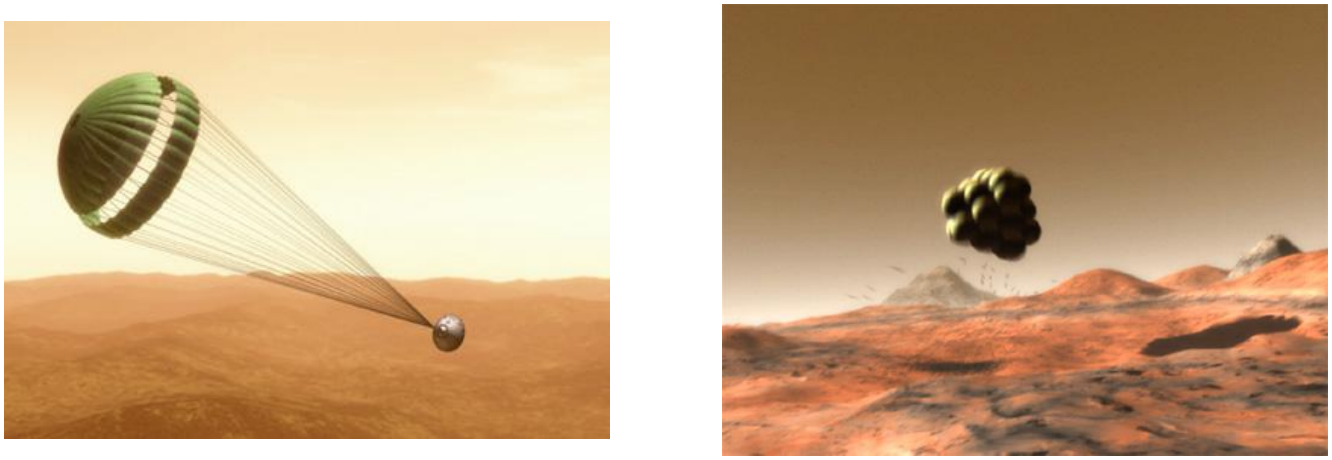


Figure 8.13: Left: the parachute system used to slow down the descent vehicle. Right: the system of airbags allowing a safe landing of the Spirit and Opportunity rovers on Mars.

The aeroshell formed a protective covering during the cruise, but also, and actually mainly, during the entry into the Martian atmosphere. The aeroshell consisted of the heat shield and the backshell. The latter carried a parachute, the electronics needed to control the usage of the pyrotechnic devices used for the separation of the various parts, an inertial measurement unit to monitor the orientation of the backshell as well as several rocket engines used to control the descent of the vehicle. The outer surface of the heatshield was recovered by an ablator (a blend featuring cork wood, binder and tiny silicate glass spheres) that dissipated the heat due to the atmospheric friction. Each descent vehicle featured four airbags with six lobes each. During the flight, the airbags were stowed along with the gas generators used for their inflation. The airbags had to be strong enough to cushion the spacecraft. They were made of Vectran, a synthetic material stronger than Kevlar (which is used in bullet-proof vests) and with a better performance at lower temperatures. The landers themselves had a tetrahedral shape consisting of a base and three side petals. Each petal was connected to the base through a hinge. Each petal hinge had a motor powerful enough to lift the entire weight of the lander. In this way, they could place the rover in an upright position regardless of the actual orientation of the lander after the impact. This operation was commanded by the rover itself which was equipped with accelerometers allowing it to determine its instantaneous orientation.

The rovers are built around a main body that hosts the warm electronics, the batteries and the computer. The ground

temperatures on the landing sites vary between about $-99°$ C and $22°$ C, whilst the atmospheric temperatures vary between about $-96°$ C and $-3°$ C. To ensure that the electronics work properly, the rover features both an active (heaters and heat rejection systems,...) and a passive (gold paint, insulation,...) thermal control. Electrical power (140 W at maximum) is provided by solar arrays. This limits the operation of the rovers to about 4 hours per Martian day (1 sol = 24 h 37.5 min). The generation of power by these arrays has decreased with time as a result of the seasonal change of the solar irradiance and the progressive dust coverage of the solar arrays.
Each rover has six wheels, each one with its own individual motor. The front and rear wheels have individual steering motors, allowing the rover to turn almost in place. The design of the suspension is such that the rover could withstand a tilt of $45°$, although the on-board computer limits the acceptable tilt to $30°$. During nominal operations, the rover moves for about 10 s (roughly 30 cm). It then stops and examines the terrain for 20 s before driving for another 10 s. In this way, the average top speed reached by the rovers is about $0.01\,\mathrm{m\,s^{-1}}$. The time delay in communication between Mars and Earth (20 min on average) limits the possibility of directly reacting to a new situation. The on-board computers therefore create maps of the surroundings and use hazard avoidance software to safely move at the Martian surface.

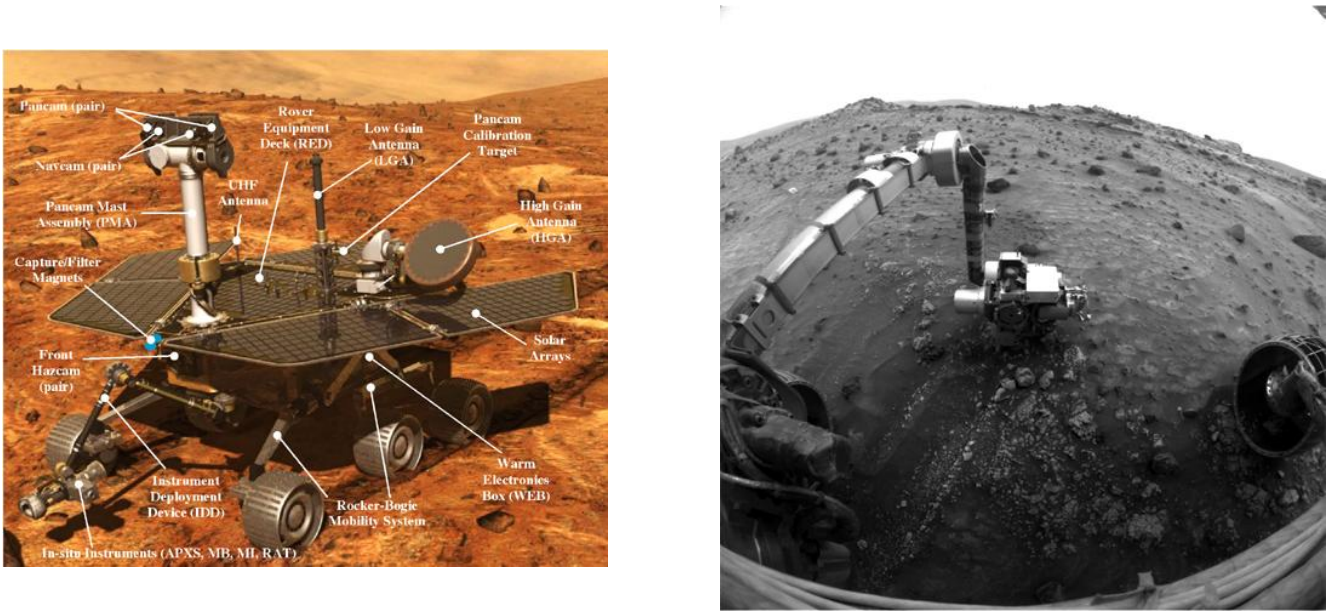


Figure 8.14: Left: a detailed view of the MER rovers. Right: the robotic arm of the Spirit rover performing measurements of the properties of the Martian soil.

Each rover features six engineering cameras for navigation and three for scientific investigation. The engineering cameras have a wide field of view of $120°$ each. Two science cameras are mounted in the panoramic camera (Pancam) mast assembly that stands about 1.4 m above the ground. Each of them is equipped with a filter wheel giving a multi-colour coverage. The Pancam allows a better point of view of the surrounding area and serves as a periscope for one of the instruments housed in the warm part of the rover. The third scientific camera is mounted on a titanium robotic arm. The latter has flexibility through three joints, allowing to perform several geological operations such as providing a close-up image with a microscopic imager, analysing the composition through spectroscopy or grinding away the outer surface (a few mm) of a rock to expose fresh material. For geolocial studies it is indeed very important to accurately determine the composition of the soil and the minerals as well as to study differences between the interior and the outer parts of the rocks since this allows to draw conclusions on the way these minerals have formed and the weather conditions they have been exposed to (including the contact with liquid water).

Communication with Earth is done either directly using the omni-directional low-gain or the steerable high-gain antenna together with the DSN or using one of the orbiters around Mars as a relais.

In 2009, Spirit got trapped in soft sand and during the year 2010, this rover was operated as a stationary science platform. In January 2011, the contact with Spirit was finally lost. Opportunity was operated as a rover for more than 14 years and drove over a distance of more than 45 km. During these 14 years, the mission also encountered a number of problems. In fact, Opportunity also got trapped in sand at several occasions and strong friction in one of the wheels forced the mission operators to drive the rover mainly in reverse motion. Episodic accumulation of dust on the solar arrays sometimes reduced the electrical power available, although wind events cleaned the arrays as was the case in March 2014. In June 2018, NASA finally lost contact with the rover. The most likely reason was a low power fault, possibly related to the global dust storm that lasted the whole summer 2018. Repeated attempts to re-establish contact with the rover failed and the mission was declared complete in February 2019.

8.2.3 Curiosity: The Mars Science Laboratory Mission

The Mars Science Laboratory (MSL), better known as Curiosity, is an ambitious NASA mission that comes as a successor to the MERs Spirit and Opportunity. For this purpose, MSL relies on a rover (Curiosity) equipped with a huge number of instruments for in-situ measurements (see below). The main scientific driver of this mission is to determine whether life ever arose on Mars. This question will be addressed by determining the abundances of C, H, N, O, P and S. These elements are necessary ingredients for any form of life as we know it on Earth. Another question that will be addressed concerns the cycle of carbon and water throughout the history of planet Mars. Curiosity will further characterize the Martian climate and weather. It will measure the radiation environment at the surface, due to high-energy radiation, cosmic rays, and solar particle bombardment. Curiosity will also study the geology of Mars, with the goal to understand the processes that created and modified the Martian crust. Last, but not least, Curiosity is a step in the preparation of human exploration of planet Mars, as it has demonstrated the technologies for landing heavy loads on the surface.

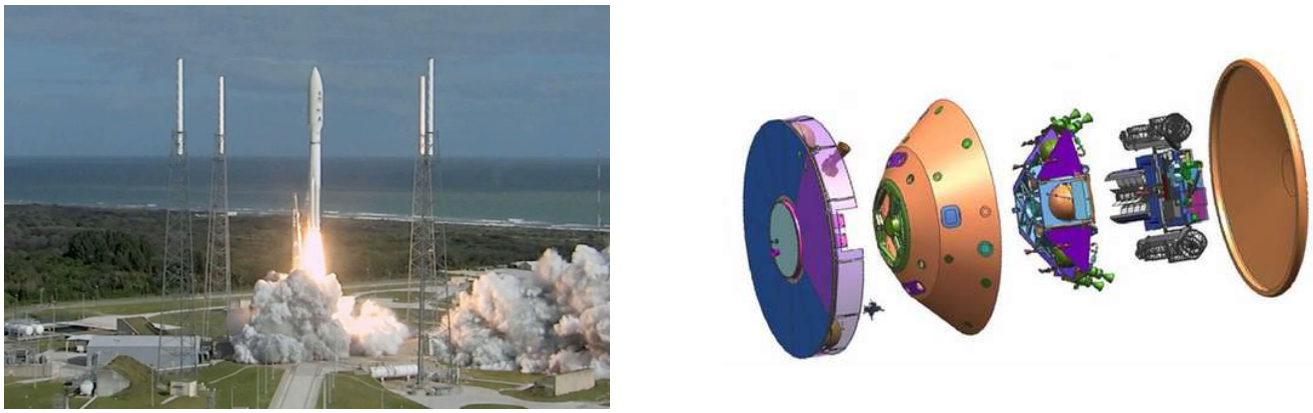


Figure 8.15: Left: the launch of the MSL mission on 26 November 2011. Right: an expanded view of the spacecraft showing (from left to right) the cruise stage, the backshell, the descent stage, the rover and the heat shell.

Curiosity was launched on 26 November 2011 by an Atlas Centaur V-541 rocket[4] which had already been used for the Mars Reconnaissance Orbiter and New Horizon missions. The Centaur stage was fired twice, once to insert the spacecraft, still attached to the Centaur stage, into a low-Earth parking orbit, and then again to accelerate the

[4]The 541 designation indicates that the fairing had a diameter of 5 m, that 4 solid-rocket boosters were used alongside the central common booster and that a one-engine Centaur upper stage was used.

spacecraft out of this orbit onto its trajectory towards Mars. The total mass of the rocket on the launch pad was 531 tons. The total mass of the spacecraft was 3893 kg, including 539 kg for the fueled cruise stage, 2401 kg for the entry, descent and landing system (aeroshell and fueled descent stage). The rover itself has a mass of 899 kg.
During the six months journey to Mars, the cruise stage performed several trajectory corrections, actively controled the temperature of the spacecraft and ensured the communication between the spacecraft and the ground. The power of the cruise stage was generated through the use of solar panels. The attitude system consisted of a star scanner used to monitor the spacecraft position relative to bright stars, and eight hydrazine thrusters used for the attitude corrections. The vehicle was further spin-stabilized about its central axis at a spin rate of two rotations per minute. The cruise stage remained attached to the spacecraft until 10 minutes before contact with the Martian atmosphere.

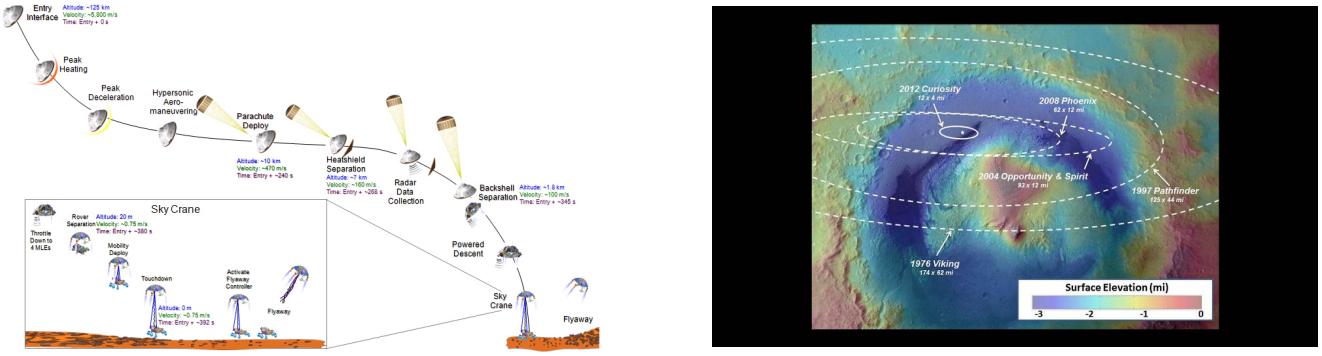


Figure 8.16: Left: the landing sequence of the Curiosity rover. Right: the landing sequence adopted for Curiosity led to a spectacular improvement on the uncertainty ellipse on the landing site with respect to previous missions.

Curiosity landed on Mars on 6 August 2012. The landing sequence is shown in Fig. 8.16. The exact timing of the events could not be completely predicted prior to the landing, because it depended on the atmospheric conditions on that day. Protected by the capsule consisting of the heat-shield and the back-shell, the probe entered the Martian atmosphere at an altitude of 125 km and a speed of 5.8 km s^{-1}. The heat-shield had a diameter of 4.5 m and the angle of attack was 70° with respect to the vertical direction. The orientation of the entry probe was stabilized by engines and the heat-shield reached a peak temperature of 2000° C. About 240 s later, the probe was at an altitude of about 10 km and had been decelerated to 470 m s^{-1}. At this moment a parachute of 20 m diameter deployed, and 28 s later, at an altitude of 7 km, the heat shield separated. The descent stage featured a radar that started working at this point and measured the velocity and altitude of the probe. About 77 s later, at 1.8 km above the ground and at a speed of 100 m s^{-1}, the back-shell and the parachute separated from the descent stage, and the 8 retrorockets started firing to further reduce the descent velocity. Shortly after, at about 20 m over the ground, 4 of the 8 rockets shut off and the sky crane came into play: nylon cords started to spool out to lower the rover from the descent stage to the ground. The wheels of the rover unfolded shortly before it touched the ground. Once the rover had touched down, the cords were severed and the descent stage flew out of the way.
The Curiosity rover is about 2.2 m tall, 3 m long and 2.7 m wide. The wheels of the rover allow it to roll over obstacles up to 75 cm high and to climb along a 45° slope. These wheels feature regular perforations that make up the letters J (. - - -), P (. - - .) and L (. - . .) in the Morse alphabet. These leave traces on the ground allowing the rover's navigation cameras to check whether the rover is moving normally or sliding. The rover can travel at a maximum speed of up to 100 m hr^{-1}, though its average speed is expected to be around 30 m hr^{-1}. Contrary to the MER rovers, Curiosity features no solar panels, but carries an RTG with 4.8 kg of ^{238}Pu. The RTG provides the power and heat for up to 14 years of operation[5]. The rover features 17 cameras, twelve of them being used for

[5]The nominal lifetime of the mission was one Martian year (i.e. 687 days). The power production of the RTG degrades by 4% per year.

