Chapter 1
Overview Space Segment

1.1 The Space Environment

Adrian R.L. Tatnall

1.1.1 Introduction

The space environment in which spacecraft have to operate is an alien world in which we would not survive for more than a few minutes without protection. Fortunately, in this respect, spacecraft are generally more robust than humans and it is possible for spacecraft to regularly operate continuously for more than 15 years. In the case of Voyager 1, launched over 35 years ago, the spacecraft continues to operate and communicate with the earth 18 billion km away. It is interesting to question how this longevity can be achieved when there is no possibility of maintenance and the environment, at first sight, appears so unattractive.

To understand these issues the space environment is considered. The answer is that whilst the space environment is different many of the sources of erosion and wear on Earth are not present in space.

Whilst the space environment is alien, it is only remote in the sense that it is difficult and costly to get into. Space is generally considered to start at the Karman line at 100 km altitude in the thermosphere (Fig. 1.1). The short trip of 100 km does represent a major challenge for rockets and the trip itself subjects the spacecraft to a totally different environment to that which it is subjected on the ground or in space.

A.R.L. Tatnall (✉)
University of Southhampton, Southampton SO17 1BJ, UK
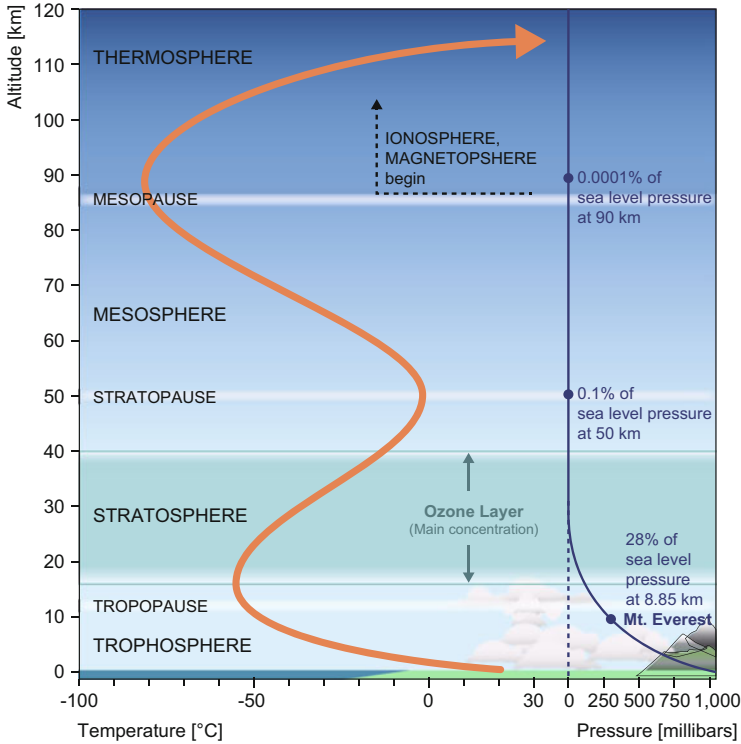e-mail: art4@soton.ac.uk

Fig. 1.1 Earth's atmosphere (adapted from Encyclopaedia Britanica Inc.)

1.1.2   Launch Vehicle

1.1.2.1   Acoustic/Vibration levels

Anybody who has witnessed a launch will testify to the noise levels that are produced. At the moment of launch the rocket motor firing and the exhaust products reflected from the ground produce a peak in the acoustic/vibration environment. As the rocket ascends the ground contribution decreases but other mechanical moving parts and unsteady aerodynamic phenomena continue to excite the structure. This excitation of the structure produces a secondary acoustic field within the structure. As the speed of the rocket increases a further secondary peak in the acoustic field occurs during the transonic flight which typically occurs just below Mach 1, the speed of sound. The overall levels experienced within the fairing of the Ariane V and Falcon 9 rocket are shown in Fig. 1.2. Acoustic noise affects lightweight structures, and antenna parabolic reflectors, solar arrays and spacecraft panels are particularly vulnerable.
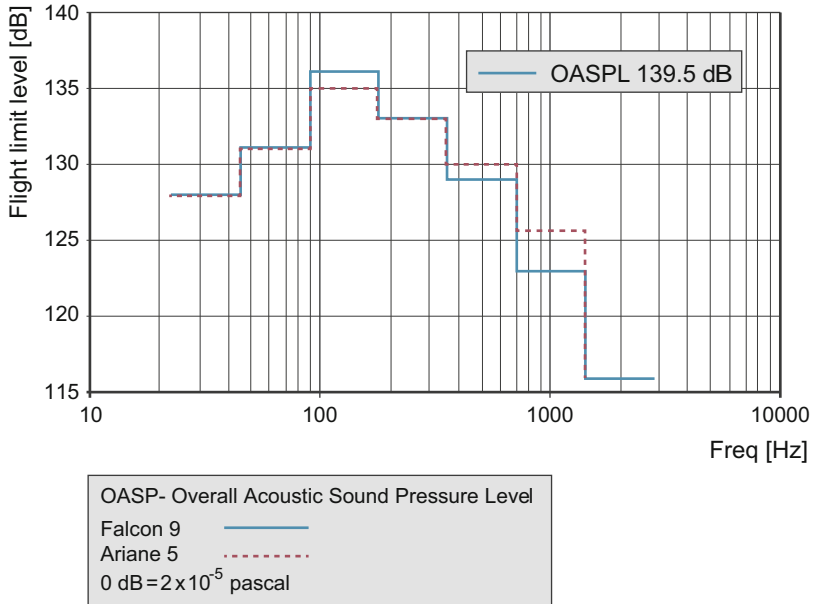
Fig. 1.2 Acoustic noise levels (data taken from Ariane 5 users manual 2011, Arianespace and Falcon 9 Launch Vehicle Payload User's Guide 2009, Space X)
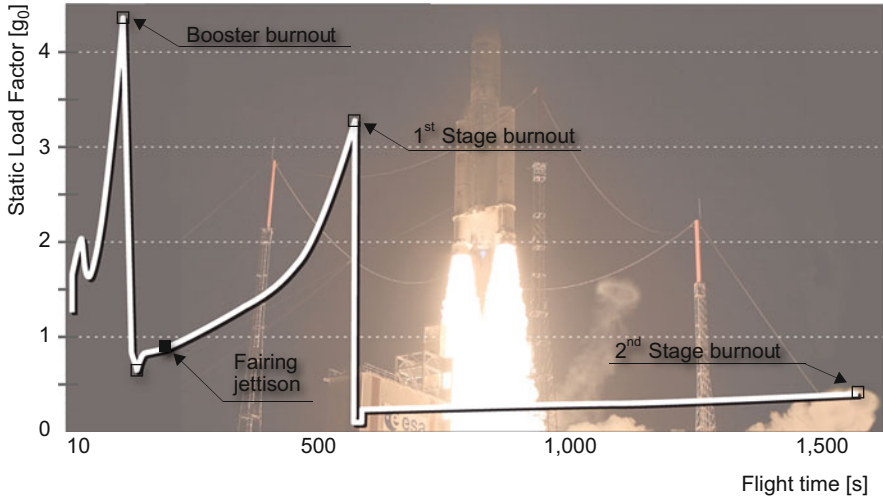
1.1.2.2 Static Acceleration

At the moment of launch the rocket has its highest mass and the acceleration is correspondingly low as the thrust produced is virtually constant. As the propellant is used the rocket acceleration increases until the solid rocket booster flame out and separation occurs. This gives the distinctive launch static acceleration profile shown in Fig. 1.3. Since the accelerations vary with time, the effect on the spacecraft is to generate quasi-static loads. These loads determine the major load bearing parts of the spacecraft structure such as the central thrust tube.

1.1.2.3 Mechanical Shock

A number of events can lead to very high accelerations being produced for very short periods of time. These shocks include

- Ignition and separation of the launch vehicle stages
- Fairing jettison
- Spacecraft separation
- Docking and landing

Fig. 1.3 Ariane 5 static acceleration profile

On Ariane and Falcon 9 the peak excitation occurs in the range 1–10 KHz and is 2,000 g_0 for Ariane 5 and 3,000 g_0 for Falcon 9. Despite these very high figures the transient nature of these loads means that they do not usually affect structural strength but they are of concern to the functioning of equipment such as relays.

1.1.3   Spacecraft Operational Environment

1.1.3.1   Vacuum

By the time a spacecraft reaches the low earth orbit at 300 km, the ambient pressure is as low as that which could be achieved by a very good vacuum chamber (about 10^{-7} Pa) on the earth and by the time a spacecraft has reached 800 km the pressure is so low that it cannot be reproduced on the ground. It is therefore important that materials which do not outgas are used in the construction of a spacecraft. Outgassing occurs because the material itself sublimates, gases are released from cracked materials, or gases that are adsorbed by the surfaces are released in a near vacuum. Whilst this will probably not cause a problem to the structural integrity of the spacecraft, it might result in a change to the surface properties and there is always the possibility that the vaporised material will condense if it impinges on colder spacecraft surfaces. It is therefore important that materials such as cadmium, zinc, PVC and many plastics with high vapour pressures are avoided. Adsorbed gases can be responsible for changing the properties of materials. Graphite is a solid lubricant commonly used on Earth but in space the adsorbed water vapour is lost

and graphite is ineffective as a lubricant. Alternatives such as molybdenum disulphide have to be used. If a material affected by a vacuum has to be used, residual contaminants can be removed by baking and the application of protective coatings or shielding.

1.1.3.2 Solar Radiation Flux

The spectrum of the radiation from the sun approximates to that of a black body at a temperature of 5,800 K. This is the temperature of the photosphere, the opaque region of the sun which is usually considered to be the surface of the sun. Our eyes have a response which is optimised for the light which is emitted by the sun which peaks at about 550 nm. This radiation is virtually constant and varies by less than 1 % from sunspot maximum to sunspot minimum although there are seasonal variations in the radiation incident on the earth as the earth moves in an elliptical orbit round the sun. Radiation we don't see, however, is much more variable. Rather than originating from the photosphere, the UV and X-rays originate from the outer regions of the sun, i.e. the chromosphere and the corona. This can be understood by noting that the temperature increases as the distance from the sun increases. By the time the corona is reached at a distance of about 2,500 km from the sun's surface, the temperature is about a million degrees and so hot that it emits radiation at X-ray wavelengths. The corona is highly variable in timescales of seconds to months and this is reflected in the variability of the UV and X-rays produced. Whilst the terrestrial weather is undoubtedly influenced by variations of the sun, as the overall variations in energy output are very small, it is not easy to distinguish these terrestrial variations from the much larger natural variability of our weather. The space weather, however, is dominated by variations in the sun output as this has a profound impact on the UV, X-rays and particles impacting the earth. UV radiation has a direct impact on the materials used in spacecraft, particularly the solar array. The absorption of UV by the cover glass used to protect the solar cell from particle radiation and the solar cell to slide adhesive can lead to darkening. This has a dual effect; it reduces the cell illumination and so reduces electrical power produced and it also raises the temperature of the cell which leads to a reduction in efficiency. Doping the cover glass with cerium oxide leads to absorption of UV and the prevention of darkening.

