MINISTRY OF DEFENCE

The UK Military Space Primer

Development, Concepts and Doctrine Centre

THE UK MILITARY SPACE PRIMER

June 2010

AUTHORISATION

The Development, Concepts and Doctrine Centre (DCDC) is responsible for providing foundational material for UK Joint Military Education. The UK Military Space Primer is a stand-alone publication intended to satisfy that remit regarding military uses of Space. Comments on factual accuracy or proposals for amendment are welcomed by the Doctrine Editor at:

The Development, Concepts and Doctrine Centre
Ministry of Defence
Shrivenham
SWINDON, Wiltshire, SN6 8RF

Telephone number:	01793 314216/7
Military Network:	96161 4216/4217
Facsimile number:	01793 314232
Military Network:	96161 4232
E-mail:	publications@dcdc.org.uk

DISTRIBUTION

Distribution of DCDC Publications is managed by the Forms and Publications Section, DSDA Operations Centre, C16 Site, Ploughley Road, Arncott, Bicester, OX25 1LP. Requests for issue of this publication, or amendments to its distribution, should be referred to the DSDA Operations Centre. Other DCDC publications, including a regularly updated CD *Joint Doctrine Disk*, can also be demanded from the DSDA Operations Centre.

DSDA Help Desk:	01869 256052
Military Network:	94240 2052

All DCDC publications (including drafts) are available to view and download on the Defence Intranet (RLI) at: www.dcdc.dii.r.mod.uk

Unclassified DCDC publications are available to view and download at: http://www.mod.uk/dcdc

Copies of The UK Military Space Primer may be purchased by members of the public. All enquiries should be addressed to:

DSDA Forms and Publications (Sales Dept.) at: desdsda-formsandpubs@mod.uk

Telephone Number: 01869 256139

Front Cover: The Illustration on the front cover shows the UK on 23 May 2001, taken as part of the Sea Wide Field of View (SeaWiFS) experiment carried on the SeaStar/OrbView2 satellite launched in 1997. The satellite was launched by a Pegasus air-launched rocket, as described in Paragraph 166. Image courtesy of NASA/Goddard Space Flight Center.

FOREWORD

Astute students of military affairs cannot have missed the emerging realisation that modern military operations are inextricably dependent on Space as a domain and on the capability that resides there. From this reality comes the need to incorporate Space training in contemporary professional military education. This document fills the void in UK military publications on the basic principles of Space and thus contributes to the educational effort. It demonstrates the links, which at times can appear tenuous, between the theoretical constraints of orbits, launch sites and spacecraft design and the practical application of military capability from and through Space.

The book was written at the UK Development, Concepts and Doctrine Centre during 2008-10. Its aim is to demonstrate not only what the underlying principles of Space Operations are, but also how those principles have direct consequences for routine operations. Furthermore, it demonstrates that theory can be reduced to digestible chunks that do not require pages of mathematics and physics to understand. For those who need such complexity, references are provided, though such readers are probably already aware of many of them. For the rest, the intention is to have a publication that is comprehensible, and which will de-mystify an area all too often cloaked in jargon and impenetrable acronyms.

The extensive cross-referencing is deliberate, to reinforce the links between principle and practice. Equally, the index should serve to direct the otherwise busy reader to the information that they seek. Any reader with particular suggestions for improvements, corrections or comments is both welcome and indeed encouraged, to make their views known by contacting the Developments, Concepts and Doctrine Centre at Shrivenham. Contact details are at Page ii.

Few readers will need to digest this book in its entirety in one session; rather it is designed to provide a reference text that can be consulted whenever necessary. It also provides a springboard for further study by those whose interests lie in one particular aspect of capability. It will support formal courses throughout Defence. Additionally it should provide background or support for those who have not had the opportunity for such training, but who nonetheless recognise the importance of Space as a military domain. Finally, it may also contribute to public understanding of the importance of military Space capability. Whichever category of reader you think you are, I commend it to you.

MP Colley

ACDS DC&D

(INTENTIONALLY BLANK)

CONTENTS

Page

Chapter 1 – Space Fundamentals

Section I – The Space Environment

The Boundary Between Air and Space

Environmental Threats to a Spacecraft in Flight

Section II – Geometry, Physics and Their Implications for Orbits

From Orbit to Ground Track

Necessary Complications

Section III – Military Implications

Generic Considerations

Characteristics of Particular Orbits

Section IV – Practical Space Vehicles

Launch Strategies

Manoeuvre in Orbit

Satellite End-of-Life Issues

Characteristics of Spacecraft

Section V – An Integrated Space Capability

Chapter 2 – Space and the Law

Chapter 3 – Military Uses of Space

Section I – Fires Through Space – Ballistic Missiles and Space Weapons

Missiles

Space Weapons

Section II(A) – Data Networks in Space - Communications

Satellite Communications – SATCOM Theory and History

Communications and the RF Spectrum

Practical SATCOM

Section II(B) – Data Networks in Space – Position, Navigation and Timing

Section III – Surveillance from Space

Reconnaissance Principles

Imaging – Visual and IR Surveillance

Other Forms of Surveillance

Surveillance and Reconnaissance from Space – Special Factors

Section IV – Surveillance of Space

Section V – Space Control

Space Control and Space Situational Awareness

Defensive Space Control

Offensive Space Control, Ballistic Missile Defence and Anti-Satellite Systems

Ballistic Missile Defence

Anti-Satellite Systems

Chapter 4 – Space and Society

Section I – Thinking Strategically About Space

Section II – Dependencies on Space

Section III – Commercial Uses of Space

Section IV – Scientific Uses of Space

Section V – Military Space Doctrine and Organisation

Section VI – Civil Space Capability and Governance

Annexes

Annex A Conic Sections

Annex B Brahe, Kepler and Newton

Annex C Full Orbital Elements

Annex D 'Many-Body' Problems and Lagrangian Points

Dynamics of the Many-body Problem

Applications

Annex E Rocket Engine Fuel Cycles

Annex F The Electromagnetic Spectrum

Electromagnetic Radiation – An Introduction

Varieties of EM Radiation

Properties of EM Radiation

Generating EM Radiation

Other EM Radiation Phenomena

Annex G RF Frequency Bands

Annex H GPS

Bibliography

Acknowledgements

Lexicon

Index

(INTENTIONALLY BLANK)

CHAPTER 1 – SPACE FUNDAMENTALS

101. **Understanding Space**. This Chapter looks at gaining and maintaining access to space and, as such says relatively little about what can be done *from* space. Such discussion, except where needed to explain a principle, is reserved for Chapter 3. To the extent that use of space is now so pervasive as to be transparent to many end users, it is true that the reader does not require detailed understanding of orbits or launch strategies to look at a satellite photograph, or use satellite communications. However, all useful applications in space are enabled or constrained by orbital dynamics and the engineering challenges posed by the space environment. Anyone who wishes to become an intelligent customer for the military use of space, whether engaged in planning, current operations or capability development, will require some understanding of these factors.

102. **Cross-referencing**. The chapters in Space Primer are rigorously cross-referenced, so the reader can start in Chapter 3 and refer back to this Chapter where required. There is also a full index at the back of the Primer. Equally, though, the flow of the explanations may be aided by reading Chapter 1 in sequence.

SECTION I – THE SPACE ENVIRONMENT

In this section you will find a discussion of:

- The boundary between conventional air platforms and space vehicles and how that affects their operation.
- The unique physical characteristics of space and the influence that has on the design and operation of space platforms.
- Some of the threats this environment poses to military operations, and their potential mitigation.

The Boundary between Air and Space

103. **From Air to Space – Aerodynamics and Ballistics**. From the earliest days of aviation, people have known that the higher you go above the Earth's surface, the thinner the air becomes. For conventional aircraft, this confers both advantages and limits. Drag becomes less, or (more usefully) you can travel faster over the surface of the Earth for a given amount of drag, but thinner air also makes aerodynamic control harder and ultimately limits the power output of air breathing engines. Control may well become the limiting factor. Once an aircraft reaches air too thin for safe control, it has passed beyond the realm of aerodynamics and is entering the realm of astronautics – if it has not yet reached space, it is certainly leaving air.

104. **Near-space and the Karman Line**. Where is the boundary between air and space?

a. **The Edge of the Atmosphere**. There is no clear natural boundary between the atmosphere and space. The atmosphere becomes gradually thinner all the way up, and even when it has become too thin to sustain life, thrust or lift, it is still measurably present. The practical thrust and lift limits for aviation have increased with time – modern airliners carry passengers at altitudes and speeds beyond the reach of military and research aircraft of yesteryear – but it is still broadly true that access becomes harder and more expensive with altitude. In rough terms, airliners typically cruise between 30 000 ft and 40 000 ft, high-performance combat aircraft might reach to between 50 000ft and 60 000ft, or even a little more, but very few aircraft have gone much above this. 100 000 ft would be a practical limit for conventional aviation.

b. **The Karman Line**. Although aviation is thus limited to below 100 000 ft, at that point the atmosphere is still far too thick for a spacecraft to achieve orbit without continuous thrust. Any declared boundary is to an extent arbitrary, but the Federation Aeronautique Internationale (FAI) has proposed a limit called the Karman Line – an altitude of 100 km – as a working boundary.[1] Below this altitude, a spacecraft would need to use thrust regularly to maintain orbit, and would face severe problems with frictional heating.[2]

c. **Near Space**. The gap between the upper practical limits of air operations, and this bottom limit of true space is sometimes referred to as **near space**. Innovative platforms might be able to exploit this domain for military use,[3] but this theme is not explored further in this publication.

Feet, Miles and Kilometres

- Aircraft altitudes are typically quoted in feet (ft), but spacecraft orbits are usually described in miles (nautical or statute) or kilometres (km). This disguises where the boundaries really are.
- Where no distinction is made, and particularly in American references, ‘miles’ are usually statute miles (sm), and miles per hour (mph) are used in preference to knots.
- There are 5280 ft in a sm. There are 6076 ft in a nautical mile (nm). For rough arithmetic, 6000 ft is a good approximation for a nm.
- An airliner cruising at 36,000 ft (a realistic figure) is therefore about 11 km above the Earth.
- A satellite in low-Earth orbit at 370 km is at about 1,200,000 ft – roughly 33 times higher.
- Geostationary orbit (where the period of orbit matches the revolution of the Earth) is at about 35,800 km – 100 times higher than a typical low-earth orbit, and over 3000 times higher than the airliner.
- The Karman Line (100 km) is at 328,000 ft or 62 sm, or about 54 nm.
- The radius of the Earth is about 6400 km or 4000 sm, or 3400 nm.
- This Primer uses kilometres (km) as its standard unit of measurement wherever possible. The reference table below may provide useful comparisons between km and the various imperial measurements that may still be encountered:

---

[1] During the 1960s, the United States Air Force (USAF) used a definition of 80 km to qualify personnel as astronauts.

[2] Even above the Karman line, re-boosting is required to maintain orbital altitude, but at or below it, the re-boosting would be so regular that it would be impractical with any currently conceivable satellite and thrust system.

[3] Near space has been little used by manned platforms. In the UK, QinetiQ have researched very-high altitude, long-endurance solar powered Unmanned Aerial Vehicles (UAVs), though as wing-borne platforms, they will always have relatively low altitude limits, and there have also been theoretical studies of airships operating in near-space.

Nautical Miles (NM)	Statute Miles (SM)	Kilometres (km)	Feet (approx)	Remarks
1	1.15	1.85	6080	Conversion factors
0.87	1	1.61	5280	
0.54	0.62	1	3281	
43.2	49.7	80	262 500	Early US altitude definition for Astronauts' wings
47.5	54.7	88	288 700	V2 rocket maximum altitude (operational trajectory)
54	62	100	328 000	Karman Line (FAI proposed definition of the boundary of space)
81	93	150	492 100	Lowest known sustained orbital altitude (Soviet 'Zenit' series)
185	213	342.6	1.12 million	International Space Station typical altitude (lower range of Low Earth Orbit)
302	348	560	1.8 million	Hubble Space Telescope altitude
3440	3959	6371	21 million	Mean radius of the Earth (add to altitudes to derive radii from the centre of the Earth)
10 900	12 550	20 200	66.3 million	Altitude of semi-synchronous orbit
19 320	22 240	35 786	117.5 million	Altitude of geo-synchronous orbit

Illustration 1.1 – Operating in Near Space. Between 1959 and 1968, 3 North American X-15 research aircraft (the picture shows the third aircraft) explored the boundaries between air and space flight. The aircraft was carried aloft by a B-52 launch platform, and then released to make a rocket-powered flight culminating in a glide landing on the dry lake-bed at Edwards Air Force Base (AFB), California. During 199 separate flights, the aircraft reached a maximum speed of Mach 6.72 (4520 mph) and an altitude of 108 km (67 statute miles, 354 330 ft). Thirteen of the flights (conducted by 8 of the 12 people who flew the X-15) qualified the pilots as USAF or NASA astronauts (above 50 statute miles (80 km) altitude); 2 (both by the same pilot) met the FAI definition (above 100 km). Two of the X-15 pilots flew again as astronauts; Neil Armstrong became the first man to walk on the Moon during the Apollo missions, and Joe Engle (arguably making more use of his X-15 experience) flew early Space Shuttle missions. (NASA Image)

105. **Aerodynamic Forces Acting on a Spacecraft in Flight**. In conventional aviation, there are normally 4 forces acting on a body in flight: lift, thrust, weight and drag. As we move into space, we need to consider how these change:

a. **Lift and Weight**. Lift is a product of the airflow around a body, which supports weight and generates forces for manoeuvre and stability. As mentioned above, as the atmosphere thins with increasing altitude, these forces diminish, becoming negligible near the Karman Line. Weight, however, remains essentially unchanged. More strictly, it is mass that is unchanged, and weight derives from mass. Weight causes a satellite to fall towards the Earth, but forward speed, which is what keeps it in orbit, postpones the resulting impact – ideally indefinitely.

b. **Thrust and Drag**. In conventional aviation, thrust balances drag (at constant speed), and for a space-craft drag diminishes (though it may not completely disappear) as it climbs to orbit. This means the need for continuous thrust in flight vanishes. The rocket that launches a satellite provides enormous amounts of thrust to lift and accelerate the payload, but it does so for a relatively short period of time. Once orbit is achieved, thrust is needed only for manoeuvre, and occasionally to maintain orbit in response to drag. Although drag is small, applied over weeks, months or years, it can still impose a significant penalty on space operations.

Environmental Threats to a Spacecraft in Flight

106. **Hostile Space**. Space is a uniquely hostile environment both for people and for machinery. Aside from any practical difficulties relating to launch and re-entry, and the unique issues of life-support for manned missions, there are other problems. This section examines in sequence 4 main factors: the radiation environment of space and the phenomena described as space weather; the consequences of drag arising from the tenuous remnants of the Earth's atmosphere; the implications of the other physical characteristics of Space such as extremes of temperature and pressure; and finally the threat posed to spacecraft by natural and artificial debris. As the mitigation of these effects is complex and inter-related, the section concludes with a brief summary.

107. **Solar Radiation and Space Weather**. Almost all life on Earth is sustained by energy from the Sun, but in space, the Sun is both a source of energy and a potent threat. The Sun emits not just visible light, but additionally infrared (IR), ultraviolet (UV), x-ray and gamma radiation, as well as various energetic charged particles.[4] All these are potentially hazardous both to life and equipment. On Earth, a combination of the atmosphere and the shielding effect of the Earth's magnetic field protect it,[5] but this protection is left behind as one ventures into space. The distribution of the various resulting hazards is too complex to explore in detail, but it should be noted that, for a variety of reasons, the radiation output of the Sun is subject to small fluctuations over time. These include:

a. The distance from the Earth to the Sun is not constant – the Earth's orbit around the Sun is slightly elliptical. Received energy is higher the closer to the source.

b. The Sun rotates on its axis, taking about 27 Earth days for a complete rotation around its equator. Some of the fluctuations in output are directional, so they may or may not affect the Earth, depending on their position on the Sun's surface and its orientation. They are also directed and distorted by the Sun's magnetic field.

c. The overall activity level of the Sun, and particularly the frequency of the various types of energetic outbursts, varies very gradually following an 11-year cycle. Within that overall cycle, the exact timing of an outburst is essentially random.

---

[4] See Annex F for a breakdown of the various categories of radiation.

[5] For example, the ozone layer high in the atmosphere protects us from excessive exposure to ultraviolet (UV) radiation.

In the same way as terrestrial weather is a description of changes in the local Earth environment, the term ‘space weather’ is used to refer to changes in the space environment. Just like Earth weather, it can be recorded and reported, and follows broadly predictable patterns, but local variations require considerable effort to forecast.

Illustration 1.2 – Space Weather Seen from Space. This photo taken from the Space Shuttle in orbit shows an aurora: a very visible aspect of space weather. Aurorae are caused by charged particles from the Sun interacting with the upper reaches of the Earth’s atmosphere. They are commonly seen on Earth at high latitudes, due to the concentrating effect of the Earth’s magnetic field. (NASA image).

108. **Concentration of Radiation Effects**:

a. **The Van Allen Belts**. Early spaceflight enabled exploration of the various concentrations of radiation, magnetic fields and charged particles that surround the Earth. The Van Allen Belts are the most widely known of these.[6] Their structure is complex. Most of their effects become noticeable above approximately 1000 km, and diminish again at higher altitudes, although some are apparent closer to Earth. The concentration of hazardous phenomena in the belts’ structure is a serious constraint on orbit selection, for both manned and unmanned missions.

b. **The South Atlantic Anomaly**. The Van Allen Belts are aligned with the Earth’s magnetic field, and thus with respect to the North and South Magnetic poles, not the Geographic poles. When the shape and orientation of the Earth, the shape of the Belts and the difference between the geographic and magnetic poles are accounted for, the result is a

---

[6] Named after Professor James Van Allen (University of Iowa) who discovered them in 1958, using data from the first US satellite Explorer 1.

concentration of effects over the South Atlantic Ocean at the point where the Inner Belt comes closest to the surface of the Earth. This is known as the South Atlantic Anomaly (SAA). There is a matching point opposite, over the Northern Pacific Ocean, where the Belt is at its furthest from the Earth's surface, but this has less effect on typical orbits, and no specific name is commonly given to it. The limits of the SAA change slowly over time as the Magnetic Poles move and, possibly, also in response to more complex geo-magnetic effects. The SAA has significant effects on satellites that must pass through it. For example, the International Space Station carries additional shielding for its crew to account for transit through the SAA and the Hubble Space Telescope is unable to make observations from within the SAA due to the effects of radiation on its sensitive instruments.

109. **The Effect of Space Weather in Space**. Different components of space weather have different effects on spacecraft. For example:

a. The impact of solar particles hitting a spacecraft's structure directly affects its motion. Although the particles are exceptionally small, they hit it continually, at significant speeds. The effect accumulates slowly, ultimately altering the craft's orbit measurably.[7] It can also affect the orientation of the spacecraft by generating asymmetric forces on specific parts of the craft.

b. Individual charged particles can penetrate electronic components, and can, for example, corrupt computer memory. Some effects are the result of single impacts, but continuous lower energy impacts can also have cumulative effects on the reliability of electronic components.

c. The continuous impact of solar particles on the spacecraft's outer surface can also cause physical damage. Delicate components such as sensors or solar panels, which must be exposed to function, can gradually degrade.

d. Electrically charged particles, most notably electrons, accumulating on the surface of the craft can transfer their charge to its structure. This can induce large currents in conducting components, or build up significant voltages across insulated parts. When these discharge, they can cause severe damage.

e. Solar activity has a direct effect on the thickness and extent of the Earth's upper atmosphere, and may thus have an impact on low altitude orbits through atmospheric drag.

110. **The Effects of Space Weather on the Earth**. Some space weather effects have direct impact on the Earth too:

a. Interaction between charged particles and the upper reaches of the atmosphere (specifically the ionosphere) affects radio communications and radar propagation, most notably in the frequency bands that depend on particular atmospheric properties to achieve long ranges.[8] Affected applications include high-frequency radio communications, long-range radar systems, as well as satellite communications and navigation systems.

---

[7] The possibility of harnessing this effect as a source of propulsion by using a sail has been explored theoretically, although a recent attempt at a live trial failed when the launch platform malfunctioned. It may become a practical system at some time in the future. The thrust generated by even a large sail is not great, and the effect diminishes with distance from the Sun, but it may have utility in specialised applications.

[8] These interactions can also be seen from the surface as aurorae.

b. Hazardous radiations, and the problems of protecting astronauts from it, are a constraint on manned spaceflight. Even within the atmosphere, increased particle densities at high altitudes could in some circumstances be hazardous to aircrew. The advent of modern airliners capable of long-range trips at high altitudes has led to some airlines monitoring the exposure of individual crew members.[9]

c. Just as unwanted charges and currents can build up on spacecraft, similar mechanisms resulting from very energetic events and involving the Earth's magnetic field can induce currents in large metal structures on Earth. Metal pipelines, power distribution grids and other cabled installations can all be affected.

Solar Storms

- In 1859, a uniquely powerful solar flare impacting the Earth, commonly referred to as the 'Perfect Solar Storm' caused widespread damage to telegraph installations, including igniting spontaneous fires as fixed wires shorted out. Aurorae, normally confined to high latitudes, were seen as far south as Rome and Havana, with equivalent effects in the Southern Hemisphere. NASA scientists have recently tried to characterise this event, using contemporary records, in an attempt to predict what a similar flare could do today.[10] It is believed that a chance alignment of the Earth's magnetic field with the magnetic field generated by the flare enhanced its effect significantly.
- In the 20th Century, 2 solar events in 1989 and 1994 caused widely publicised effects, principally in Canada. The 1989 event caused significant power cuts in Quebec, with damages and losses estimated to run into hundreds of millions of dollars.[11]

Singular Effect of the Aurora Borealis on the Telegraph Wires

Chicago Press and Tribune Aug 30, 1859

NEW YORK Aug 25 – The Superintendant of the Leyland Steamship Company's line telegraphs as follows in relation to the effect of the Aurora Borealis last night.

"I never, in an experience of fifteen years working telegraph lines, witnessed anything like the extraordinary effects of the Aurora Borealis between Quebec and Farther Point last night."