Table 8.1: Comparison of some key features of the MER and MSL rovers.

	Spirit	Curiosity
Mass (kg)	185	900
Mass of instruments (kg)	6	85
Wheel diameter (cm)	25	50
Height (m)	1.5	2.2
Length (m)	1.6	3
Width (m)	2.3	2.7

Figure 8.17: Left: an artist view of the Curiosity rover on Mars. Right: comparison between the wheels of the rovers Sojourner, Spirit and Curiosity.

the navigation. The mission is designed to analyse samples scooped from the soil and/or drilled from rocks, with the goal to detect chemical building blocks of (microbial) life on Mars. For this purpose, i.e. to collect samples and transfer them to the instruments for analyses, the rover has a sophisticated robotic arm that has a reach of 2.2 m.

Curiosity carries ten scientific instruments, some of which are described hereafter. Mastcam consists of two cameras mounted on a mast that extends above the rover deck. These cameras take colour images and videos (10 frames per second) of the Martian terrain. These images are used to study the landscape, the soils, to monitor weather phenomena, and to support the driving and sampling operations of the rover. Mastcam further features a series of filters to record monochromatic images.

The Mars Hand Lens Imager (MAHLI) is located at the turret of tools at the end of Curiosity's robotic arm. This camera is designed to acquire close-up, high-resolution (down to details of 12.5 μm) views of rocks and soil. The camera has an adjustable focus to look at objects at distances between 2.1 cm and infinity. A calibration target is located on the rover near the base of the arm. This includes colour references, a metric bar graphic, a Lincoln penny, and a stair-step pattern for depth calibration. MAHLI carries white light and UV sources. The UV light is used to induce fluorescence to help detect carbonate.

The ChemCam instrument uses so-called laser-induced breakdown spectroscopy (LIBS). A brief laser pulse (at a wavelength of 10 670 Å) focused on a target provides sufficient energy to ablate a pinhead-size sample of material in electronically excited states. This results in a short-lived plasma with a temperature near 10 000 K that produces characteristic emission lines. The flash of light emitted by the plasma enters the telescope and travels along a fiber optics link to the three spectrographs which are hosted inside the main body of the rover and cover the UV,

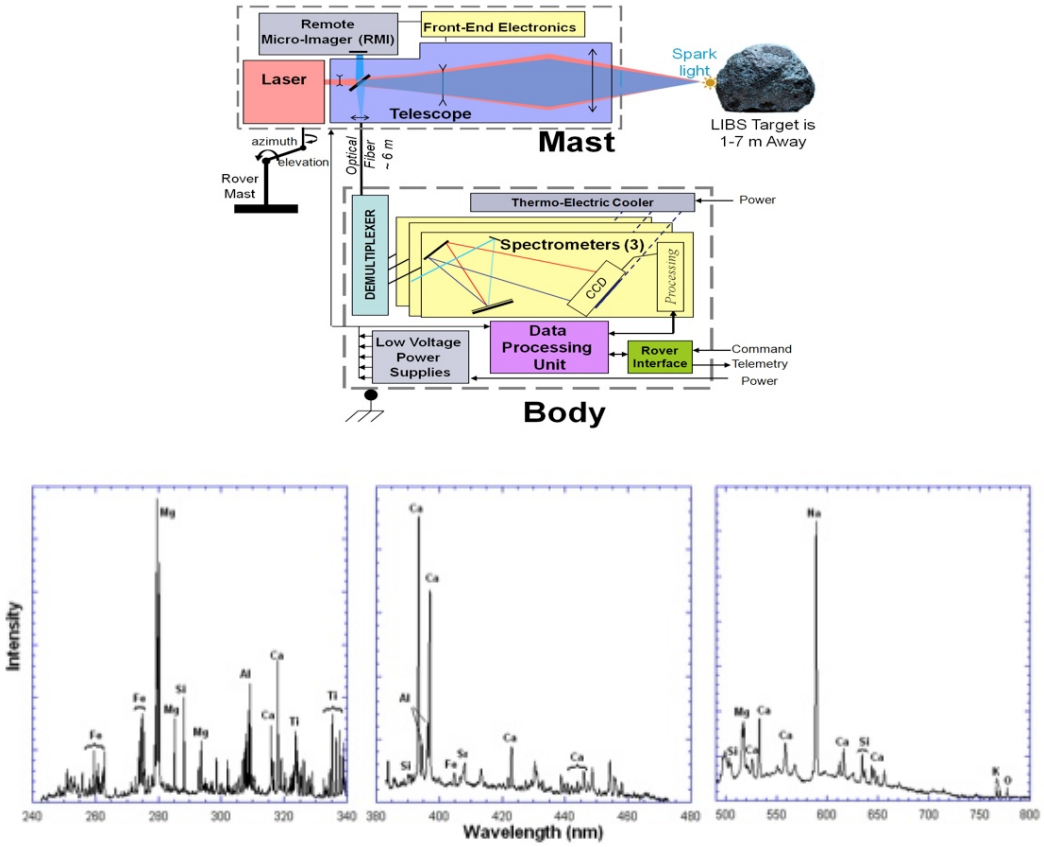


Figure 8.18: Top: block diagram illustrating the way the ChemCam instrument is operated. Bottom: example ChemCam spectra, illustrating the wavelength domains covered by the three spectrographs.

visible and near-IR wavelength bands from 2400 to 9050 Å (see Fig. 8.18). Each observation consists of 30 laser pulses at 3 Hz. The spectrum is recorded after each pulse. The first (usually 5) spectra are discarded due to dust contamination and the remaining ones are then combined to increase the signal to noise ratio. Repeated laser pulses at the same location provide profiles into the surface of the target. This instrument can operate from a distance between 1 and 7 m. The laser and the camera of ChemCam are mounted on the mast and can be tilted or rotated to view the target rock. The ChemCam spectrometers are further used for passive visible range spectroscopy of the surface and observations of atmospheric absorption to study the properties of dust aerosol or ice particles.

The Sample Analysis at Mars (SAM) instrument suite consists of a mass spectrometer, a gas chromatograph and a tunable laser spectrometer. The robotic arm delivers solid samples into one of 75 sterile single-use cells. The cells are mounted on a rotating platform. The cell with the sample to be analyzed is moved under an oven that evaporates the sample at temperatures up to 1100° C. SAM is designed to measure abundances of C, methane, H, O and N. The mass spectrometer separates elements and compounds by mass for identification. The gas chromatograph heats samples until they vaporize and separates then the resulting gases into various components for analysis. The laser spectrometer measures abundances of various isotopes of C, H and O in the Martian atmosphere.

Underground water ice can be diagnosed via the energy of neutrons that escape from the planet's surface. The bombardment of the surface with cosmic rays liberates neutrons at some depth in the soil. If there exists liquid or frozen water, hydrogen atoms slow the neutrons down and they escape with a reduced energy and velocity. The DAN instrument (Dynamic Albedo of Neutrons) is a pulsing neutron generator that sends a beam of neutrons

into the Martian ground and measures the speed of the reflected neutrons at the surface. If water is present, these reflected neutrons should be slower.
The Alpha Particle X-ray spectrometer (APXS) measures abundances of chemical elements in rocks and soils. To do so, it is placed on the robotic arm allowing a close contact with the sample. The AXPS exposes the sample material to α particles and X-rays produced by the radioactive decay of curium ^{244}Cm. The X-rays and α particles ionize the atoms that make up the material and the subsequent emission of X-rays is recorded by the APXS, allowing it to establish the chemical composition. Similar instruments were already used on the Sojourner, Spirit and Opportunity rovers. On the MER rovers, the elements detected by the APXS in rock and soil samples are typically Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Ni, Zn, and Br.

The NASA scientists selected the Gale crater with its central peak (known as Mount Sharp) as the site to be explored by Curiosity. The Gale crater has a diameter of 150 km, and is suspected to have been filled up with water about 3 billion years ago. The action of the water should be responsible for the shape of Mount Sharp. Alternatively, Mount Sharp could also result from wind erosion. Indeed, the winds along the outer crater walls revert their direction in the morning when the crater ground heats up and in the evening, when it cools down. Over about 100 million years, this effect could have piled-up an important quantity of material, giving rise to Mount Sharp.

8.2.4 Mars Express

The ESA mission Mars Express was launched on 2 June 2003 by a Soyuz Fregat launch vehicle. The launch mass of the spacecraft was 1120 kg including 113 kg for the orbiter and 60 kg for the small lander Beagle 2 named after the ship Charles Darwin used to travel around the world.
The Mars Express orbiter successfully entered a highly eccentric orbit around Mars on 25 December 2003. Following some orbital manoeuvres, the spacecraft reached its nearly polar (86° inclination) operational orbit in January 2004. The initial operational orbit had a period of 7.5 hrs, a pericentre altitude of 259 km and an apocentre altitude of 11 560 km. Mars Express operations are currently approved until December 2020.
The science goals of the orbiter are

- to image the entire surface of Mars with a resolution of 10 m/pixel and 2 m/pixel over selected areas. This is done with the High Resolution Stereo Camera (HRSC) that provides 3-D images, revealing the topography of Mars. This camera is operated close to pericentre, i.e. when the orbiter is closest to the surface of the planet. By the end of 2018, Mars Express had imaged 98.8% of the Martian surface. Out of these, 80% were imaged at a resolution better than 20 m/pixel.

- to produce a map of the mineral composition of the surface at a resolution of 100 m. This is done by analysing the visible and infrared light (between 5000 and 52 000 Å) reflected by the Martian surface with a mapping spectrometer (OMEGA).

- to map the composition of the atmosphere and determine its global circulation. This is done by two spectrometers: one operating in the UV and IR to measure the absorption of the light by the Martian atmosphere, allowing to search for ozone (in the 2500 Å band) and water vapor (at 13 800 Å); the other one operating in the near IR (12 000 to 45 000 Å) looking for molecular absorption and emission bands.

- to characterize the interaction of the atmosphere with the solar wind by measuring ions, electrons and energetic neutral atoms. Indeed, the constant bombardment of the atmosphere by the solar wind is likely to be responsible for the progressive loss of the Martian atmosphere which is not protected by a strong magnetic field unlike the Earth's atmosphere.

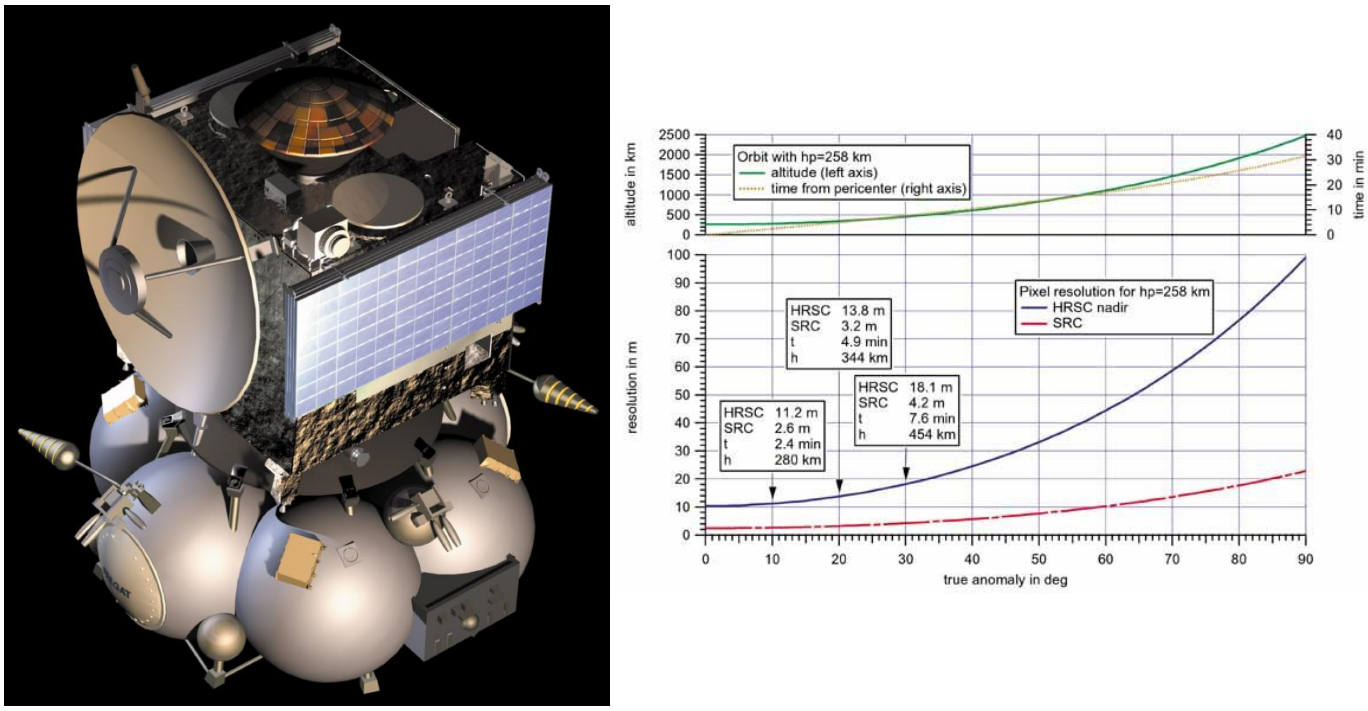


Figure 8.19: Left: artist view of the Mars Express spacecraft connected to the fourth stage of the Soyuz Fregat launcher. Right: the linear resolution of the HRSC camera as a function of the altitude of the spacecraft or its anomaly. The true anomaly yields the position of the probe along its (highly eccentric) orbit with 0° corresponding to the pericentre position.

- to probe the sub-surface structure to a depth of a few km. This is done with the MARSIS sub-surface sounding radar instrument. This instrument consists of a 40 m long antenna (two 20 m long booms) that sends low-frequency radio waves towards the planet. A fraction of these waves penetrates into the ground and is then reflected by the sub-surface interfaces between layers of different material.

The Beagle 2 lander was planned to determine the geology and the chemical composition at the lander's site, to search for life signatures and study the climate. Unfortunately, the lander was lost as it failed to establish contact with the orbiter or Earth-based radio telescopes upon its landing in December 2003. It was only in 2015 that the lander was located on high-resolution images taken by the Mars Reconnaissance Orbiter. These images revealed the lander within the anticipated landing area, but in a partially deployed configuration. This partial deployment of the solar panels was probably responsible for the loss of the lander as a full deployment would have been needed to expose the antenna.

8.2.5 Manned missions to Mars?

Whilst manned missions are obviously more risky and far more expensive than robotic probes, it is nevertheless true that they open up new possibilities. So far, the only body of the Solar System that has been visited by humans is the Moon. The Apollo missions in the late sixties - early seventies had not only a major political impact, but they also allowed to perform scientific experiments. Some of them were done in real time (soil mechanics investigation, solar wind composition experiments with the collection of solar wind samples), others were deployed by the crews and monitored afterwards from Earth by radio telemetry. The latter category includes seismic experiments that detected about 40 active zones of moonquakes and showed a periodic activity correlated with the lunar tides, laser

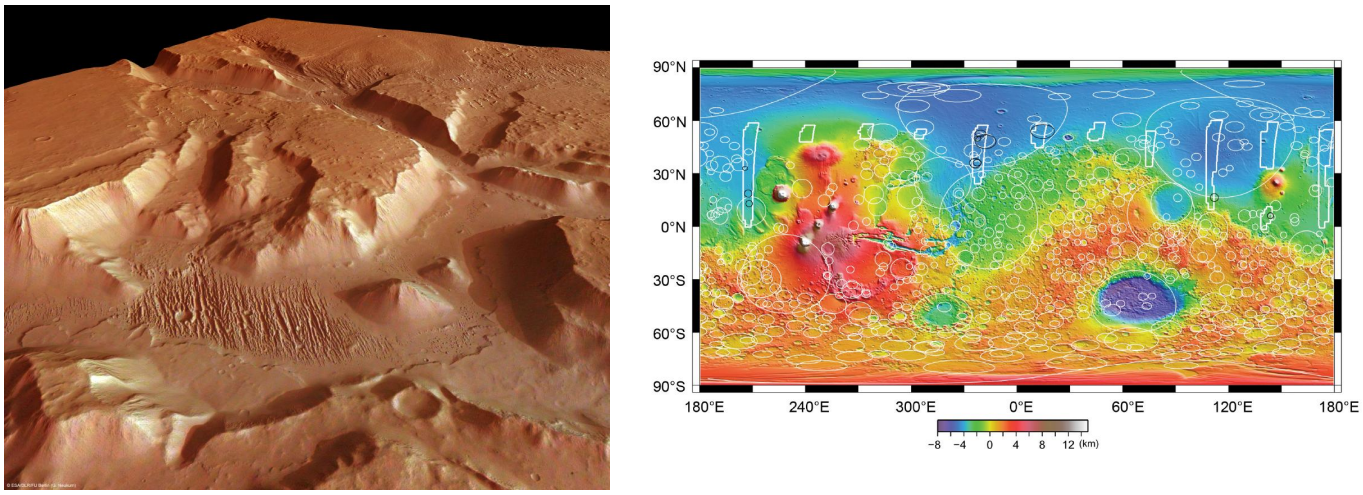


Figure 8.20: Left: HRSC 3-D image of the Aeolis Mensae region (6° South, 145° East). The valleys most likely originate from tectonic activity, possibly eroded by subsequent fluid or mass wasting activity. Right: location of a series of buried basins detected by the MARSIS sounding radar experiment (black contours).

ranging retroreflectors that allowed precise measurements of the Earth-Moon distance, heat flow measurements[6] that consisted in drilling holes of about 2 m depth and measuring the temperature gradient along the holes (corrected for the varying solar illumination of the surface) and the installation of portable magnetometers that allowed to detect strong variations of the Moon's magnetic field on local and regional scales. Of course, another major aspect was the return of 382 kg of lunar material that allowed to show that the chemical composition is different from that of the Earth and to prove the absence of minerals formed by the action of water (see also Chapter 9).