Additionally, the solar radiation flux is responsible for the radiation pressure created by absorbed or reflected photons. It is this source which provides the force for "solar sailing", a means of controlling or propelling a spacecraft. An experimental spacecraft of the Japan Aerospace Exploration Agency (JAXA) called IKAROS successfully flew a spacecraft to Venus using solar sailing (Osamu Mori et al. 2009).
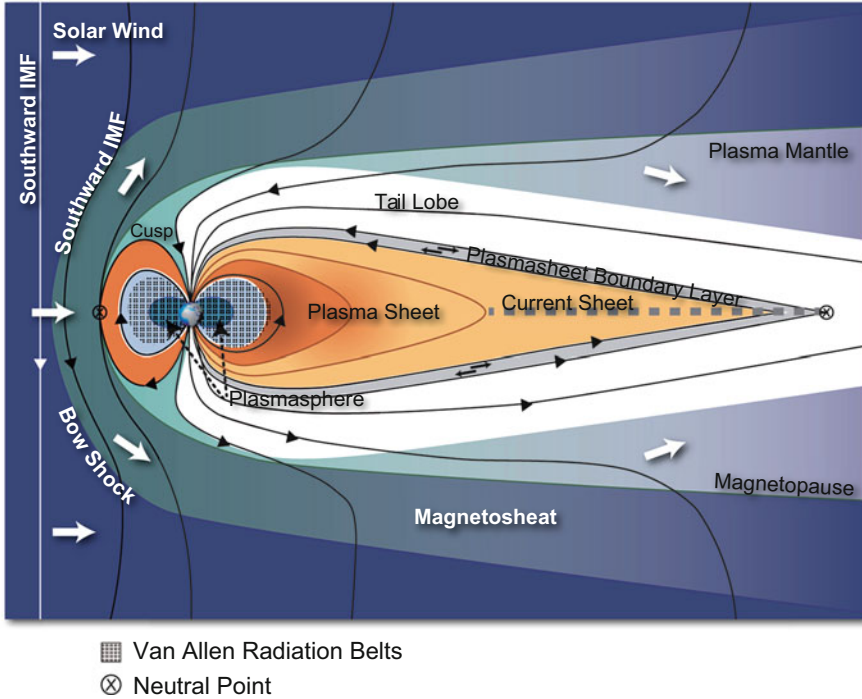
Fig. 1.4 Earth's magnetosphere (Reiff 1999)

1.1.3.3 Particle Radiation

There is a continual stream of high energy particles emitted by the sun. These particles are mainly protons and electrons with an energy of 1.5–10 keV. They are moving at speed of 400–800 km/s. This constitutes the solar wind that pervades our solar system and extends to at least 100 AU from the sun. Despite the high speed of the particles in the solar wind, their density is only 5–10 atoms/cm^3 rising to a few hundred atoms/cm^3 during a time of high solar activity. This provides a negligible pressure on any spacecraft that is impacted by the solar wind. A far greater pressure comes from the light pressure of photons as described above. Whilst the pressure of the solar wind is negligible, the consequences of the solar wind have to be considered, however, because it does have a major impact on the environment of the earth. The plasma of the solar wind interacts with the earth's dipole magnetic field to form the magnetosphere shown in Fig. 1.4. The magnetosphere's distinctive asymmetric shape is due to the pressure of the solar wind. On the sunward side, the magnetosphere extends out to a distance of approximately 10 Earth radii under quiet conditions, whilst in the anti-sun direction it extends several hundred Earth radii. The shape and extent of the magnetosphere depends on the strength and orientation of the magnetic field of the solar wind. This determines the reconnection
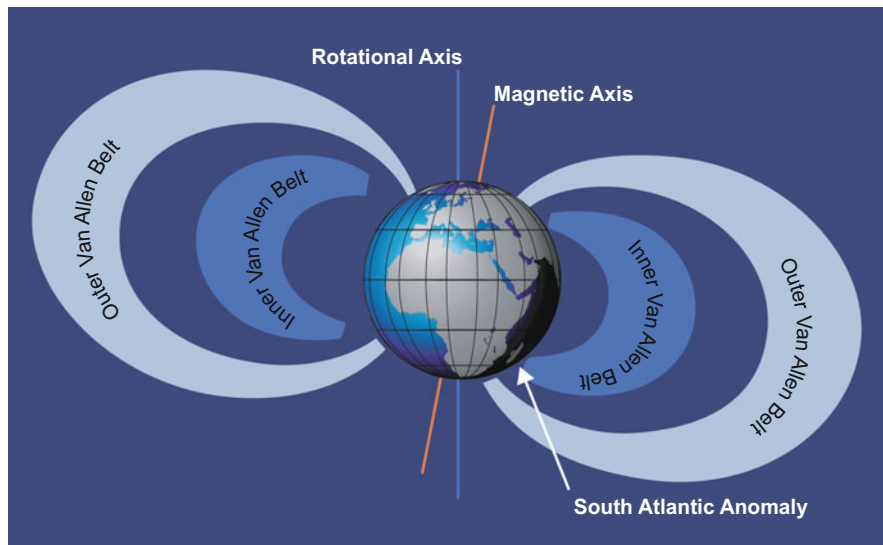
Fig. 1.5 Radiation belts inner and outer Van Allen radiation belts

process of the earth's and the sun's magnetic field that allows energy and momentum to be transferred from the solar wind into the magnetosphere. It may also be the acceleration mechanism for the very high energy particles that can be found in the radiation belts within the magnetosphere.

1.1.3.4 Radiation Belts

In 1958 the existence of a belt of trapped charged particles around the earth was confirmed by Explorer 1 and 3 using instrumentation designed by James Van Allen who had predicted that the belts would exist. The belt detected was the inner radiation belt. In the same year the Soviets—S. N. Vernov and A. E. Chudakov—discovered the outer radiation belt. These belts are doughnut shaped and extend from 1,000 km to 60,000 km above the earth. The outer belt is predominantly made up of electrons with a peak at 15,000–20,000 km, whereas the inner belt consists largely of high energy protons that peak at 3,000 km. Proton energies range from 0.01 to 400 MeV and electron energies from 0.4 to 4.5 MeV. They are shown as two distinct belts in Fig. 1.5 but in practice there is no real gap between the belts and they are highly variable depending on solar activity. The location of the radiation belts follows the magnetic field of the earth and this means that they are not symmetrically placed with respect to the earth. The axis of this field is offset and tilted with respect to the earth's rotation axis and so this leads to a location over the South Atlantic where the magnetic field is anomalously low. As a result the radiation belts are closer to the earth over this region which is commonly known

as the South Atlantic Anomaly. A satellite in a low earth orbit is more likely to encounter energetic particles and hence suffer damage in this part of the world.

A geomagnetic storm is caused by a solar wind shock wave interacting with the earth's magnetic field. This leads to measurable changes on the earth's surface of the earth's magnetic field. Accompanying these changes are increases in charged particles in the radiation belts. These particles are subject to the magnetic fields and perform three types of motions. All particles spiral around the field lines, move down field lines and bounce from one hemisphere to another and drift around the earth. This last motion eastward for electrons and westward for protons produces a current known as the ring current which can be measured by observing the associated magnetic field on the surface of the earth. It can lead to a decrease in the magnetic field measured at the earth's surface of >1 % during a major geomagnetic storm.

The origin of the particles in the radiation belts is solar, terrestrial or cosmic. Particles of solar origin are injected into the outer belts during magnetic storms. It is believed that the protons of the inner belt originate from the decay of neutrons produced when high-energy cosmic rays from outside the solar system collide with atoms and molecules of Earth's atmosphere.

Radiation effects include total dose effects, e.g. CMOS problems, lattice displacement damage which can damage solar cells and reduce amplifier gain, single event effects and additional noise in sensors and increased electrostatic charging. The charging of a spacecraft relative to the surrounding plasma is not so much of a problem as the possibility of increased discharges that can damage equipment and lead to the generation of electromagnetic interference. This has traditionally been thought to be more of a problem in GEO (Geostationary Earth Orbit) than LEO (Low Earth Orbit) where the plasma is low in energy and high in density, but nevertheless in LEO and particularly over the polar regions high levels of spacecraft surface charging can occur.

1.1.3.5   Atmosphere

When the sun is active and the amount of UV emitted increases, this is reflected in increased heating and expansion of the atmosphere so that the atmospheric drag on spacecraft increases also in region above the actual atmosphere. The change in height of the International Space Station (ISS) is shown in Fig. 1.6. Reductions in altitude caused by atmospheric drag are compensated by boosts using the station's thrusters. The number of boosts required depends on the atmospheric drag and the height variation permissible. For a spacecraft like GOCE which has to be maintained at a constant height and is at a very low altitude in order to measure small changes in the gravity field, this means thrusters have to be used for long periods of time. This is why electric propulsion thrusters are appropriate. Objects and spacecraft which are not controlled, such as debris, will lose height more quickly when the sun is active and the atmosphere has expanded.
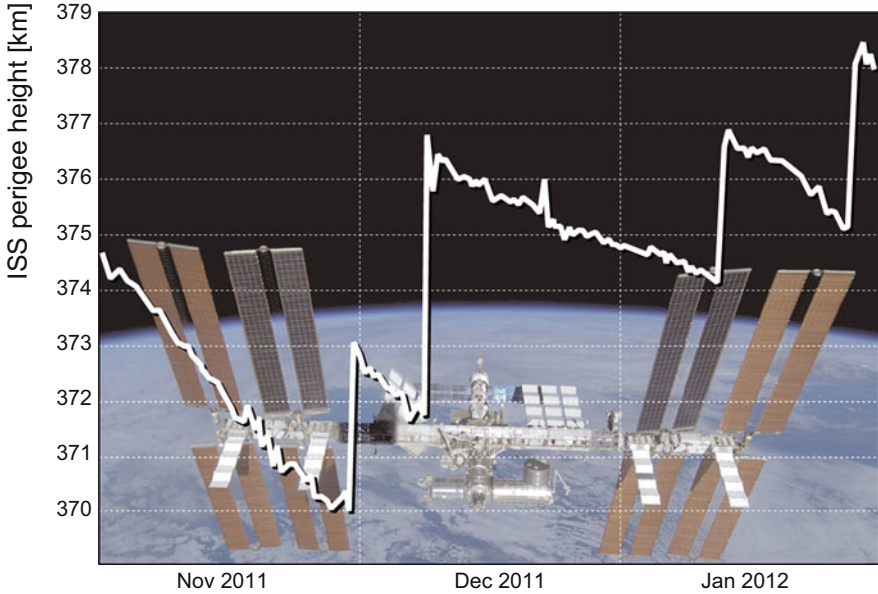
Fig. 1.6 ISS Perigee height change

The effect of solar activity on space debris is illustrated by Fig. 1.7 which shows that the reductions in debris occur at times of solar maximum when atmospheric drag is at its greatest, e.g. 1989. There have been recent suggestions based on the recent low activity of the sun that the sun is entering a new phase which will be characterised by low solar activity. This will have implications for satellites and debris. The effect of drag on a spacecraft provides a force give by

$$F_{D} = \frac{1}{2}\rho A C_{D} V_{r}^{2}\left(\frac{-V_{r}}{|V_{r}|}\right) \tag{1.1}$$

where V_{r} is the velocity vector relative to the atmosphere, ρ the atmospheric density, A the area of the vehicle perpendicular to the flight direction and C_{D} is the coefficient of drag which is typically ~2.5. Spacecraft with high area and low mass are particularly likely to be affected.