"The line was in most perfect order, and well skilled operators worked incessantly from eight o'clock last evening till one o'clock this morning to get over in an intelligible form about four hundred words of the report per steamer *Indian* for the Associated Press. At the latter hour, so completely were the wires under the influence of the Aurora Borealis, that it was found utterly impossible to communicate between the telegraph stations, and the line had to be closed" The same difficulty prevailed as far South as Washington.

Illustration 1.3 – An Early Account of Space Weather. This transcript of an excerpt from the *Chicago Press and Tribune* in 1859 provides one of the earliest accounts of the effects of space weather on the Earth. This report makes clear that the installation concerned did not experience the most severe effects of the Solar Storm (see above); apparently the apparatus at least did not catch fire. However the effects plainly rendered the telegraph inoperative.

---

[9] Generally, the statistical risk to aircrew from radiation is assessed as slight, and routine monitoring of exposure is intended to allow management action if any individual appears likely to reach a safe limit in a given time period. Concorde was fitted with radiation detection and monitoring equipment, which *in extremis* could warn the crew if a transient solar event was causing hazardous levels of exposure and prompt a descent from the routine cruising altitude.

[10] See *http://www.jpl.nasa.gov/news/news.cfm?release=140* for details of this research.

[11] See *http://science.nasa.gov/headlines/y2003/23oct_superstorm.htm* for details of the 1989 and 1994 events, cross-referenced to the 1859 research effort alluded to above.

111. **Orbital Decay**. We have already noted that the Earth's atmosphere does not end abruptly – rather its density tapers off gradually. Orbital decay is caused by drag, even from this tenuous remnant of the atmosphere, slowing the spacecraft down. For bodies in orbit near the Earth's surface, this has serious implications.[12] Reducing speed leads unavoidably to a reduction in altitude, taking the spacecraft down into (relatively) thicker atmosphere. This in turn increases the drag more, feeding a vicious cycle until either the spacecraft is boosted to regain altitude, or the orbit decays catastrophically and the spacecraft re-enters the atmosphere. Several factors affect the rate of decay in different ways:

a. The shape or thickness of the atmosphere is a complex and dynamic variable. It changes with latitude, and also with time, being affected principally by solar activity, which at these high altitudes can thicken or thin the atmosphere.

b. The shape of the spacecraft is also important. Spacecraft designed for atmospheric re-entry are very carefully shaped to exploit drag when required, but that apart, the more drag the shape of a satellite produces, the quicker its orbit will decay. This too is exploited by designers – the final stage of a launch rocket will often initially enter orbit along with its payload, but its shape ensures that its orbit decays quicker than the payload, achieving separation between the 2 objects and quickly removing the unwanted item from orbit. Satellites intended for notably low orbital altitudes may also incorporate elements of aerodynamic design to minimise drag and orbital decay.

c. The faster an object is moving, the more drag it experiences, but it will be shown that the speed of a satellite is dictated by the size and shape of its orbit, so designers do not have the option of reducing speed to minimise drag for a given orbit.

d. Remember drag is purely a function of atmosphere. NASA placed the Lunar Reconnaissance Orbiter unmanned probe in orbit around the Moon during 2009, in an orbit about 50 km above the Moon's surface. The intended primary mission will last for at least 12 months. Such an orbit would be completely impractical above the Earth, but poses no significant drag problem above the airless Moon.

e. If it is necessary to keep a satellite in orbit at low altitude for an extended period, it will require occasional boosts to overcome drag.[13] This carries with it a fuel penalty, which will either impact on launch mass or limit the satellite's operational life.

All these factors combine to make orbital decay roughly, but not exactly predictable. Thus it needs to be monitored to account for fluctuations in the atmosphere. This also has implications for tracking space objects, as the size and shape of their orbits can change. Calcultaed tracking solutions may only be useable for a matter of days, particularly for items in very low orbits.

112. **Temperature, Pressure and Weightlessness**. Space is essentially a cold near-vacuum, where orbiting craft are free of the effects of gravity. This poses a series of unique engineering problems:

a. A spacecraft is regularly exposed to the Sun's un-moderated radiation on one side of the craft and the extreme cold of space on the other, setting up thermal stress across its

---

[12] Although some drag occurs at all altitudes, above 600 km (330 nm) it is essentially negligible compared to other effects.

[13] For example, the International Space Station (ISS), which orbits at about 340 km altitude, requires regular re-boosting – the Space Shuttle has regularly done this during its periodic visits. The task has also been undertaken by the ESA *Jules Verne* cargo vessel and Russian *Progress* supply vessels while they have been attached to the ISS.

structure. Depending on its orbit, a satellite may also pass through the Earth's shadow on an hourly basis, again undergoing thermal stress.

b. The spacecraft payload will undoubtedly generate heat, which will probably have to be managed. Losing heat into the vacuum of Space is technically challenging.

c. In the very thin atmosphere of space, some exotic effects become apparent. Exposed liquids will boil away,[14] with severe implications for the lubrication of moving parts. Moving surfaces in contact on Earth sometimes rely on a very thin air cushion to separate them, but this too vanishes in the vacuum of space. The resulting effect is known as 'cold welding' (although strictly it is a consequence of the vacuum, not the temperature). Metallic structures may become brittle over time, and composite materials may tend to lose gases from their internal structure. While this may not be catastrophic for the material, the gas emitted has a tendency to degrade the optical properties of surfaces such as mirrors, sensor windows and solar panels, at least on a temporary basis.

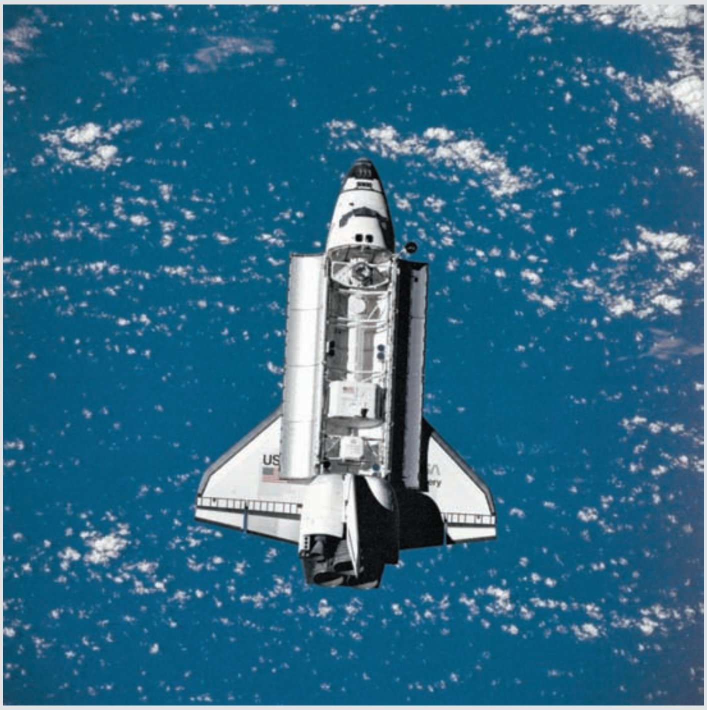


Illustration 1.4 – Cooling a Spacecraft. Cooling a spacecraft (in this case the Space Shuttle *Discovery*) in orbit is always challenging for engineers. The only way to lose heat overboard from a spacecraft is by radiation. This image of the Space Shuttle, taken from the Mir space station, shows the payload bay doors in their open position. The inside of each door includes four large radiator panels, which can clearly be seen. This is why the Space Shuttle always flies in orbit with the doors open. If for any reason the doors failed to open after launch, the mission would be aborted very quickly, as the Shuttle would be unable to operate much of its onboard equipment with the doors closed. (NASA Image).

113. **Space Debris**. It is possible to collide with both natural and man-made objects in orbit. Each poses a similar threat to satellites if a collision occurs, but the numbers, sizes and distribution of the objects vary between the 2 populations. These objects include:

---

[14] This is the extreme case of the observable effect on Earth that boiling temperatures reduce with altitude – food, for example, needs to be cooked for longer in the thinner air at high altitude. In space, the pressure is so low that boiling occurs spontaneously.

a. **Natural Debris**. The Earth, and anything orbiting it, collides continually with meteorites and micro-meteorites which originate from elsewhere in the Solar System. Their size varies from substantial to negligible. The Earth's surface is protected from small meteorites by the atmosphere – shooting stars are particles, typically no bigger than gravel grains, burning up as they enter the atmosphere. In space, there is no such protection. Since the debris in this case originates outside Earth orbit, there is little scope for warning of small particles approaching. Although the statistical risk of a substantial impact disabling a spacecraft is low, structures must be robust enough to absorb the small impacts that are much more likely.

b. **Man-Made Debris**. Since the earliest days of the space programme, Earth orbit has become cluttered with debris of various kinds. Some is intentional, such as dead satellites that have finished their missions and debris from launch activity. Other instances are accidental, including tools lost by astronauts, components that have failed and tiny flecks of paint and such-like shed by space vehicles. Since these objects are all in some kind of orbit around the Earth, ground-based sensors can potentially track them, at least down to a certain size. Thus it might be possible to manoeuvre a satellite to avoid a collision, though again for objects too small to track, the structure must be made robust.[15]

It is important to remember that if something detaches from an orbiting body, it initially follows the same orbit as its source, varied only by the event that caused the breakup. This means that accidental debris may take weeks, months or even years to separate slowly from the source, while an event like an explosion will create a cloud of orbiting objects that will slowly disperse. Depending on altitude and the nature of the objects, they may remain in a stable (though unwanted) orbit for extended periods.[16]

Space Debris – Examples

- The altitude at which debris is created at has a huge effect on its duration and the hazard it poses. Four ostensibly similar incidents, each with vastly different long-term outcomes, illustrate this point:
  - On 11 January 2007, China conducted a test impact of a missile into a redundant weather satellite. This satellite was in a relatively high orbit, at about 850 km. The resulting debris cloud has spread since then.
    - The impact generated about 150 000 pieces of debris of 1cm size or greater. About 2500 of them are being tracked from the ground, and as of May 2008, only 22 of those have re-entered the Earth's atmosphere.
    - The debris orbits now vary between 200 and 4000 km altitude.
    - Much of the debris will survive in orbit for years or even decades.
  - About a month after the Chinese test, a Russian booster rocket exploded at relatively low altitude. It had failed to place a commercial communications satellite in the required orbit during 2006 and had remained derelict and uncontrollable in a low orbit until its unused fuel became unstable and exploded.

---

[15] NASA sustained substantial damage to the windshield of a Space Shuttle during an early mission in 1983, caused by a fleck of paint estimated to be 0.2 mm across. Unfortunately, the impact speed was probably somewhere around 20 000 mph, and at this speed, even tiny fragments possess significant energies. See Illustration 1.5 for the outcome.

[16] See Paragraph 387d for further discussion.

Because of the orbit, however, most of the debris will re-enter the atmosphere relatively quickly.

- On 20 February 2008, the USA deliberately destroyed a reconnaissance satellite at 250 km altitude. More than half the debris re-entered the atmosphere within 45 minutes, and more than 99% re-entered within a week.
- In February 2009, an obsolete Russian satellite collided with one element of the Iridium communications constellation. This was once again a relatively high altitude collision (about 790 km), which will have long-term debris implications. There has been significant open-source modelling of the collision to try to quantify them, but the true picture will take time to emerge.

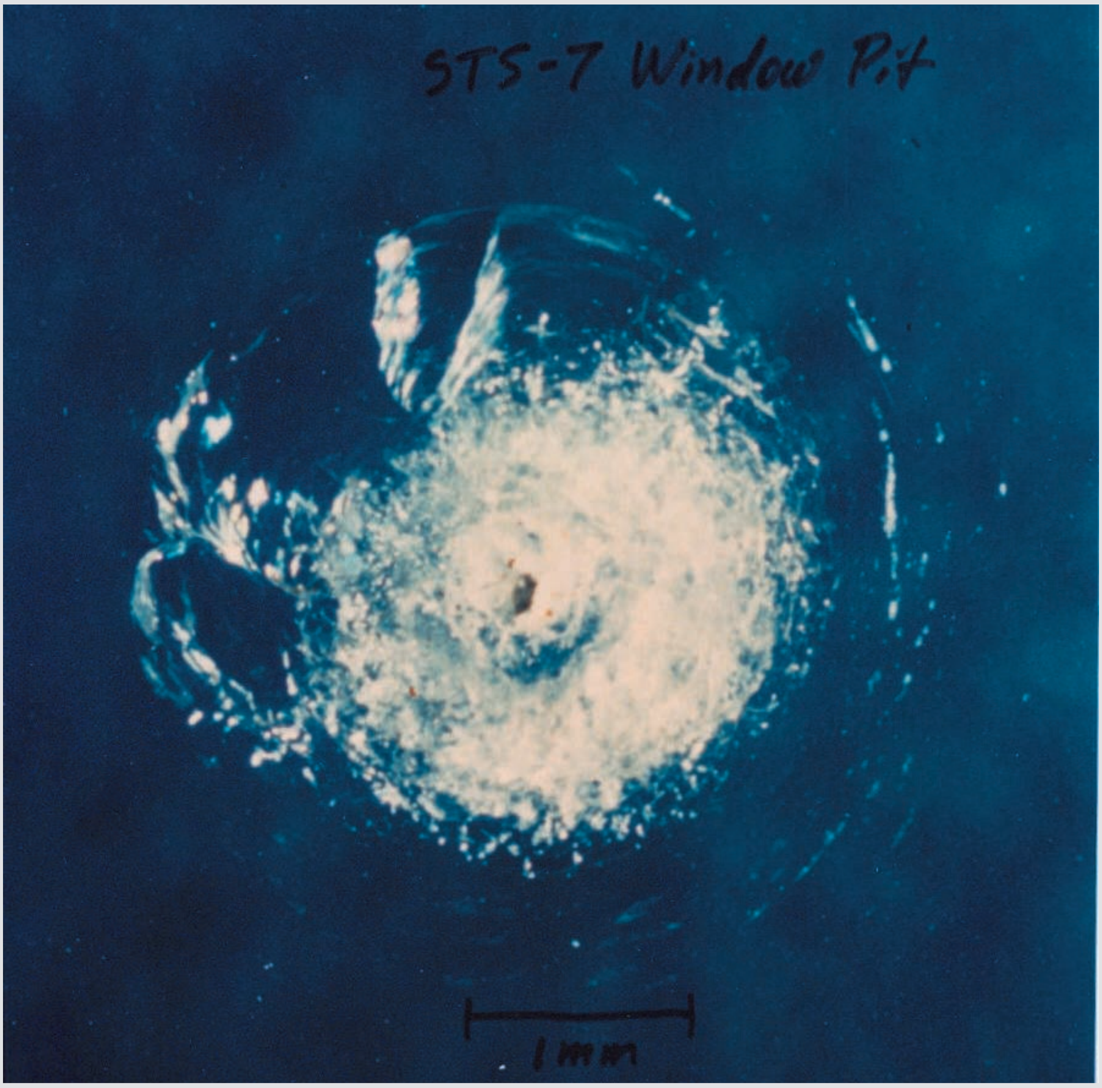


Illustration 1.5 – Debris Impact Damage. This picture shows the impact damage to the windshield of the Space Shuttle *Challenger* during Flight STS-7 in 1983, caused by a fleck of paint. A similar, but more serious, impact affected the Space Shuttle *Endeavour* on Flight STS-59 in 1994; that impact penetrated the windshield to about half its depth. Consequently, the Space Shuttle now flies in orbit facing ‘backwards’ when possible, so that any high-speed impacts occur on the more robust rear face of the vehicle. Other Space Shuttle debris impacts have occurred, and their increasing frequency verifies the observation that the amount of debris in low-earth orbit continues to increase. (NASA Image)

114. **Mitigation for Spacecraft**. Engineers attempt to mitigate all the effects of the space environment that would otherwise hamper a spacecraft's operation. Different threats require different kinds of mitigation:

a. The outer structure of a spacecraft will be constructed to resist space weather effects. The outermost layer is often a blanket of metallic-coated plastics, which will have very carefully tailored thermal properties to allow it to reflect incoming radiation and radiate excess heat. Additionally, both the blanket and the underlying hard structure must resist physical and electrical hazards. The spacecraft may also be made to spin slowly to even out heating and cooling.[17]

b. The only way to lose excess heat from the spacecraft payload is ultimately by radiation into space – an onboard environmental control system can use conduction or convection to move heat around the craft, but there is no fluid medium to support these processes into the vacuum of outer space. The ideal for an engineer is a passive system where the properties of the structure maintain the required temperature automatically. Complex payloads are likely, however, to require both heating and cooling in differing areas, with a resulting penalty of weight, power consumption and complexity. See Illustration 1.4 for one solution to this problem.

c. The electrical properties of the spacecraft components and materials will be accounted for in its design, with the incidence of electrically isolated conductors being minimised.

d. Computer hardware and software can be designed to mitigate the consequences of isolated failures.

e. One of the reasons that spacecrafts are normally assembled in 'clean-rooms' is to prevent the incorporation of dust or any other debris that might float to an inconvenient location once on orbit. See Illustration 1.6.

---

[17] The so-called 'rotisserie' method.

Illustration 1.6 – Mitigating Onboard Debris and Contamination. Engineers assemble a spacecraft (in this case NASA's *Dawn* Interplanetary Probe) in clean conditions to avoid inclusion of any contamination or debris that might migrate in weightless conditions to an undesirable location. In this picture, the craft has yet to be covered in its thermal blanket. (NASA/courtesy of nasaimages.org).

SECTION II – GEOMETRY, PHYSICS AND THEIR IMPLICATIONS FOR ORBITS

In this section you will find:

- An explanation of the geometry of circles and ellipses to enable later discussions of orbits.
- A recap of gravity and Newtonian dynamics – to explain how a satellite can gain and maintain orbit.
- A discussion of the various kinds of orbits.
- The implications of Kepler's laws, which flow from the principles outlined above.
- Why it takes 6 numbers to describe an orbit exactly (the orbital elements).
- Definitions of apogee and perigee.
- Factors affecting the ground track of a satellite over the earth.
- The implications of the Earth not being exactly spherical.

Orbits and Geometry

115. **Geometry – Circles and Ellipses**. All practical Earth orbits are elliptical. It is therefore useful to understand something about the geometry of ellipses.

116. **Drawing an Ellipse**. The standard method of drawing an ellipse on paper uses 2 pins on a drawing board and a loop of string or thread. Rotate a pencil inside a loop of thread that also runs round the pins and the result will be an ellipse – see Illustration 1.7. The ellipse is symmetrical along 2 axes, one passing through the pinholes, the other perpendicular to them and half way between them.

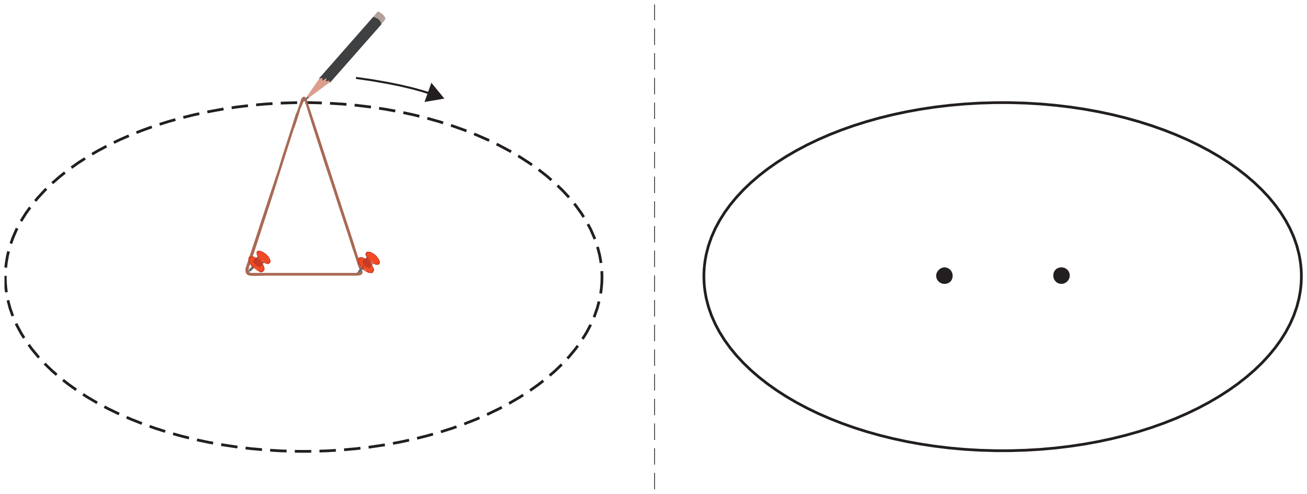


Illustration 1.7 – Drawing an Ellipse. This shows one method of constructing an ellipse, using a loop of thread or string and 2 pins. The size and shape of the ellipse will depend on the total length of the thread loop and the pins' distance apart. The position of the pins will mark the *foci* of the resulting ellipse; see Paragraph 118 below for a definition of foci.

117. **Changing Size and Shape**. Some experimenting with the thread and pins will reveal that the further apart you place the pins, the more elongated the ellipse becomes – conversely, the closer together the pins are, the more circular it becomes. Eventually, if you stuck both pins in one hole (or used one pin), the ellipse would be a circle – a circle is thus a special case of an ellipse (in the same way as a square is a special kind of rectangle).

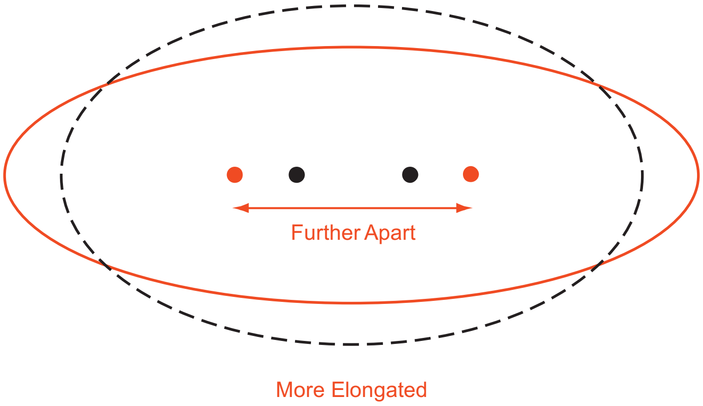


Illustration 1.8a – Increasing Distance Apart. Starting from the black ellipse, if the pins are moved further apart (to the red positions), the ellipse becomes more elongated. It is said to become more *eccentric*.