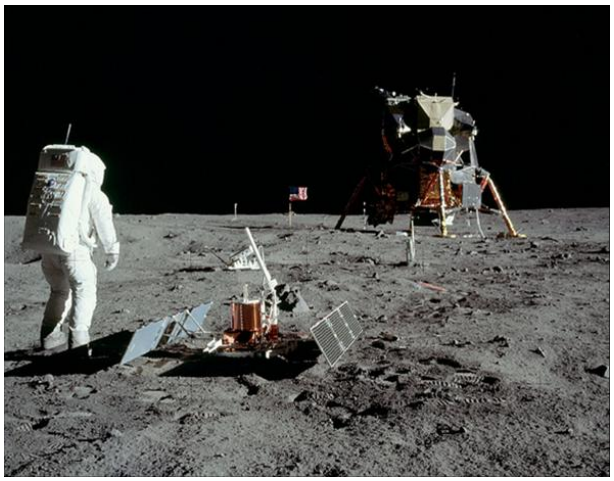
Ever since the beginning of spaceflight, there have been a number of projects for manned missions to Mars[7]. A major difficulty associated with these projects stems from the limitations of the rocket equation. Indeed, a manned mission to Mars needs to accelerate a huge quantity of material (for example 4350 tons in the case of the abandoned Deimos project) to the velocity required to reach Mars. Ionic propulsion might be a solution, but needs further developments to reach higher thrusts.

Landing a heavy manned space vehicle on the Martian surface is also a big technological challenge. In fact, the Martian atmosphere is not sufficiently dense to slow down a massive ($\sim$ 40 tons) capsule and allow it to land safely by the sole use of parachutes. A controlled descend by means of a propulsion system is possible, but requires large amounts of propellant, and hence additional mass. Two-way missions will have to include a return vehicle that can take-off from Mars.

During a trip to Mars, there will be no chance of rescue. There are several risks such as a failure of the spacecraft, illness of the crew, impacts of micrometeorites, solar eruptions, etc. All spare parts required to repair vital parts of the spacecraft need to be available on board. Owing to the long communication delays between the spacecraft and the ground control, the crew of a mission to Mars will have to be autonomous when it comes down to diagnose a problem and take decisions on how to solve it. In case of emergencies, there is no time to iterate with the

[6] This is the measurement of the rate at which the body loses its internal energy to space.

[7] We do not consider the Mars One project promoted by the Dutch businessman Bas Lansdorp. Mars One officially aims at sending humans to Mars by 2032 with a one-way ticket only. The passengers would have to establish a permanent human colony on Mars. Every step of their journey would be documented for a reality TV program. The TV rights are supposed to provide most of the funds. Lansdorp claimed that only 6 billion USD would be needed for this project. In 2013, more than 200 000 people applied to become the first astronauts. After two phases of downselection, 100 candidates were retained for the third phase initially announced for the end of 2016, but currently (July 2019) still pending. Meanwhile, the Mars One company was declared bankrupt in February 2019 and its current status is uncertain.



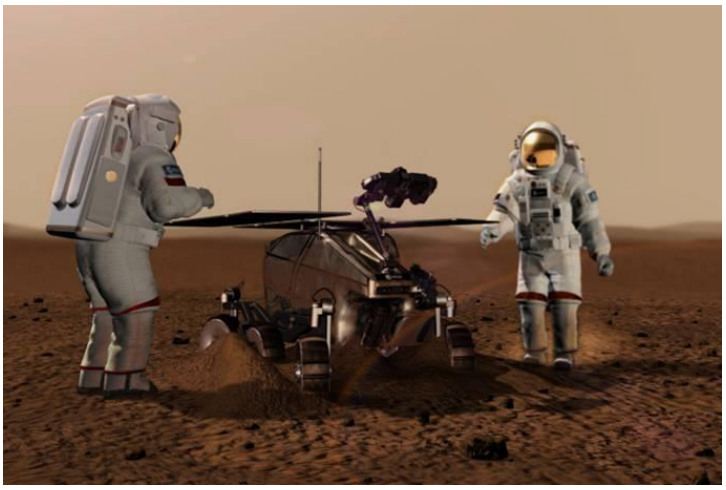


Figure 8.21: Between 1969 and 1972, the Apollo missions were the first (and so far only) manned flights to another body of the Solar System. Whether or not there will be manned missions to Mars one day is currently unclear.

mission control as such procedures will take hours or even days and weeks. To assist the crew, the spacecraft itself would need to perform periodic monitoring of its health and functionalities. It is also compulsory that the crew can quickly have an overview of all system's components and spare parts. In particular, it is important to know their location and properties (physical, chemical, etc.), and how to assemble them into something useful for solving a problem.

Another major issue is the duration of the mission and its impact on the physical and mental health of the spacecraft crew. Indeed, with conventional propulsion means, a one way trip to Mars takes between 6 and 9 months and the crew has to be completely autonomous with a sophisticated recycling system for water, oxygen and waste. During this journey, the human body is exposed to an unusual environment: low gravity, high radiation doses as well as psychological challenges (e.g. stress and anxiety related to the distance from Earth and communication delays). All these problems represent high risk factors and will have to be addressed before a manned mission to Mars becomes possible. On the mental side, a healthy interaction balance among the crew is mandatory. This is impacted by factors such as gender balance, command hierarchy and the nature and frequency of crew interactions.

The absence of gravity (during the cruise) and the lower gravity on the Martian surface lead to muscle atrophy, bone degradation, and accelerated aging of the body. On the ISS, astronauts spend 2.5 hours each day exercising to keep their muscles active and strong. The rate of loss of bone mass on Earth is typically a few percent per year. In space, this rate can be 1 – 2 % per month. A manned mission to Mars would ideally carry a centrifugal wheel to simulate gravity. A 9 m diameter wheel rotating at 10 rotations per minute can produce 0.5 g. The low gravity environment also impacts the heart activity and blood circulation. On Earth, the human blood circulatory system has to work against gravity that pulls the blood down to the lower half of the body and the heart pumps it back again to the upper part. In the absence of gravity, blood tends to accumulate in the upper body. The human body reacts by modifying the blood pressure and the heart activity.

Extended exposure to ionizing radiation (both during cruise and on Mars) damages cells, alters DNA and induces mutations, ultimately leading to cancer. Astronauts on a Mars mission would exceed their nominal lifetime radiation dose after about 1.5 or 2 years in space. New ways of shielding spacecraft, including light-weight plastics that can reduce the radiation from galactic cosmic rays are currently being investigated. Another option is active shielding by means of superconducting magnetic shields that deflect particles and prevent the harmful rays from penetrating the spacecraft.

To illustrate the complexity of making such a mission self-sufficient, let us briefly consider the recycling of water and air. At any time during the mission, the spacecraft itself must provide clean water and air for the passengers. On Earth, an adult typically needs 0.84 kg of oxygen and 14 kg of drinking water every day. Whilst the quantity of oxygen cannot be reduced in space, the consumption of drinking water has been reduced to 2.5 kg.
In the case of the ISS, water is first supplied via containers brought by servicing missions. Re-supplying the ISS crew with water only via service missions would be impractical, extremely expensive and risky, not to speak about the difficulties to implement such a scenario for a mission to Mars. Recycling efforts are thus mandatory and only lost water is replaced by bringing containers of fresh water from the Earth to the ISS. The purification of wastewater occurs in several steps that combine physical and chemical processes. Particles as well as organic and inorganic impurities are filtered out. Finally, a catalytic oxidation reactor removes volatile organic compounds and kills bacteria and viruses. The system needs to be as efficient as possible. For instance, the passengers of a manned spacecraft lose water when they exhale or sweat, thus contributing to the ambient humidity. This humidity needs to be condensed and returned to the general water supply. Nevertheless, the system currently used onboard the ISS achieves a recovery rate of 'only' 85% and the losses must be compensated by supplying fresh water.
The recycling of water and air in a spacecraft are highly interconnected. Fuel cells, as used on-board the Space Shuttle and Apollo capsules, use hydrogen and oxygen from cryogenic tanks to provide electrical power. A by-product of these devices is water. Conversely, oxygen can be generated via electrolysis combining water and electricity to produce H_2 and O_2. Because of its flammability, storing hydrogen on a spacecraft is dangerous and the H_2 is thus vented into space. Alternatively, it can be combined with CO_2 in the Sabatier reaction ($CO_2 + 4\,H_2 \rightarrow CH_4 + 2\,H_2O$) to reclaim water, the resulting methane being vented into space. CO_2 removal from the cabin air can also be done with granules of a synthetic rock (zeolite). The air is blown through the zeolite filter which retains CO_2 and water. The zeolite is regenerated by heating it and exposing it to empty space. For manned missions on the Martian surface, the Sabatier reaction as well as the use of plants for producing oxygen through photosynthesis are possible avenues to produce water and oxygen.

Last but not least, the cost is also a major issue... Moreover, there are large uncertainties on this cost. For instance, the Mars Direct mission concept proposed in 1991 aimed at sending men to Mars for a total cost of about 50 billion USD. The mission would last 900 days (360 days for the trip and 550 days on Mars). A recent re-assessment of the cost of a two-way trip to Mars led to an estimate of about 210 billion USD.

Around the world in 80 days?

Recently, an alternative to the above scenarios has been proposed: a new-type of ion engine that could possibly shorten the one-way journey to 39 days. This new engine, called VASIMR (Variable Specific Impulse Magneto-plasma Rocket, see Sect. 5.2.2), is currently under development. The working principle of this engine is based on three stages. In the first stage, argon is ionized by a radio frequency generator. The resulting plasma is then heated in the second stage up to temperatures of several million degrees. This is done by means of a strong magnetic field produced by a supraconducting magnet. Finally, strong magnetic fields channel the plasma out of the engine, thus producing the thrust, which is about 100 times higher than for other ion engines. For the travel to Mars in 39 days, however, the engine would need 1000 times more power than can be provided by solar arrays. Therefore an on-board nuclear reactor is needed and this technology is still to be developed.

Humans on Mars in 30 years time?

Whether or not these projects will eventually become more than mere science fiction depends to a large extent on the development of new technologies and on the efficiency of international collaborations.

In this context, the International Space Exploration Coordination Group (ISECG) gathers a number of space agencies (NASA, ESA, Roscosmos, JAXA, CNES, DLR,...)[8] who have elaborated a report called *The Global Exploration Roadmap* where they present their view on the future of space exploration beyond low-Earth orbit with manned missions to Mars as the ultimate goal. The 2013 edition of this report identified three mission themes as precursor steps to a manned Mars mission: (1) exploration of near-Earth asteroids, (2) extended duration crew missions in the lunar vicinity, and (3) human missions on the Moon. The first goal was to use a robotic mission to capture a small (diameter < 10 m) asteroid and guide it to a stable orbit in the lunar vicinity where it could be visited and explored by astronauts. The connection with manned missions to Mars is not obvious, but the presence of theme (1) was most likely motivated by the ISECG's wish to enhance public-private partnerships (see also below and Chapter 9) and to integrate the official objectives of the administration of the former US President Obama into their plans. The second theme allows testing and validating life support systems before travelling to distant destinations. The third theme, finally allows demonstrating operation concepts and enhanced crew autonomy for surface exploration. The 2018 update of this report no longer refers to the human exploration of asteroids, but emphasizes the role of the Deep Space Gateway as a platform for future deep space missions.
The ISECG report stresses the importance of robotic probes in preparing for human missions. A series of robotic missions are identified that can help characterize the environment, identify hazards and assess resources. The roadmap also highlights the role the ISS can play in developping countermeasures against cardiovascular, musculoskeletal and neurological challenges that might threaten human missions to Mars.
However, it has to be stressed that the ISECG document does not contain any firm commitments, nor does it provide a firm schedule for most of the planned activities. Finally, private companies, such as Elon Musk's SpaceX, who announced in 2016 his intention[9] to send a Red Dragon capsule of 6 tons to Mars by 2018, might also play a role. However, they will probably not be able to raise the necessary funds without substantial input of public money.

8.3 The Cassini-Huygens mission to Saturn and its moons

Among the most fascinating results of the Voyager missions were the impressive images taken by these probes when they flew by Saturn and its moons. Saturn with its rings is not only one of the most spectacular planets of the Solar System, but it is orbited by a large number of satellites, including Titan, one of only a handful of moons in the Solar System having its own atmosphere[10]. Titan is an important object since it has been considered as a primitive (frozen) version of the Earth. These considerations stimulated a new mission carried out in collaboration between NASA and ESA (acting as a junior partner by providing the Huygens probe to land on Titan).

The Cassini-Huygens mission was launched on 15 October 1997 by a Titan IV-B/Centaur rocket. No existing launch vehicle could have sent the spacecraft with its mass of 5650 kg directly to Saturn. Therefore, the trajectory was chosen in such a way to take advantage of the gravity assists by Venus (twice), Earth (once) and Jupiter (once). This rather complex journey took the spacecraft through extreme temperature regimes, from three times hotter than at Earth during the Venus flybys to 100 times colder at its final destination. During its trip across the inner Solar System (out to 2.7 AU), the 4 m high-gain antenna of the spacecraft was pointed at the Sun acting as a Sun-shield to prevent the instruments aboard the spacecraft from overheating[11]. In addition, the spacecraft had several passive and active thermal control systems to maintain the instrument temperatures within an acceptable range.

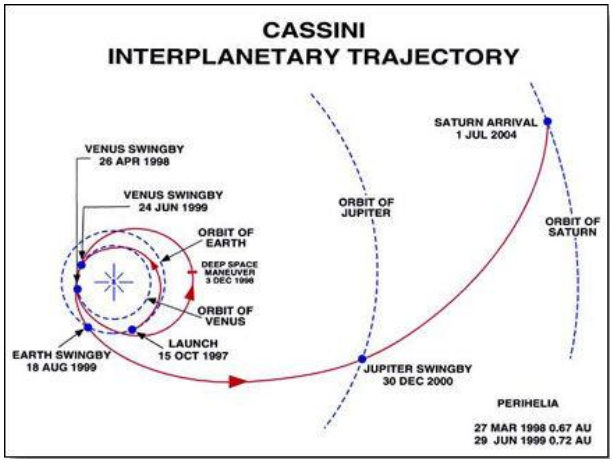
[8] China and India were not member of the ISECG when the 2013 report was published, but have since then joined this group and contributed to the 2018 update of the roadmap.

[9] This announcement was little more than a clever public relation manoeuvre and the timeline was obviously more than optimistic.

[10] Triton, Neptune's largest moon, also has an atmosphere and there have been claims that Io, Jupiter's most active satellite also features a very thin atmosphere.

[11] During that phase of the mission, the communication with the spacecraft was done through the low-gain antennas.

Figure 8.22: A schematic illustration of Saturn, its rings and its moons.



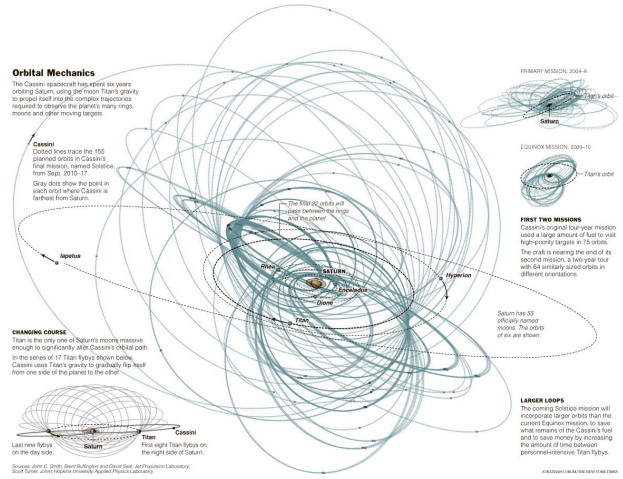


Figure 8.23: Left: the journey of the Cassini-Huygens spacecraft across the Solar System towards Saturn. Right: Cassini's orbit in the Saturn system over time.