The lifetime of a spacecraft as a function of altitude and the mass area ratio m/A is shown in Fig. 1.8 for a representative atmosphere. In order to meet regulations imposed from January 2010 that all spacecraft in LEO must be deorbited within 25 years, this has resulted in a number of proposals to achieve this by deploying a structure that will greatly increase the area of the spacecraft and hence increase atmospheric drag.
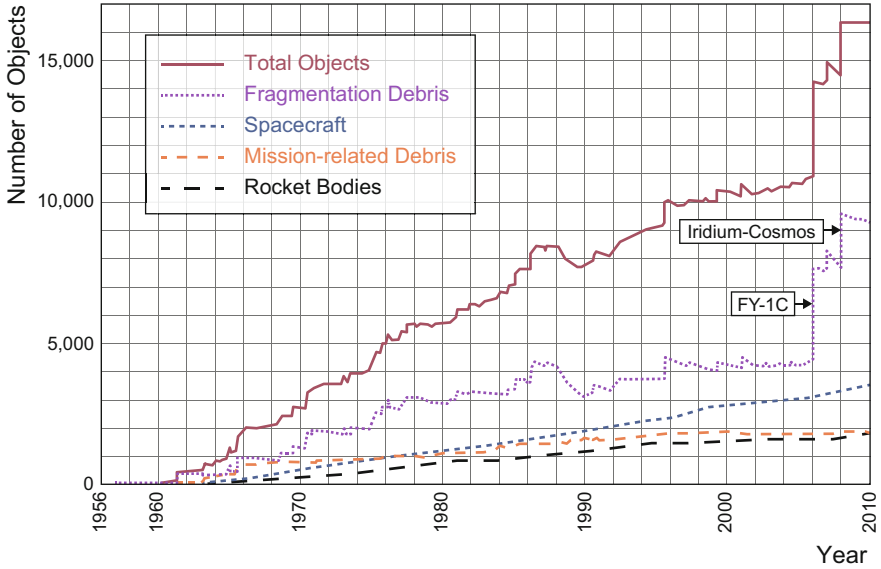
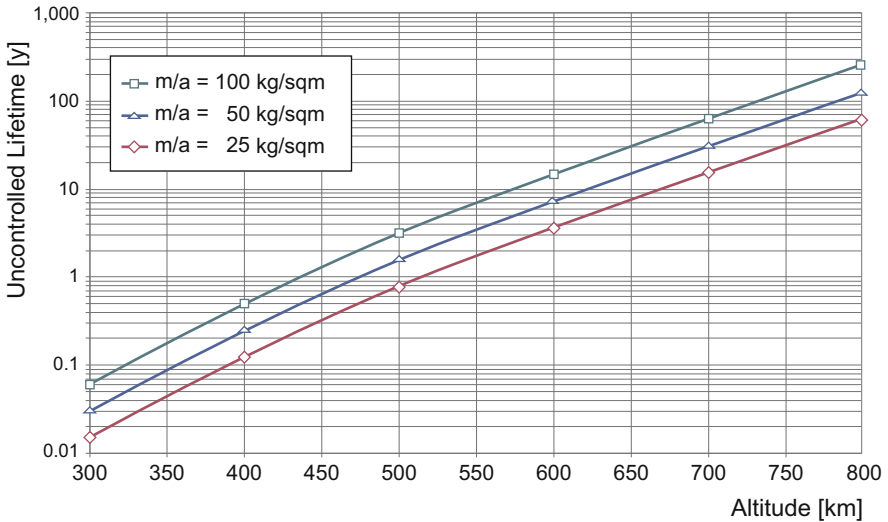**Fig. 1.7** Objects in the earth orbit by object type (NASA)

Fig. 1.8 Spacecraft lifetime as a function of altitude and mass/area ratio [m/a] (Kahn 2012)

1.1.3.6 Debris

Ever since the earth has been in existence it has been impacted by material. The mass flux of this material is currently about 10^7–10^9 kg/year. Much of this material is dust-sized objects called micrometeoroids that have a mass of less than 1 g. Their

velocities relative to spacecraft averages about 10 km/s and so whilst they are not likely to cause catastrophic damage to spacecraft they do contribute to the weathering process and can modify material properties. Objects larger than 1 g do exist and over the lifetime of the earth it has been hit by many objects of over 1 km in diameter. It is thought that 65 million years ago a 10 km meteorite hit the Yucatan peninsula in Mexico and produced a crater 180 km in diameter and probably caused the mass extinction of the dinosaurs. Of more concern to spacecraft is the increase in natural debris that occurs when the earth moves through the debris of a comet and the earth undergoes a meteor shower. The Olympus communications spacecraft was damaged by one of the Leonid meteoroids in 1993 and subsequently suffered an electrical failure.

Manmade debris is a growing problem as illustrated in Fig. 1.7. Whilst the problem has been gradually growing since man first started launching satellites, it has been exacerbated and highlighted by some recent events that have contributed to large amount of debris. One of these was the Chinese destruction of satellite FY-1C in an anti-satellite test in 2007 and another was the 2009 satellite collision between Iridium 33 and Kosmos-2251. The impact of both these events is shown in Fig. 1.7.

There are several contributors to the debris population:

- Launch and operational debris
- Space vehicle breakup (~215 events to date, 57 of them deliberate)
- Explosions
- Collision induced (5 to date—latest Iridium 33/Kosmos 2251, 10 Feb 2009)
- Upper stage breakup (largest contribution—Breeze M in 2007, 2010, 2011, 2012)
- Shedding of spacecraft surfaces (paint, MLI, etc.)
- Liquid metal coolant droplets
- Sodium–potassium (Na K) droplets from RORSAT reactor cores
- Solid propellant motor firings
- ASAT operations (Fengyun-1C, 11 Jan 2007; USA 193 21 Feb 2008)

The only current debris sink for the low earth orbit is the atmosphere, although there are a large number of innovative solutions being considered. These include electromagnetic methods, momentum exchange methods, remote methods, capture methods, and modification of material properties or change-of material state. If it is possible to remove debris and to enforce a requirement that satellites should have a lost mission lifetime of no more than 25 years, Fig. 1.9 illustrates that it could be possible to stabilise the debris environment. 90 % PMD (post-mission disposal) means that in 90 % of satellites the 25 year rule was implemented and ADR 2020/02 and ADR 2020/05 means that 2 and 5 objects per year respectively were removed after 2020 (ADR-active debris removal). Although the scenario with 90 % PMD and the removal of 5 objects per year suggest stabilisation, it does not take into account the possibility of unpredictable events such as the loss of Envisat, an 8 tonne Earth Observation satellite in April 2012. It is still in one piece but it is not controllable and constitutes a space debris threat as there is a distinct possibility

LEO Environment Projection (averages of 100 LEGEND MC runs)
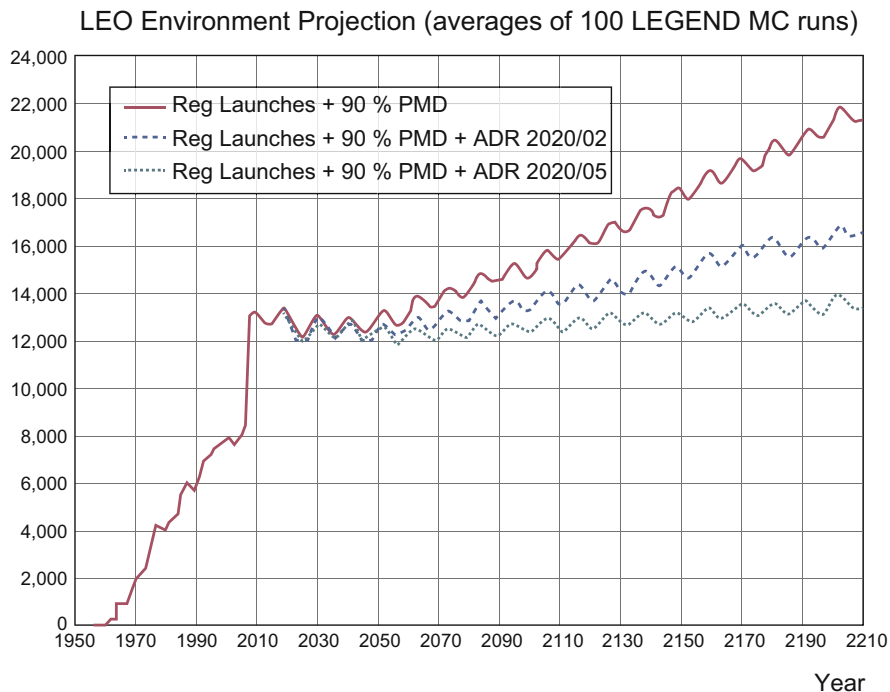
Fig. 1.9 Debris as a function of active debris removal and 25 year rule (Liou 2011)

that it will be struck by other debris. An analysis of space debris at Envisat's orbit suggests there is a 15–30 % chance of collision with another piece of junk during the 150 years it is thought Envisat could remain in orbit. Should this collision happen a very large debris cloud will be produced in a widely used region of space.