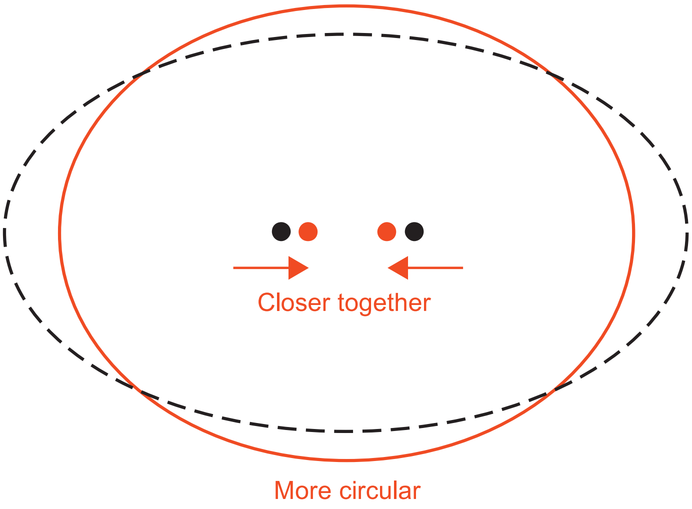


Illustration 1.8b – Decreasing Distance Apart. If instead the pins are moved closer together, the ellipse becomes more circular; its eccentricity is reducing.

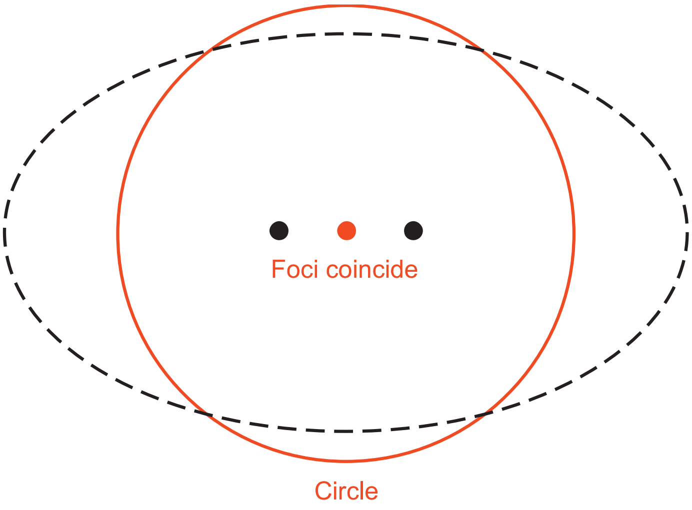


Illustration 1.8c – Ellipses and Circles. Eventually, if the 2 pins are so close together that they coincide, the ellipse will become circular, demonstrating that a circle is a special case of an ellipse.

118. **Size and Shape – Eccentricity and Axes**. We summarise the geometrical terms used in regards to ellipses, as outlined in Illustrations 1.8a-1.8c.

a. The 2 pinholes used in the ‘string method’ above are each known as a *focus* (plural *foci*) of the ellipse. Where the ellipse is a circle, both foci are in the same place – in the centre of the circle – they are said to be *coincident* or *collocated*.

b. The measure of elongation is called *eccentricity*. A circle has an eccentricity of 0. More elongated ellipses have increasing eccentricity up to, but not reaching 1.

c. The line joining the 2 pinholes in Illustrations 1.8a-1.8c, and extended across the ellipse at its widest point, is called the *major axis*. Similarly, the line at right angles to this, crossing the ellipse at its narrowest, between the pinholes, is the *minor axis*. For practical reasons, an ellipse is often defined in terms of half of these axes (the semi-major and semi-minor axes). This is exactly the same as describing a circle by reference to its radius instead of its diameter.

d. There is a fixed relationship between the lengths of the major or semi-major axis, the minor or semi-minor axis and the eccentricity of an ellipse. Any 2 values define the third.

e. Both the ellipse and the circle (along with some other curves) are what are known as conic sections, and you may sometimes encounter the term. All conic sections have eccentricity – ellipses are just the conic sections where the eccentricity is less than 1. A summary of how conic sections are derived is at Annex A.

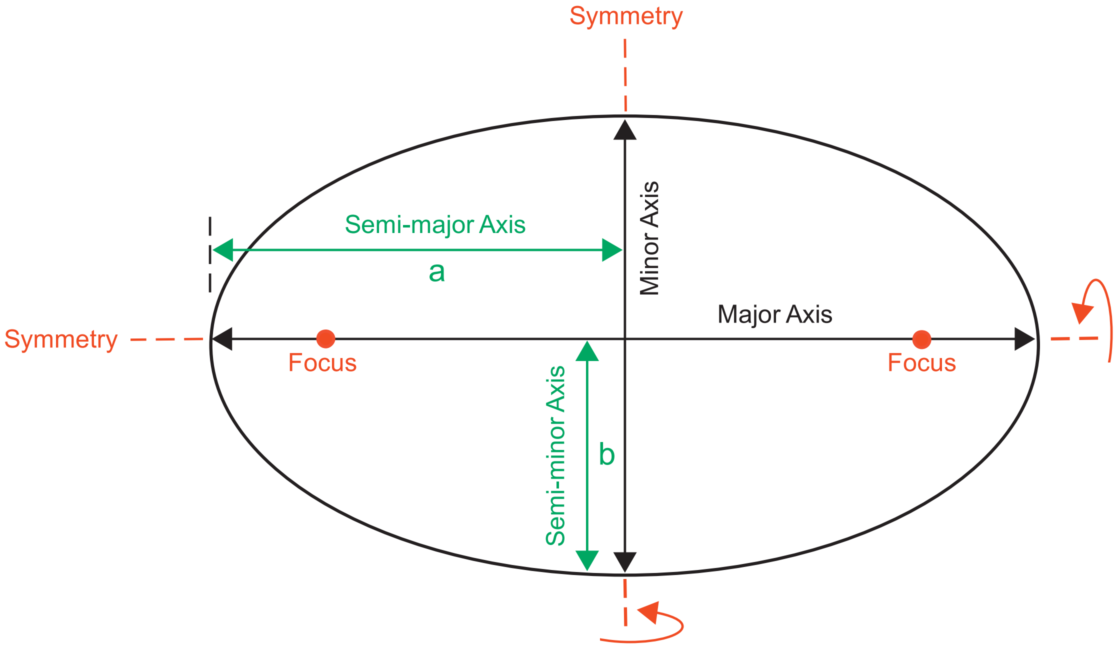


Illustration 1.9 – The Parts of an Ellipse. This diagram shows the principal named features of an ellipse. The axes of symmetry are also marked. The eccentricity of the ellipse can be calculated from the proportions of the axes. If *a* and *b* are the lengths of the semi-major and semi-minor axes respectively, then the eccentricity, usually denoted as ε or *e*, is given by $e=\sqrt{[(a^2-b^2)/a^2]}$.

119. **Planes and Great Circles**. An ellipse is a 2-dimensional figure; it always lies in a plane.[18] This was implicit in the drawing experiment above where the ellipse was drawn on a flat board, which only possesses 2 dimensions. Ideas about planes will be used regularly in describing orbits, so it is worthwhile to state a few simple properties.

a. **Two Dimensions only**. Geometric planes possess 2 dimensions only. This may not correspond directly with the idea of a shape with length and breadth, but not thickness, though that can be a useful model. They are, however, usually infinite in extent; they may be drawn in diagrams as being rectangles or squares, but in principle they extend forever.

b. **Intersections**. Planes can be parallel with each other, in which case they do not intersect, but otherwise, they intersect with each other defining a straight line. The angle between the planes can also be measured, using lines at right angles to the line of intersection.

---

[18] This is the geometrical meaning of 'plane' – a flat surface.

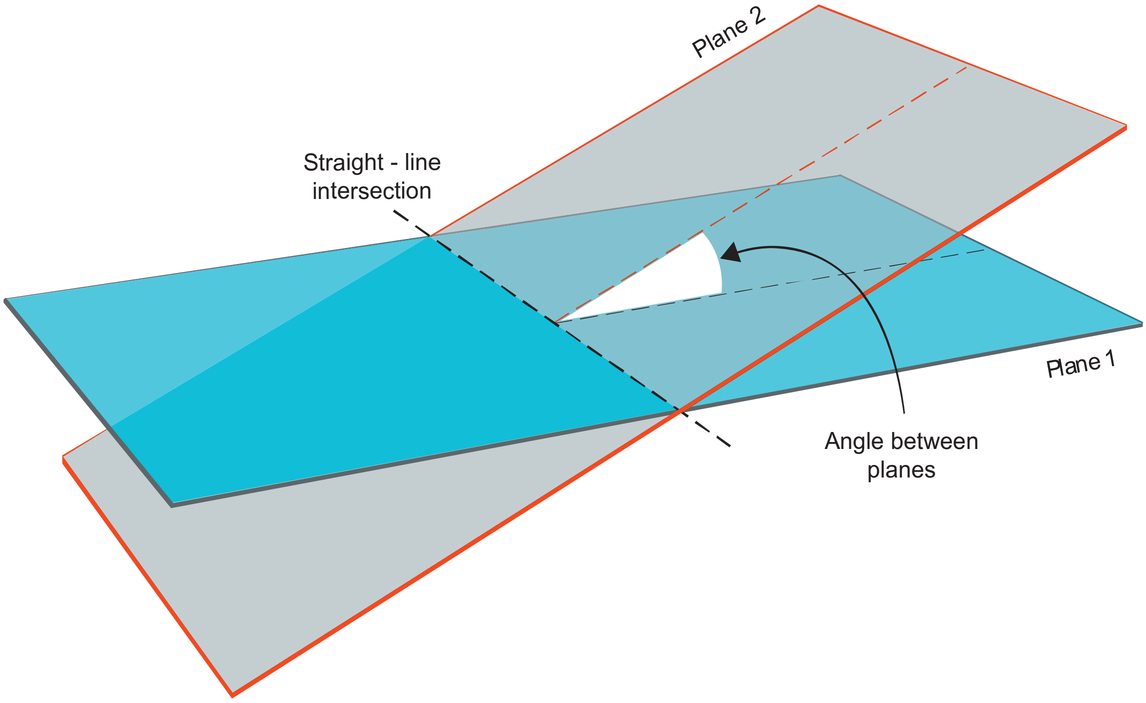


Illustration 1.10a – Intersecting Planes. Planes are in theory infinite in extent, but here they are shown as limited areas – in this case as rectangles. The planes intersect to define a straight line (shown as a black broken line) which is also in theory infinite in length. The angle between the planes (shown in white) is conventionally measured at right-angles to the line of intersection.

c. **Cross-sections**. Where a plane intersects with a solid body, the intersection is still 2-dimensional (just like the plane), but it is bounded by the solid, creating a 2-dimensional shape known as a **cross-section**. For regular solid bodies, the shape of the cross-section is sometimes predictable; any cross-section through a sphere is a circle, for example.

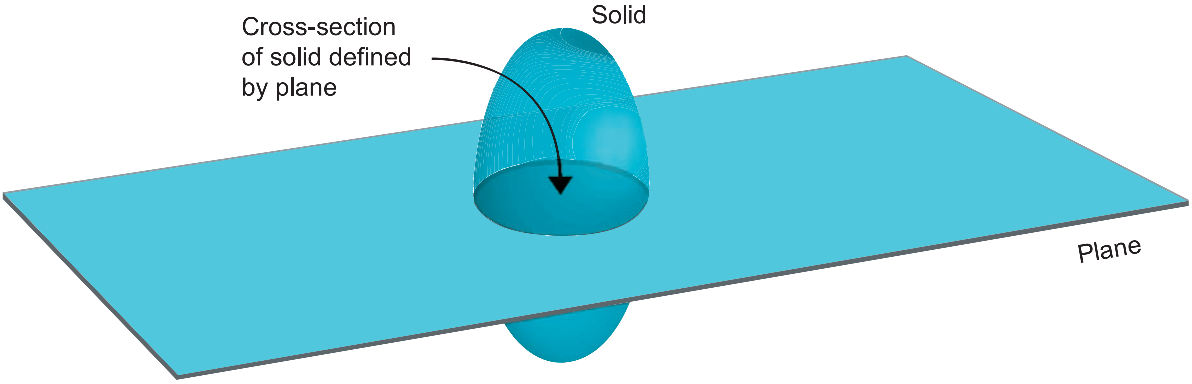


Illustration 1.10b – Cross-sections. A plane and a solid intersect in two dimensions. The 2-dimensional shape defined by the intersection (in the case illustrated above, a circle) is the cross section.

d. **Defining a Plane from a Cross-section**. The definition of a cross-section can be used to define the plane. Thus for example, if you think of the Earth as the solid body, the equator defines a cross-section through it. It also defines a plane in space – referred to as the **equatorial plane**. It is this plane, and various angles and lines relating to it that will be used when examining orbital elements.

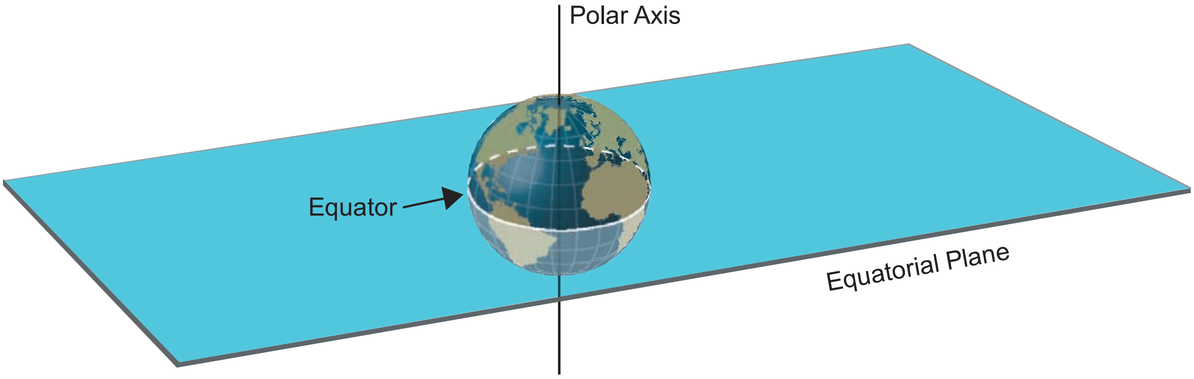


Illustration 1.10c – The Equatorial Plane. In the same way as a solid body and a plane define a cross-sectional shape, we can start with the solid body (in this case the Earth) and the cross-section (in this case the Equator), and use those to define the plane that would produce the cross-section. This plane is the **equatorial plane**.

e. **Orbital Planes**. Anything moving in an elliptical path is moving in 2 dimensions only. If viewed from edge-on it will not (cannot) move out of the plane of the ellipse unless something is propelling it (since thrust cannot be sustained indefinitely, this is implicitly on a temporary basis). Thus a satellite's orbit also defines a plane – the **orbital plane**.

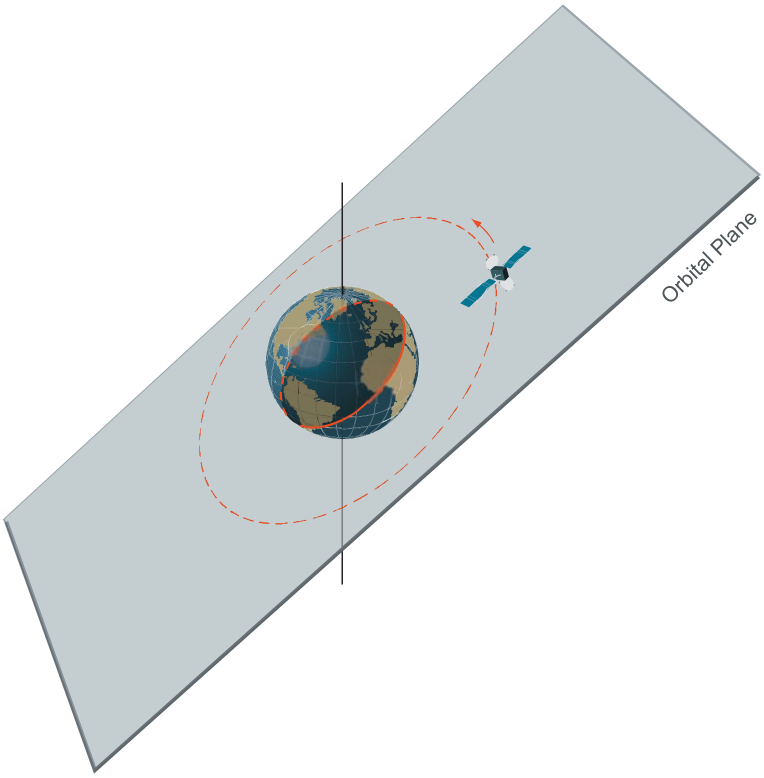


Illustration 1.10d – The Orbital Plane. The orbit of a satellite around a body also defines a plane. In this case it is known as the **orbital plane**.

f. **Great Circles**. Finally, it was noted before that a cross-section through a sphere is always a circle, but for spheres in general, a distinction is made between different kinds of cross-sections. If the cross-section includes the centre of the sphere, the resulting circle is known as a **great circle**; if not, it is known as a **small circle**. Thus the equator is a great circle on the surface of the Earth, as is a circle passing through both geographic poles and running up and down two diametrically opposite lines of longitude. The Tropics, and the Arctic and Antarctic circles are small circles.

Dynamics

120. **Force, Mass and Acceleration**. In 1687, Sir Isaac Newton (1642-1727) codified three laws that govern the behaviour of bodies at rest and in motion.[19] These remained the best descriptions of motion until the 20th century, when Einstein showed that they break down in extreme circumstances. However, even up to the speeds involved in spaceflight, most of the corrections Einstein introduced are small enough to ignore.[20]

a. **Newton's First Law of Motion**. Newton's first law states that 'a body continues its state of rest or uniform motion in a straight line, unless a force is applied to it'. Although this is easy to state, it appears to contradict everyday experience, at least for moving objects. A bullet fired from a gun generally loses speed. Even if it doesn't hit anything, it will gradually slow down, limiting the weapon's range. Cars and other vehicles need to keep their engines running to progress along the road. The explanation of this is, of course, that there are hidden forces continually affecting moving bodies on Earth – air resistance and friction in our examples. In the vacuum of space, we approach the Newtonian ideal and motion can continue unchecked.

b. **Newton's Second Law of Motion**. Newton's second law states that 'the rate of change in a body's momentum is proportional to the force applied'. This again is easy to state, but needs a little exploration. If a body's mass is constant, the law simplifies to 'force changes velocity' or 'force equals mass times acceleration'. Two consequences follow from this: firstly, a typical rocket consumes its fuel so quickly, and so much of its mass is fuel, that the 'constant mass' assumption breaks down. Detailed calculations must allow for the change of mass of a rocket during its acceleration. Secondly, we need to think a little more about 'velocity'; see sub-paragraph d.

c. **Newton's Third Law of Motion**. Newton's third law states that 'for every action, there is an equal but opposite reaction'. The most obvious consequence of this Law for spaceflight is again in respect of rockets and propulsion. These work by accelerating something – usually hot gas or other combustion products – away from the rocket. This is the 'action'. The equal but opposite 'reaction' is the acceleration of the rocket in the opposite direction. Notice that this reaction is completely independent of what happens to the exhaust – including whether it is discharged into the atmosphere, or into the vacuum of space – so the rocket motor works equally well in each case.

d. **Velocity**. We mentioned above the need to think more about velocity. We need to bear in mind that velocity comprises speed and direction. Changing either is changing velocity, and we call any change in velocity acceleration. Once again, everyday experience intrudes. We think of a car having a 0-60 mph acceleration time, (achieved by using the accelerator), measured in a straight line. For the physicist, engineer or rocket scientist however, changing direction at constant speed is equally an acceleration, which needs a force to achieve it. So too is decelerating – it is a change in speed, and therefore an acceleration. To pursue our car analogy, the accelerator gives us acceleration, but so do the

---

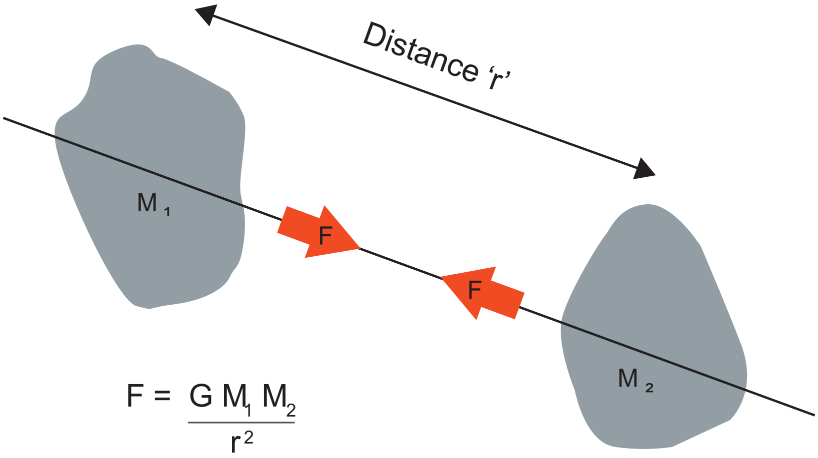
[19] See Annex B for the historical context of Newton's work.