The Cassini-Huygens spacecraft reached Saturn on 1 July 2004. During the Saturn orbit insertion, the spacecraft approached the planet from below the ring plane and crossed through the large gap between the F and G rings, approaching the planet down to an altitude of 20 000 km. To ensure that the spacecraft would be captured by the planet's gravity, it had to be slowed down significantly. This was achieved by turning Cassini's main engine to face the direction of travel and the thrust acted as a brake. Cassini performed several hundred round trips in the Saturn system, including more than 100 close flybys of Titan. The latter not only allowed studying Titan, but were also used for gravity assist manoeuvres to progressively change the orbit of the spacecraft including changes of the orbital inclination with respect to Saturn's equator from a near equatorial orbit to an inclination of $64°$ and back again (see Fig. 8.23). The mission was initially foreseen to last for four years (2004 – 2008), but was then extended twice (Cassini Equinox Mission from 2008 – 2010 and Cassini Solstice Mission since 2010) and ended on 15 September 2017 when the spacecraft entered Saturn's atmosphere, to avoid any microbes, that might have survived aboard the spacecraft, from contaminating Titan or Enceladus.

Cassini-Huygens was a rather large (6.7 m high, 4 m wide) three-axis stabilized spacecraft hosting a large number

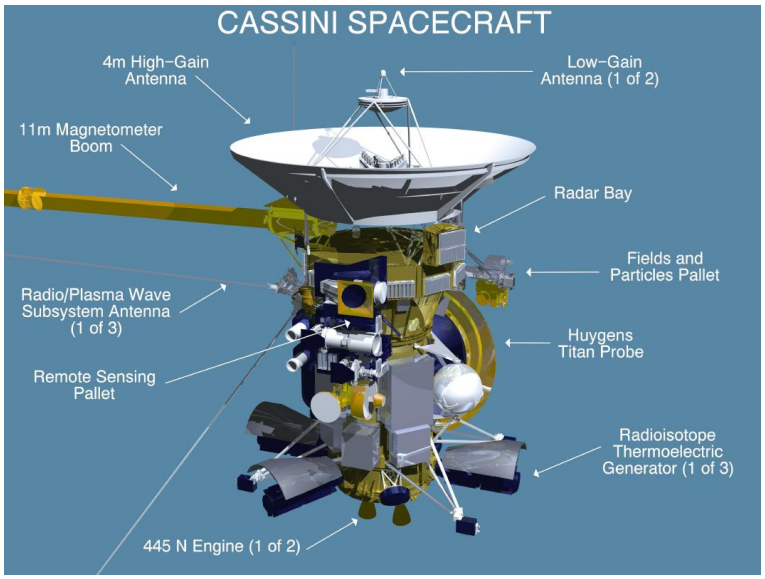


Figure 8.24: Schematic illustration of the Cassini-Huygens spacecraft and its payload components.

of instruments. The large structure of the spacecraft not only provided the mechanical support for all the equipment, it also provided the thermal conductivity and served as an equipotential and electrical grounding reference. Since solar arrays cannot provide enough power in the outer Solar System, the electrical power was provided by three RTGs.
The instruments on board the orbiter can be divided into three broad categories:

- optical remote sensing: this included an infrared (9.0 – 1000 μm) spectrograph (CIRS) featuring a 50 cm telescope for the analysis of the composition of the atmospheres of Saturn and Titan; an imaging science subsystem (ISS) featuring a wide angle, low resolution and a narrow angle, high spatial resolution CCD camera; an ultraviolet imaging spectrograph (UVIS) allowing to study the composition of Titan's and Saturn's atmospheres through occultation experiments; a visual and infrared mapping spectrometer (VIMS) that mapped reflected and emitted light from atmospheres, rings and surfaces over 352 contiguous wavelength channels from 3500 Å to 5.1 μm.

- microwave remote sensing: this included a synthetic aperture radar imager operating at 13.78 GHz to produce maps of the surface of Titan; the radio science subsystem (RSS) that sended radio waves to Earth across the rings of Saturn allowing to study the composition and properties of the rings.

- field, wave and particle detectors: this included a plasma spectrometer (CAPS) that measured the energy and electrical charge of ions and electrons over the ranges from 1 eV to 50 keV and 0.7 eV to 30 keV respectively; a cosmic dust analyzer (CDA) that measured the speed, size and chemical composition of dust particles[12]; a magnetometer located on an 11 m long boom; a magnetospheric imaging instrument designed to study the energetic particles in the planetary magnetosphere by remote imaging of the hot ions population and localized in-situ measurements; a radio and plasma wave detector consisting of an electric field detector plus a magnetic field search coil to measure the radio waves produced by the interaction of the Solar wind with Saturn and Titan.

After thirteen years and 293 revolutions around Saturn, Cassini plunged into Saturn's atmosphere on 15 September 2017. This mission has provided a host of amazing results and deeply changed our view of Saturn and its moons. Among the highlights of the mission, one can cite the observation of a giant storm that erupted in the

[12]The dust grains impact a collection surface inside the instrument and are vaporized. The resulting plasma is then analyzed.

northern hemisphere in 2010 and lasted several months, the observation of vertical structures of up to 3 km size in the rings, the discovery of active jets containing water ice from Enceladus, the study of the prebiotic chemistry in Titan's atmosphere where CH_4, N_2 and sunlight react to form increasingly complex hydrocarbons that eventually form the haze that shrouds Titan, etc... Whilst the bulk of the results of the mission concern Saturn and its moons, one should not forget that the spacecraft has already performed measurements during its flyby of Jupiter. For instance, the CDA instrument analysed nano dust particles that are continuously released by the Jovian system.

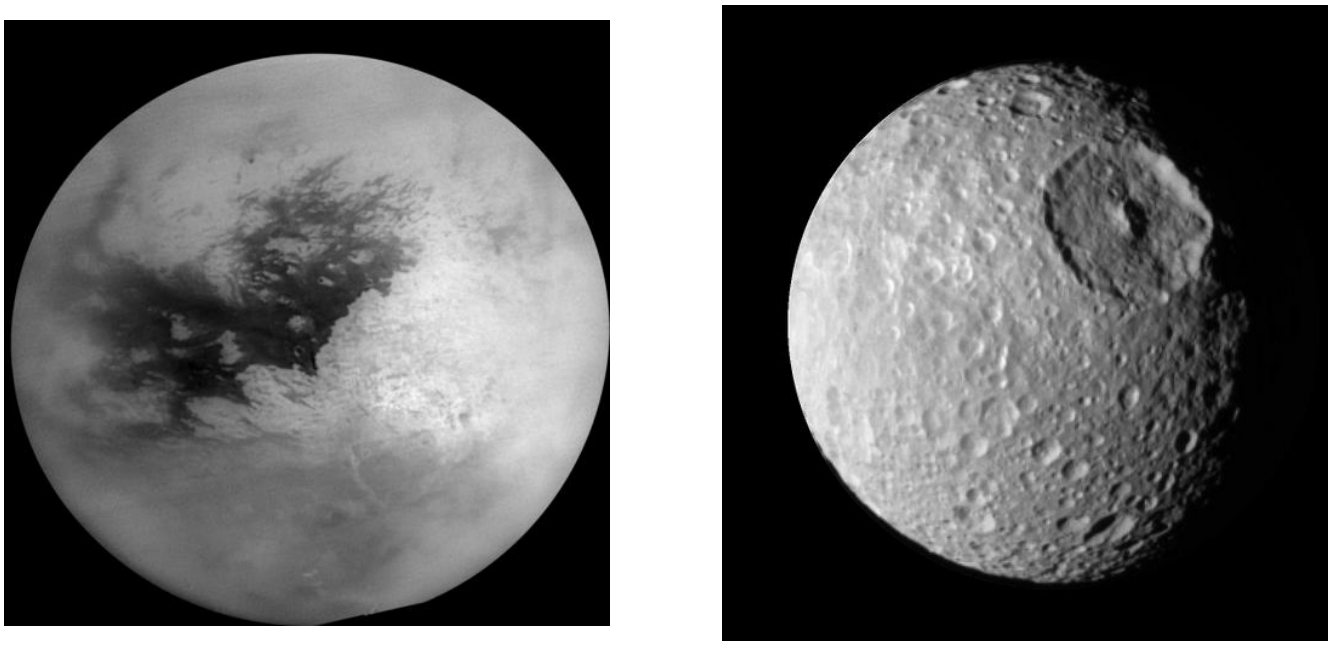


Figure 8.25: Left: Cassini mosaic of Titan. The images were taken with the narrow angle camera through a filter sensitive to polarized IR light at distances ranging from 226 000 to 242 000 km. The resolution of the image is about 1.3 km per pixel. Cassini has also detected cryogenic volcanos on Titan. Right: this picture could come from George Lucas' Star Wars movies, but it actually shows an image of Mimas, one of the satellites of Saturn as observed by Cassini.

The Huygens experiment consisted of two parts: the probe itself and the probe support equipment that remained attached to the Cassini orbiter. The latter hosted the electronics necessary to track the probe and provided the data from the probe to the orbiter. The probe itself was protected by a front shield (100 kg, i.e. about one third of the total mass of the probe) covered with a special thermal protection material (similar to the Space Shuttle tiles) for the atmospheric entry phase (when the plasma in front of the probe reached a temperature as high as 12 000° C) and an aft cover (also covered with thermal protection material) that hosted the parachutes for the descent into the atmosphere of Titan. After landing on the surface the probe was exposed to a temperature of $-200°$ C.
On 17 December 2004, the Cassini-Huygens spacecraft was placed on a direct collision trajectory with Titan. On 25 December 2004, the Huygens probe was released from the Cassini spacecraft by a spin/eject device (using pyro bolts). The probe was ejected at a spin rate of 7.5 rounds per minute and a relative speed of $0.35\,\mathrm{m\,s^{-1}}$. During its 20 days cruise to Titan, the probe was spin-stabilized and had no possibility to correct its attitude. Three days after the separation, Cassini's orbit was corrected to avoid a collision with Titan and ensure instead a flyby at an altitude of 60 000 km. After separation from the orbiter, the Huygens probe received its power from five $LiSO_2$ batteries. A critical aspect of the experiment was thus to ensure that the probe would reach the surface of Titan before running out of power.

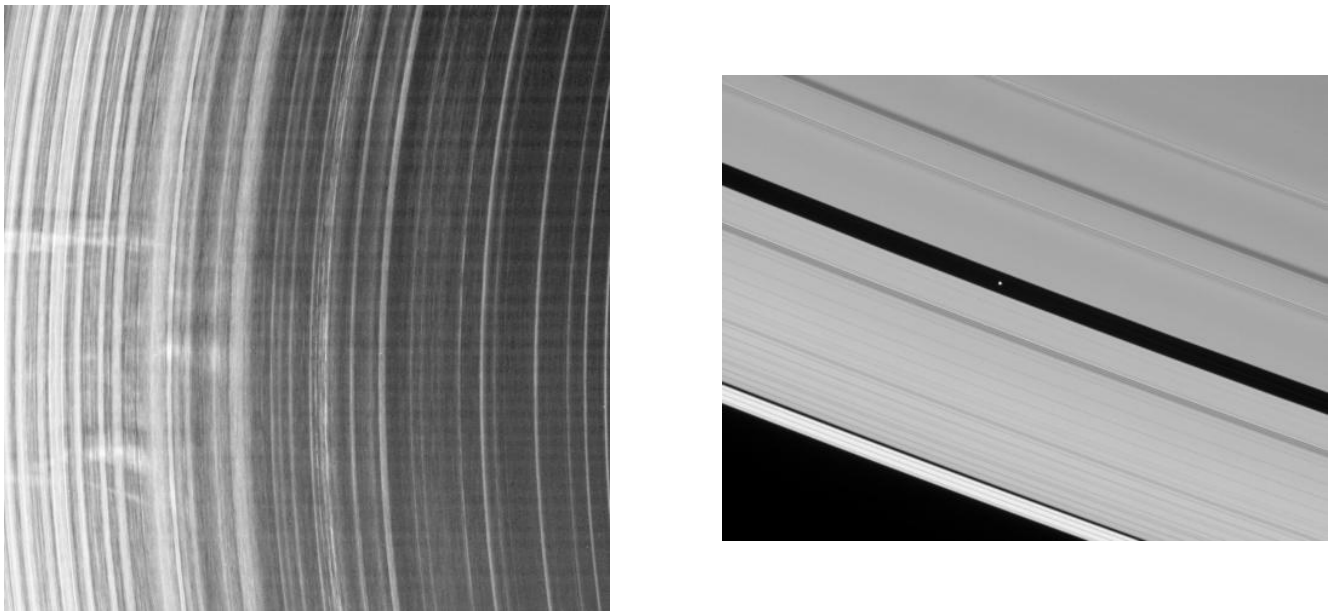


Figure 8.26: Left: composite image of the B ring. Two images were taken 45 seconds apart, and the first image is superposed in negative colours illustrating the motion of the radial spokes. The top-most spoke was 2500 km long. Spokes are long, finger-like features that stretch across the B-ring. They are short-lived (a few hours) and are likely due to Saturn's magnetic field interacting with tiny dust particles in the ring. Right: the moon Pan travels across the Encke Gap. A similar situation is seen for the satellite Daphnis that cruises through the Keeler Gap raising edge waves in the ring material.

Figure 8.27: Artist view of the landing sequence of the Huygens probe on Titan. The probe decelerated from an initial velocity of 6.1 km s^{-1} to 7 m s^{-1} by the effects of atmospheric friction and by the usage of three parachutes.

On 14 January 2005, the Huygens probe descended through the atmosphere of Titan and landed safely on its ground. The transmission of the data started a few minutes after the probe had entered the atmosphere (at an altitude of 1270 km) and ended 4.5 hours later when the Cassini orbiter dissapeared below Titan's horizon. The descent phase took 2 h 27 min, 7 minutes more than estimated before. The descent of the spacecraft was followed by the Cassini orbiter that had turned its high-gain antenna towards Titan as well as by radio telescopes all around

the world.
The Huygens probe carried several scientific instruments:

- The aerosol collector and pyrolyser (ACP) collected aerosols by means of a deployable sampling device that was first operated from the top of Titan's atmosphere down to an altitude of roughly 40 km. The device was reactivated between about 23 and 17 km above the ground. At the end of each collecting period, the filter was retracted into a pyrolysis furnace where the sample was first analysed at ambient temperature (about 0° C) and then at higher temperatures of 250 and 600° C. Finally, the products of the pyrolysis were analysed by a gas chromatograph and mass spectrometer designed to measure the chemical composition and determine the isotope ratios of the major gaseous constituents.

- The descent imager and spectral radiometer (DISR) was an optical remote sensing instrument consisting of a Sun sensor (to measure the spin rate of the probe), several upwards and downwards looking photometers, visible and IR spectrometers (covering the wavelength range from 4800 Å to 1 μm), a side-looking imager as well as two down-looking imagers (one with medium resolution, the other with high resolution imaging capabilities).

- The Doppler wind experiment was designed to determine a height profile of the direction and the strength of the winds on Titan with a precision of 1 m s^{-1}. For this purpose, the instrument was supposed to use the Doppler shift of the Huygens radio relay signal as measured by Cassini. This could not be done however due to a configuration problem of the bandpass of one of the Cassini receivers.

- The Huygens atmospheric structure instrument (HASI) was a multi-sensor instrument that measured the physical and electrical properties of Titan's atmosphere. For this purpose, HASI featured a three-axes accelerometer, a temperature sensor, a pressure sensor, a microphone and an electric field sensor.

- Finally, the Surface Science Package (SSP) featured a suite of sensors to determine the physical properties of the surface at the landing site. Among other things, these sensors allowed to measure the deceleration at impact, the temperature and thermal properties of the ground and featured also an acoustic sounder to sound the surface prior to landing.

The main constituents of the atmosphere of Titan are N_2 (97%), CH_4 (2%) and Ar. The temperature on Titan is such that it allows the simultaneous existence of three states of solid, liquid and gazeous methane. The data from the descent phase showed that the atmosphere of Titan was filled with haze to an altitude of about 30 km. The Huygens probe showed that the orange colour of Titan is due to aerosols between altitudes of 300 and 800 km. The probe further revealed a mix of complex hydrocarbons (benzene, nitriles,..) in Titan's atmosphere. The surface of Titan at the probe's landing site consisted of a thin frozen crust of about 10 cm thickness with a lower density medium beneath.