1.1.3.7   Gravity and Magnetic Fields

In addition to the environmental torques that can be provided by atmospheric drag and solar radiation, there are also gravity gradient torques and magnetic torques. The former are due to the differential gravity forces between the top and the bottom of the spacecraft and can be used to maintain a spacecraft earth pointing to about $\pm 5°$. Magnetic torques are caused by the earth's magnetic field acting on the residual magnetic dipole moment of the spacecraft. It can be utilised to provide a control torque by generating a controllable magnetic dipole moment on the spacecraft that interacts with the earth's magnetic field and generates a torque. In addition to being lightweight they require no expendable resources. They do require a significant external field and so can only be used for low earth orbiting missions.

1.2 Space Systems Engineering

Adrian R.L. Tatnall

1.2.1 Definition of System Engineering

System engineering requires skills that are traditionally associated with both art and science. Good system engineering requires the art of technical leadership including creativity, problem solving, knowledge and communication skills but it also requires the science of systems management or the application of a systematic disciplined approach. In this section the systematic disciplined approach balance is considered in more detail with the emphasis on the methodology of systems engineering but the main goal of system engineering is to get the right design. This can only be done using skills that cross traditional boundaries between the arts and the sciences.

The definition of systems engineering is an interdisciplinary approach governing the total technical effort to transform requirements into a system solution. The European Standard for Space System Engineering is described in the European Cooperation for Space Standardisation (ECSS) with the document number ECSS E ST 10C. The system can be any integrated product or processes that provide a capability to meet a stated objective. This inevitably means that a system can be a subsystem of a larger system and/or a system of systems. A spacecraft is a system but it is one element of the space mission that will include the launch vehicle and the ground segment and may include other systems such as GPS (Global Positioning System) and a data relay system. The ground segment itself is a combination of systems that is responsible for spacecraft operations and the processing of the data. It is therefore often necessary to consider products at a number of different levels.

The boundaries of the system engineering discipline and its relationship with production, operations, product assurance and management disciplines are given in Fig. 1.10 taken from the ECSS-E-ST-10C.

System engineering encompasses the following functions:

- Requirement engineering, which includes requirement analysis and validation, requirement allocation, and requirement maintenance.
- Analysis, which is performed for the purpose of resolving requirements conflicts, decomposing and allocating requirements during functional analysis, assessing system effectiveness (including analysing risk factors), complementing testing evaluation and providing trade studies for assessing effectiveness, risk, cost and planning.

A.R.L. Tatnall
University of Southhampton, Southampton SO17 1BJ, UK
e-mail: art4@soton.ac.uk
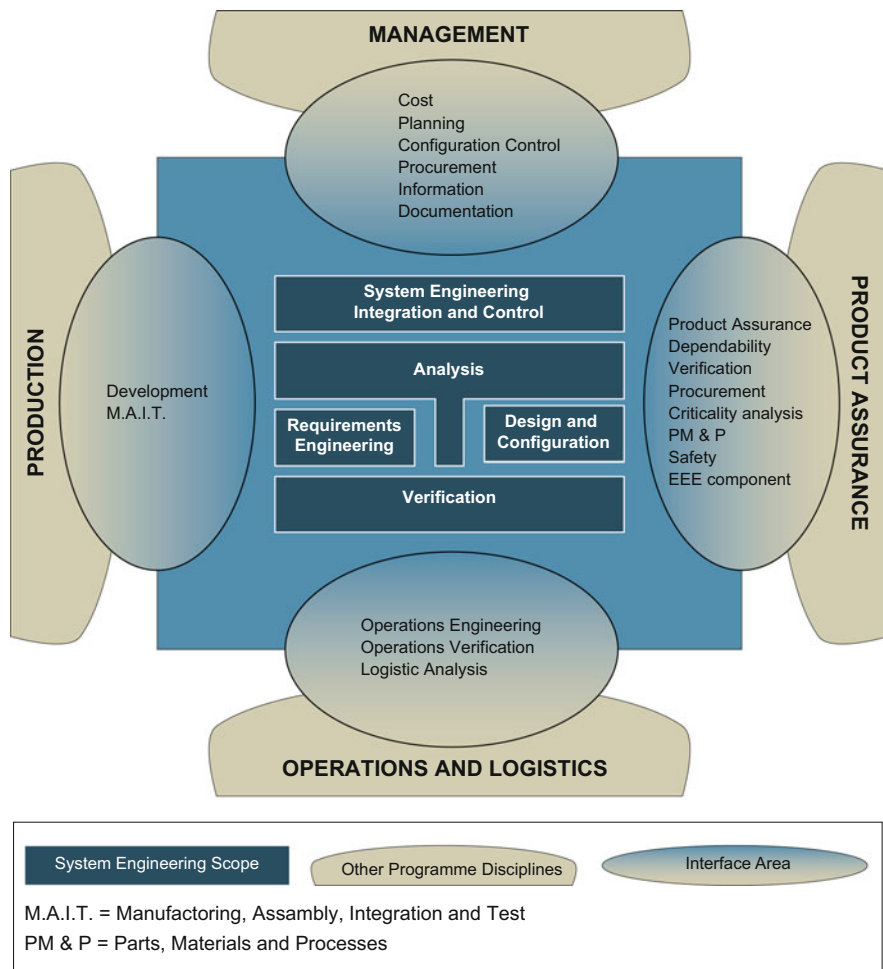
Fig. 1.10 System engineering boundaries (reproduced from ECSS E ST 10C, credit: ESA)

- Design and configuration which result in a physical architecture, and its complete system functional, physical and software characteristics.
- Verification, whose objective is to demonstrate that the deliverables conform to the specified requirements, including qualification and acceptance.
- System engineering integration and control, which ensures the integration of the various engineering disciplines and participants throughout all the project phases.

These functions require the techniques defined in Table 1.1 to be used.

Table 1.1 System engineering techniques (Fortescue et al. 2011)

Requirements identification/analysis	Concept selection
System specification	Budget allocation
Options identification	Performance analysis
Mission assessments	System optimisation
Trade-offs	Interface specification
Feasibility assessment	System definition
Cost comparison	Cost estimation

1.2.2 Objectives and Requirements

The starting point for the mission is the mission statement: a document established by the customer, which reflects the user needs. It is often a single line that describes the mission, e.g. John F. Kennedy in 1960 said that we would put a man on the moon by the end of the decade. The mission objectives are derived from this statement and qualitatively define what this mission should accomplish.

The mission requirements are the top level requirements on all aspects of the mission. They are usually quantitative in nature, specified by the customer or user, and they are an assessment of the performance required to meet the mission objectives. For the spacecraft system design these requirements are translated into engineering parameters. This translation can be complex, depending on the particular application. The requirements drive the rest of the design and determine all aspects of the mission. They are the single biggest cause of project problems.

For a communication spacecraft the translation between the user requirements and the engineering requirements is relatively straightforward since the user coverage and data requirements can readily be used to define the satellite parameters. On an Earth Observation and Science spacecraft, however, they can be considerably more complex. For example, on the GRACE (Gravity Recovery and Climate Experiment) mission (Wiese et al. 2012), the user requirements on geophysical parameters, such as an ice sheet changes into an instrument specification, have to be translated into measurements of the gravity field and ultimately to the measurements of changes in the speed and distance between two identical spacecraft. In this case the process involves assumptions about other related parameters such as the level of processing required and GPS data. At the start of the design process it may not be clear to what extent the requirements are driving the design and so it is an essential part of the space system engineering process that the requirements are re-evaluated when there is a clearer understanding of the impact they have on the spacecraft design.

This iterative process is essential to ensure that the most relevant and realistic requirements are used for the spacecraft design. There are plenty of examples where the engineering requirements have become "tablets of stone" at the start of the design and the overall system has suffered because of unwillingness to question them as the design has evolved. It is always necessary to define how much quality is needed, or how much "science" is enough in order to hold down mission costs and
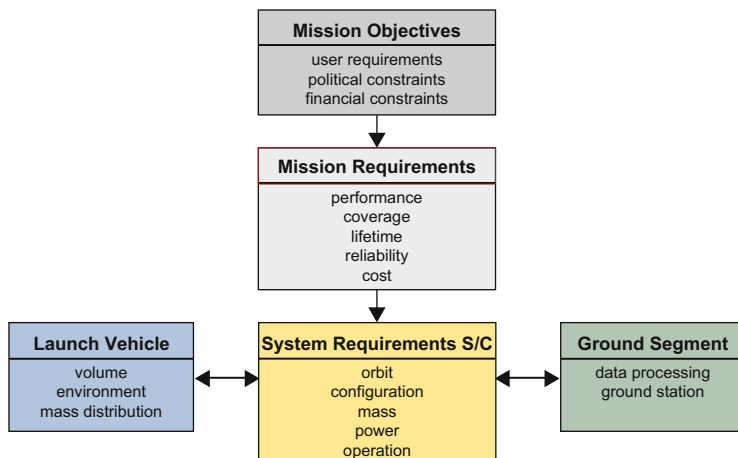
Fig. 1.11 Objectives and requirements of a space mission

avoid unnecessarily restrictive requirements. Whilst Augustine's law that "the last 10 % of performance generates one-third of the cost and two-thirds of the problems" is an oversimplification it does encapsulate the problem of overspecification. In other examples technological constraints such as the inability to space qualify critical parts or processes that may dictate a revision of requirements. The importance of the requirements should not be underestimated. Relatively little of the total project budget is spent on requirement analysis and initial design but it does determine the cost commitment for the rest of the programme. The later the change in a requirement the greater the cost impact on the programme as a whole.

Figure 1.11 shows how it is necessary to expand these top-level requirements into specifications covering the entire range of system and subsystem engineering parameters. It also shows the importance of establishing, in parallel, budget data.

Table 1.2 is a checklist of the full range of parameters that are likely to be specified in later, more detailed phases of a programme.

There are many systems options that have to be considered in the early design phase of a mission. These include the type of orbit, the launcher, the propulsion system, the type of spacecraft configuration and the attitude control concept.

The choice of orbit for an astronomy mission is a good example of the kind of choices that have to be made. This highlights some of the key points that must be taken into account in concept selection and optimisation.

Figure 1.12 is a tree diagram showing the possible orbits about Earth and Sun which could be adopted for an astronomy mission. The mission names of spacecraft flown or due to be flown for the different orbits are shown.