[20] There are circumstances in orbit, and affecting some space applications, where it is necessary to make relativistic corrections in equations, but these are not discussed further here.

brake pedal and the steering wheel. As soon as the force is removed, constant direction, constant speed motion resumes.[21]

121. **Gravity**. His laws of motion alone would have secured Newton's reputation, but he followed them up with his law of universal gravitation. Whether or not he discovered it watching an apple fall to the ground, his insight was to conceive of a universal attractive force that any body continuously exerts on any other.

a. Gravity is a fundamental property of a body's existence. It attracts anything else possessing mass towards the body. The more massive the bodies, the greater the forces they generate.[22] The force diminishes with distance apart, following an inverse square law. This means at double the distance, the force is diminished by a factor of four, at 3 times the distance, it is one-ninth its original strength.



$$F = \frac{G M_1 M_2}{r^2}$$

Illustration 1.11a – Gravity Described. Gravity is a fundamental property of matter, which generates an attractive force between two bodies. The magnitude of the force is determined by their masses and their distance apart. The force always acts directly between them, along a straight line connecting them. The equation shows that the magnitude of the force varies directly with the masses and as the inverse square of their distance apart. 'G' is a constant, the numerical value of which depends on the units used for the other quantities. Do not confuse 'G', as used here, with 'g' used to denote the acceleration due to gravity on or near the Earth.

b. Gravity is a relatively weak force; for example, it takes a body the size of the Earth to make us weigh as much as we do! Nevertheless, it is continuously present. It is a common mistake to assume that a weightless astronaut has moved far enough away from the Earth to get beyond the reach of gravity. This is not the case. He still possesses mass, and is being attracted towards Earth by the resulting pull of gravity, diminished only slightly by distance.[23] The weightless experience is due to his forward motion keeping him in orbit and the fact that the spacecraft that surrounds him is moving in exactly the same way.

c. Since one body is often much more massive than the other in calculations involving gravity – such as the Earth compared to an individual person – it is common to ignore the fact that *both bodies attract each other*. Things fall towards the Earth due to gravity, and

---

[21] For once, aviation provides a useful model here – acceleration in an aircraft is understood to encompass the force exerted in a turn, even though in many cases the aircraft's speed remains unchanged throughout the turn.

[22] But the greater the mass, the greater the force required for acceleration (Newton's Second Law of Motion). This is why all bodies accelerate at the same rate under the influence of gravity, regardless of their mass.

[23] At the altitude of typical manned orbits, the astronaut's total distance from the centre of the Earth is not greatly changed.

we can see the resulting acceleration each time we drop something, but we ignore the fact that the falling object is attracting the Earth towards it with the same force. We will make this same assumption extensively when considering bodies in orbit.

Illustration 1.11b – Gravity on Earth. We experience gravity continually, keeping us bound to the surface of the Earth. This tends to hide the facts that firstly gravity is a relatively weak force, and secondly that gravity not only attracts us towards the Earth, but also attracts the Earth towards us.

d. Where we are dealing with a physically large but symmetrical body, such as the Earth, we assume that *all the forces it generates originate at its centre*.[24] (See Illustration 1.11c). For a small body, particularly one with a notably asymmetric shape (like many spacecraft), the designer must calculate its centre of gravity (the origin of the term may now be clearer). Forces on the craft are assumed to act through that point.

---

[24] We will look later at the consequences of the Earth not being truly spherical, and thus not symmetrical from certain viewpoints.

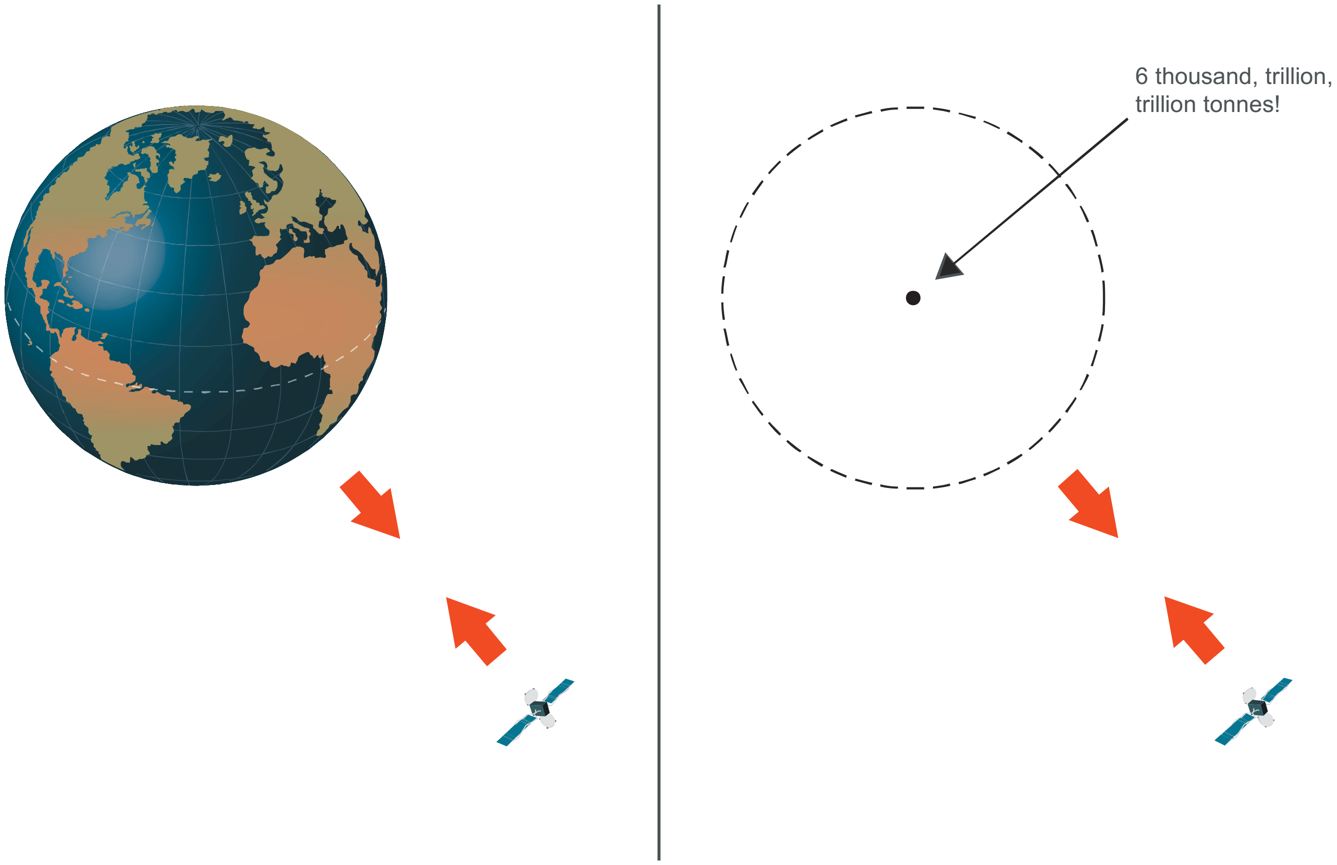


Illustration 1.11c – Centres of Gravity and Point Masses. It can be proven that where a symmetrical body has significant size (like the Earth), its gravitational attraction can be assumed to originate at its centre. Although this may seem a trivial point, it has significant implications for orbital motion; since the centre of the Earth is the apparent origin of the gravitational attraction, stable orbits must include that point in the plane of the orbit.

122. **Falling Towards the Earth – Newton's Thought Experiment**. Having established the theoretical background, Newton pondered its consequences. Some 270 years before the first artificial satellite, he understood how one could be made to work. His argument went like this:

a. Firstly, all objects falling vertically under the Earth's gravity accelerate at a constant rate, if air resistance is ignored.[25] If we imagine gravity to act strictly downwards (i.e. vertically), we can analyse the horizontal and vertical motions of a body entirely separately. It will fall vertically towards the centre of the Earth whether it is moving horizontally or not, and one motion does not influence the other. This remains true until it has moved far enough horizontally for the apparent direction of gravity (still towards the centre of the Earth) to have changed appreciably.

---

[25] This is Galileo's alleged experiment dropping cannonballs of varying weight off the Leaning Tower of Pisa.

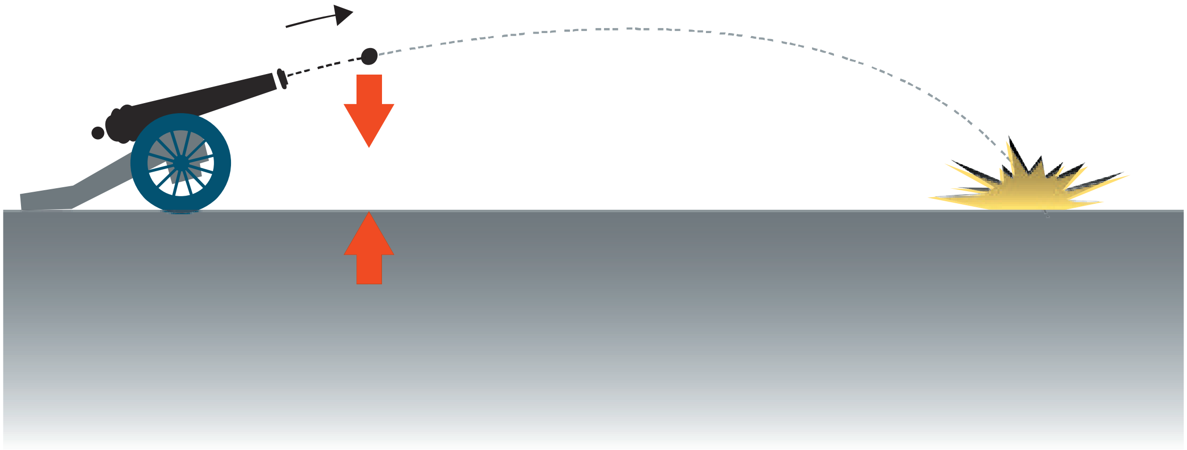


Illustration 1.12a – Gravity is Unaffected by Horizontal Motion. Gravity acts vertically (towards the centre of the Earth), regardless of horizontal motion. Thus the cannonball illustrated is being attracted downwards, even while it is moving upwards and forwards from the muzzle of the cannon. Ultimately, gravity wins the fight and the cannonball impacts the ground somewhere.

b. Imagine a cannon at the top of a high mountain. If you train the barrel horizontally, and if the ground surrounding the mountain is flat, when the cannon is fired, the shot will move forwards away from the cannon, and will simultaneously accelerate (fall) towards the Earth. Constant forward (horizontal) speed and constant vertical acceleration give it a curved path. The faster the shot emerges from the cannon (for example by using a bigger charge), the further forward it will travel before it hits the ground, increasing the range achieved.[26] The total time for the shot to reach the ground will not change; it accelerates vertically towards the ground at the same rate whatever its forward speed. It is the constant flight time at greater forward speed that equates to the greater range.

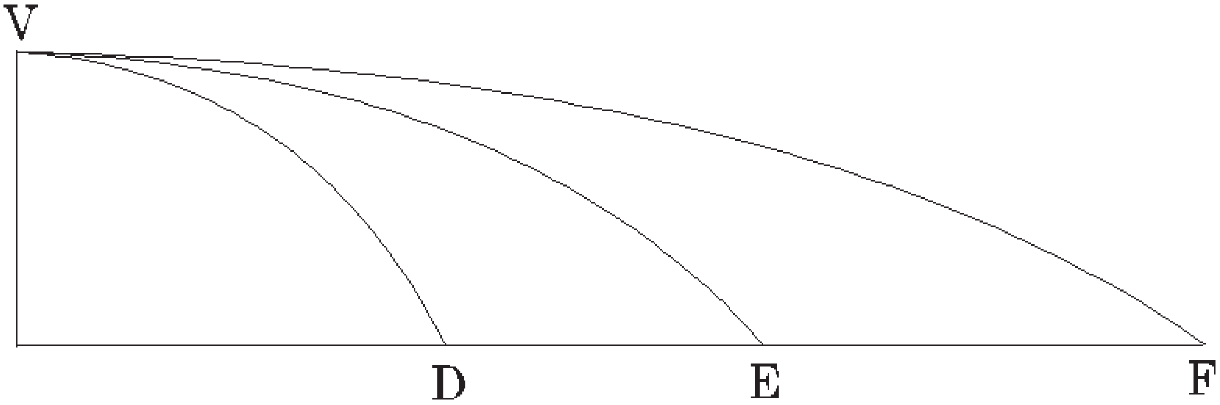


Illustration 1.12b – Parabolic Paths. This drawing by Isaac Newton, dating from the late 1680s, shows idealised cannon trajectories with increasing initial velocities over a flat surface. We show it here so that it is consistent with the more famous drawing at Illustration 1.12c from the same source.

c. Imagine now that we can achieve a truly colossal speed and range (by cannon standards). We would reach the stage where we could no longer assume that the terrain below the mountain was flat. While the cannonball was in flight, the Earth would be curving away below it, and would have to start allowing for this in predicting or measuring

---

[26] Air resistance would intrude in the real world and would have to be allowed for, but the 'more speed, more range' relationship would still apply.

our range. The range would be where two curved lines – the curved path of our cannonball and the curved surface of the Earth – intersect. Bigger and bigger charges would increase the radius of our curved path by increasing the forward speed, while the curvature of the Earth and the downward fall of the shot due to gravity would remain a constant.

d. Eventually, if we could make the shot go fast enough, the radius of curvature of its flight would equal that of the Earth. Although the shot would be falling towards the Earth, in exactly the same way as if it fell out the front of the barrel without any forward speed at all, the Earth would be curving away ahead of it at the same rate, and its altitude would thus remain constant. We thus have a body moving at a constant speed, but changing its direction continuously (i.e. accelerating), with the necessary force supplied by gravity. It is now orbiting the Earth (in a circular orbit).

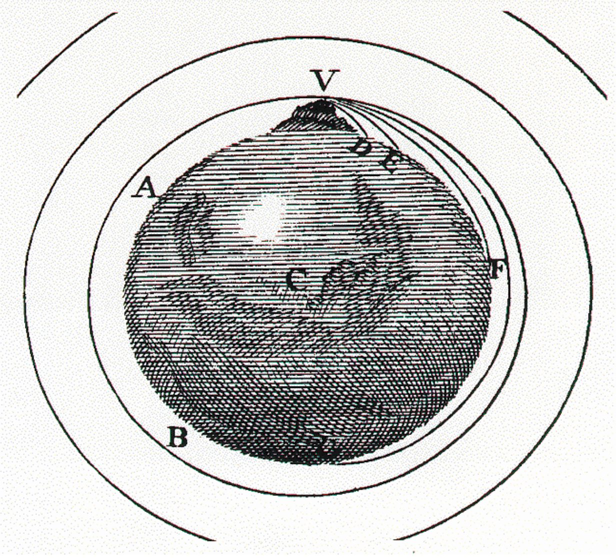


Illustration 1.12c – Newton's 'Cannon on the Mountain'. The illustration from Newton's publication *"A Treatise of the System of the World"* (c. 1689), drawing on ideas previously explained in the *Principia* (1686). The cannon fires shots of increasing velocity (compare the first 3 shown as impact points D-F with Illustration 1.12b above) until eventually, the curvature of the shot's path matches the curvature of the Earth, and although the cannonball is falling freely, the Earth is curving away below it at the same rate and it never reaches the ground. It has achieved orbital motion.

e. An explanation with numbers may be easier to follow (although it is the same principle being explained):

(1) In the first second of free fall, starting from rest, a body near the Earth's surface will fall just under 5m vertically. This is true whether it is moving horizontally at the time or not.

(2) The Earth's surface is curved, such that over 8 km in any given direction, if the Earth were smooth, it would fall away by about 5m.

(3) Suppose an object was moving horizontally at about 8 km per second as it began to fall. Over the 8 km distance travelled in a second, the object would drop 5m, but a smooth Earth would curve away below it by about 5m because of its shape. Thus at the end of that second, it would still be the same height above the ground.

(4) Assuming we can combat air resistance along both the horizontal and vertical directions (for example by moving the body into outer space and changing the numbers very slightly), the body will continue to fall towards the Earth indefinitely.

(5) We cannot, of course, orbit the Earth in one-second straight lines joined together; in the real world the motion is a smooth curve and you would have to analyse it in infinitesimal segments.[27] The principle above would still hold true, however.

f. If you repeat this analysis with accurate measurements inserted, the required orbital velocity for an orbit just above the Earth's surface translates from 8 km per second to just under 29 000 kilometres per hour (about 18,000 mph), and one orbit of the Earth would take about 84 minutes.

g. Since all falling bodies accelerate towards the Earth at the same rate, the speed required for orbit does not depend on the orbiting body. It is purely a function of the radius of the Earth and the height of the mountain (which dictate the curvature required) and the magnitude of the Earth's gravitational attraction. All of these are constants. We are not, however, getting 'something for nothing' here. A bigger cannonball will need a bigger charge in the cannon to accelerate it to the speed required.

h. Newton understood that this was a purely hypothetical experiment. He knew that the cannon was totally impractical for the velocity required, and that air resistance would decelerate the shot, but his fundamental concept was completely sound.

123. **Higher Speeds**. The curious reader might also wonder what would happen if you increased the speed of the shot beyond the minimum required for orbit? The answer is that initially the orbit would cease to be circular and would become elliptical, and that the speed and altitude in orbit would no longer be constant. The analysis to prove this is beyond our scope here (although we will see that the concept has great practical use). If we kept increasing the speed, the shot would eventually travel fast enough to overcome local gravity altogether and escape from Earth orbit (and the path would cease to be elliptical, becoming parabolic and then hyperbolic).[28] Along the correct path, the speed where the orbit transitions to a parabolic (i.e. an open, not closed) path is known as the escape velocity (although strictly it is a speed, not a velocity). The escape velocity from a given body depends on the body's mass (so for example you need to go faster to escape from the Earth than from the Moon), and on the starting location, so for example you need to go faster to escape directly from the surface of the Earth than from an existing closed orbit. In practical terms, even interplanetary probes, which escape from the Earth's gravity with the intention of being captured by a different body, do not usually escape from the Sun's gravity.[29]

124. **The Origins and Limitations of Kepler's Laws**.

a. **Johannes Kepler**. A German mathematician, Johannes Kepler (1571-1630), was the first person to analyse orbital motion rigorously. His analysis concerned the orbital motion

---

[27] Anything with the word *infinitesimal* in it suggests calculus to a mathematician, and Newton co-invented calculus, as well as formulating the laws of gravity and motion, so he was well placed to perform the detailed analysis alluded to here.

[28] The four shapes we have just described – circle, ellipse, parabola and hyperbola are the conic sections we alluded to in Paragraph 118 e. Note they occur in increasing order of eccentricity; further discussion of their geometry is in Annex A.

[29] Escape velocity from the Earth's surface is about 11.2 km per second, reducing to about 10.9 km per second from low-Earth orbit (roughly 24 000mph, compared with 18 000 mph for orbit). At the distance of the Earth's orbit from the Sun, the escape velocity for the Sun's gravity (i.e. to break out of the Solar System altogether) is about 42 km per second. This speed *has* been achieved by deep-space probes such as the NASA *Pioneer* and *Voyager* series.

of the planets around the Sun, but the principles governing these are exactly the same as those governing satellites orbiting the Earth, or indeed any other celestial body. Kepler was analysing older records of planetary positions, as observed from Earth, to try and find the shape of their orbits. His breakthrough was to realise that only an elliptical track fitted the observations to theory and allowed accurate prediction of future positions.[30]

b. **Kepler's Laws and Implicit Assumptions – the Restricted Two-body Problem**. Having arrived at this conclusion, Kepler was able to make deductions about its implications. He codified these as three laws (Kepler's Laws) governing orbital motion. As we commonly use them now, certain assumptions are implicit. The two most important assumptions are that one body is much more massive than the other; this allows us to consider the large body to be stationary.[31] The other is that the system only contains two bodies, so we ignore any influences from other celestial bodies. These are generally good approximations for any system we are likely to encounter; they are described as defining the *restricted two-body problem*.

125. **Kepler's Laws Summarised**. Within these constraints, Kepler's laws state:

a. The small body will orbit the big body following an elliptical path. The centre of the big body will lie at one focus of the ellipse. Remember that the ellipse could be a circle.[32]

b. The speed of the small body will vary in a specific way as it traverses the ellipse.

c. There is a fixed linkage between the size of the orbit (actually the length of the major axis) and the time taken for one orbit (for any given big body).

We can draw useful conclusions from each of these laws.

126. **Kepler's First Law**. We have already explored the geometry of ellipses and the definition of a focus.[33] The Earth's centre will be at one focus for every satellite orbiting it. The other focus, which will potentially be different for each satellite, will be un-occupied. It is a purely hypothetical construct which might lie inside or outside the Earth's surface.

a. The radius of the Earth (about 6400 km) puts a minimum limit on the size of an orbit, as the Earth needs to fit completely inside the ellipse. This minimum is the orbit we envisaged in Newton's thought experiment.[34] From our discussion of atmospheric drag,[35] we also need to add some clearance between the Earth's surface and the orbit – at least 150 km. This is a constraint of the atmosphere, however, not of Kepler's laws.

b. Finally, and perhaps most importantly, since the focus lies in the plane of the ellipse, the plane of the ellipse/orbit must include the centre of the Earth. This is where the Earth's attractive force (gravity), which sustains the orbit, originates.[36] So an orbit could, for example, lie directly above the equator, or equally could pass over both poles. The plane

---

[30] See Annex B for the historical context of Kepler's discoveries, but note that Kepler pre-dated Newton, and consequently was working without the theoretical background of gravity and Newton's laws of motion.

[31] See Paragraph 121c.

[32] See Paragraph 117.

[33] See Paragraphs 116-118.

[34] See Paragraph 122.

[35] See Paragraph 111.

[36] See Paragraph 121d.

cutting the Earth at one of the tropics, however, does not include the centre of the Earth, so a satellite cannot maintain orbit in this plane.

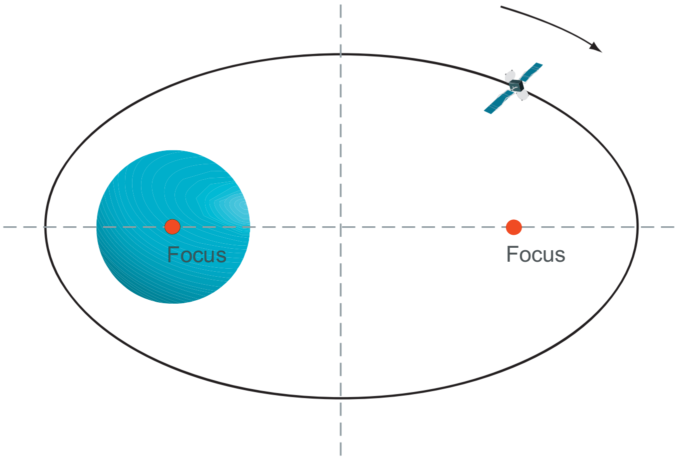


Illustration 1.13 – Kepler's First Law. An arbitrary satellite in an elliptical orbit. Note that the diagram is 'perspective-neutral'. So long as the orbit is clear of the planet's surface (and in reality clear of the atmosphere if applicable), it can lie over the poles, around the equator or at any other orientation, subject only to the *centre* of the planet lying in the plane of the orbit. If the orbit were more circular, the empty focus might lie below the surface of the planet, but that is of no consequence. If the orbit were truly circular, both foci would lie at the centre of the planet.