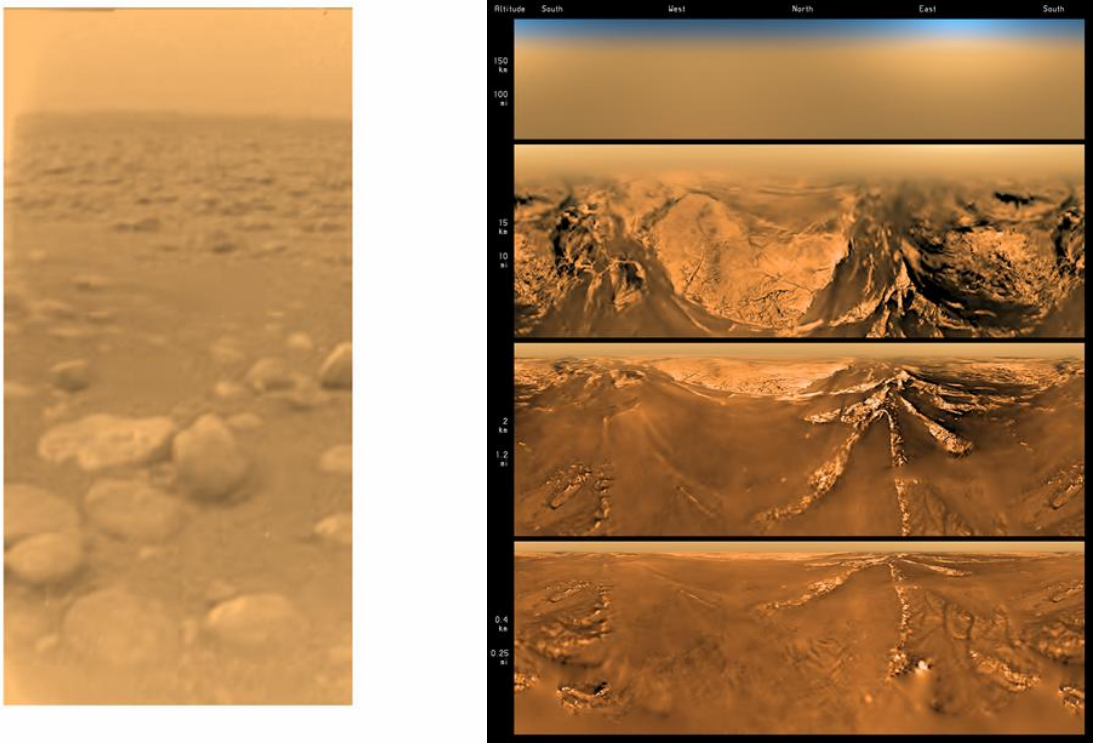


Figure 8.28: Left: a colour view of Titan's surface taken by Huygens. The surface is darker than originally expected consisting of a mixture of water and hydrocarbon ice. Right: panoramic view of the surface of Titan as taken from different altitudes (150, 15, 2 and 0.4 km respectively from top to bottom) during the descent of the Huygens probe to the moon's surface.

Chapter 9

From space exploration to space exploitation?

We choose to go to the Moon in this decade and do the other things,
not because they are easy, but because they are hard.
John F. Kennedy, *US Congress Speach 25 May, 1961*

More than 50 years after NASA was created, its goal is no longer just to reach a destination in outer space but rather to develop the capabilities that will allow Americans to explore and expand their economic horizons beyond Earth.
Linda Dawson, 2017, *The Politics and Perils of Space Exploration*

An entire eighth continent worth of natural resources sits at the edge of the Earth's gravity well, waiting for the right combination of vision, capital and inititive to unlock its wealth.
Brad Blair in Kornuta et al. 2019

Sixty years after it all begun, space exploration has reached a crossroads, and the direction to follow is not clear, to say the least. The financial crisis, political decisions, and other man-made terrestrial problems impact the funding of scientific research activities. Manned spaceflight is desperately trying to re-invent itself with ever changing destinations. In parallel, private companies, motivated by economic considerations, are progressively invading what used to be the playground of the big space agencies. So what's next? Are humans going to return to the Moon anytime soon? Will there be manned missions to other destinations in the Solar System? If yes, for what purpose? Will future space activities be ruled entirely by economic considerations?

9.1 Back to the Moon?

In the context of the Cold War, the Apollo programme was mainly driven by the political competition between the USA and the USSR. Following US President John F. Kennedy's famous speech, the entire US space programme was aiming at the goal to send humans to the Moon and bring them back to Earth. Between July 1969 and December 1972, six Apollo missions landed on the Moon and twelve astronauts left their footprint on our natural satellite. The Apollo programme was a major technological driver. Many new techniques were developed and were later used in everyday's life applications.
The scientific return of these missions was however comparatively modest: beside the installation of instruments to perform seismic studies of the Moon, the astronauts brought 381.7 kg of rocks back from the Moon. These

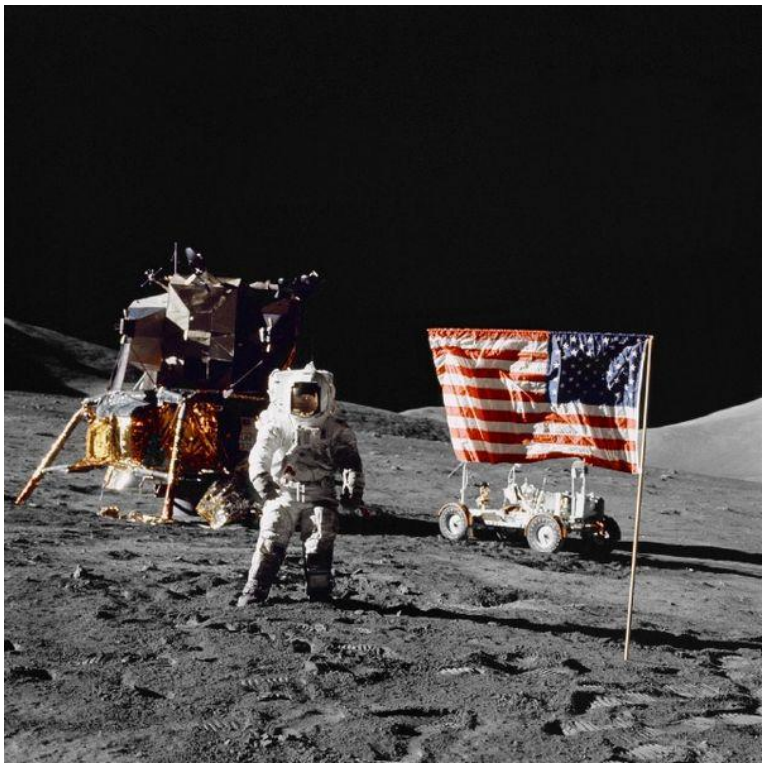


Figure 9.1: Astronaut Harrison Schmitt during the Apollo 17 mission in December 1972. Schmitt is currently the last astronaut who walked on the surface of the Moon.

rocks were shown to be extremely old (3.2 to 4.6 billion years) compared to rocks found on Earth. These samples revealed the existence of geochemical components indicating that the Moon was once completely molten, supporting the hypothesis that the Moon formed during an impact of a large astronomical body on the Earth, 4.6 billion years ago. Once, the Moon's mantle had cooled down and volcanic activity had decreased, the surface was only altered through impacts of smaller bodies (asteroids, comets and meteoroids). In fact, the continuous impacts of micrometeoroites grind the surface rocks into a layer of fine dusty powder known as regolith that covers the Moon with a thickness of several meters.

Shortly after Apollo 11, the political and popular support for the Apollo programme decreased. The USA were struggling with the war in Vietnam and the aftermath of the race riots that had erupted in the sixties. US President Richard Nixon's administration thus cancelled three lunar missions (Apollo 18 – 20), and rather focused on the development of the partially reusable Space Shuttle system which was designed for LEO operations. Following the two accidents of the Shuttle (Challenger in 1986, Columbia in 2003), it was decided to retire the Shuttles after completing the assembly of the International Space Station[1]. In January 2004, US President George W. Bush announced that the US would return to the Moon by 2020 (*Constellation* program) and build a manned lunar outpost with the goal to use this as a basis for a mission to planet Mars. Six years and 9 billion USD later, his successor, President Barack Obama decided to stop this project because it was considered 'over budget, behind schedule and lacking in innovation'. In parallel, Obama's administration led the way towards a deeper involvment of private companies in space activities, known as *New Space*. Yet another turnaround occured when the current occupant of the White House, Donald Trump, made a human return to the Moon (by 2024) one of the priorities

[1]The last flight of a Shuttle took place in 2011.

of the US space policy. Several other countries or space agencies (China, Japan, India, Russia, ESA, Germany,...) are also currently developing plans for unmanned and, in some cases, manned missions to the Moon. And private companies are also considering missions to our natural satellite.

9.2 Future manned missions to the Moon... and beyond

To return to the Moon and travel beyond, NASA initiated the now cancelled *Constellation* programme relying on a (up to 10 times) re-usable space vehicle called *Orion* capable of carrying 4 to 6 astronauts. The Orion spacecraft looks similar to the Apollo commando module and was expected to be launched by an *Ares I* rocket (a first successful test of Ares I was conducted in October 2009). To support missions of longer duration, the Orion vehicle should have had larger fuel tanks and should have carried advanced solar arrays to produce electrical power instead of the fuel cells used aboard the Apollo and Space Shuttle spacecraft. In parallel to the Ares I rocket, NASA was expected to develop the *Ares V* launcher with a heavier lift capacity (188 tons into LEO, 71 tons to the Moon) than any past or existing rocket.

For a typical mission to the Moon, Ares V was expected to put the *Altair* lunar surface access module as well as a transfer stage into orbit around Earth. About 90 minutes later, four astronauts would be put into Earth orbit on board an Orion capsule carried by an Ares I rocket. After a rendez-vous and docking manoeuvre in LEO, the transfer stage would provide the acceleration needed to take the Orion vehicle and the Altair module to the Moon. The transfer stage would then be jettisoned. Once in orbit around the Moon, the Orion spacecraft would be configured for automatic flight and all four astronauts would transfer to the Altair module. The design of the Altair lander was intended to be quite similar to the Apollo lunar module, although it should have been about five times larger than the latter and was designed to land near the lunar poles which are the favorite sites for the construction of a future lunar base that could constitute a first step for a journey to Mars.
Although, the Obama administration stopped the *Constellation* programme, some of its components (such as the Orion module) remain in the current plans of NASA for future manned spaceflights. The Ares launchers were replaced by the concept of the Space Launch System (SLS)[2]. The new programme to return to Moon, called *Artemis* should use the SLS and the Orion capsules. Recently, NASA estimated the cost for a 2024 return to the Moon to 30 billion USD beyond the amounts already spent on SLS and Orion[3].
On the longer term, the Moon could be used as a platform for launches to other destinations in our Solar System. Starting from the Moon, where the gravity is reduced to 1/6 of the Earth value makes the development of long-term flights easier and the absence of an atmosphere removes some design constraints. In response to the objectives of the Trump administration, NASA is promoting the *Deep Space Gateway* project aimed at building a space station in lunar vicinity, i.e. either around the Moon on a wide near rectilinear halo orbit or around the L_2 point of the Earth – Moon system on a halo orbit. The DSG could host four astronauts, though it would not be occupied permanently. The DSG would be used as a staging platform for exploration missions to the lunar surface and eventually to other destinations in the Solar System.

9.3 The Moon, a giant fuel reservoir?

The main question that comes up when dealing with the issue of future missions to the Moon is what would be the purpose of such missions. Nowadays, the cost of such an endeavour needs to be weighted against the expected

[2]The development of the SLS is currently behind schedule. A first unmanned flight is scheduled for the second half of 2021 at the earliest.

[3]The cost of the *Constellation* programme was evaluated to 104 billion USD and the Apollo program costed 25 billion USD (in 1960s economic conditions).

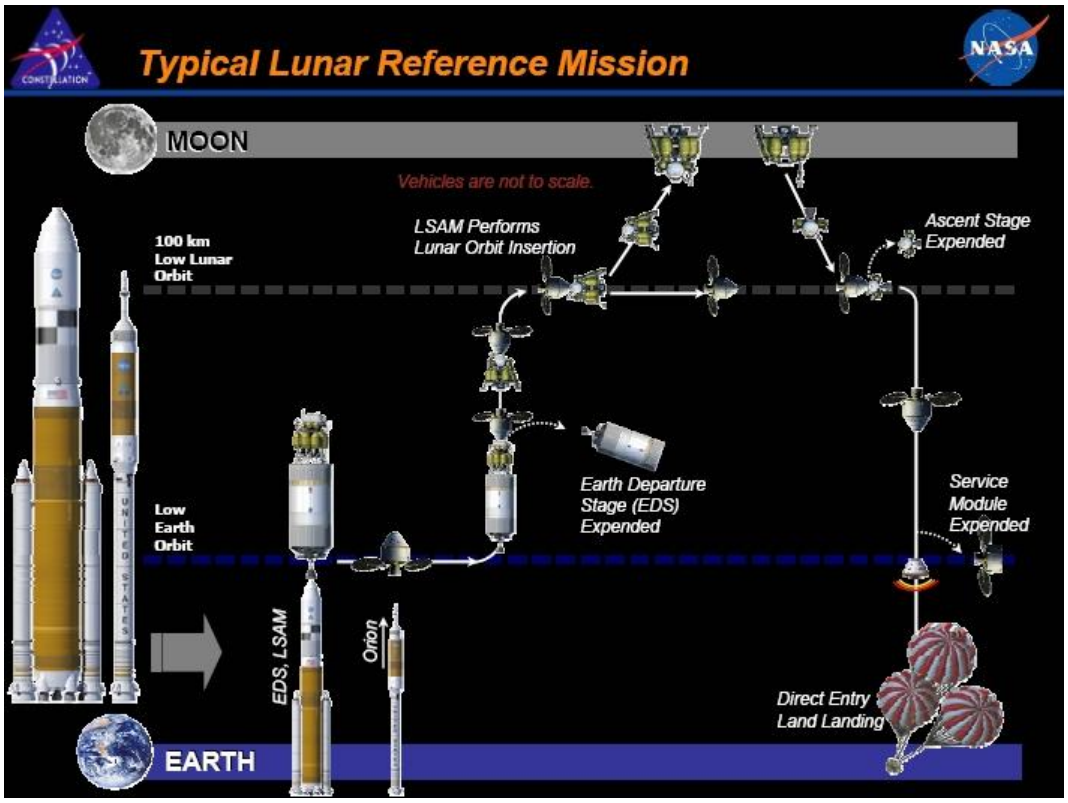


Figure 9.2: The scenario for a manned mission to the Moon within the *Constellation* programme.

return. For instance, to make a man survive on the surface of the Moon costs of order 1 million USD per minute if all resources have to be brought from Earth!
These costs could be reduced if some vital resources, such as water, were directly available on the Moon itself. Unfortunately, the rock samples brought back by the astronauts of the Apollo programme were found to be extremely dry with almost no water in their molecular structure. Indeed, the temperature near the lunar equator varies between -173° C on the dark side and $+127^\circ$ C on the sun-lit side, leaving no chance for finding water in the absence of an atmosphere. However, near the lunar poles, there exist some permanently shadowed deep craters where the temperature is always as low as -240° C and where the Clementine and Lunar Prospector probes actually detected evidence for the existence of water ice in the 1990s. One of the main objectives of current and future robotic missions is therefore an assessment of the actual quantities of water ice that are available on the Moon. Indeed, there is a huge uncertainty on the actual amount of water ice. When the cruise stage of the LCROSS mission impacted the Cabeus crater, a plume of material was created and its composition was analysed, revealing the presence of $5.6 \pm 2.9\%$ of water. Yet, no water ice could be detected in this crater with a ground penetrating radar. Current estimates of the fraction of water ice in permanently shadowed regions range between 1 and 30% (see below).