It is clear that the choice of orbit for this class of mission is determined by a large number of factors but there is often an overriding consideration. For example NASA's major observatories—Hubble Space Telescope (HST) and Gamma Ray Observatory (GRO)—had to be in a circular low earth orbit (LEO) in order that they

Table 1.2 Checklist of system requirements [Adapted from Fortescue et al. (2011)]

Mission requirements	Environmental requirements
Launch windows	Ground activities
Orbit (Transfer; Operating Nominal and Back-up)	Launch and ascent conditions
	Transfer and operating orbit environment (Reentry, Descent)[a]
Operations mass	Structural/thermal inputs, loads, ranges
Launch and early orbit phase	Environmental protection
Operational phase	Cleanliness/Contamination
End-of-life	Electromagnetic compatibility (EMC)
Lifetime	DC magnetic fields
(Retrieval/repair/re-supply)	Radiation
Autonomy	Spacecraft charge
Reliability/availability	Atomic oxygen[a]
Ground segment	Autonomy
Physical requirements	**PA requirements**
Axes definition	Reliability
Configuration constraints dimensions	Availability
Mass properties	Maintainability
Internal torques	Safety test philosophy
Disturbances	Parts, materials, processes
Power/energy	Cleanliness
	Storage, handling, transport
	Configuration management
	Software
Performance requirements	**AIV programme requirements**
Orbit maintenance	Schedule
Ranging accuracy	Model philosophy
Timing accuracy	Safety test philosophy
Pointing accuracy	Ground segment equipment requirements
Measurement accuracy	Facilities usage
Stability	
Pointing range	
Slew rate	
Data rate	
Data storage capacity	
Onboard processing	
Link budget margins	
Telemetry/telecommands	
Strength/stiffness	
Thermal control	
Reliability	
Cost constraints	

[a]For some missions only

could be launched/serviced by Space Transportation System (STS)/Shuttle and Tracking and Data Relay Satellite System (TDRSS) could be used for data retrieval. As far as the science is concerned these orbits are far from ideal. They suffer from regular eclipse periods and the scope for uninterrupted observation is very limited. Without the constraints of a Shuttle launch, two of ESA's astronomy missions,
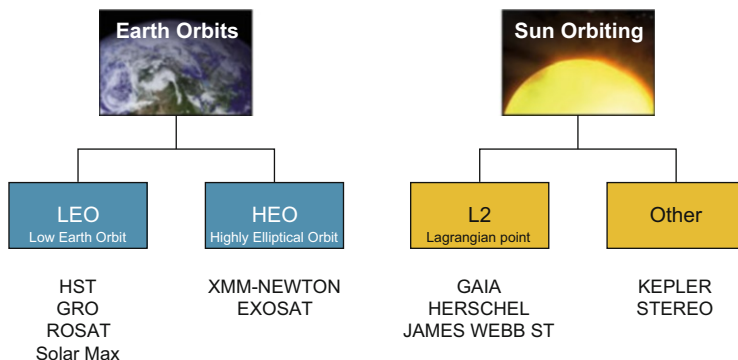
Fig. 1.12 Orbit options for astronomy missions

Integral and the X-ray Multi-Mirror Mission (XMM-Newton), selected highly elliptical orbits (HEO) which can provide long periods of uninterrupted observation away from trapped radiation in the earth's proton and electron belts. More recent missions, such as GAIA, HERSCHEL and First/Planck, have selected orbits around a spot about 1.5 million km from Earth in the direction away from Sun known as the L2 Lagrangian point. In this orbit advantage can be taken of the fact that the benign thermal and radiation environments are ideal for long-distance observations. In addition, by careful choice of the particular orbit around the L2 point, it is possible to have continuous solar power and a continuous communications link. Other spacecraft, such as NASA's Kepler spacecraft, are in orbits around Sun trailing Earth so that a star field can be observed continuously for several years. The importance of the various factors varies with each mission and the current technology.

It is now a common feature of spacecraft that they reuse existing designs of spacecraft equipment. This can offer very significant saving compared to new developments, e.g. the satellite bus used for Venus Express was almost a copy of that used for Mars Express and this in turn was based on the Rosetta bus. The reuse of existing designs and hardware must be treated with caution. Qualification by similarity is a legitimate process but there have been notable failures in the past that have been down to this approach. Examples include the first Ariane V failure because of software inherited form Ariane IV and the loss of the Mars Observer because of the over reliance on hardware qualified for near Earth missions.

## 1.2.3 Design Drivers and Trade-offs

The purpose of the satellite bus to is to provide the support required for the payload to ensure that it can operate in the required orbit and environment. This makes the payload in most cases, the single most significant driver of the satellite design.

Fig. 1.13 GOCE spacecraft (Credit: ESA)

Power, heating and cooling, structure, power and communication are all provided to ensure that the payload can operate satisfactorily and relay its data back to ground. The propulsion, Attitude and Orbit Control Subsystems (AOCS) and the mission analysis provide the means of getting the payload into the right position to make its measurements. In the case of GOCE, the Gravitational Ocean Composition Explorer, shown in Fig. 1.13, the spacecraft has to fly at a constant, very low altitude of 260 km in order to measure very small changes in the gravity field (Wiese et al. 2012). The effect of the residual atmosphere is very significant and so a main design driver is to minimise air drag forces and torques. As a consequence the satellite body has an octagonal prism shape with two long, fixed solar array wings fitting the launcher fairing dynamic envelope. This requires triple-junction GaAs solar cell technology to generate the maximum power. It also requires an electric propulsion system to ensure the orbit altitude is maintained with the most efficient use of the propellant.

Whilst there may well be a key technological design driver, in a typical space mission there are a number of factors that need to be considered to determine the optimum mission. A trade-off study is an objective comparison with respect to a number of different criteria and is particularly useful if there are a number of possible design solutions. It is common to make use of trade-off tables to "score" the alternative options in early concept studies. Major evaluation criteria for such trade-offs include:
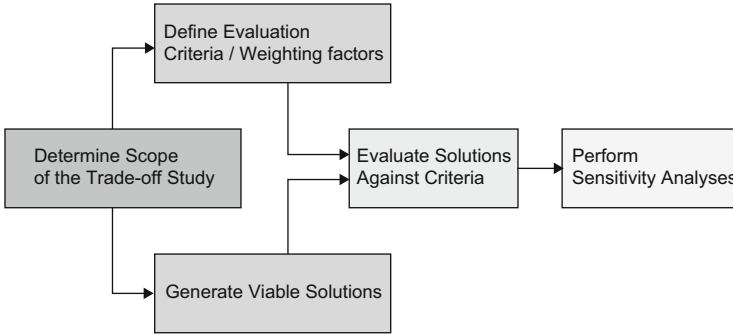
Fig. 1.14 Trade-off process

- Cost, this is generally a dominant factor
- Satisfaction of performance requirements (for example image quality in an astronomy mission)
- Accommodation of physical characteristics, notably mass, size and power which, in turn, impact on cost and feasibility
- Availability of suitable hardware technology and timescales for any predevelopment
- Compatibility with launcher, ground segment and other system elements and the complexity of interfaces
- Flexibility to encompass alternative mission options
- Reliability and availability

Evaluation criteria should be selected that discriminate between the options. If some of these criteria are considered more important than others then a weighted trade-off can be performed. The process is shown in Fig. 1.14 adapted from the National Airspace System (NAS) system engineering manual (NAS 2006). Regardless of whether a trade-off is weighted or not it should only be used as a guide. It is impossible to guarantee that a trade-off is entirely objective and that the evaluation criteria are exhaustive and independent. Cost, for example, is influenced by all the criteria above and its use as an independent parameter is highly questionable. Numeric results are useful but may well give a false sense of accuracy and so should be used carefully.

Whilst some factors can be evaluated numerically many other factors that need to be considered rely on engineering judgement. In addition, quantitative values attributed to factors, can often not be made with sufficient confidence to allow a particular solution to be selected from a number of options. In this case there are often a number of viable solutions possible.

1.2.4 Concurrent Engineering

Concurrent engineering (CE) is a relatively new design tool developed to optimise engineering design cycles. It relies on the principles that all elements of a products lifecycle should be taken into account in the early design phase and that the design activities required should occur at the same time or concurrently. Whilst system engineering has always recognised the value of this approach, the enabling factor for the CE approach has been the rapid development of information technology (IT). Concurrent engineering has enabled design iterations to be performed much quicker and it has enabled the designer to be more closely involved in the design process. ESA's concurrent engineering facility at ESTEC (Netherlands) has achieved the following:

- Studies have been performed in 3–6 weeks rather than 6–9 months
- Cost has been reduced by a factor of 2
- Overall improvement in the quality of the studies by providing consistent and complete mission designs

There are many concurrent facilities around the world and they have become an integral part of the early design phase of a space mission.

1.3 Fundamentals of Space Communications

Felix Huber

1.3.1 Introduction

Radio communication with a spacecraft has to deal with the fact that there are large distances between transmitter and receiver, possible low elevation angles (Fig. 1.15) resulting in a substantial attenuation by the atmosphere and large Doppler shifts due to the orbital velocity of the satellite. Moreover, the ionosphere reflects or absorbs certain frequencies that are thus unusable for space communications.
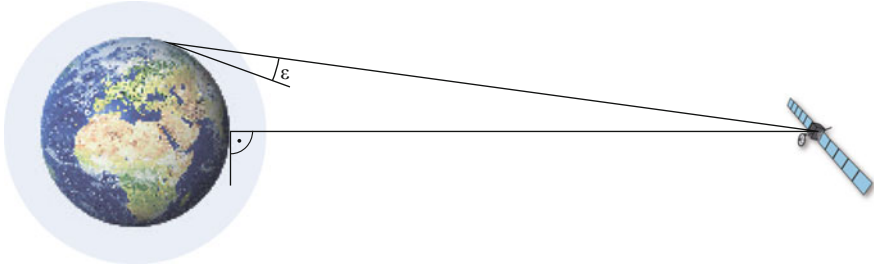
Two aspects of communications have to be considered:

- Baseband—user aspect
- Carrier—service aspect

———————————

F. Huber
German Aerospace Center (DLR), German Space Operations Center (GSOC),
Oberpfaffenhofen, 82234 Wessling, Germany

Fig. 1.15 Communication between ground and space at different elevations ε

Both aspects can be handled more or less separately.

1.3.2 Baseband

Signal sources can be discrete values such as switch on–off or pressure and temperature values on the satellite side ("telemetry") or telecommands in case of a ground station. The path of the signals is shown in Fig. 1.16. Non-digital signals have to be converted into a serial digital signal first in a process called source coding that is described in the next section.