127. **Kepler's Second Law**. Kepler's second law allows you to calculate the speed of a satellite at any instant on an elliptical orbit. The detail says that if you draw a line connecting the centre of the satellite with the centre of the Earth, then wherever the satellite is in its orbit, and for any given unit of time, the line will sweep out a constant area of the ellipse. In practice this means that for an elliptical orbit, the satellite will travel slowest when it is furthest from the Earth, and fastest when it is closest. The line joining the centres is commonly called the radius vector.

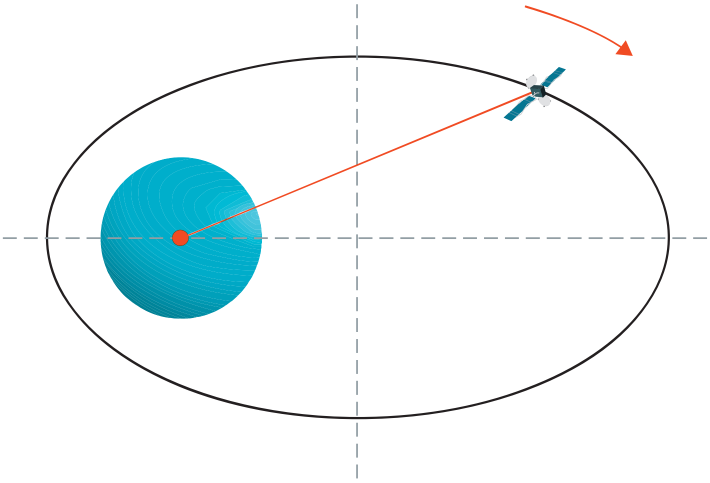


Illustration 1.14a – Kepler's Second Law. The arbitrary satellite in orbit, with the radius vector marked in red. As the satellite moves around its orbit, the radius vector sweeps out an area within the orbit. Kepler's Second Law says that this area is constant for a given orbit and given time. Since the length of the radius vector changes during an elliptical orbit, this implies that the speed of the satellite must change too.

a. We are not usually interested in the area swept out by the radius vector *per se*. What we are generally interested in is the speed of flight of the satellite. The speed varies greatly in a highly eccentric orbit, with the satellite moving slowly at one end and quickly at the other. Note, however, that the satellite does this solely under the influence of gravity, without any other thrust or braking, in the same way as a pendulum moves quickly at the bottom of its swing and slowly at the extremities without external interference. The ‘swept area’ is just a convenient method of working out what the speeds are.

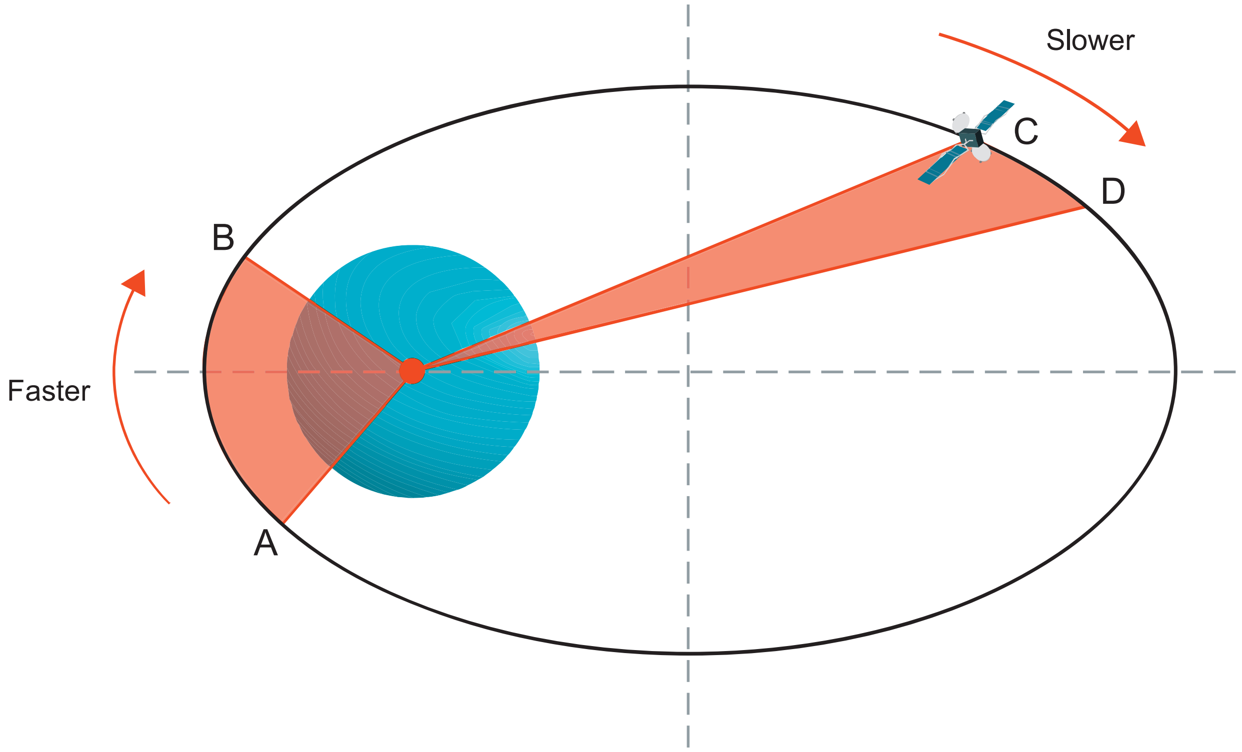


Illustration 1.14b – Kepler’s Second Law. Imagine that the satellite had travelled from A to B in its orbit above in 15 minutes, and that it is now at point C. Kepler’s Second Law says that its speed around the orbit is changing such that in 15 minutes from point C, it will have reached whatever point D in the orbit makes the two shaded segments have the same area. That implies that from A to B, it was moving faster than from C to D, since the distance AB is plainly greater than CD, to compensate for the radius vector being much shorter around A and B. It also implies that the speed is changing continually, since our choice of 15 minutes is entirely arbitrary. It would still be true if we had used 15 seconds, or even 1/15th of a second as our time interval, though plainly the satellite will cover less of its orbit in such a short period of time. Thus at A, the satellite’s speed is relatively high, and increasing, and at D it is relatively slow and decreasing. The maximum and minimum speeds will occur at perigee and apogee (see Paragraph 129) respectively.

b. The second law also implies that the speed of a satellite in a circular orbit is constant. A useful analogy here is the minute hand of a clock. The tip moves at a constant speed and the minute hand sweeps out identical areas every 5, or 15 or 30 minutes (or any other interval you choose).

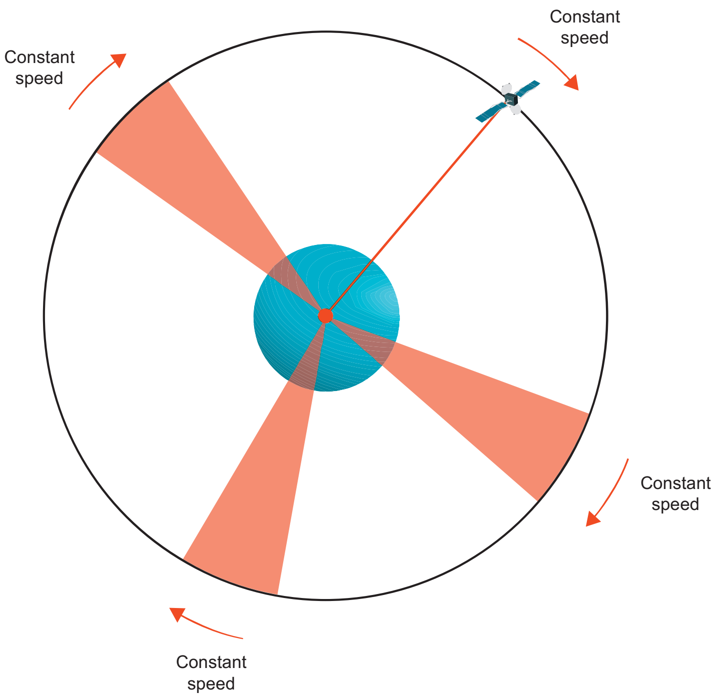


Illustration 1.14c – Kepler's Second Law. The speed of a satellite must be constant in a circular orbit, since this is the only way that all the shaded segments (and any others you care to add) will have the same area. Thus you cannot, for example, change the speed in a circular orbit so that the satellite lingers over a particular area of interest, or hastens over a less useful area.

128. **Kepler's Third Law**. The third law is the most 'mathematical' of the laws. It states that the cube of the semi-major axis varies with the square of the period (the time taken for one orbit).

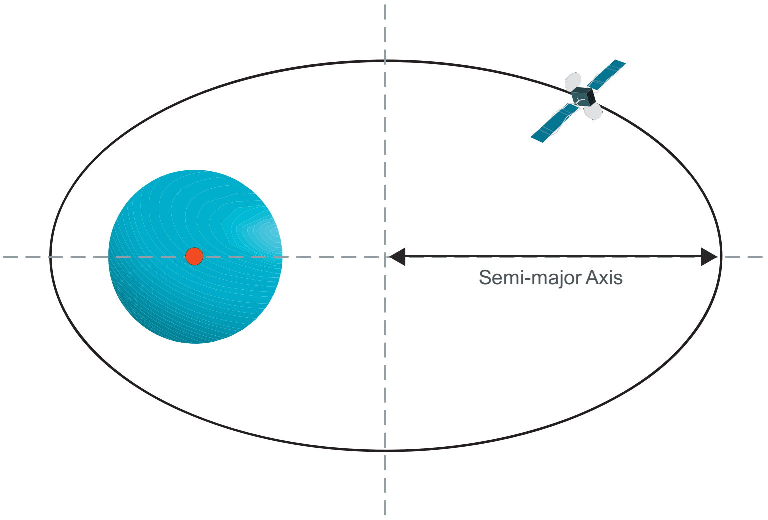


Illustration 1.15a – Kepler's Third Law. Kepler's third law says that there is a fixed relationship between the length of the semi-major axis of an orbit (see Illustration 1.9 to recap the axes of an ellipse) and the period. So for a given planet, if two satellites have orbits with the same length of semi-major axis (they could still have radically different shapes), they have the same period (This is shown in Illustration 1.15b). Additionally, 2 satellites in the *same* orbit, where the semi-major axis is plainly the same, must have the same period, or more practically, there is only one period for any given orbit.

a. The most immediate practical implication of this is that there is a fixed relationship between the 2 quantities; it does not matter what the relationship actually is. So any two satellites in the same orbit must have the same period, and following from the second law must travel at the same speed at any given point on it.

b. Note that the rule also implies that if the semi-major axis gets bigger, the period gets longer. Equally, for circular orbits, if the radius gets bigger, the period gets longer. Our 84 minute orbit above, even if it were practical, would be the fastest one could orbit the Earth. By the time we enlarge the orbit to allow for the atmosphere, the period has increased, and as we keep increasing altitude, orbits keep taking longer. We cannot speed the satellite up to get round quicker; that would change the shape of the orbit rather than achieving a faster period of rotation.

c. Finally note the use of the semi-major axis. This means that eccentricity does not alter the period. An elliptical orbit will have the same period as a circular one, if the semi-major axis of the elliptical orbit is the same length as the radius of the circular one.

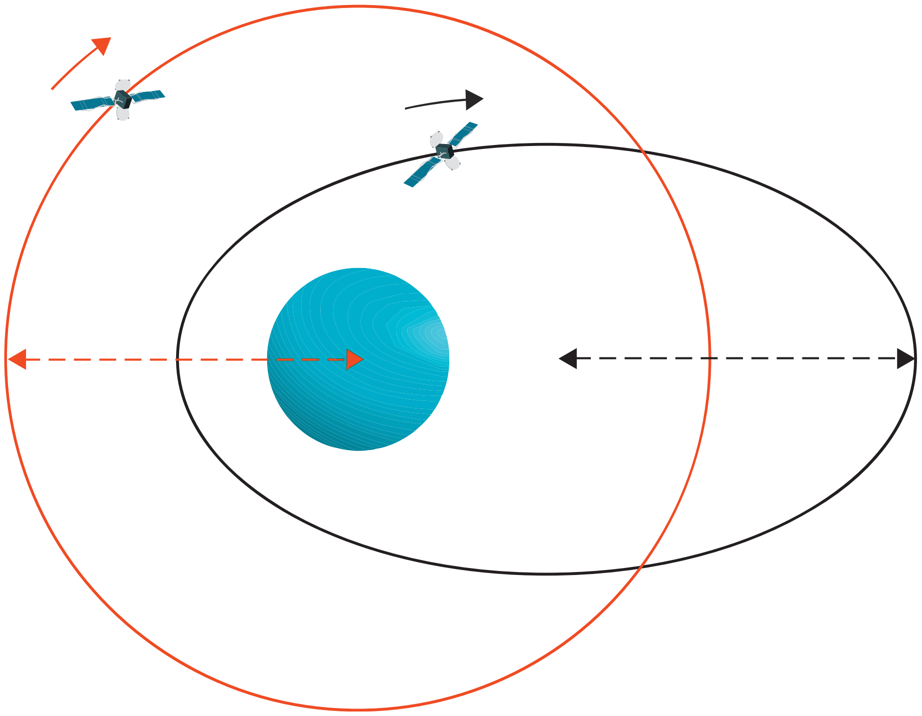


Illustration 1.15b – Kepler's Third Law. In the diagram above, the red dotted line (radius of the circular orbit) and the black dotted line (semi-major axis of the elliptical orbit) are the same length. This implies that the period of both satellites will be the same. (It does not imply that they move at the same speed at any instant. As we saw above, due to Kepler's Second Law, the satellite in the circular orbit will move at constant speed, while the satellite in the elliptical orbit will change speed continually. It is the *total time for one orbit* that will be the same for both).

129. **Apogee and Perigee**. Apart from implying that the Earth must fit inside the orbit, Kepler's laws say nothing about the altitude of a satellite as measured from the surface of the Earth. For many applications, however, this is a very important quantity. In an elliptical orbit, and assuming a spherical Earth, the satellite's altitude will vary continuously. As we saw above, the satellite has its highest speed when it is closest to the centre of the Earth (and thus at the same instant closest to the surface). This point is called the perigee of the orbit. The perigee lies on the major axis of the orbit,

as does the point where the satellite is at the greatest altitude. This latter point, where the speed is lowest, is known as the apogee.[37]

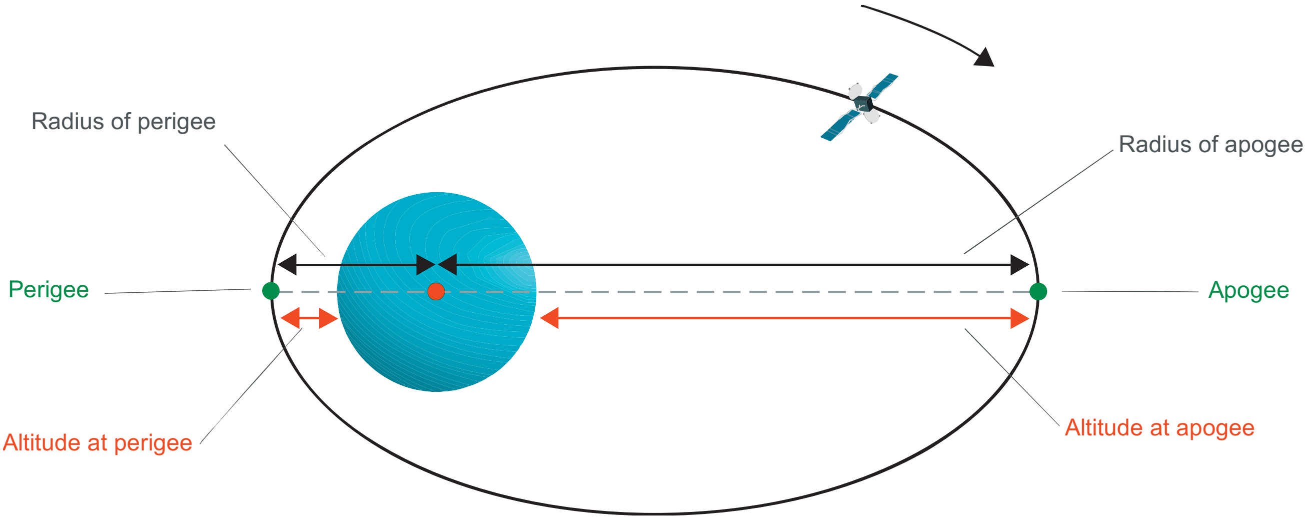


Illustration 1.16 – Apogee and Perigee. The point in a satellite's orbit nearest to the surface of the Earth is known as *perigee,* and the point where it is furthest away is *apogee.* The position of the perigee in space is very important for describing an orbit in a rigorous manner. When the orbit is circular, as we note below, there is no apogee or perigee, so alternative methods have to be used to describe the orbit. Finally, although the term *radius/radii* suggests a circle, rather than an ellipse, the terms *radius of apogee* and *radius of perigee* are understood in the sense shown above.

130. **Orbital Elements**. We have discussed so far the various characteristics of orbits, and it should be obvious that there are an essentially limitless number of possible orbits for a satellite around the Earth. There is therefore a clear need for a standard method of describing an orbit explicitly. We describe the standard method briefly here; those seeking further information should refer to Annex C. To characterise an orbit around the Earth takes 6 numbers. You will see below that for analytical purposes at our level, they are of varying importance.

a. **Describing the Orbit in Space**. Two numbers describe the size and shape of the ellipse in isolation, without locating it around a planet or placing the satellite on it. The convention is to use the length of the **semi-major axis** and the value of the **eccentricity**, as we described above.

b. **Fixing the Orbital Plane**. Two numbers describe where the plane of the orbit lies in space. Because the Earth rotates while the satellite orbits, the definition needs to be located in space, rather than being tied to the surface of the Earth. The elements that do this are known as **inclination**, and **right ascension of the ascending node**. For our purposes, it is more important to understand inclination; see below for more details.

c. **Placing the Orbit on the Plane**. The fifth parameter orients the ellipse on the plane. It is known as the **argument of perigee**.[38] Roughly, it says where on the orbital plane you can find the major axis.

---

[37] The '-gee' portion of the name ties these terms to Earth orbit only. Other terms are used for other bodies, for example 'perihelion' and 'aphelion' for bodies orbiting the Sun. The generic term '-apsis' (plural 'apsides'), and thus *periapsis* and *apoapsis*, which deliberately does not define the orbited body, may also sometimes be encountered.

[38] 'perigee' ties this element to the Earth (see Paragraph 129 and footnotes). The term 'argument of apsides' (*apsides* being a generic term) may thus used for satellites around other bodies, if a specific term for that body is not in common use.

d. **Placing the Satellite on the Orbit**. Finally, you need to define where to find the satellite on the orbit at a given time. The parameter is known as the **true anomaly**.

e. **The Orbital Elements**. These numbers together are called the **orbital elements**. A full set of all six defines an orbit uniquely and fixes the position of the satellite at any given instant.[39]

f. **Un-defined Elements**. There are some circumstances where the elements we have described are undefined. For example, a circular orbit has no major axis (or rather every axis is the major axis) and consequently no perigee. Thus the argument of perigee is undefined. There are ways to circumvent this by defining alternate elements, but we will not examine these further here.

g. **Useful Elements**. To make useful deductions about an orbit, and its implications for the satellite's payload, we need usually only consider the orbit's size and shape (semi-major axis and eccentricity), and its inclination. These elements are always defined, even for circular orbits.

h. **Two-line elements (TLEs)**. There are standardised ways of writing down the 6 Keplerian elements – a sort of shorthand in common use in space operations. The most common is to use the format standardised by the North American Air Defense Command (NORAD) and by NASA. This format dates back to the earliest days of spaceflight, and was originally intended to standardise how orbital elements were formatted on 80-column punched-cards for entry into early computers. The standardisation included the order the elements were listed in and units used. Because the six elements (and the subsidiary information) would display on two lines of printed text, they became known as two-line elements, usually abbreviated to TLEs. The description has stuck, and TLEs are still a widely used format for publishing orbital data, even though punched-card readers are now museum curiosities.

131. **More About Inclination**. Inclination is the angle between the plane of the orbit and the plane of the equator.[40] It is measured in degrees between 0° and 180°. An orbit that lies directly above the equator, with the satellite moving in the same direction as the Earth's rotation (though of any size and eccentricity) has an inclination of 0°. An orbit that lies over the poles, crossing the equator at right angles has an inclination of 90°. Inclinations between 90° and 180° indicate that the satellite is orbiting in the opposite direction to the rotation of the Earth. Orbits matching the Earth's direction of rotation (i.e. up to 90º inclination) are said to be **prograde** orbits; those in the opposite direction (i.e. inclinations greater than 90º) are said to be **retrograde**.

---

[39] An analogy is that it takes 6 values to characterise an aircraft's position and motion exactly, for example latitude, longitude, height, speed, track and rate of climb or descent.

[40] Do not confuse this quantity with the inclination of the Earth's axis of rotation to the plane of its orbit around the Sun. That inclination (23.5º) gives rise to the seasons, and the varying lengths of day and night at different times and places on Earth. When we are considering Earth-orbiting satellites, however, we only relate the plane of the orbit to the plane of the Earth's equator, ignoring the relative location of the Earth-Sun system. The location and visibility of the sun from the satellite may be important, for example for power generation purposes via solar panels, but it does not affect the orbital dynamics directly.