Over the last decade, the interest for Moon missions has also been motivated by economic considerations aiming at the exploitation of the resources available at the lunar surface. In this context, Crawford (2015) distinguished three levels in the exploitation. A first possibility is to use lunar material to support lunar exploration activities. This would include for instance the extraction of Si to build solar panels in situ or the extraction of water. A second possibility is to exploit lunar resources to facilitate space exploitation. An example here would be the use

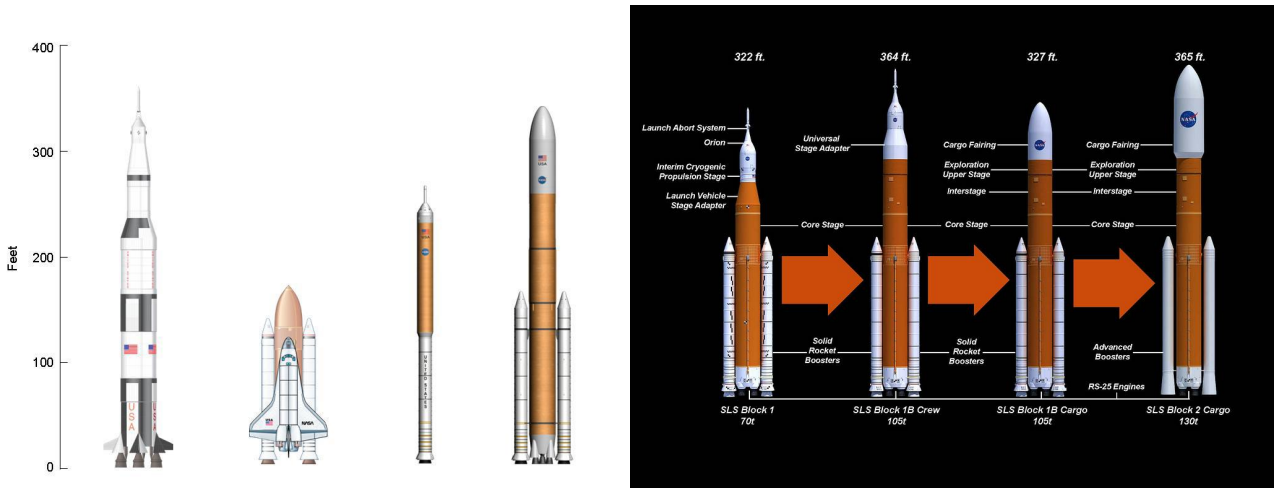


Figure 9.3: Left: Comparison between some previous NASA launch vehicles and the launchers initially foreseen in the Constellation program. From left to right, Saturn V, Space Shuttle, Ares I and Ares V. Right: foreseen evolution of the Space Launch System. The masses indicated at the bottom correspond to the payload that can be put into LEO.

of metals (Fe, Al, Ti,...) extracted from the regolith to build and repair spaceships. The last category includes activities where economically valuable lunar material would be brought back to Earth. One of the most frequently quoted examples in this respect is the extraction of ^{3}He from the lunar surface.

9.3.1 Much ado about nothing: the ^{3}He pipe dream

In the first decade of this century, NASA set up an advisory committee for the implementation of the Moon programme chaired by former Apollo 17 astronaut Harrison Schmitt. Schmitt acted as a leading proponent of a mining activity on the Moon for ^{3}He. ^{3}He is a non-radioactive isotope of helium with two protons and one neutron in its nucleus. Its abundance in the solar neighbourhood is thought to be about 10^{-4} times the abundance of ^{4}He. Some scientists claim that ^{3}He could be the perfect fuel source to satisfy the continuously growing power demands of mankind. According to the ^{3}He lobby, this isotope could ease energy production through nuclear fusion. In 1951, scientists started research with the goal to use controlled fusion to produce energy. One of the main problems is that fusion based on deuterium and tritium releases a large number of highly energetic neutrons that impact on the reactor walls and render them radioactive.

$$^2\mathrm{H} + {}^2\mathrm{H} \rightarrow {}^3\mathrm{He} + \mathrm{n}$$
$$^2\mathrm{H} + {}^2\mathrm{H} \rightarrow {}^3\mathrm{H} + \mathrm{p}$$
$$^2\mathrm{H} + {}^3\mathrm{H} \rightarrow {}^4\mathrm{He} + \mathrm{n}$$

This problem increases the safety costs for such a reactor. Fusion based on ^{3}He and deuterium would produce much less neutrons (about 1% of the number generated by the deuterium - tritium reaction) and the by-product of the reaction, a proton, could be controlled by electric fields and could even be used to generate electricity directly[4].

$$^2\mathrm{H} + {}^3\mathrm{He} \rightarrow {}^4\mathrm{He} + \mathrm{p}$$
$$^3\mathrm{He} + {}^3\mathrm{He} \rightarrow {}^4\mathrm{He} + 2\,\mathrm{p}$$

According to some estimates, about 15 to 20 tons of ^{3}He would be needed to cover the yearly electricity consumption of the USA. However, the abundance of ^{3}He on Earth is very low and it cannot be extracted easily. On

[4]However, since deuterium and ^{3}He need to be mixed together, side reactions (between deuterium nuclei) will occur that do produce neutrons, so that the reaction would not be totally clean either.

the contrary, the absence of a lunar atmosphere allows the solar wind to directly impact the lunar surface and to deposit particles, including ^{3}He, directly into the regolith. Since the mineral ilmenite ($FeTiO_3$) was found to be most efficient in retaining solar-wind ^{3}He, the concentration of ^{3}He is expected to scale with the concentration of ilmenite, which is highest in Mare Tranquilitatis and Oceanus Procellarum (e.g. Crawford 2015, see also Fig. 9.4). The total estimated quantity of ^{3}He in the lunar soil is however quite poorly constrained and ranges between about 2×10^5 and 10^6 tons. Apart from the USA, China, India and Russia have also stated their interest in ^{3}He.

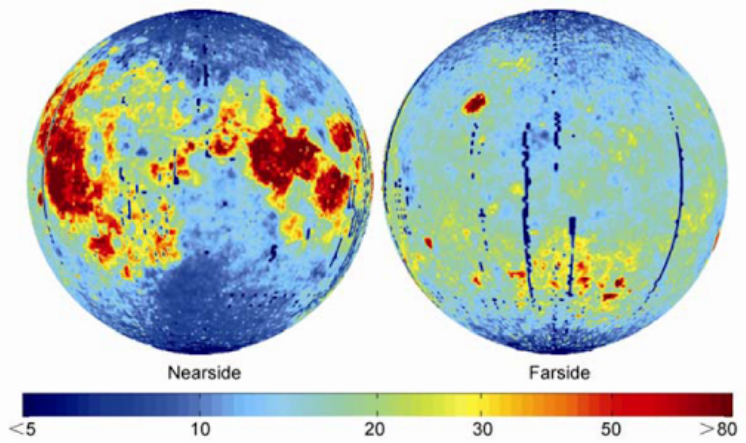


Figure 9.4: Map of the estimated concentration of ^{3}He on the lunar surface (in parts per billion per m^2) as traced by the presence of Ti. Observations were taken with the Chinese Chang'e-1 probe (Fa & Jin 2007).

One may wonder whether or not exploiting this ^{3}He reservoir is indeed as economically attractive as claimed by the supporters of this project. First, to extract ^{3}He from the regolith, the latter must be heated to about 700°C. This is very energy-consuming (about 10^9 J per m^3 of regolith) and according to Crawford (2015), this would already consume $\sim 5\%$ of the energy that can eventually be produced by fusion. Next, to extract 1 ton of ^{3}He, one would need to process about 150 million tons of lunar soil. This would require a vast effort with lots of machines. Assuming that this mining activity would be done to a depth of 1 m, one would need to process an area of 100 km^2 to collect 1 ton of ^{3}He. Finally, the material needs to be brought back to Earth and, in addition, one also needs to consider the energetic costs of extracting deuterium from sea water. This is not to speak about the fact that there currently exist no operational fusion reactors and experts consider that production of nuclear fusion power is still decades away. And, last but not least, nobody has yet started to develop the equipment needed to extract ^{3}He from the lunar regolith. Therefore, the economic interest of lunar ^{3}He is more than questionable. As an alternative, Crawford (2015) estimates that a much more efficient and lasting solution for energy production would be to cover a near equatorial lunar region with solar panels to collect solar light, convert it into microwave energy and beam this energy to Earth...

9.3.2 Other lunar resources?

Beside ^{3}He, other economical exploitation of lunar resources have been proposed. These include extraction of oxygen (O_2), metals (Fe, Ti, Al,...), rare earth elements, thorium and uranium, etc. Whilst all these elements are present in some form on the Moon, their concentration is far from uniform. On Earth, the volcanic and hydrologic activities led to the concentration of interesting metals in mineral deposits, but the situation is more complex on the Moon. Lunar highlands represent the original crust of the Moon which contain anorthite ($CaAl_2Si_2O_8$) as main mineralogical component. Lunar maria are mainly composed of basaltic lava and include, beside anorthite, also

Mg, Fe and Ti-bearing minerals such as ilmenite ($FeTiO_3$). The most interesting places to exploit several types of resources would thus be the maria. However, as for ^{3}He, thermo- or electrochemical processes to extract the most common resources (O_2, Ti, Al, Si) from lunar regolith all require enormous amounts of energy (either thermal or electrical). In the same context, the superficial concentration of Th and U were mapped by orbiters with γ-ray spectrometers able to measure the lines produced by the radiactive decay of these elements. The concentrations that were determined are well below the threshold for making them economically interesting.
Others claim that the economically most interesting lunar resource would be water. Water is extremely precious in Earth orbit[5]. It is also crucial for human exploration of the lunar surface (including obviously direct consumption, but also for the purpose of radiation shielding). Water can also be decomposed via electrolysis to produce liquid hydrogen and oxygen which can be used for life support as well as for rocket propulsion. Kornuta et al. (2019) promote the idea that a lunar water extraction activity should be done solely by robotic systems. Their business model is based on the observation that on Earth, the mix of liquid hydrogen and liquid oxygen costs about one USD/kg, but once transported into LEO, its cost is about 4 000 USD/kg. These authors estimated an annual demand of 450 tons of lunar derived propellant and an annual revenue of 2.4 billion USD for an initial investment of 4 billion USD. Rather than excavating, which would require heavy infrastructure, the idea would be to heat the regolith and extract water by sublimation. Heat could be either applied directly on the surface via concentrated sunlight or heating elements or subsurface via heaters placed in boreholes. The water vapor would then be captured by a dome-shaped tent (see Fig. 9.5) and vented through openings into cold traps where it freezes. The water ice would then be transported to a processing plant where electrolysis would take place. Electrolysis requires about 4.4 kW/kg. The total power needed for this mining activity would be about 2.8 MW, 70% of which would be in the form of electrical power, the rest would be thermal power. This is a factor 30 more power than the needs of the ISS. A challenge is to transport this power generated outside the permanently shadowed regions (epecially if power would be generated by solar arrays) to the extraction site which is likely located in deep shadow. Options include wired tranmission, though this requires deployment efforts, power beaming and heliostats (see Fig. 9.5). Nuclear fission could provide an alternative to solar energy, but would require the design of space-qualified reactors.

Figure 9.5: Left: schematic view of a mining tent for the extraction of water from the lunar soil. Middle: artist view of the power beaming used to concentrate sunlight onto a permanently shadowed extraction site. Right: schematic view of a robotic tanker combined with a pipeline near the crater wall to transport the extracted water. From Kornuta et al. (2019).

Propellant storage would be another challenge especially for liquid hydrogen where cryogenic storage facilities are required. Several critical aspects, requiring the development of dedicated infrastructure elements were identified by Kornuta et al. (2019): mining and processing, propellant storage, power supply, robotic systems, communication and transportation (both in and from space and on the lunar surface, involving notably the development of rendez-vous capabilities).
Again there are huge uncertainties on the economic interest of such a project mainly because of the uncertainties

[5]Delivering 1 kg of water from the Earth to the ISS costs several 1 000 USD.

regarding the amount of water available (currently estimated between 1 and 30%). The power needed to extract water via sublimation strongly increases when the water ice content gets lower. Moreover, permanently shadowed regions are difficult to exploit because of the very low temperatures. The low temperatures are notably problematic for the use of liquid lubricants for joints and bearings in mechanical devices. Neutron detectors suggest that hydrogen enriched material also exists close to but outside the permanently shadowed regions. These regions could offer an alternative.
Another major problem comes from the lunar regolith which is sharp and abrasive and could damage the robots. The exhaust gas from the descent stage of the Apollo LEM blew at least one ton of regolith at each landing. Some of the finest particles were accelerated to velocities of up to $3\,\mathrm{km\,s^{-1}}$, thus exceeding the escape velocity of the Moon ($2.38\,\mathrm{km\,s^{-1}}$). Such particles can thus travel over very long distances, leading to a pollution problem at each take-off or landing.
Moving around on the lunar surface is needed to transport and place the extraction equipment, to move the water from the extraction site to the processing center, and to move the propellant to a storage facility and to the landing site for distribution to customers. Pipelines are probably not a good solution as they would have to be dismantled, moved, and reassembled each time the extraction site changes. Robotic tankers are an option, but would have trouble climbing along the crater walls. The best solution is probably a combination of these options: the robotic tanker would remain in the crater and piping would be used to move the water out of the crater (see Fig. 9.5). An important bottleneck of the whole scenario concerns the transportation of the extracted fuel to the customers in orbit. The cost of this operation can be appreciated via the amount of fuel required. For this purpose, we introduce the gear ratio, defined as the ratio between the mass of propellant needed for one roundtrip and the mass of the payload that can be delivered. According to Kornuta et al. (2019), to get to the Earth-Moon L_1 point from the surface of the Moon and back, requires a gear ratio near 2. Likewise, bringing material from the Moon to LEO requires a gear ratio of 6.

9.4 *New Space* or the 21st century gold rush

As stated earlier in this chapter, former US President Obama paved the way towards a deeper involvment of private companies in terms of human spaceflight towards low-Earth orbits. This led to a clear shift in space exploration over the last decade. Private companies are playing an increasingly important role. The best-known example is certainly Space-X founded by Elon Musk in 2002. The company employed 160 people in 2005, but this number has grown with time to more than 3500 around the year 2017. In 2008, Space-X launched the first privately funded liquid-propellant rocket (Falcon I) into the Earth's orbit. In 2012, Space-X became the first private company to successfully launch a space capsule (Dragon) that docked on the ISS to deliver cargo. In December 2015, they were the first to successfully land back the first stage of a rocket after launching and sending a payload into orbit. In May 2020, they finally launched the Crew Dragon spacecraft with two astronauts onboard. This was the first manned spaceflight that was entirely operated by a private company. Whilst these accomplishments are certainly impressive and constitute a clear challenge for traditional launchers such as Ariane V, one needs to keep in mind that Space-X, despite being a private company, is fueled by substantial amounts of public money. Beside large sums of private capital, Space-X benefits from public subsidies and contracts with the US government.
These actors of the *New Space* pursue an extremely aggressive public relation policy. In February 2018, Space-X used the first launch of its new Falcon Heavy launcher to create an enormous buzz by putting a Tesla roadster[6] into a heliocentric orbit. Beside this event being an enormous public relation operation, it is also a dangerous first: never before had humanity spent the equivalent of several hundred million USD to deliberately create a piece of (promotional) space junk.
Beside the prospective of human spaceflight, the New Space phenomenon also attracts private companies who are

[6]Elon Musk is chief executive of both Space-X and Tesla Motors.

Figure 9.6: Left: artist view of the Interceptor micro-spacecraft flying by an asteroid (from the website of the Planetary Resources company, http://www.planetaryresources.com). Right: same for the Deep Space Industries concept of a propellant refinery working on a carbonaceous asteroid (from the website of Deep Space Industries, http://deepspaceindustries.com).

eager to make money with un-manned probes in space mining. The best examples are Planetary Resources (PR) and Deep Space Industries (DSI). Both companies were launched in 2012 - 2013 to raise funds for missions to near-Earth asteroids with the goal to use them as giant reservoirs of natural resources, including water, hydrocarbons, but also rare metals such as platinum[7]. These companies used a very aggressive marketing, based on the statement that *harnessing valuable minerals from a practically infinite source will provide stability on Earth, increase humanity's prosperity, and help establish and maintain human presence in space.* For this purpose, they first proposed to use small, low-cost commercial robotic spacecraft (such as the Interceptor, see Fig. 9.6) to explore thousands of asteroids and determine their properties and suitability for mining. To reduce costs, they planed to *incorporate recent innovations in commercial microelectronics, medical devices, and information technology, in ways not traditionally used by robotic spacecraft* and they were willing to *aggressively accept mission risk where appropriate*. In a second step, these companies proposed more complex, sample return missions and later on fully autonomous mining spacecraft (see Fig. 9.6) that could process thousands of tons of material per year. This included so-called 'biomining'[8] where a swarm of CubeSats would be used to inject a fluid with genetically engineered bacteria into the asteroid. After several years, these bacteria would then alter the chemical states of the asteroid's material rendering it more suitable for the subsequent extraction of metals. The potential customers could be space-borne (e.g. re-fueling of commercial satellites with propellant to increase their lifetime, tanking up manned missions to Mars with water and propellant) or on Earth.

Again, the question comes up whether such activities would be economically viable. To answer this question, let us consider the US scientific mission Osiris Rex launched in September 2016 and designed to bring $\sim 1\,\mathrm{kg}$ of material back from a carbonaceous asteroid in 2023. The cost of this mission (development and launch) is about 1 billion USD. From this example, we see that if the goal is to bring back material to the Earth in an economically viable way, it needs to be a highly valuable mineral and the extraction needs to be done in situ. It would be a huge waste of resources to bring back material containing only a few percent of the precious mineral. The best candidate materials would be platinum, scandium, palladium and thulium which are used for building supraconducting devices and cost up to 40 000 USD per kg. However, extracting these minerals would imply assembling complex robotic factories in space, representing a huge investment that would take many years to be recovered. As for the Moon, the economically most interesting resource might once more be water (20% of the mass for carbonaceous asteroids).