1.3.2.1 Source Coding

Sensors convert physical properties such as pressure or temperature into a normalised electrical voltage for example 5 V. Next, the voltage is sampled at discrete time intervals (sampling) and converted into a binary number by an analog-to-digital converter ADC (discretisation) (Fig. 1.17). If the signal is to be recoverable without losses, it has to be sampled at a speed twice as fast as its bandwidth (not highest frequency!): this is called the Nyquist theorem. The number of steps that the ADC can create (quantisation) has an influence on the rounding errors that occur when the nearest value has to be chosen. This quantisation noise can be made smaller with smaller steps at the cost of a higher data rate that is needed for transmission, thus a trade-off has to be found. The resulting binary numbers are transmitted as a stream of binary pulses, referred to as pulse code modulation (PCM).

If the signal has a bandwidth higher than what the Nyquist theorem allows for, it has to be filtered before being fed to the ADC or frequencies outside the allowed bandwidth will be mapped onto the desired range. This phenomenon is called aliasing and causes a heavy distortion of the signal.
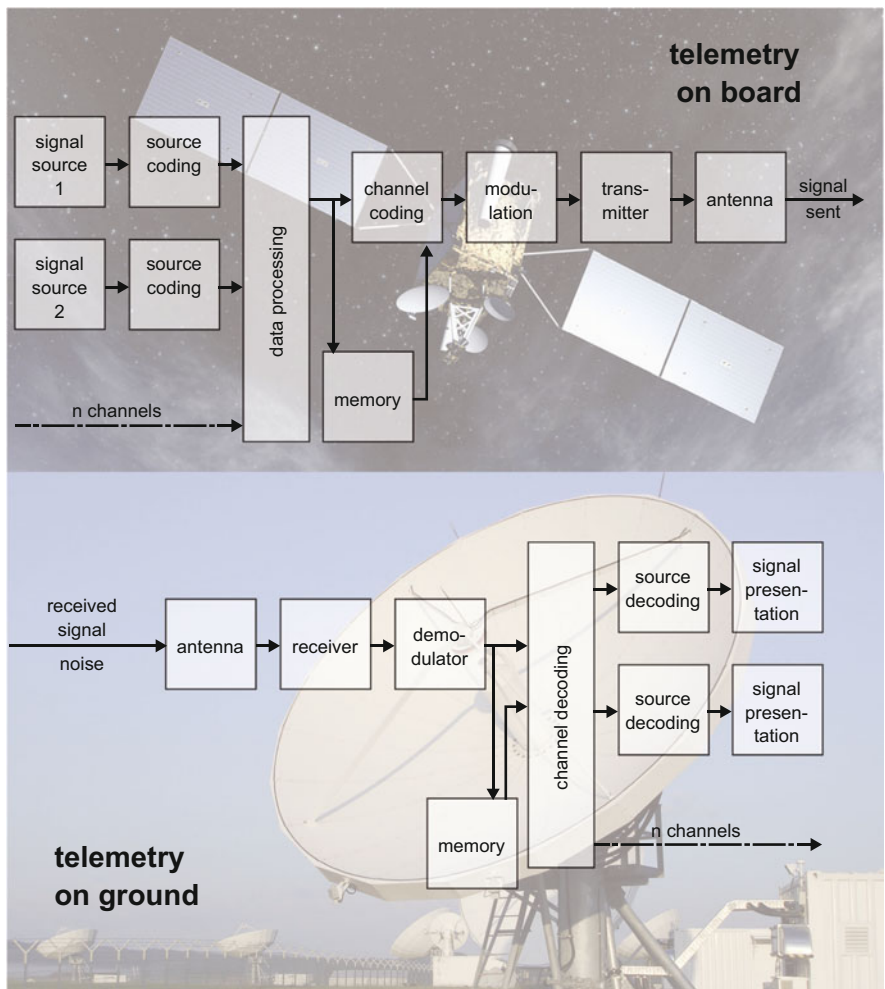
Fig. 1.16 Telemetry signal processing

In a next step, various sources have to be combined (Multiplexing), formatted (range indication, sequential numbering) and perhaps stored for later transmission.

1.3.2.2 Channel Coding

Before the digital data can be sent over the air, precautions have to be taken for errors that can occur during the transmission. In this process of channel coding check sums are added to the data packets (CRC cyclic redundancy check, "inner checksum") and the resulting serial data stream is run through a convolutional coder
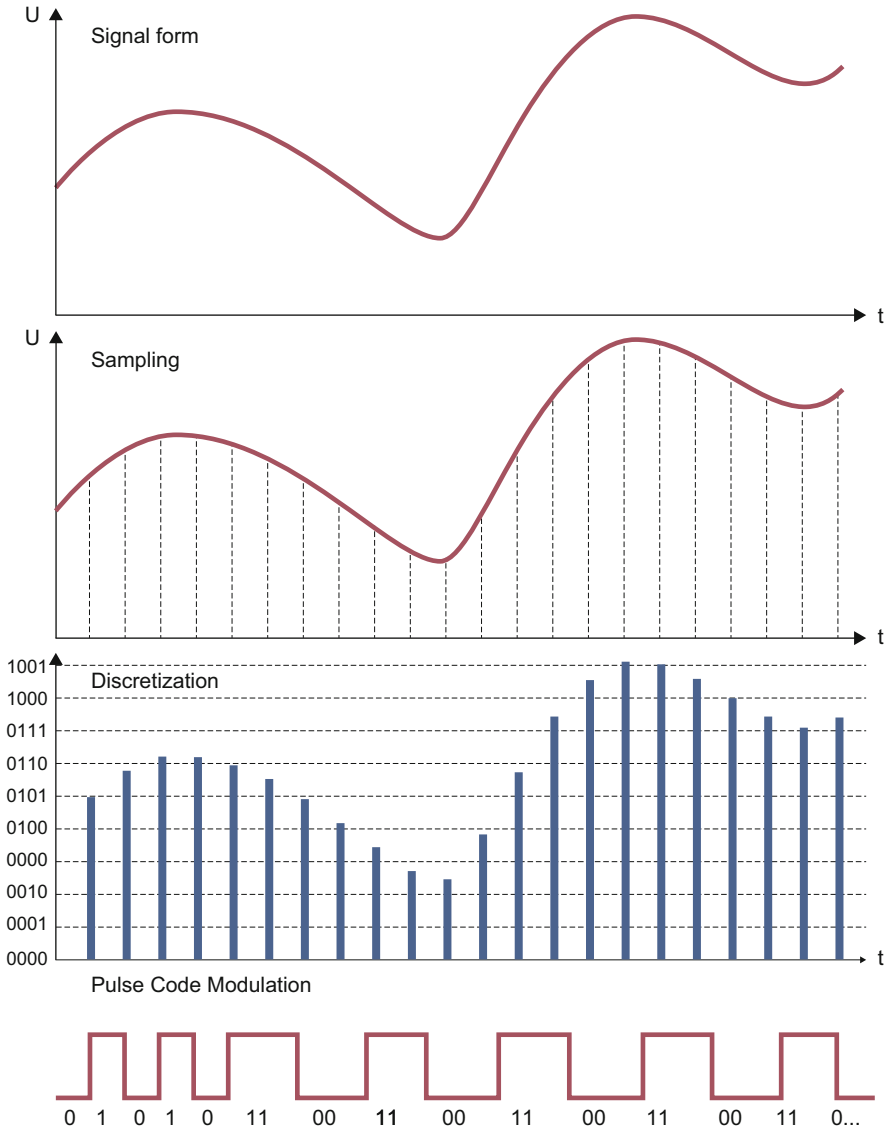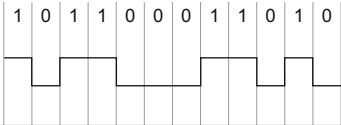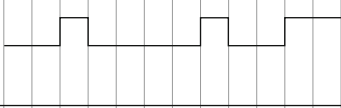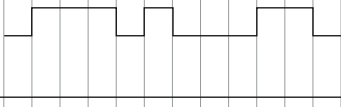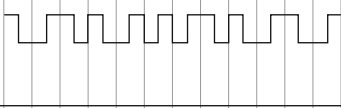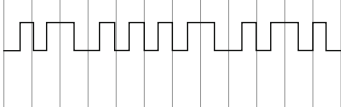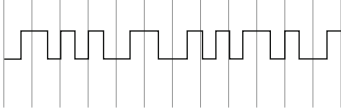
Fig. 1.17 Source coding of a signal

("outer checksum") that adds more check bits in order to recover the distorted bits later after reception. While the CRC only allows for the detection of bit errors, the convolutional coding has enough information to correct bit errors without requesting a retransmission of the data, hence the name "forward error correction" (FEC). This error correction capability is achieved at the cost of a lower net bit rate,

Code	Logic Wave-form Levels	Code Waveforms	Code Definitions

Non Return to Zero - Level
1 „ONE" is represented by one level
0 „ZERO" is represented by the other level

Non Return to Zero - Mark
1 „ONE" is represented by a change in level
0 „ZERO" is represented by NO change in level

Non Return to Zero - Space
1 „ONE" is represented by NO change in level
0 „ZERO" is represented by a change in level

Bi-Phase - Level [1]
1 „ONE" is represented by a „ONE" level with transition to the „ZERO" level
0 „ZERO" is represented by „ZERO" level with transition to the „ONE"level

Bi-Phase - Mark [1]
1 „ONE" is represented by NO level change at the beginning of the bit period
0 „ZERO" is represented by a level change at the beginning of the bit period

Bi-Phase - Space [1]
1 „ONE" is represented by a level change at the beginning of the bit period
0 „ZERO" is represented by NO level change at the beginning of the bit period

1 The Bi Φ codes may be derived from the corresponding NRZ codes by inverting the level for the last half of each bit interval.

Fig. 1.18 Channel coding according to IRIG 106

referred to rate ½, rate ¾, etc. encoding. It should be noted that the bit error rate could also be lowered by using a smaller bit rate, but the FEC achieves a better lowering of bit errors when compared to the net bit rate; this is called the coding gain.

The final PCM is sent out as a sequence of equally length pulses. However, this sequence of pulses, called non-return to zero (NRZ), can create a DC offset of the average voltage fed to the transmitter that cannot be handled by the system (Fig. 1.18). Therefore, the PCM code has to be converted into a DC-free signal, for example by a bi-phase coding: The signal is multiplied with a square wave carrier of half the bit rate which removes the DC offset, however, at the cost of a higher bandwidth.

It should also be noted that there are two definitions of bi-phase depending on whether a real multiplication is used as an exclusive or logic with the latter being the inverted signal of the first method.