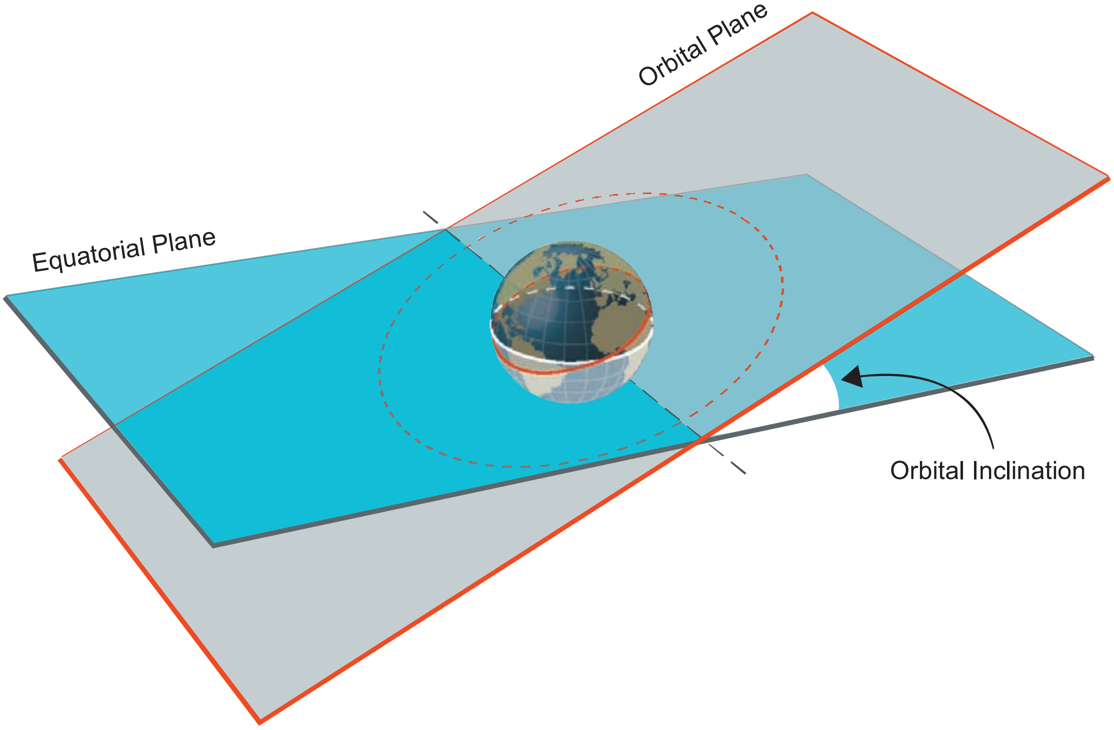


Illustration 1.17 – Orbital Inclination. The diagram above is the same basic drawing as Illustration 1.10a above, where we showed the intersection between two planes and the angle between them. The blue plane is the Equatorial plane of the Earth, the grey plane is the Orbital plane of the satellite. The orbital inclination is the angle between the 2 planes.

From Orbit to Ground Track

132. **The Rotating Earth**. The Earth's rotation has important practical implications for satellite launch (see Paragraph 164g), but once aloft, none of the dynamic principles we have examined depend on it. Gravity would exist whether the Earth rotated or not, so even if the Earth were stationary on its axis, it would still be possible to place objects in orbit around it. However, the Earth does rotate, and we need to think how that affects applications of spaceflight. The main effect is when we examine the path of a satellite over the Earth's surface. The problem is similar to that of a mapmaker, *viz* how to project a curve or series of curves onto a flat surface. We will build up our analysis in stages:

a. Imagine a satellite in a circular orbit over the equator, orbiting in the same direction as the Earth rotates. We saw (Paragraph 122f) that, for an orbit close to the Earth's surface, the orbital period would be slightly greater than 84 minutes. Since it takes the Earth 24 hours, rather than 84 minutes, to rotate on its axis, the satellite would overtake the Earth below it, (in the same way as the minute hand on a clock overtakes the hour hand about once every 65 minutes). If we plotted the ground track on a map, it would be a straight line, drawn over the equator. If we could somehow monitor the position of the satellite continuously, it would move smoothly along the line.

Illustration 1.18a – Ground Track for a Circular, Equatorial Orbit. The satellite above is in orbit over the equator. For a low-earth orbit, its period will probably be in the range 90-120 minutes (for clarity, the satellite in the illustration is drawn at a much greater altitude). Its path will lie above the equator, overtaking the Earth smoothly.

b. Now think again about an inclined circular orbit. Let us say, for example, that the inclination was 23.5°.[41] Imagine now looking down on the Earth and the satellite from above the equator as the satellite crossed it. The inclination means that the satellite now flies over land north of the equator. The plane of its orbit includes the centre of the Earth, so the ground track will move gradually north until it reaches latitude 23.5°N. It will then reverse, cross the equator heading south, reach latitude 23.5°S and return to the equator. In general, the latitude limits of the ground track are set by the inclination; if an orbit has inclination ‘x’, its ground track will cover ground between ‘x’°N, and ‘x’°S. If the inclination of the orbit is greater than 90º, i.e. a retrograde orbit in the opposite direction to the Earth’s rotation, the limit is given by (180º - inclination). So for an orbit at 135º inclination, the overflight limit is given by (180-135) = 45º North and South.

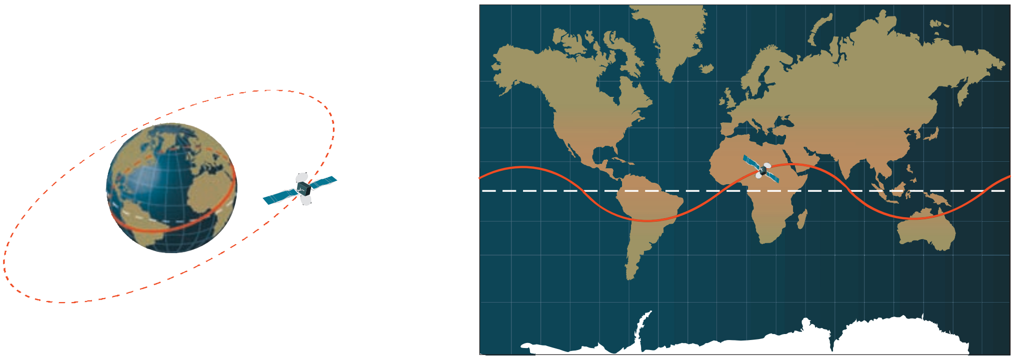


Illustration 1.18b – Ground Track for a Circular, Inclined Orbit. The satellite illustrated above is in a circular, inclined orbit. It will spend half its time over the northern hemisphere, and half over the southern hemisphere (a consequence of the plane of the orbit passing through the *centre* of the Earth, in accordance with Kepler’s First Law). The inclination of the orbit (see Illustration 1.17) limits the latitude of the points it can overfly. In the example shown, the satellite orbit is inclined at 20º to the Equatorial Plane, and the ground track is restricted to lie between latitudes 20ºN and 20ºS.

c. How this translates to a path over the ground depends on the period of the orbit. Assuming that it is still relatively short compared to the Earth’s rotation, then in one orbit it will make constant progress from west to east (as in the equatorial example above), while

---

[41] The latitude of the tropics.

simultaneously moving north and south between the tropics. On most maps, this would be depicted as a line making an s-shaped curve between the tropics. After one complete orbit, the satellite has been once around the Earth, while the Earth has completed some part of its daily rotation. It will thus cross the equator at a different place to its first crossing, and the next s-curve will begin from a different place. We therefore need to redraw the path for second and subsequent orbits.

d. The path or ground track has a constant shape, but begins from a different location each time. A computer display might only show the current and possibly the next orbit in the series, re-plotting these after each orbit. If the period of the orbit divides exactly into the Earth's period of rotation, the ground track will eventually repeat, and it may be possible to show it as a closed cycle on the map without too much clutter. We will look in Chapter 3 at how this can be exploited for several applications. Such orbits, where a whole number of orbits occur in a single day, or small multiple of days, and the ground tracks thus repeat in a closed cycle, are sometimes known as *resonant orbits.*

e. For high-inclination orbits, we may run into practical difficulties plotting orbits in the polar regions, due to inherent distortions in many commonly used maps. When the orbit is directly over the poles, east-west motion is due solely to the Earth's rotation, and the ground-track will appear on the map as a series of almost north-south lines, gently curved at high latitudes. It may not be obvious that the orbits are in fact continuous, depending on how the map depicts the polar regions.

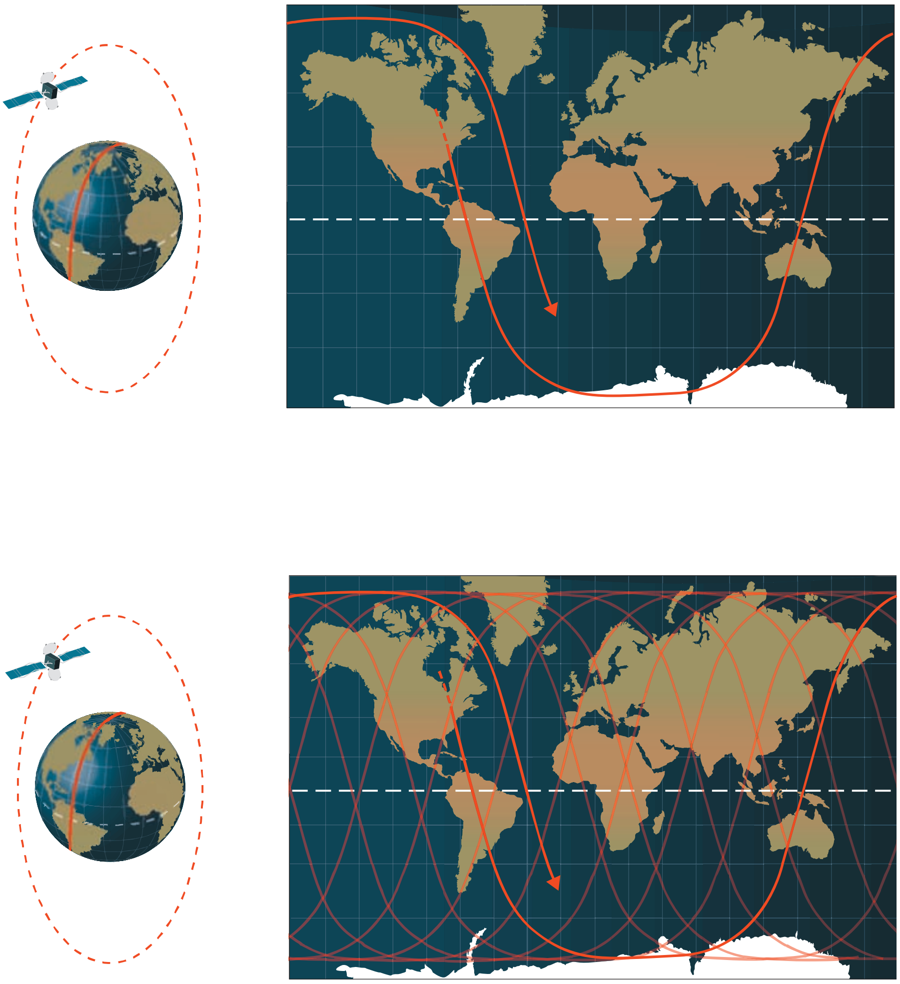


Illustration 1.18c – Ground Track for a Circular, Highly-inclined (near-Polar) Orbit. The upper illustration shows a single orbit of a satellite in a circular, highly-inclined orbit. The satellite orbits the Earth, overflying the two polar regions; in the time it takes it to do this, the Earth rotates on its axis by some amount. Thus, on the subsequent orbit, the ground track is displaced. The lower illustration shows several consecutive orbits overlaid on the Earth. Eventually, the orbital path may repeat. The orbital altitude (and thus the period) can be adjusted to control whether and how often this happens. If the orbit was less steeply inclined, you would see the same overlapping tracks, but constrained in latitude by the angle of inclination, as in Illustration 1.18b.

f. We run into other problems describing inclined elliptical orbits. Until now, we have assumed that the satellite is moving faster than the surface of the Earth is rotating, and thus overtakes it at a more or less constant rate. In an elliptical orbit, with the speed of the satellite changing continuously, the ratio between the Earth's rotation and the satellite's rotation is not a constant, and this affects the ground track. If the inclination is greater than 90°, we have already noted that the satellite orbits in the opposite direction to the Earth's rotation, which plainly will have a major impact on the apparent ground track. These features do, however, have useful practical applications.

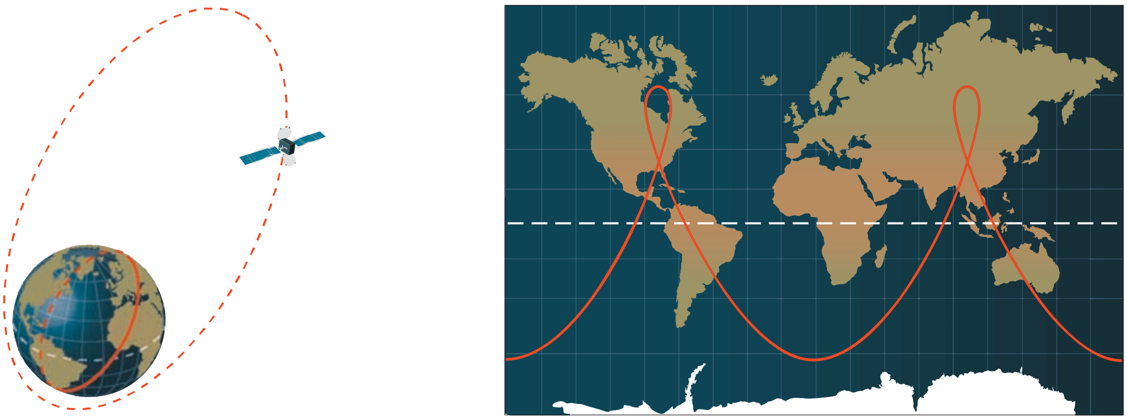


Illustration 1.18d – Ground Track for an Elliptical, Inclined Orbit. In accordance with Kepler's Second Law, the speed of a satellite in an elliptical orbit will change throughout its path. This greatly changes its path over the Earth during its orbit; it may at times be rotating faster than the Earth, at other times slower, so its path may at times appear to change direction, and its apparent motion seen from the surface may at times be almost stationary, at other times rapid. In the example shown above, the two tight loops over Canada and Central Asia will take longer than the wide arcs over the southern oceans. This feature has great practical utility. The limitations of latitude relating to inclination outlined above (sub-paragraph b and Illustration 1.18b) still apply, however.

g. Finally, we shall shortly see how satellites can be made to hold station at points over the Earth's equator. These plainly do not have a ground track; rather they have a location which can be plotted in a conventional fashion.

133. **Footprint**. The other property we may wish to plot for a satellite is its footprint. This is the area under the satellite, and thus potentially visible to it. We shall look more closely at this in Chapter 3, but for now note that when it is plotted on a map, it may be in addition to its ground track. It may reflect the distance from the satellite to the horizon, or it might reflect some characteristic of the payload, such as a line of sight of a sensor system. Measured from the ground looking up is the related quantity, the ground-elevation angle. Depending on its value, it may indicate whether the satellite passes directly overhead (when the angle is 90°), or reaches a given angle above the horizon (lesser values, but greater than zero). See also Paragraph 144a and Illustrations 1.22a & b.

134. **Revisit time**. A frequent question with obvious practical importance is how often a satellite over-flies a given point on the Earth's surface. For a practical mission, this often dictates the required orbit. The answer is, however, a complex trade-off between frequency and coverage.

a. Firstly, the higher the orbit altitude, the greater the period, and the greater the minimum interval between revisiting any point, though the longer the duration of each over-flight too.

b. Secondly, if the orbit is synchronised to the Earth's rotation, (i.e. the period is some simple fraction or multiple of 24 hours), then some points are regularly over-flown, while others are never under its track. If this is not the case, the ground track deliberately does not repeat, allowing the satellite to survey different terrain each orbit and over time potentially to achieve global coverage.

Such a satellite is still constrained by its inclination. A satellite in an inclined orbit has its ground track restricted in latitude by its inclination, so for example, in a 20°-inclined orbit, the ground track is constrained to lie between 20°N and 20°S. If the inclination is increased to cover more of the Earth's surface, given that the satellite is moving at a constant speed, the interval between revisits must increase. We will look in more detail at counters to this when we consider manoeuvring in orbit later in this section.

Revisit Time Calculations – A Vignette[42]

- The Indian IRS 1A satellite was designed to provide global earth-resources monitoring from relatively low altitude. To achieve the 'global' aim, it worked from a polar orbit (any lower inclination would leave it unable to survey the areas at a higher latitude than its inclination). The actual inclination used was 99º, so there is in fact a very small 'blind spot' around the North and South poles (remember from Paragraph 132b above that you can treat 99º inclination as being equivalent to (180-99) = 81º for ground overflight calculations) – the satellite will not overfly points beyond 81º N or S on the Earth's surface.

- The chosen orbit circles the Earth just under 14 times a day. An exact 14 orbits per day would yield a period of 102.8 minutes (24 hrs x 60 mins/14 orbits = 102.8 mins). IRS 1A is flown at an altitude to give a period of 103.2 minutes. The 0.4 minutes discrepancy is deliberate.

- Imagine the satellite crossing the equator over a given point. One orbit later, the satellite crosses the equator again in the same direction, having overflown both polar regions in the meantime.

- In the 103.2 minutes that this takes, the Earth has rotated by just over 1/14th of a full revolution (about 25.8º) on its axis. So the second crossing point is 25.8º west of the first. At the equator, this is equivalent to 2872 km displacement, and each successive equatorial crossing is a further 2872 km to the west of its predecessor.

- After 14 orbits, the equatorial crossing point is near the original point again, but since the period has been chosen to be very nearly, but not exactly a simple fraction of 24 hours, the following day's pattern is very slightly displaced from its predecessor. For the figures given, the daily displacement is actually about 130 km, or 1.17º of longitude. It is no accident that the IRS 1A sensor has a field of view about 145 km wide from its working altitude.

- Each daily cycle of 14 orbits thus mimics its predecessor, but shifted about 130 km westward at the equator. Given the 2872 km gap between adjacent orbits, and with the satellite moving 130 km westward per day, it takes 22 days for the satellite to cross this gap. After 22 days, which equates to 307 orbits, the satellite is exactly over a previous crossing point, and the cycle repeats. Since the overlap between daily cycles was calculated with the field of view of the sensor in mind, the satellite has flown within survey range of every point on the Earth's surface, less the small missed areas at the poles.

135. **Visualising Orbits**. At this point, you may be relieved to learn that while the concept may seem abstract, it is a relatively simple task given modern computers to generate maps and diagrams showing overflight paths, pass times, footprints and similar. There are several commercial software packages available that can generate a variety of presentations. All that is required is access to the orbital elements for the satellite in question. We will look in Chapter 3 at how this information can be obtained for non-cooperative targets.

---

[42] With acknowledgements to George Joseph, *Fundamentals of Remote Sensing* (2nd Edition) (Hyderabad: Universities Press (India) 2005), Chapter 8.

Necessary Complications

136. **Deviations from the Ideal**. Our explanation of orbital motion so far has depended on several assumptions; some of these assumptions are not completely justified, and we now need to analyse the implications of that. We will look at three significant deviations from the ideal: the non-spherical Earth, the effect of atmospheric drag on elliptical orbits, and the effect of mass-concentrations in the Earth on geo-stationary orbit.

137. **The Oblate Earth**. Having derived the theory behind orbits using the assumptions that the Earth is spherical, and that gravitational forces originate at its centre, we need to look at the consequences of those assumptions not being quite true.

a. **The Earth's Figure**. The exact shape of the Earth is referred to as its *figure*. It is a very complex and subtly irregular shape when seen in detail.

b. **Oblateness**. The Earth is fatter around the equator than elsewhere.[43] This is *oblateness.* The oblateness is not great; the radius of the Earth measured at the equator is about 20 km more than the average measured across the whole of the Earth's surface.[44] Nonetheless, the extra Earth-mass in the equatorial bulge causes gravitational effects on satellites. Although the bulge is relatively small, the effects it produces can be very significant. One useful way to visualise this is to consider a spherical Earth with the extra mass of the equatorial bulge concentrated like a lifebelt. This is not a representation of the true shape of the bulge, but may help explain its effects (Illustration 1.19a)

c. **Implications for Inclined Orbits**. When a satellite is in an inclined orbit and is north or south of the equator, the apparent point of origin of gravity changes (Illustration 1.19b). The satellite is affected by the normal attractive force towards the centre of the Earth, and an extra, much smaller, attractive force towards the bulge. This force is directed out of the plane of the orbit. A simple analysis would suggest that this would simply pull the plane of the satellite's orbit down towards the equatorial plane, reducing the inclination. However, because the satellite is rotating around the Earth at great speed, it is effectively a gyroscope and thus behaves in a similar way to all gyroscopes.

---

[43] The resulting effects are sometimes said to be due to the 'J2' term. This relates to an equation accounting for several different perturbations of an orbit. The 'Earth-oblateness' term is the largest of these, and is conventionally denoted 'J2' in the equation.

[44] This flattening results from a combination of gravity pulling the Earth into a spherical shape and the Earth's rotation on its axis distorting that ideal. It is possible to derive the theoretical extent of the equatorial bulge based on assumptions about the density of the Earth at different depths and its period of rotation and size, but the Earth does not conform exactly to such models, suggesting that our understanding of its density distribution is still incomplete.

Illustration 1.19a - The Oblate Earth. The consequence of the Earth's equatorial bulge is that gravity does not appear to originate directly from its centre. One way to think about this is to imagine the gravity originating from a spherical Earth, and an extra component arising from the bulge around the equator.

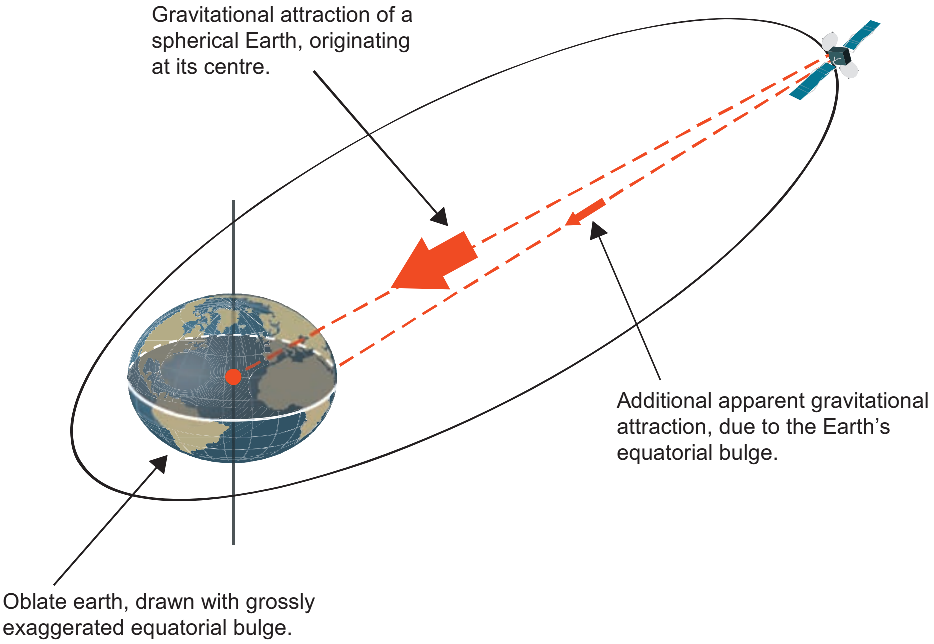


Illustration 1.19b – Apparent Gravity. The illustration grossly exaggerates the proportions of the bulge and the relative effects, but shows the two components. When added together, they constitute a small out-of-plane force. The result affects every satellite unless it is orbiting directly over the poles; in this last case, the extra force would actually be 'in-plane'.