[7]Platinum is up to 30 times more abundant in some asteroids than on Earth.
[8]This technique is already used on Earth for extracting Cu, Au and U.

In the case of PR and DSI, the question about the economic viability was answered negatively. Although these companies had the support of ultra-rich investors (including Google co-founder Eric Schmidt and filmmaker James Cameron for PR, they failed to raise the funds they were counting on and both companies were acquired by others with no space mining activity in their mid-term agenda.

9.5 It's a lawyer's world

In the 1960's, Ralph J. Cordiner, chairman of the General Electric company, predicted that there would be three basic stages of space activities: (1) exploration , (2) economic development, and (3) mature economic operations. But, is it actually possible to economically exploit the Moon, or outer space in general?

The UN **Outer Space Treaty** sets the basic rules of international space law, defining outer space as the common heritage of mankind. The treaty was first signed by the USA, the UK and the USSR and entered into force in October 1967. Today, it has been ratified by about 100 countries, including all major countries with a space activity. This treaty prohibits the placing of nuclear or other mass destruction weapons in orbit or in outer space. It further limits the use of the Moon and other celestial bodies to peaceful purposes (no installation of military bases or military manoeuvres) and **explicitly forbids any government from claiming sovereignity of any celestial body**. The activities of non-governmental entities require authorization and supervision by the appropriate state party to the treaty and the state party bears the international responsibility for space activities carried out by individuals or non-governmental entities from their country. Clearly, a commercial exploitation of the Moon and its resources was not considered when the treaty was set-up. In the late 1970's there has been an attempt to set up a specific Moon Treaty, but this attempt was not successful as none of the leading space powers supported it.

Legal questions thus concern the priority rights to mining claims and the right to own and sell the extracted resources. Space mining lobbyists consider that the Moon and the asteroids belong to a group of people (humanity) and may be used by each member of the group, but cannot be appropriated by anyone. This situation is analogues to international waters where fishing is allowed. Likewise, the lobbyists for Moon mining activities claim that the Moon could eventually become an independent country, and if this happened, it would have ^{3}He or other lunar resources to trade.

In November 2015, the US Congress adopted the so-called Commercial Space Launch Competitiveness Act (**SpaceAct**) which allows US citizens to own, transport and sell any space resource obtained during commercial exploration. This legislation, based on a *finders, keepers* rule, opens the avenue to private investors to perform mining activities in outer space for their own benefit. The SpaceAct could thus be seen as a violation of the UN Space Treaty. However, US lawyers consider that the Space Treaty only applies to nations as such and does not explicitly apply to their citizens: state-owned enterprises would not be able to appropriate parts of outer space, but private entities could. This development attracted a lot of interest around the world. In February 2016, Luxembourg adopted a similar legislation (SpaceResources.lu) with the clear goal to attract private companies interested in asteroid mining activities (Deep Space Industries indeed announced a joint venture with the government of Luxembourg in May 2016). In December 2016, the Japanese Space Agency JAXA also signed an agreement with the private company ispace.inc to lay down plans for mining activities on the Moon. This was followed in February 2020 by the United Arab Emirates adopting a legislation very similar to that of the US and Luxembourg. In May 2020, the US government announced the so-called **Artemis Accords** which consist in bilateral agreements between the US and their partners who must adhere to the US policy on lunar exploration and resource exploitation. The US notably introduce the concept of **safety zones** where an actor occupies a site and withdraws its resources. What is totally unclear is how these safety zones would be defined and how conflicts would be solved. There is a high risk that this will lead to a first come, first served approach which clearly violates the Outer Space Treaty.

In August 2020, the US have set-up a **Space Force** to *defend space from those who will seek to undermine their goals in space*. In this context, the word goals is to be understood as the *totality of a nation's ability to exploit*

the space domain in pursuit of prosperity and security. Beside the obvious military aspects, which by themselves probably constitute a dangerous violation of the Outer Space Treaty, this includes also economic exploitation of space. The Trump administration thus defines the Space Force as the *protector of US national and private activities throughout the Solar System*.
There is obviously a strong *conquest of the far west* mentality, although the economic profitability of such an endeavour remains to be established (see above). In this context, Elvis & Milligan (2019) recently argued that the part of the Solar System accessible to human exploitation should be limited to one eighth, with the remainder left as space wilderness (i.e. areas where humans would be visitors who do not remain). This argument is actually not motivated by ethical or ecological considerations. Instead, it is based on the assumption of a 3.5% annual growth rate for space economy, which corresponds to a doubling time of 20 years. Taking the estimated quantity of iron in the asteroids of the Solar System as a reference[9], the one eighth limit would be reached after 400 years. This would then leave humanity another 60 years (before exhausting the remaining seven eighths!) for a transition to a new economic system. How exactly the one eighth should be defined (e.g. in terms of surface area or mass) and by whom are controversial issues. Elvis & Milligan (2019) cite the example of the asteroids where, at first sight, mass would be the best quantity to define the limit. Yet, since the asteroids Ceres, Vesta, Pallas and Hygeia account for about half of the mass of the asteroids of the main belt, Elvis & Milligan (2019) argue that for logistic reasons these four objects will likely become prime targets of space mining activities. For these four objects, it might thus be better to set the limit in terms of surface area.

Beyond these legal considerations, let us stress that mining the Moon for ^{3}He or any other resources will not go without changing the lunar landscape. Indeed, according to Crawford (2015) to produce 10% of mankind's annual electricity consumption, at least 500 km^2 of the lunar surface would have to be processed every year. Obviously, the destructive power that humans have developed strongly argues against space exploitation and colonization. However, supporters of the economical exploitation of space argue that *economic growth requires the exploitation of resources and inevitably leads to environmental pollution*. In their opinion, *economic growth requires the exploitation of resources and must lead to environmental pollution*. They argue that *ethical objections to space exploitation fail when the survival of human species is at stake, and colonization of space becomes a moral duty* (Abylkasymova & Szocik 2019). Yet, we need to ask ourselves whether humanity has the moral right to do this, instead of dealing with and protecting the resources available on our home planet Earth.

9.6 Afterword

The space mining bubble has bursted for now, and some of the above plans might sound like science fiction, but wasn't that also the case of the visions of many pioneers of space exploration less than a century ago? The role that private companies will play on the medium term in space exploration and exploitation remains currently unclear. However, unlike scientific exploration missions, funded by national or international space agencies, private companies aim at creating returns upon their investment in the short term, implying a radical change in philosophy. Is outer space becoming the playground of the *financial industry*?
Obviously, space exploration is at a turning point of its existence. With the multiplication and diversification of the actors in space, some weaknesses of the Outer Space Treaty are becoming apparent. For instance, there exists no overarching authority able to enforce compliance with treaty regulations and private companies usually do not share the scientists' preoccupations on planetary protection, etc.
To conclude, we will thus end these lectures as we started them with a remarkable quote from the science fiction parody 'Mars Attacks': *You wanna conquer the world, you're going to need lawyers, right?*

[9] Elvis & Milligan (2019) estimate that one eighth of the iron ore in the asteroid belt would be more than a million times the known iron reserves on Earth.

Chapter 10

Exercises

10.1 The rocket equation

1.) Establish the rocket equation for a rocket that takes off vertically, accounting for gravity losses. How can the $\Delta\, v$ be increased? Establish also the energy loss and discuss how it can be reduced.

Answer: $\Delta\, v = v_{ej}\, \ln \frac{m_{init}}{m_{final}} - \int_{init}^{final}\, g\, dt$; reduce the duration of the engine operation; $G = \frac{g\, m_{final}\, v_{ej}}{\dot{m}} \left[1 + \frac{m_{init}}{m_{final}} \left(\ln \frac{m_{init}}{m_{final}} - 1\right)\right]$; reduce G by increasing $\dot{m}$.

2.) Consider a multistage rocket consisting of N stages. The initial mass of a stage is given by $m_{i,n} = m_{f,n} + m_{p,n}$ where the subscripts i, f and p indicate the initial, final (burnout) and propellant masses. The final mass of the stage is given by $m_{f,n} = m_{s,n} + m_{e,n} + m_{u,n}$ where the subscripts s, e and u indicate the structural, engine and useful masses, with $m_{u,n} = m_{i,n+1}$. We further define $u_n = \frac{m_{u,n}}{m_{i,n}}$, $r_n = \frac{m_{i,n}}{m_{f,n}}$ and $s_n = \frac{m_{s,n}+m_{e,n}+m_{p,n}}{m_{s,n}+m_{e,n}}$. The overall performance of the rocket is given by $u = u_1\, u_2 \ldots u_N$. Establish the value of r_n that renders u maximum, if s_n is independent of r_n.

Answer: $u = \left[s\, \exp \frac{-\Delta\, v}{N\, v_{ej}} - 1\right]^N\, \frac{1}{(s-1)^N}$

3.) Same question, but considering this time the more realistic assumption that the parameters $a_n = \frac{F_n}{g_0\, m_{f,n}}$, $\kappa_n = \frac{F_n}{g_0\, m_{e,n}}$ and $\zeta_n = \frac{m_{s,n}}{m_{p,n}}$ are independent of r_n.

Answer: $u = \left[(1 + \zeta - \frac{a}{\kappa})\, \exp \frac{-\Delta\, v}{N\, v_{ej}} - \zeta\right]^N$

4.) Compute the velocity increment that a rocket launched from Kourou (latitude $5.2°$), Cape Canaveral (latitude $28.5°$) or Baïkonour (latitude $51.6°$) can gain in being launched due East. Also, compute the overall performance u from these three sites for a launch onto a geostationary transfer orbit, assuming a three-stage rocket with a performance given by the answer of question 3 above.

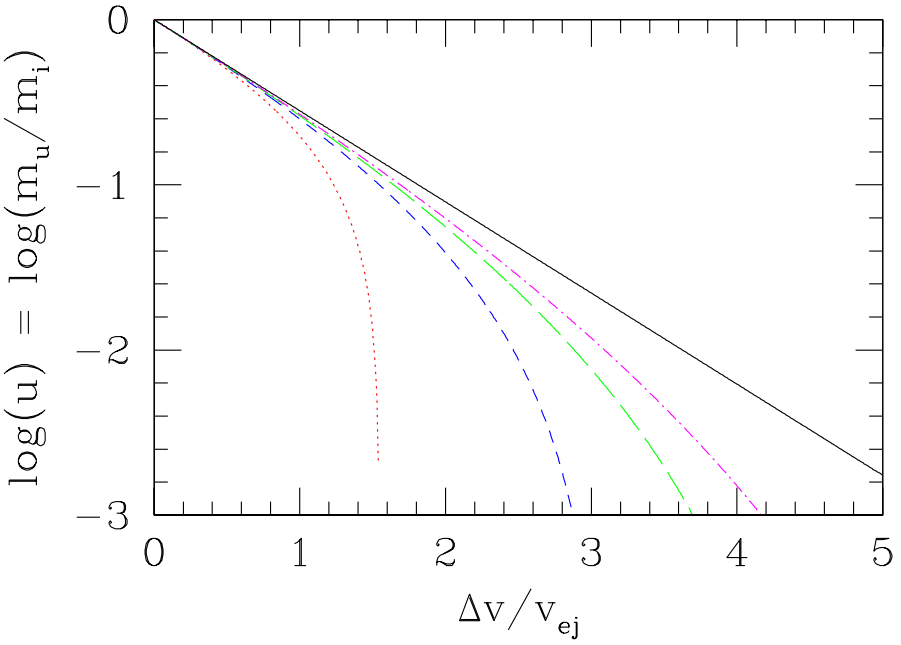


Figure 10.1: u as a function of the ratio $\Delta\, v/v_{ej}$ for $s = 4.7$. The different lines correspond to N = 1, 2, 3, 4 and ∞.

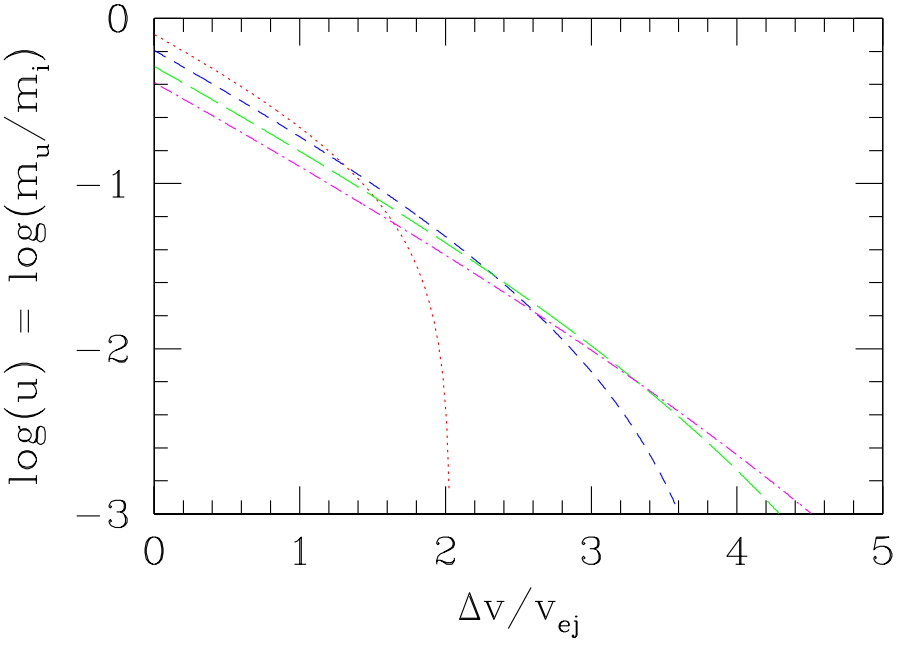
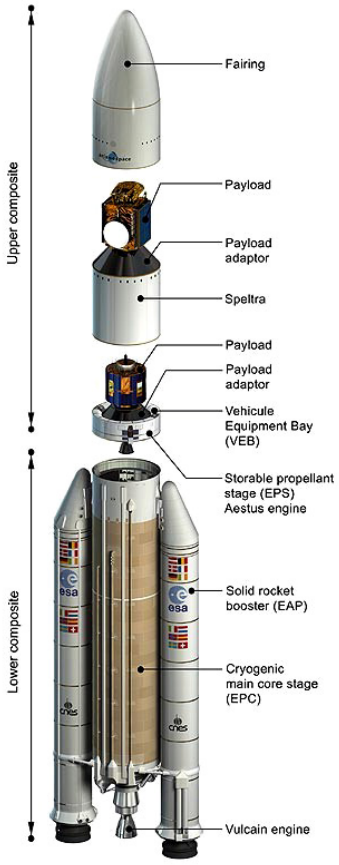
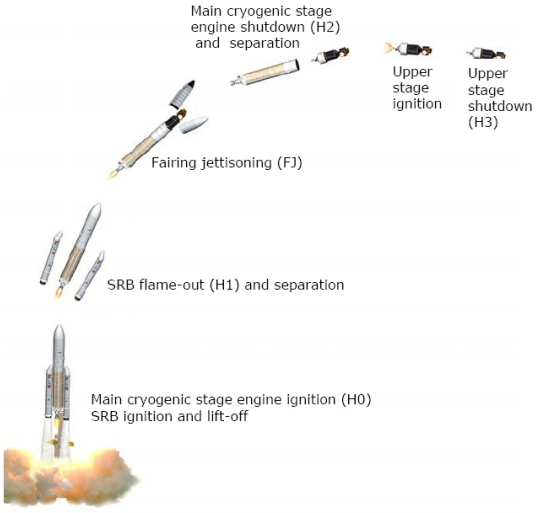


Figure 10.2: u as a function of the ratio $\Delta\, v/v_{ej}$ for $\zeta = 0.12$, $\kappa = 40$ and $a = 8$. The different lines correspond to N = 1, 2, 3 and 4.

10.2 The Ariane V launcher



1.) The cryogenic stage of the Ariane V launcher contains 132 tons of liquid oxygen and 25 tons of liquid hydrogen at a temperature of $-252.8°$ C. The densities of liquid oxygen and hydrogen are $1.14\,\mathrm{g\,cm^{-3}}$ and $0.07\,\mathrm{g\,cm^{-3}}$ respectively. Given that the diameter of the Ariane V rocket is 5.4 m, what is the minimum height of the tanks of both propellants?
2.) The chemical reaction $2\,H_2 + O_2 \rightarrow 2\,H_2O + Q$, where Q = $12.792\,\mathrm{kJ\,g^{-1}}$ (per unit mass of water) is limited by the quantity of one of the propellants. Which one? What do you think is the other exceeding propellant used for?
3.) The rate at which oxygen enters the combustion chamber is about $221\,\mathrm{l\,s^{-1}}$. What is the production of energy over one second? Compare this energy with the potential energy of a mass of 10 000 kg lifted to an altitude h. What would be the value of h corresponding to the same energy?
4.) Figure 10.3 illustrates a typical launch sequence of the Ariane V rocket. The left panel shows the successive steps starting with take-off and ending with orbit insertion. The right panel yields the acceleration during the various steps of the operation. Why does the acceleration go to zero after burnout of the second stage and why is it different from zero at the separation of the boosters that make up the first stage?
5.) During the operation of the second stage, the thrust $-\dot{m}\vec{v}_{ej}$ is essentially constant. Why does the acceleration increase?
Why is the shroud ejected at an altitude of about 100 km?