Another critical effect is that if there is an imbalance of zeroes and ones over a certain period of time, a temporary DC offset is generated and the centre frequency will shift causing signal losses due to the bandpass filtering at the receiver. A self-synchronising scrambler is therefore used to smear out periodic patterns that can occur in the data stream. This process, which is also referred to as energy dispersal, creates a uniform spectrum by toggling the PCM bits with a pseudo random number pattern using linear feedback shift registers (LFRS) that have mathematical properties of almost pure randomness. Since the mathematical law of the random numbers is known, the receiver can undo the process of randomising and recover the original bits.

Hence, the transmitted serial signal has no block structure anymore and therefore synchronisation markers have to be added so that the receiver can determine the start of a frame. These synchronisation words are known as Barker codes and have a pattern that has a low cross-correlation since the Barker data pattern could also occur anywhere in the data stream and should not trigger the frame detection. Since the frame length is fixed and known to the receiver, it can check for the regular appearance of the Barker codes and determine the start of a frame.

1.3.2.3   Baseband Shaping

The rectangular pulses occupy a large bandwidth due to their steep edges. If this spectrum is bandwidth limited due to filtering in the signal path, the shape of the pulse gets distorted and spreads over its bit cell time into adjacent cells causing bit errors. This phenomenon is called Inter Symbol Interference (ISI). In order to prevent ISI, the signal would have to be filtered with a bandpass filter with a brickwall characteristic. However, such an ideal filter would have an impulse response that spreads over $\pm$ infinity with a non-causal behaviour and cannot be reached in reality. A more practical approach uses the shape of a raised cosine as the filter transfer function which has also in infinite pulse response, but the corresponding $\sin(x)/x$ shape decays faster at the cost of twice the bandwidth. In practical implementations, a linear mixture of both extremes is used, described by a roll-off factor α, where $\alpha = 0$ corresponds to the rectangular filter transfer function and $\alpha = 1$ to the raised cosine shape with an occupied bandwith of $(1 + \alpha) \cdot$ symbolrate. Typical implementations use α between 0.2 and 0.5.

This filtering can be performed at the analog base band signal where the raised cosine shaping has to be approximated by real circuits or in the numerical domain where the pre-calculated impulse responses are superimposed over several bit cells. Since we have a non-causal filter, the output of the bits has to be delayed using a history shift register in order to make signal causal again. This superimposing of the bits is spread over ± 4 to ± 8 bits, meaning the ideal impulse response is cut off after a certain time, leading to a negligible distortion of the signal.

The resulting signal has zero crossings independently of α always after the bit cell time T, meaning that there is no ISI (see Fig. 1.19). Since we have a symmetric filter, the resulting pulses are also symmetric. A superposition of randomly selected
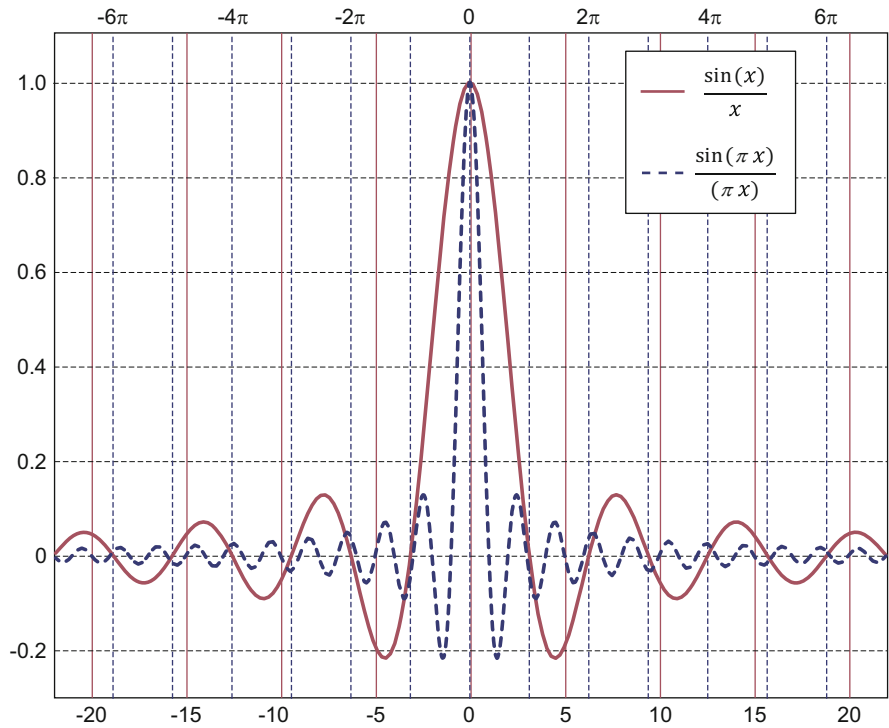
Fig. 1.19 Impulse response of a raised cosine filter

bits leads to a pattern that has the shape of an eye, hence the name eye pattern. It can be used to judge the quality of the received signal: the eye pattern has to be wide open in the centre where the detection of the bits takes place. No zero crossings should occur in the centre; this indicates ISI (see Fig. 1.20).

One more optimisation can be performed in order to maximise the signal-to-noise ratio: The transmitter filter and the receiver filter should have the same conjugate complex transfer function. Since the raised cosine function is symmetric, these transfer functions are the same. However, the eye pattern requires the raised cosine shape at the bit detection in order to avoid ISI; therefore, the filtering is shared between the transmitter and receiver by using the square root of the raised cosine. Such a root-raise-cosine filter system is optimal for both ISI and noise.

1.3.2.4   Modulation

Modulation is the process of applying a (coded) signal onto a higher frequency carrier. A radio frequency signal has the form:
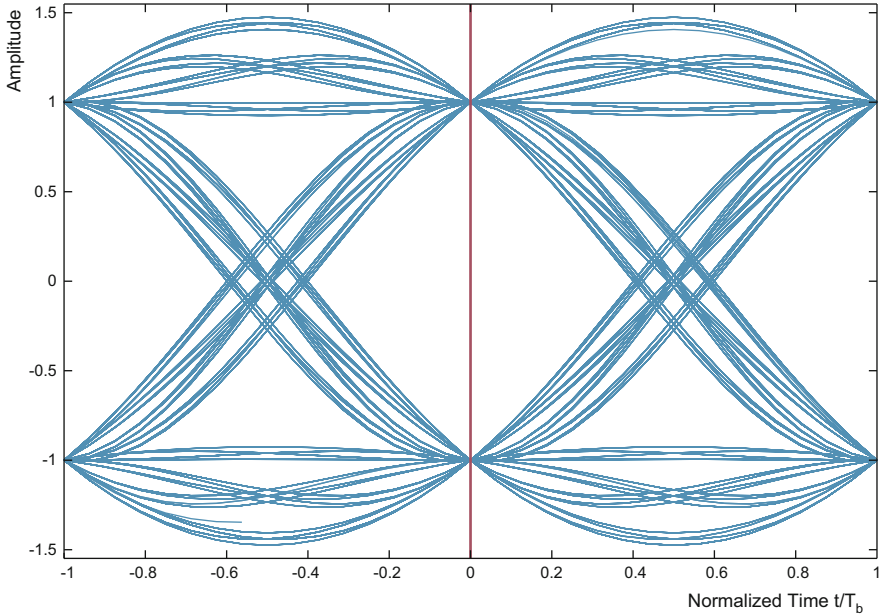
Fig. 1.20 Eye pattern of shaped signal

$$U(t) = A_C \cos{(\omega_C t + \varphi_C)} \tag{1.2}$$

with the three parameters amplitude A_C, frequency ω_C and phase φ_C that can be influenced by the modulating base band signal. If the base band signal is of analogue type, these changes are named amplitude modulation, frequency modulation and phase modulation, respectively.

The modulation of the carrier converts its single frequency into a band of frequencies that is at least twice as wide as the modulating signal (AM) or even more (wide band FM) (Fig. 1.21).

In the case of a digital PCM signal that has only discrete values, the modulated signal also only takes on discrete values. In this case one speaks of "Keying" instead of modulation (Fig. 1.22) as in the beginning of radio communications the Morse code was generated by pressing the transmission key (beep beep beep, beeeeep, beep beep ...). One should note the phase jump after the first bit of FSK. These jumps cause side lobes in the spectra and should be avoided by proper design of the frequency switching circuit.

AM ⇨ Amplitude shift keying ASK
FM ⇨ Frequency shift keying FSK
PM ⇨ Phase shift keying PSK

In the case of PSK, several options are possible, depending on the number of phase values that the signal takes. In the case of only two (0°, 180°) it is called phase

Modulation signal
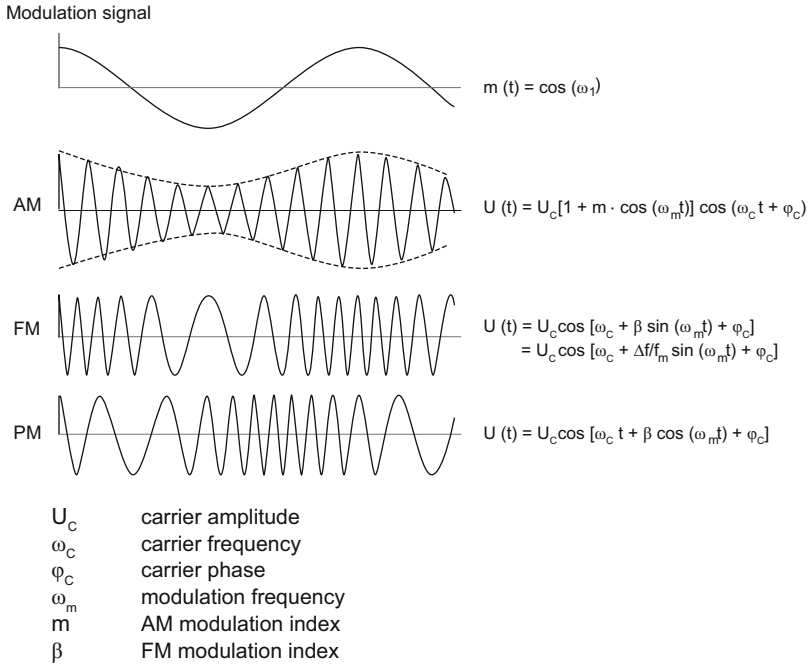
$m(t) = \cos(\omega_1)$

AM $U(t) = U_C[1 + m \cdot \cos(\omega_m t)] \cos(\omega_C t + \varphi_C)$

FM $U(t) = U_C \cos[\omega_C + \beta \sin(\omega_m t) + \varphi_C]$
      $= U_C \cos[\omega_C + \Delta f/f_m \sin(\omega_m t) + \varphi_C]$