138. **Precession**. If a force is applied to a gyroscope in an attempt to change its plane of rotation, the gyroscope resists the force, and instead responds as if the force had been applied 90º further around its rotation – this is known as *precession*. A satellite responds to the attraction of the Earth's equatorial bulge in the same way. The inclination of the orbit is not reduced, and instead two different effects emerge:

139. **Nodal Regression**. This effect causes the *right ascension of the ascending node* (see Paragraph 130b and Annex C) to regress. The equatorial bulge applies a pull towards the equatorial plane to the satellite, but instead of the inclination reducing (the simplistic analysis of what might happen), gyroscopic precession makes the orbital plane of the satellite rotate. This is measured by looking where the orbital plane crosses the equatorial plane (the *ascending node*). That direction

rotates in space due to the effect of the equatorial bulge. Three quantities govern how quickly this occurs: [45]

a. **Inclination**. At 90º inclination, no regression occurs, since there is no force outside the plane of the orbit. As the inclination reduces towards 0º, the rate of regression *increases*. You can persuade yourself that this should be so, since at lower inclinations the satellite spends longer closer to the equatorial bulge, so the effect is greater. For truly equatorial orbits (inclination = 0º), the analogy breaks down, but since in an equatorial orbit there is no ascending node, regression can thus be ignored.

b. **Altitude**. By a similar argument to the one above, at lower altitudes the satellite is closer to the bulge. Since this is a gravitational effect, which reduces as $1/r,^2$ the closer the satellite, the greater the effect. The rate of nodal regression thus *increases* at lower altitudes.

c. **Eccentricity**. Finally, eccentricity affects the rate of regression, although the analysis of how it does so is complex. Many practical orbits are deliberately circular (eccentricity = 0), so it can be ignored as a factor. As we explain in Paragraph 146 (footnote 59), orbits near the Earth's surface – known as Low-Earth Orbits (LEOs) – cannot deviate too much from circularity, or else the Earth would intrude. Where regression is a useful effect (we will see momentarily that it can be exploited deliberately), it is induced and controlled by manipulating altitude and inclination. In many cases it is sufficient to assume that the orbit is circular and account only for the first two factors above.

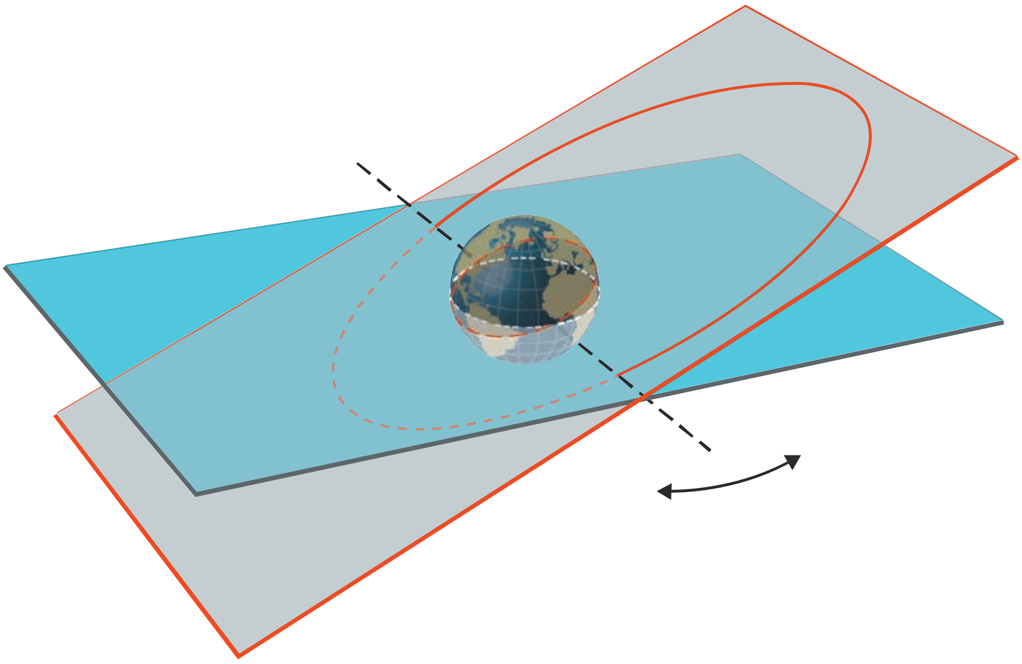


Illustration 1.20a – Nodal Regression. Because of its orbital motion, a satellite responds to the gravitational attraction of the equatorial bulge in the same way as a gyroscope would. Thus instead of dragging the orbital plane towards the equatorial plane (the simplistic analysis of the extra pull), the effect is that the orbital plane (and thus the line of nodes) rotates around the Earth. This effect is true both for circular and elliptical orbits, as long as they are not at right angles to the equatorial plane (i.e. it does not apply to polar orbits).

---

[45] Nodal regression can easily approach 10º per day for a low-inclination, low altitude orbit, so the effect is far too big to ignore. Readers wishing to delve deeper into the values and how they vary will find summary graphs in the reference texts mentioned in the Bibliography.

140. **Perigee Rotation**. It is harder to derive a simple explanation of why Earth-oblateness causes the major axis of the orbit to rotate in the orbital plane, but it does so.[46] This becomes apparent to the user when the perigee location (the point on the Earth's surface under the satellite at perigee) moves (this also, of course, means that the apogee location moves). Two factors affect the rate of rotation: the inclination of the orbit and the eccentricity. It is possible to solve the equation for a rotation rate of zero, at which point it emerges that for inclinations of 63.4º and 116.6º (approximately), regardless of eccentricity, the perigee location is fixed. We will see shortly that this has great practical implications for Molniya orbits,[47] which have inclinations of 63.4º.

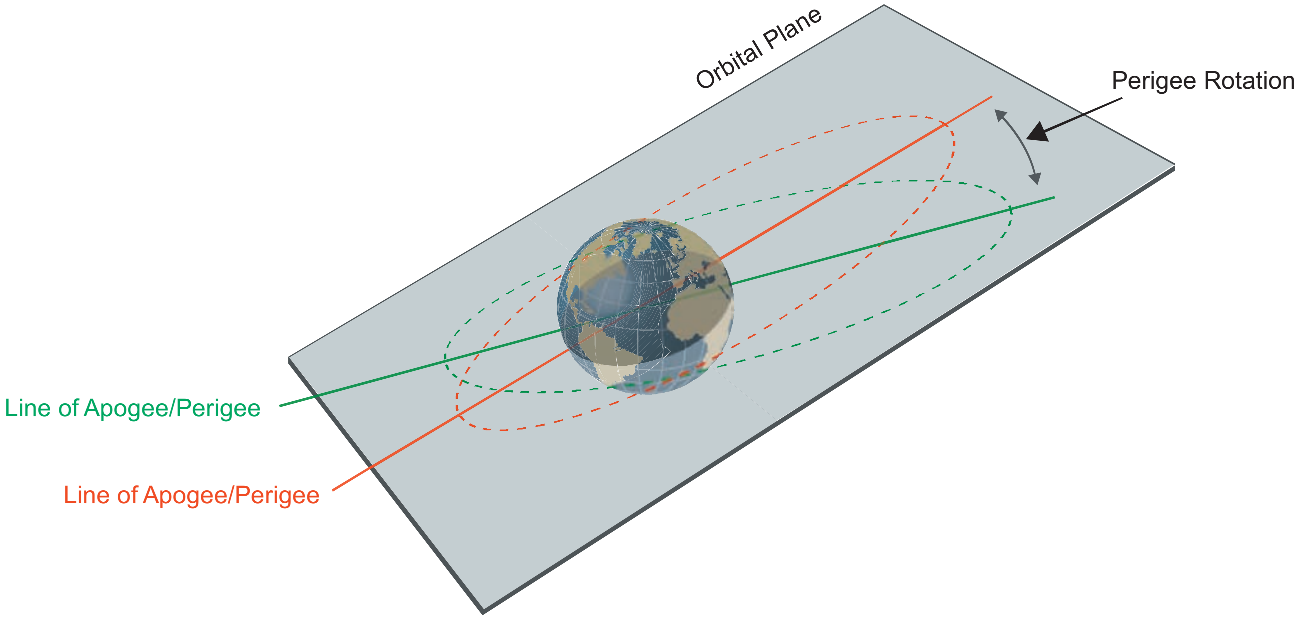


Illustration 1.20b – Perigee Rotation. For an orbit with significant eccentricity, the contribution of the Earth's equatorial bulge varies notably due to the changing distance between the Earth and the satellite, and the changes in its orbital speed. The resultant effect is that the location of the apogee and perigee change within the orbital plane. What is really happening is that the major axis is rotating within the plane, dragging the apogee and perigee with it. For complex reasons, the effect disappears at inclinations of 63.4º and 116.6º.

141. **Circularisation of Eccentric Orbits**. We derived the principle of orbital motion assuming that satellites orbit above the atmosphere, and thus escape atmospheric drag, but for some orbits we know this is not quite true. Drag is governed by several factors: the size and shape of the satellite; the thickness of the atmosphere (the lower the altitude, the thicker the atmosphere and the greater the drag); and the speed of the satellite (the faster the satellite, the more the drag). These factors combine to affect satellites in highly-elliptical orbits. Their perigees are at low altitude (thicker atmosphere), and due to Kepler's Second Law this is where their speed is highest (i.e. most drag). In Paragraph 123 we noted that if a satellite is placed in orbit above the minimum orbital velocity, the orbit will be eccentric; the greater the excess speed, the more eccentric the orbit (until the satellite reaches escape velocity). Robbing the satellite of speed through drag at perigee is akin to injecting it into orbit at successively slower speeds, and the eccentricity gradually reduces; this is *circularisation*. Although satellite operators will try to ensure that this effect is gradual enough not to inhibit the operation of a satellite over its planned life, remember for example that the thickness of the tenuous atmosphere in space near the Earth is governed by unpredictable factors such as solar

---

[46] Even reference textbooks often resort to saying that perigee rotation occurs, without any very convincing explanation *why*. They invariably provide a description of the equation governing the rotation rate, however, and graphs illustrating how the value changes for different inclinations and eccentricities. Rates of up to 20º per day are possible, so again the effect cannot be ignored.

[47] See Paragraph 151a.

activity (see Paragraph 109e). Orbital elements, in this case the eccentricity of the orbit, will thus again change over time.

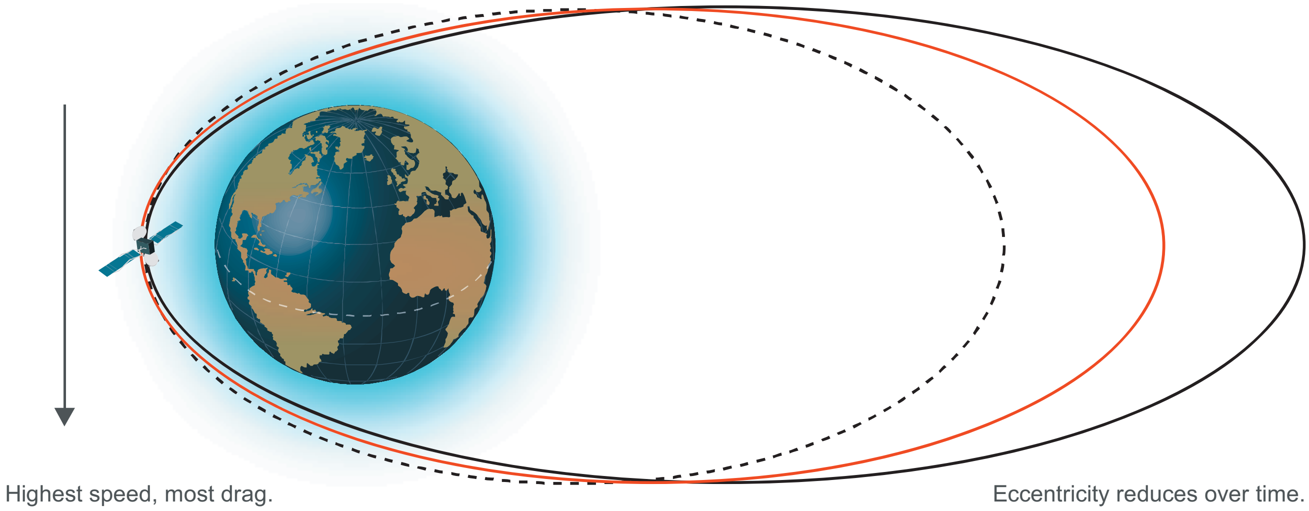


Illustration 1.21 – Circularisation of Elliptical Orbits. Drag affects an elliptical orbit most notably when the satellite is at perigee (highest speed at the thickest part of the atmosphere). The net result is that the a little speed is lost each time, which reduces the subsequent altitude at apogee. This means the eccentricity of the orbit gradually reduces.

142. **Mass Concentrations and the Himalaya Anomaly**. For most practical purposes, the Earth can be assumed to be a uniform solid.[48] We have mentioned that it is not absolutely spherical, but we ignored any subtle variations in its composition. Any effects caused by its slightly uneven make-up (such as Earth relief features) are usually swamped by other perturbations of the orbit. There is one instance where this is not true. In geo-stationary orbit, where the satellite holds a fixed position over the Earth's surface, small effects from large relief features can be detected. Over time, they apply a consistent effect on the satellite. While the satellite is active, the resulting deviations are cancelled out during routine station-keeping, but once the satellite's life comes to an end and it is boosted into a graveyard orbit, no further station-keeping is undertaken. Under these circumstances, the derelict satellites tend to accumulate where the gravitational pull takes them. The Himalayas give rise to one concentration, and the Rocky Mountains may give rise to another. Equally there are areas where few satellites linger. The full implications of this effect are not yet known, although the dangers of collisions occurring at the concentration points are obvious.

143. **Many-body problems**. Finally, remember that we stated above the conditions that denote the 'restricted two-body' assumption.[49] Our subsequent analysis depended on those assumptions. They break down, however, where three bodies, for example the Earth, the Moon and a spacecraft, form particular configurations. Further implications of these 'Lagrangian' points, which have some practical uses, are discussed in Annex D.

---

[48] This turned out not to be the case for the Moon, as NASA discovered during the build-up to the Apollo landings. The Moon's crust contains significant variations in density, which give rise to mass-concentrations (**'mascons'**). These played havoc with the orbital dynamics of the various Apollo vehicles and associated probes and made stable lunar orbits very elusive. Even today, our understanding of the Moon's composition is incomplete; at the time of the Apollo missions, the astronauts simply had to apply corrections to their flight-paths when unforeseen deviations occurred, using up precious fuel in the process.

[49] See Paragraph 124b.

SECTION III – MILITARY IMPLICATIONS

In this Section:

- Drawing on Section II, how satellite orbits are matched to specific missions.
- A review of the useful characteristics of specific classes of orbit:
  - o Low-Earth orbits
  - o Medium-Earth orbits
  - o Geo-synchronous and geo-stationary orbits
  - o Highly-elliptical orbits (including Molniya and Tundra orbits)
  - o Sun-synchronous orbits.

Generic Considerations

144. **Perspective, Field of View and Overflight**. We will frequently be interested in operating a satellite to view or to broadcast to the surface of the Earth. It is easy to assume that from their great height, satellites have an unobstructed view of the Earth beneath them, but in fact they are more constrained than one might imagine.

> a. **Distance to the Horizon**. Since the first construction of a raised lookout position, man has known that height allows one to see further. If we ignored terrain, a person of average height, standing on a smooth Earth could only see about 5 km to the horizon. From 500 ft, the horizon is 44 km distant, while from an airliner cruising at 36 000 ft, the horizon is about 370 km away. Unfortunately, we can already see this is not a ‘straight-line’ relationship. For a satellite in orbit at 400 km altitude (an increase of 36 times over the airliner), the horizon is about 2130 km away (an increase of not quite 6 times) – see Illustration 1.22a. At any instant, this is as far as the satellite can see. It is also seeing the edges of its field of view through the full thickness of the atmosphere. Whilst its view is significant, it is not limitless. Even if you extend the example to geo-stationary altitude (35 800 km), with the satellite over the Equator, the line of sight touches the Earth at about latitudes 81ºN and S.[50] North or south respectively of those extremes, even a geo-stationary satellite does not clear the horizon.
>
> b. **Time of Overflight**. In addition to the limits imposed by range, the satellite also has a view constrained by its speed. Taking our 400 km altitude example above again, and with the satellite directly overflying the target to achieve the longest possible look, it would come into view 2130 km ahead of the satellite, and would be visible for 4260 km of the satellite’s track. At a typical orbital speed of 8 km per second,[51] this equates to 532 seconds of view, or just less than 9 minutes. For any target off the line of flight, the time in view is shorter, until at the extremes of range, the target is only momentarily in view.

---

[50] This case is shown in Illustration 3.5 on Page 3-16.

[51] For an accurate calculation, one would have to allow for the rotation of the Earth relative to the satellite’s flight too, but 8 km per second is still a realistic figure.

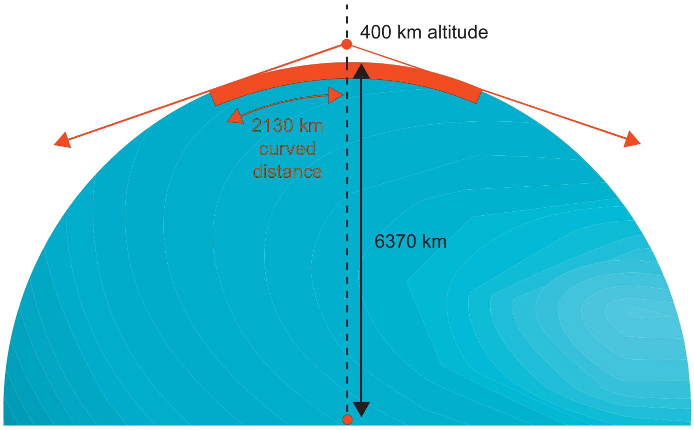


Illustration 1.22a – Distance to the Horizon from LEO. The diagram shows a satellite in LEO at 400 km altitude, drawn at scale distance. Drawing the tangent lines to the horizon shows that its field of view is limited to points within about 2130 km of its track. (The example assumes a spherical Earth and does not correct for atmospheric refraction). Note also that something directly below the satellite is seen through 400 km of atmosphere, while something at the extreme distance of the horizon is seen through more than 2200 km (the straight-line distance is greater than the curved distance over the ground's surface).

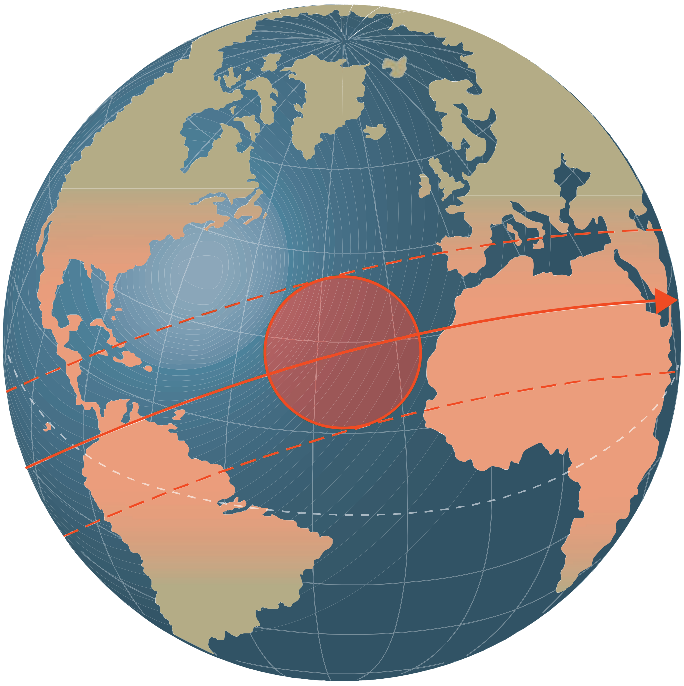


Illustration 1.22b – Field of View and Overflight. The practical implication of Illustration 1.22a. This shows, very roughly, the 2130 km field of view of a 400 km altitude satellite, with a notional ground track added. At its extremities (dotted lines), the satellite will only have a fleeting view of the point on the ground (and vice-versa), and that would be through the thickest part of the atmosphere. For any practical purpose such as surveillance, or communicating with the ground, the useful swathe of ground track is much narrower.

c. **Practical implications**. The 2 constraints above limit the satellite's view of the ground and the view of the satellite from the ground. Both these factors affect operations – they limit what a sensor on the satellite can see from orbit, and also the ability of a ground station to communicate with the satellite. Both are unavoidable constraints of orbital motion. It is possible to trade one off against the other, and we will examine how this can be done when we look at specific orbits and roles.

145. **Useful Orbits**. In military terms, satellites are launched to achieve concrete goals, or to provide a capability that would otherwise be lacking. It follows that, just as on land we aim to secure the most useful terrain for our objectives, we must put our satellite in the right place too. What constitutes useful terrain varies from one mission type to the next. In a similar fashion, orbit types lend themselves to specific roles too. The constraints and advantages are firmly rooted in the factors we have reviewed so far. The divisions between the various bands or families of orbits are broadly arbitrary, just as the limits of space itself are, though we will note a few exceptions. Although we will use altitude to separate the various classes of orbit, it is useful to remember that *there is a fixed linkage between the size (altitude) and the period of an orbit.*[52] Frequently it is the period that is the key to the orbit's utility.