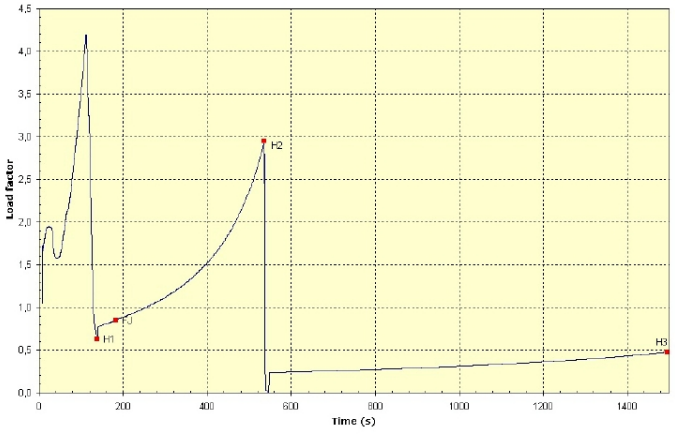


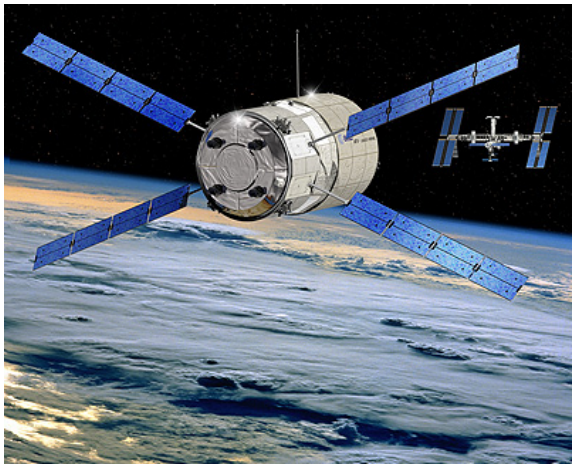
Figure 10.3: Left: typical launch sequence of the Ariane V rocket. Right: the acceleration on the rocket as a function of time during launch.

10.3 Hydrazine thrusters

Most spacecraft rely upon conventional hydrazine (N_2H_4) thrusters. Consider a spacecraft with a total mass of 8000 kg. The catalytic decomposition of hydrazine occurs through the following reactions:
$3\,N_2H_4 \rightarrow 4\,NH_3 + N_2$
followed by
$2\,NH_3 \rightarrow N_2 + 3\,H_2$
for about 40% of the NH_3 produced by the first reaction. The three gaseous components (NH_3, H_2 and N_2) are then ejected through the nozzle.
1.) What are the masses of the various gaseous products ejected per unit mass of hydrazine?
2.) Considering that the gas is ejected into space at a velocity of $3\,\mathrm{km\,s^{-1}}$, what is the amount of hydrazine needed to change the velocity of the spacecraft by $0.35\,\mathrm{m\,s^{-1}}$?

10.4 Orbital manoeuvres

1.) ESA has developed an Automated Transfer Vehicule (ATV) to deliver consumables (such as food, oxygen, water and propellant) to the International Space Station (ISS). After launch by an Ariane V from Kourou, the ATV first evolves on a circular orbit at an altitude of 300 km. The ATV then uses a Hohmann transfer orbit leading it to the point of encounter with the ISS (revolving on a circular orbit at an altitude of 350 km above sea level, with an inclination of $51.6°$ with respect to the equator).
a.) What is the ideal window for launching the ATV?
b.) What is the duration required for the transfer arc?
c.) What is the value of the angle ϕ between the ATV and the ISS at the time of the start of the transfer manoeuvre?



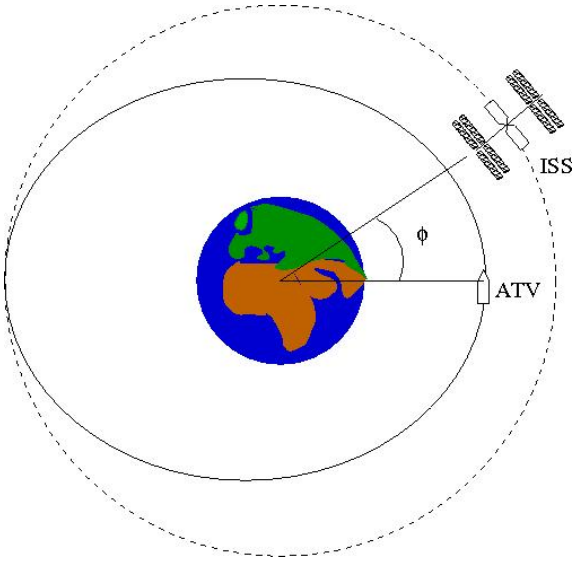


Figure 10.4: Left: artist impression of an Automated Transfer Vehicule approaching the ISS. Right: Schematic view of the ATV transfer orbit. Note that the various distances and radii are not to scale.

2.) Consider a spacecraft launched on an Earth escape orbit with $V_\infty = 5.1\,\mathrm{km\,s^{-1}}$ aligned with the Earth's orbital velocity ($29.8\,\mathrm{km\,s^{-1}}$).

a.) Determine the values of the semi-major axis a, the period and the eccentricity of the orbit.
b.) In a so-called **Delta V Gravity Assist** manoeuvre, the spacecraft brakes at aphelion. Suppose that the velocity at aphelion is reduced by $0.55\,\mathrm{km\,s^{-1}}$. Compute the value of the energy and the angular momentum after the manoeuvre.
c.) What is the velocity vector of the spacecraft relative to Earth when it flies by the Earth (at an altitude of 320 km) on its way back to perihelion?
d.) What are the change in the direction of the velocity vector and the gain in tangential velocity (in the heliocentric frame of reference) after the gravity assist manoeuvre? Compare this value to the maximum possible velocity gain.

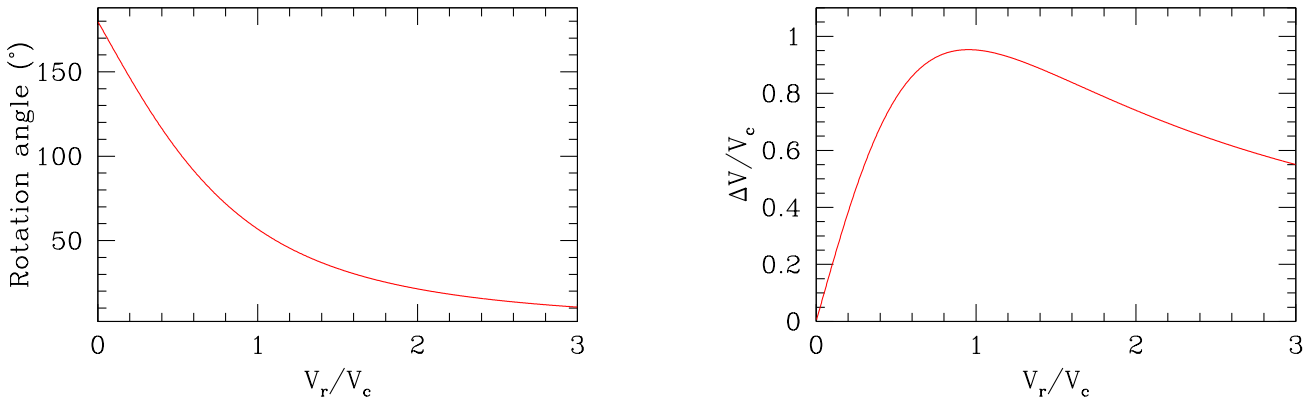


Figure 10.5: Left: the variation of the angle of the velocity vector during a fly-by manoeuvre. Right: the maximum velocity increase in terms of $V_c = \sqrt{\frac{G\,M_c}{R_c}}$ during the fly-by.

10.5 Space debris

1.) What is the velocity of a space debris on a circular orbit at an altitude between 400 and 1000 km?
2.) What is the kinetic energy liberated in the collision of a sphere of aluminium ($\rho_{\mathrm{Al}} = 2.70\,\mathrm{g\,cm^{-3}}$) at an altitude of 700 km with a diameter of 5 cm with a spacecraft if the collision is head-on? Compare this number to the kinetic energy of a car of 1000 kg that hits an obstacle at rest. What would be the velocity of the car if the energies were the same?

10.6 Missions to Mars

1.) For a lander on a distant planet, it is of fundamental importance to know what is the density of the atmosphere around the planet. On Mars, the average temperature is -55° C, the average atmospheric pressure is 6.35 mbar and the composition of the atmosphere is 95.49% CO_2, 2.7% N_2, 1.6% Ar, 0.13% O_2 and 0.08% CO. What is the average density of the Martian atmosphere? How does this compare to the average density of the Earth's atmosphere ($T_{aver} = 15^\circ$ C, $p_{aver} = 1013$ mbar = 1 atm; 78.07% N_2, 21% O_2, 0.9% Ar, 0.03% CO_2)? What is the range of densities on Mars due to the different temperatures (ranging from -133° C at the pole during winter time to 27° C during daytime in summer?
Note: $R = 0.0821\,\text{atm}\,\text{mol}^{-1}\,\text{K}^{-1}$

2.) Mars Express sends its data to the ground station at a rate of 100 kilobit per second. If the distance between Mars and Earth is 300 million km, how long does it take to downlink an HRSC image of 200 Mbytes? How long does it take to transmit the full 5 Gbits of data collected each day by the orbiter?

3.) A radar on board an orbiter can be used to establish a topographic map of the planet. Express R in terms of the phase shift ϕ and z in terms of H, R and θ. What is the accuracy with which a seismic motion (Δz) can be measured if the wavelength of the radar is 6 cm and $\theta = 20^\circ$?

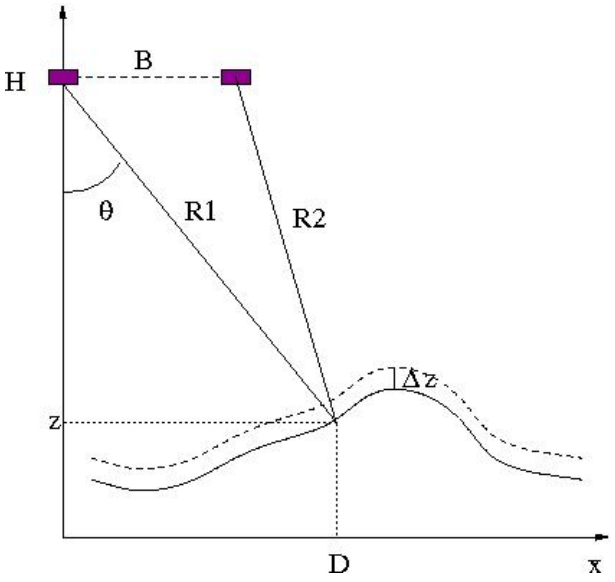


Figure 10.6: Schematic view of the altimetric measurements with a radar on board a space probe.

Bibliography

[1] Abylkasymova, R., & Szocik, K. 2019, Space Research Today, 206, 44

[2] Appourchaux, T. 2005, in *Payload and Mission Definition in Space Sciences*, eds. V. Martinez Pillet, A. Aparicio & F. Sanchez, Cambridge University Press, 185

[3] Ball, A.J., Garry, J.R.C., Lorenz, R.D., & Kerzhanovich, V.V. 2007, *Planetary Landers and Entry Probes*, Cambridge University Press

[4] Balogh, A. 2005, in *Payload and Mission Definition in Space Sciences*, eds. V. Martinez Pillet, A. Aparicio & F. Sanchez, Cambridge University Press, 233

[5] Barcons, X. 2005, in *Payload and Mission Definition in Space Sciences*, eds. V. Martinez Pillet, A. Aparicio & F. Sanchez, Cambridge University Press, 89

[6] Biddle, W. 2009, *Dark Side of the Moon, Wernher von Braun, the Third Reich and the Space Race*, W.W. Norton & Company Inc., New York

[7] Blamont, J. 2008, *Planetary balloons*, *Experimental Astronomy*, 22, 1

[8] Buisán, D. 2004, *Lift-off: European Space Agency physics and chemistry exercises based on real space data for secondary schools*, ESA Publication Division, BR-223

[9] Coradini, A. 2005, in *Payload and Mission Definition in Space Sciences*, eds. V. Martinez Pillet, A. Aparicio & F. Sanchez, Cambridge University Press, 323

[10] Crawford, I.A. 2015, *Lunar Resources: A Review*, *Progress in Physical Geography*, 39, 137

[11] Dawson, L. 2017, *The Politics and Perils of Space Exploration*, Springer Praxis Books

[12] Edgington, S.G., & Spilker, L. 2014, *Space Research Today*, 191, 80

[13] Ehrenfreund, P., Hertzfeld, H., & Howells, K. 2013, *Space Research Today*, 188, 30

[14] Elvis, M., & Milligan, T. 2019, *Acta Astronautica*, in press, arXiv1905.13681

[15] Eversberg, T. 2013, *Hollywood im Weltall, Waren wir wirklich auf dem Mond?*, Springer Spektrum

[16] Fa, W., & Jin, Y.-Q. 2007, *Icarus*, 190, 15

[17] Fraeijs de Veubeke, B. 1960, *Elementary Problems of Overall Rocket Performance* in *Rocket Propulsion*, Elsevier Publishing Company

[18] Frankel, C. 2016, *L'Eldorado des Asteroïdes*, *Ciel & Espace*, 545, 28

[19] Giménez, A. 2005, in *Payload and Mission Definition in Space Sciences*, eds. V. Martinez Pillet, A. Aparicio & F. Sanchez, Cambridge University Press, 1

[20] Harrison, R.A. 2005, in *Payload and Mission Definition in Space Sciences*, eds. V. Martinez Pillet, A. Aparicio & F. Sanchez, Cambridge University Press, 125

[21] Klinkrad, H. 2015, *The Space Devris Environment and Associated Risks*, ULg FSA departemental seminar, 22 March 2015

[22] Kminek, G., & Rummel, J.D. 2015, *Space Research Today*, 193, 7

[23] Kminek, G., Conley, C., Hipkin, V., & Yano, H. 2017, *Space Research Today*, 200, 12

[24] Koon, W.S., Lo, M.W., Marsden, J.E., & Ross, S.D. 2001, *Celestial Mechanics and Dynamical Astronomy*, 81, 63

[25] Kornuta, D., Abbud-Madrid, A., Atkinson, J., et al. 2019, *REACH - Reviews in Human Space Exploration*, 13, 100026

[26] Langevin, Y. 2005, in *Payload and Mission Definition in Space Sciences*, eds. V. Martinez Pillet, A. Aparicio & F. Sanchez, Cambridge University Press, 17

[27] Lebeau, A. 1998, *L'espace, les enjeux et les mythes*, Hachette

[28] Mark, C.P., & Kamath, S. 2019, Space Policy 47, 194

[29] Moura, D. 2011, *SSA/SW Space Situation Awareness - Space Weather*, CSL departemental seminar, 12 December 2011

[30] Nazé, Y. 2013, *Voyager dans l'Espace*, CNRS Editions

[31] Perryman, M. 2005, in *Payload and Mission Definition in Space Sciences*, eds. V. Martinez Pillet, A. Aparicio & F. Sanchez, Cambridge University Press, 219

[32] Reichl, E. 2011, *Raumsonden*, Motorbuch Verlag

[33] Reichl, E. 2011, *Trägerraketen*, Motorbuch Verlag

[34] Reichl, E. 2012, *Zukunftsprojekte der Raumfahrt*, Motorbuch Verlag

[35] Schoenmaekers, J., Jehn, R., Landgraf, M., Khan, M. 2008, *ESA Bulletin*, 133, 11

[36] Shahar, K., & Greenbaum, D. 2020, Nature Astronomy 4, 208

[37] Spudis, P.D. 2004, in *Moon*, World Book Online Reference Center, World Book InC. (`http://www.worldbookonline.com/wb`)

[38] Stark, J.P.W., & Swinerd, G.G. 2003, in *Spacecraft Systems Engineering, 3rd Edition*, eds. P. Fortescue, J. Stark, & G. Swinerd, John Wiley & Sons Ltd., 111

[39] Taylor, J.W.R. 1971, *Raketen und Lenkgeschosse*, Delphin Verlag, Stuttgart

[40] Wiens, R., Clegg, S., Maurice, S., & Gasnault, O. 2016, *Space Research Today*, 195, 21

[41] Wikipedia, The Free Encyclopedia, `http://en.wikipedia.org/wiki/`

Several images and many information are taken from the ESA (http://www.esa.int) and NASA websites (http://www.nasa.gov) as well as from Space.com (http://www.space.com) and Aerospaceweb (http://www.aerospaceweb.org).