PM $U(t) = U_C \cos[\omega_C t + \beta \cos(\omega_m t) + \varphi_C]$

U_C carrier amplitude
ω_C carrier frequency
φ_C carrier phase
ω_m modulation frequency
m AM modulation index
β FM modulation index

Fig. 1.21 Analog modulation forms

Digital Signal | 1 | 0 | 0 | 1 |

ASK

FSK

PSK

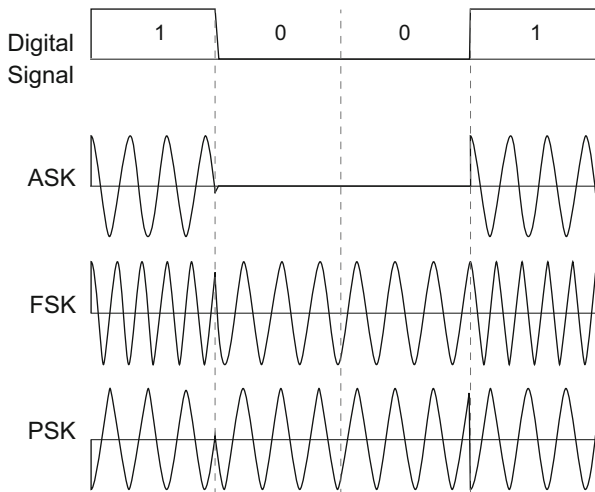**Fig. 1.22** Keying of a digital signal

Table 1.3 Space communications frequencies

Band	Frequency range [MHz]	Application
VHF	150	Voice
P-Band/UHF (lower part)	300–3,000	Military satellites
L-Band/UHF (upper part)	1,215–1,850	GNSS, Satellite telephony
S-Band	2,025–2,400	TT&C
C-Band	3,400–6,725	Future LEOP TT&C
X-Band	7,025–8,500	Payload, Deep space
Ku-Band	10,700–14,500	TV, routine TT&C
Ka-Band	18,000–35,000	Relay
V-Band	37,500–50,200	Intersatellite links

reversal keying or binary PSK (BPSK), with four values it is called quadrature PSK (QPSK), since two carriers (sine and cosine) are used. A combination of AM and PSK is also possible called Quadrature Amplitude Modulation QAM that allows for more bits to be transmitted within a symbol at the expense of a higher noise sensitivity. All *n*-ary PSK modulation schemes suffer from a *n*-ary ambiguity that has to be resolved by the data synchronisation mechanism or by using differential phase encoding.

PSK is used in space communications for its noise immunity and better bits/ energy ratio despite its high effort in the electronics.

1.3.3 Carrier

The space communication frequencies are shown in Table 1.3.

1.3.3.1 Elements of a Space Link

Power Amplifier

Transmits the signal with an average power P_T. Peak power levels can cause a distortion of the signal that has to be accounted for by lowering the input signal (input back-off).

Antenna

Directs the signal into the desired direction by the use of dipoles, horns and reflectors. Since the dimension of the antenna is in the order of the radio signal's wavelength, a Fresnel diffraction occurs that creates a main beam and unwanted side lobes. These side lobes direct energy to unwanted areas in the case of a transmission and collect additional noise in the case of a receiving antenna (Fig. 1.23).
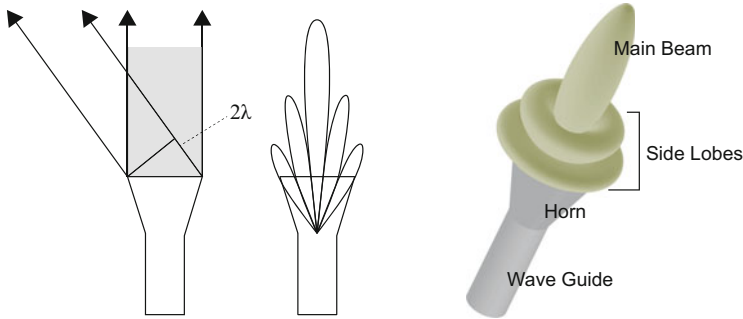
Fig. 1.23 Antenna side lobes

The angle at which the power is at a level of half the maximum value (-3 dB) is called half power beam width and is given approximately by $70° \, \lambda/D$ (in degrees), D: Aperture diameter. The maximum value of the main beam as compared to a theoretical point-like isotropic radiator is called the antenna directivity and its practical value including the efficiency η called "antenna gain" is given by:

$$G = \frac{4\pi A}{\lambda^2} \cdot \eta = \frac{4\pi A_{\text{eff}}}{\lambda^2} \tag{1.3}$$

A: (effective) Aperture area. In the case of a dipole, an effective aperture area can be defined as an area perpendicular to the electrical field lines that still has an influence on the field by "capturing" the filed lines onto its surface. The gain can be seen as the solid angle into which the antenna concentrates the signal compared to a full solid angle.

The product of transmitter power PT and antenna gain G is called EIRP (equivalent isotropic radiated power) and is the power that an isotropic transmitter would have to transmit in order to create the same power flux density at the receiver.

In the case of a receiving antenna, the effective "capture" area for the incoming signal can be calculated from the above equation if the antenna gain is known.
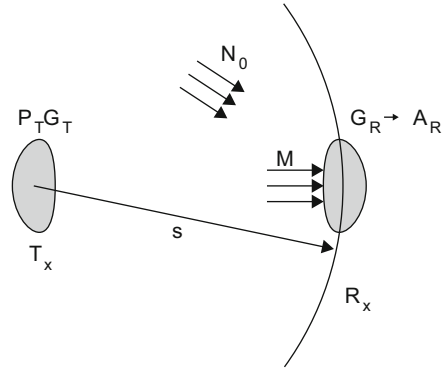
Noise

All warm bodies transmit thermal electromagnetic noise according to Planck's equation. These sources can be Sun, Moon and Earth's ground, atmosphere and clouds, but also galactic sources. In the case of the radio frequency bands, the spectral density of this noise is constant and given by

$$N_0 = kT_S \tag{1.4}$$

where k is Boltzmann's constant and T_S is the system noise temperature as the sum

of all natural noises as seen by the antenna and additional artificial signals such as other transmitters or devices (human made noise). Since this noise is purely stochastic, it cannot be removed from the received signal and this limits the sensitivity of the system.

Receiver

The receiver itself creates additional noise in its amplifiers which further decrease the system sensitivity. In a properly designed system, only the first (low-noise) amplifier contributes substantially to the system noise.

1.3.3.2 Link Budget Equation

The performance of the radio link is given by the ratio of the wanted signal (Carrier) and unwanted signals (noise) that has to have a certain value in order to recover the data bits error free (Fig. 1.24).

The transmitter of power P_T and antenna gain G_T is assumed to be in the centre of a sphere of radius s. P_T is the isotropic radiated power and thus uniformly illuminates the sphere's surface $4\pi s^2$. If the transmit antenna has a gain G_T, the power flux density at the receiver is therefore:

$$M = \frac{P_T \cdot G_T}{4\pi s^2} = \frac{\text{EIRP}}{4\pi s^2} . \tag{1.5}$$

If the receiving antenna has a gain G_R, its effective aperture area is

$$A_R = \frac{\lambda^2 G_R}{4\pi} \tag{1.6}$$

and it "captures" at total carrier power C of $M \cdot A_R$.

$$C = M \cdot A_R = \frac{P_T \cdot G_T}{4\pi s^2} \cdot \frac{\lambda^2 G_R}{4\pi}. \tag{1.7}$$

Reordering and adding an additional attenuation L_A due to rain, snow, etc. leads to:

$$C = P_T G_T \left(\frac{\lambda}{4\pi s}\right)^2 L_A G_R = P_T G_T L_s L_A G_R \tag{1.8}$$

where L_S is called free space loss even though no energy is lost but only diluted over a growing sphere surface area. The reordering of the terms was only done in order to match their position to the transmission path: transmitter, space and receiver.

If the bit rate of the signal is R, the time per bit is $1/R$ and the received energy per bit is C/R. Thus, we finally have the sought ratio of bit energy versus noise power density:

$$\frac{E_b}{N_0} = \frac{P_T G_T L_s L_A G_R}{k T_s R}. \tag{1.9}$$

This Link Budget Equation is usually given in a logarithmic scale using the pseudo unit deci-Bel dB:

$$\frac{E_b}{N_0} = \text{EIRP} + L_S + L_A + \frac{G_R}{T_S} + 228.6 - 10\log R \tag{1.10}$$

where 228.6 is the logarithm of Boltzmann's constant and G_R/T_S, the so-called figure of merit, describes the quality of the receiving system. It describes the radio link on an overall power level but does not take into account the type of modulation and nature of the noise coming from other possible transmitters.

Depending on the modulation and coding used, the required E_b/N_0 varies from 1–2 dB for turbo coded PSK to 8–10 dB for uncoded FSK. An additional 3 dB are needed in the case of a non-coherent demodulation.

In the case of orthogonal signals, other transmitters would not even affect the bit detection even though they bring noise power into the receiver as long as they don't saturate the amplifiers. This is a permanent source of faulty system design.

References

References to Sect. 1.1

Kahn, M., ESA, *private communication*, 2012

Liou J.C. "An active debris removal parametric study for LEO environment remediation", Advances in Space Research Volume 47 issue 11, Pages 1865–1876, June 1st, 2011

Mori, O., Sawada, H., Funase, R., Morimoto, M., Endo, T., Yamamoto, T., Tsuda, Y., Kawakatsu, Y., Kawaguchi, J., "First Solar Power Sail Demonstration by IKAROS" International Symposium on Space Technology and Science, July 9th, 2009

ozone layer: layers of the atmosphere. [Art]. Encyclopædia Britannica Online. Retrieved 05 July, 2014, from http://www.britannica.com/EBchecked/media/99826/The-layers-of-Earths-atmosphere

Reiff P.H. "Live from the Sun" part of the "Passport to Knowledge" series, 1999

References to Sect. 1.2

ECSS-E-ST-10C Space Engineering-Space engineering System engineering general requirements

Fortescue, P., Swinerd, G., Stark, J. (editors), "Spacecraft Systems Engineering", Fourth Edition, Wiley, 2011

National Airspace System (NAS) System Engineering Manual version 3.1, Federal Aviation Administration ATO Operations Planning, June 2006

Wiese, D. N., Nerem, R. S., and Lemoine, F. G.: "Design considerations for a dedicated gravity recovery satellite mission consisting of two pairs of satellites", Journal of Geodesy, 86:81–98, doi:10.1007/s00190-011-0493-8, 2012