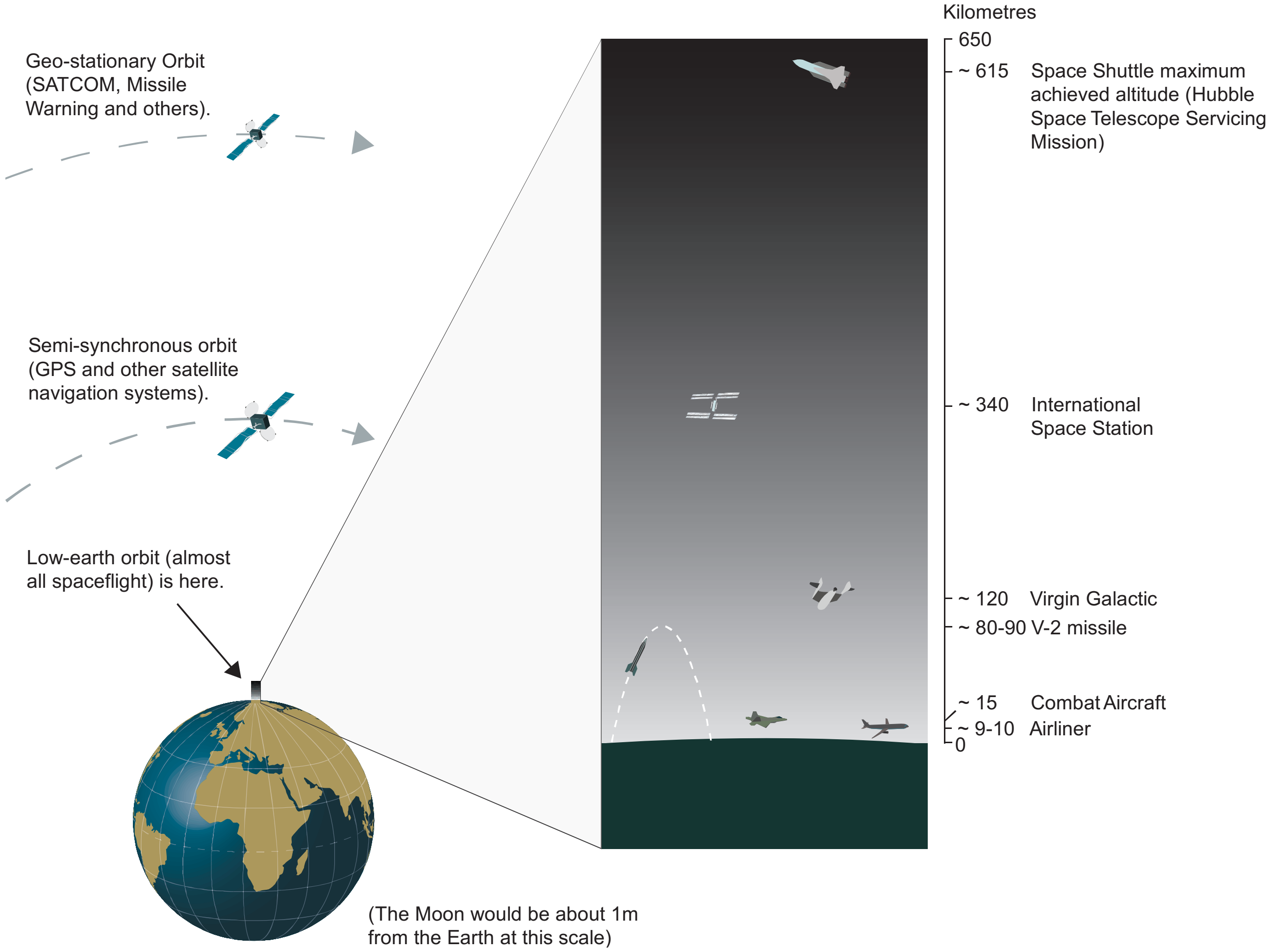


Illustration 1.23 – Orbital Altitudes Drawn to a Common Scale. The diagram above shows on the left various orbital altitudes drawn to a common scale with the Earth. Almost all satellites (and space debris) is in low-earth orbit, which is much closer to Earth than most people realise. One very descriptive (and accurate) illustration is that 'if the Earth was a peach, low-earth orbit would be almost entirely in the fuzz'. The box on the right expands the lower portion of low-earth orbit, to put it in context with other aspects of aviation.

Characteristics of Particular Orbits

146. **Low-Earth Orbits**. LEO describes the band of orbits between about 100 km altitude, and about 1200 km.[53] In this band, periods are typically in the range between 90 and 120 mins (1.5 to 2

---

[52] Kepler's Third Law - Paragraph 128.

[53] There are few rigid boundaries between orbital 'bands'. Some authors would regard altitudes up to 2000 km as 'LEO'.

hrs). All sustained manned spaceflight is conducted in LEO,[54] partly because of the weight/performance penalty in lifting large structures into higher orbits, and partly to avoid extended exposure to harmful radiation. Note finally that all true LEOs are approximately circular.[55] These facts allow us to make several deductions:[56]

a. LEOs are close to the Earth's surface, so favour anything reliant on close range, such as optical reconnaissance or exploiting weak signals.

b. LEOs are close to the Earth's surface, so have a relatively limited horizon/perspective, and fast overflight times.[57] Consequently, they are also either dependent on multiple ground-stations or satellite relay for communications and control; they can only see a given ground station for a short time. The impact of the Doppler effect on communications is also maximised by the high passing speed.[58]

c. LEOs are close to the Earth's surface, so are relatively easy to launch into with larger payloads.

d. LEOs are continually affected by atmospheric drag, particularly at lower levels. Unsustained lifetimes (i.e. without re-boosting) could typically be as short as a few days at the bottom of the band, though this extends to years or even decades near the top of the band.

147. **Medium Earth Orbits**. A strict definition of Medium Earth Orbits (MEO) is difficult to formulate, however one historical definition that may still be useful is that LEOs are up to 1000 km altitude, and that MEOs lie between the top of the LEO band and just less than 36 000 km altitude. Effectively, LEO lies up to, but not including the first Van Allen Belt,[59] and MEO from there out to, but not including, geostationary orbits at 36 000 km. In practice, many MEO satellites sit in circular orbits at about 20 200 km altitude.

a. Many satellites in MEO are manoeuvred so that their period is 12 hours (which implies the 20 200 km altitude). This means that they complete exactly two orbits in a day or one orbit in half a day (giving the orbit its alternative name of a 'semi-synchronous orbit').

b. Because they are much higher than LEO satellites, they have a much bigger footprint. A relatively small number of satellites spaced out around the same orbit will ensure that one is permanently above the horizon anywhere under the orbit ground track. Add in further orbits, similarly populated, but separated around the Earth, and you can ensure that from anywhere on Earth, a given number of satellites will be continuously visible. Fundamentally, this is why they are used by satellite navigation systems (see Chapter 3, Section II(B).)

---

[54] The manned orbital altitude record was set by the crew of the Gemini XI mission in 1966. They flew in an elliptical orbit with a maximum altitude of 1370 km and a minimum of 160 km; so technically, they intruded briefly into what might be regarded as MEO. Their mission lasted just under 3 days.

[55] Imagine an elliptical orbit straddling the LEO band, with its perigee at 100km altitude and its apogee at (say) 850km. Although you might think a factor of x8.5 a significant distortion from circularity, it is the distance from the centre of the Earth that matters. Add in 6400km for the radius of the Earth; the difference between 6500km and 7250 km is much less significant, and the eccentricity in this example is actually about 0.05.

[56] See Paragraph 369 for analysis of these deductions as they relate to surveillance from space.

[57] See Paragraph 144.

[58] The Doppler Effect is further described in Annex F, Paragraph F18.

[59] See Paragraph 108a for details of the Van Allen Belts.

c. Remember that MEO is a relatively hostile environment for a satellite, due to the Van Allen radiation belts. This has had serious implications for some satellites.

Telstar 1

- Telstar 1 was an early active communications satellite – i.e. it could re-broadcast a signal, rather than simply reflect it from a ground transmitter. It was intended to relay both television and telephone signals.
- Telstar 1 was launched into an elliptical MEO, intended to give periodic passes over the North Atlantic for transatlantic relay, on 10 July 1962. It became operational on the same day. Unfortunately for the users, the orbit passed through the Van Allen belts.
- The day before Telstar 1's launch, the USA had performed a high-altitude nuclear blast for test purposes (see Illustration 2.2 in Chapter 2). This energised the Van Allen belts to a far greater extent than had perhaps been anticipated. Further nuclear trials followed, including a Russian detonation in October 1962. Telstar 1's internal components began to fail under the intense bombardment from the energised belts. After a temporary failure during December 1962–January 1963, Telstar 1 failed permanently on 21 February 1963.
- Satellites in MEO are above measurable atmospheric drag. As of March 2008, Telstar 1 is still in orbit, nearly 46 years after its launch, and 45 years after it failed.
- Exo-atmospheric nuclear detonations were banned by the Partial Test Ban Treaty of 1963.

148. **Geo-synchronous and Geo-stationary Orbit Theory**. We looked in the previous paragraph at satellites whose altitude gave them a period of 12 hours. They orbited at greater altitude than LEO satellites, which had typical periods of 90-120 minutes. If we go even further from the Earth than semi-synchronous orbit, the period of an orbit increases beyond 12 hours. Eventually, at about 35 800 km altitude, the orbital period is 24 hours. Thus the satellite orbits the Earth in the same time as the Earth rotates on its axis. This is known as geo-synchronous orbit.

a. Note that we have not yet said anything about the satellite being stationary, as seen from the Earth, just that the Earth and the satellite rotate at the same rate. Any orbit at this altitude will have a 24-hour period, whatever its inclination. Thus it could be in an orbit over the poles, over the equator, or anywhere in-between. In practice, however, it is most likely to be over or near the equator.

b. We noted above that a satellite in an inclined orbit passes over ground north and south of the equator limited by the inclination. If we raise that orbit to geo-synchronous altitude, the satellite will still cross the equator regularly, moving north and south, but the Earth will keep pace with the satellite as it orbits. It therefore always crosses the equator at the same point. As seen from the ground, the satellite will move smoothly north and south almost along a line of longitude, constrained in latitude by its inclination.[60]

c. The lower the inclination, the smaller the range of longitudes that the satellite will traverse. If the orbit has inclination 0°, the satellite will keep station on the point directly below it. Thus from the Earth, it will appear stationary. These (highly-prized) orbits are

---

[60] The ground track of a geo-synchronous (inclined) satellite is actually a 'figure-of-eight', centred over the equator. The satellite appears to have a greater or lesser forward speed, as seen from the Earth, depending on whether it is also moving north or south at the time, and this leads to it 'overtaking' the Earth and then 'dropping back' as its latitude changes. However, the change in apparent longitude is not great for most practical orbits.

known as **geo-stationary orbits** (GEO).[61] Note, however, that they are only possible over the equator. If the satellite passes over any location above or below the equator, it will keep moving north then south in accordance with Paragraph 132b above, spending equal amounts of time north and south of the equator.

149. **Geo-synchronous and Geo-stationary Orbit Applications**. When we looked at the applications of LEO, we noted that one of their disadvantages was the relatively fast passage of the satellite over a ground location. GEO completely overcomes this problem; the satellite has a fixed view of the Earth, and Earth has a fixed view of the satellite. Both features have attractions for the user.[62]

a. **Fixed Earth, as Seen From the Satellite**. This characteristic has obvious utility. The satellite has achieved persistence, for a specific area on the Earth's surface. This could be useful for surveillance-based roles. The drawback is that we needed to go to 35 800 km altitude to achieve this, so perception of detail is hampered by range. A satellite at geo-stationary altitude also has a large field of view. It cannot quite see half of the Earth's surface, but compared with LEO it has a much bigger footprint. A relatively small number of platforms spaced around geo-stationary orbit provide global coverage of temperate latitudes.[63] The classic roles for GEO satellites thus include SATCOM, satellite broadcasting, missile warning and some other forms of low-resolution surveillance, including meteorological (weather) data gathering.

b. **Fixed Satellite, as Seen From the Earth**. We noted earlier the short time that a satellite in LEO could communicate directly with a ground station. Since a geo-stationary satellite is apparently motionless as seen from the Earth, if it is visible at all, it will always be visible. Thus, there is no need for schedules of passes over a location, and no need to transfer communication between ground stations. Additionally, an antenna on a ground station does not need complicated steering to keep the satellite in view. This explains the popularity of geo-stationary orbit for communications, including broadcasting.

150. **Polar Orbits**. We have mentioned the problems caused for satellites and for ground users by the Earth's atmosphere, particularly at high latitudes when the satellite is orbiting near the equator. For some applications, it is essential to let the satellite see the Earth's entire surface, often under broadly constant conditions. We know we cannot place a satellite to see the Earth's entire surface at once, but we can arrange it so that the ground track covers the whole of the Earth's surface within a reasonable time frame. This can be done from a polar orbit.[64]

a. There are many potentially useful polar orbits, with the payload dictating their selection. We will look at them in more detail in Chapter 3. They all exploit the fact that an orbit passing roughly over the poles can see a different swathe of the Earth on each north-south or south-north passage. Depending on whether and how the period divides into 24 hours, the ground track can be adjusted to give coverage of the whole surface of the Earth in a given time.

b. There are polar orbits that revisit given points on the Earth's surface at fixed times of day. We look in greater detail at this in Paragraph 152.

---

[61] GEO is traditionally applied as an abbreviation to geo-stationary orbits and not normally to geo-synchronous orbits. Beware of any particular author's deviation from this convention, however.

[62] See Paragraphs 323(b-e) (Communications) and 369 (Surveillance) for some implications of this.

[63] See Paragraph 144a and Illustration 3.8a on Page 3-15.

[64] See Paragraphs 132-134 and the vignette on Page 1-39 for an illustration of this.

151. **Highly Elliptical Orbits**. We have not looked much at numerical values of eccentricity, and yet again definitions are arbitrary, but remembering that orbit eccentricities lie between 0 (circular) and less than 1,[65] anything with an eccentricity greater than 0.5 would probably be thought of as a Highly Elliptical Orbits (HEO). Two examples of HEO are in common use; the Molniya[66] and the Tundra orbit.

a. **Molniya Orbits**. When we looked at geo-stationary orbits, we mentioned that, because they are located over the equator, visibility of the satellite from the ground is poor at high latitudes. This had severe implications for the (then) USSR during the early days of communication satellites; most of the USSR landmass was at high latitudes where signal coverage was thus limited. The Molniya orbit was an ingenious way of overcoming this problem. A Molniya orbit is an HEO.[67] Molniya orbits have eccentricity of about 0.7, and exploit the varying speed of a satellite in an elliptical orbit to linger over a given area.[68] They are inclined at 63.4° (we will see momentarily why this value is critical), and are oriented so that the apogee (the point where the satellite is highest and is moving slowest) is over the desired point of interest on the ground.[69]

(1) **Avoiding Perigee Rotation**. We mentioned previously that the non-spherical Earth resulted in the location of the perigee (and apogee) of an orbit rotating in the orbital plane, and that this effect disappeared at an inclination of 63.4°.[70] This is the inclination of a Molniya orbit; it ensures that the satellite comes to apogee over the same point on the Earth's surface repeatedly.

(2) **Semi-synchronous Advantages**. The period of an elliptical orbit depends only on the semi-major axis.[71] For a HEO, with a long semi-major axis, periods are relatively long – in fact Molniya orbits are a special case of a semi-synchronous orbit (i.e. an orbit with a 12 hour period).[72] This ensures that their ground track repeats every 2 orbits.[73] The eccentricity, and resulting slow speed at apogee, means they are useable for communications for at least 8 hours in each orbit.

(3) **Molniya Orbits and Surveillance**. Although this slow-moving characteristic is advantageous for communications applications, it would be a drawback for reconnaissance. While the satellite achieves persistence as it moves slowly at apogee, it does so at the greatest possible range from the target area. This is acceptable for some wide-area surveillance applications, which we will explore further in Chapter 3, but it has disadvantages for detailed reconnaissance.

b. **Tundra Orbits**. Tundra orbits are similar to Molniya orbits in shape and inclination, but sit at higher altitude, reaching out to GEO distances. They thus have a 24 hour, rather

---

[65] See Paragraph 118e.

[66] The term 'Molniya' – literally 'lightning' in Russian – relates to the name of the first series of communications satellites to employ this orbit, but over time the name has come to refer to the orbit generally.

[67] Note that in this context the 'E' refers to 'elliptical' or 'eccentric', not to the 'Earth', so 'HEO' is not a progression of LEO and MEO to high altitude.

[68] Kepler's Second Law – Paragraph 127.

[69] Illustration 1.18d, used to illustrate HEOs, actually showed the ground track for a Molniya orbit.

[70] It would also disappear at an inclination of 116.6º, but there are launch advantages in launching a **prograde**, as opposed to **retrograde** orbit. See Paragraph 164g.

[71] Kepler's Third Law – Paragraph 128.

[72] See Paragraph 147a.

[73] Orbits where the period is some simple fraction of the Earth's rotation are sometimes said to be **resonant** orbits. Their cumulative ground tracks will tend to be closed paths over the Earth's surface, their complexity depending on the fraction in question. See also the vignette example on Page 1-40.

than a 12 hour period. Their ground track is a closed figure of eight, with much more time spent in one half than the other. For most of their orbit, they therefore linger over a fixed area on the surface. For example, the Sirius satellite radio system uses a constellation of 3 satellites spaced around a tundra orbit to broadcast to the continental USA.

Illustration 1.24 – Tundra Orbit Ground Path. A Tundra orbit is another example of a HEO with a special ground track. The period of the satellite is about 24 hours, but unlike GEO the orbit is inclined, so that it migrates north and south during its period. The eccentricity also means the speed of the satellite varies greatly during an orbit. The illustration shows the satellite ground track for the Sirius XM ® radio system, which broadcasts to North America. The satellite takes about 16 hours of its period to traverse the tight path over North America, and the other 8 hours to traverse the broad path over South America and the Pacific. Three satellites share the same orbit, spaced out so that at any instant at least one and probably two are well placed to broadcast over American customers. The high altitude at that part of the orbit ensures that the signal comes from nearly overhead, minimising the effect of terrain on customer coverage.

152. **Sun-synchronous Orbits**. In the same way as a Molniya orbit was a special case of a semi-synchronous orbit, Sun-synchronous orbits are special cases of polar orbits. They are also another example of exploiting the effects of the Earth's equatorial bulge, in this case to make sure the satellite overflies points on the surface of the Earth at consistent times of day.

a. Picture a satellite in a (truly) polar orbit as it crosses the equator. At that instant, it will be a specific time of day on the Earth's surface below it. Let us say for illustration that it is midday local time. In this instance, since it is midday, the Sun, the satellite and the point on the ground must be directly aligned. One orbit later, (when it next crosses the equator), because the orbital plane is fixed in space, it will once again be local midday underneath it.[74] On the other side of the orbit, relative to the Sun, the time will be 12 hours later. For a short period of time, say for a few days, this would remain roughly true, based on the orbital path being fixed in space with respect to the Sun, so that the angle between the satellite, the Earth and the Sun would stay about the same from one orbit to the next, and the time crossing the equator would be roughly constant.

b. Over a longer period of time, as the Earth rotates around the Sun, the orientation relative to the Sun would change significantly. If, say, the orbit crossed the equator at midnight and midday, three months later it would do so at dawn and dusk. Three months later still, it would be back to a midnight/midday crossing, and three months after that it would be back to dawn/dusk. Thus we would not achieve our desired aim of a constant overflight time.

---

[74] Usually, such orbits are described by the local time at one of the points of equatorial crossing. The local time elsewhere under that half of the orbit will be slightly different.

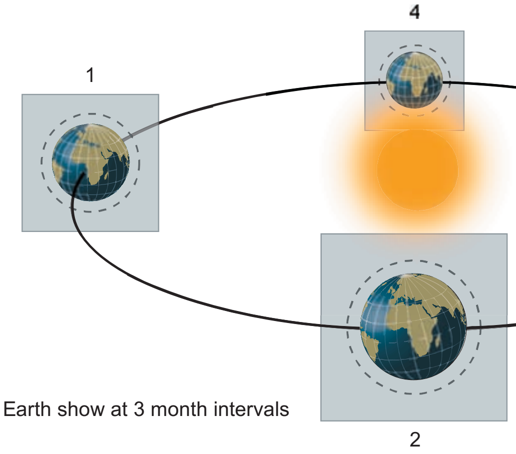


Illustration 1.25a – A True Polar Orbit. If a satellite orbited exactly over the Earth's poles, the orbital plane would stay fixed in space. Thus, taking the example above and starting at Point 1, the satellite overflies the equator twice per orbit, once at a point experiencing local midday, and half an orbit later at local midnight. Three months later (Point 2), the orbital plane remains in the same orientation, but the Earth/satellite combination has moved in space one quarter of the way round the Sun. The satellite is now following the sunrise/sunset line. Three months later (Point 3), it is back orbiting over the midday/midnight line, and three months later still (Point 4), it is back at the sunrise/sunset line. The cycle then repeats. This might meet some needs, but for many satellites, consistent overflight at the same local time each day has advantages.

c. Sun-synchronous satellites exploit precession (specifically nodal regression)[75] to keep a fixed orientation relative to the Sun. Each Earth day, the Earth and the satellites orbiting it complete 1/365th of a rotation around the Sun – i.e. almost exactly 1° per day. To keep a satellite in near-polar orbit in the same orientation with respect to the Sun, it is necessary for its orbital orientation to change by the same amount. If the satellite is in an exactly polar orbit, i.e. with inclination of 90º, there is no precession.[76] Inclined orbits precess, however, and the rate they precess at is dependent on their period (which in turn depends on their altitude) and their inclination. For a circular orbit at 600-800 km altitude (a useful compromise for Earth observation at a practical range without excessive atmospheric drag and orbital decay), and a resulting period of about 100 minutes, an inclination of 98° will yield the desired rate of precession in the correct direction. Such an orbit is also almost polar, with the consequent advantage of good coverage of the Earth's surface.

---

[75] See Paragraph 139.
[76] See Paragraph 139a.

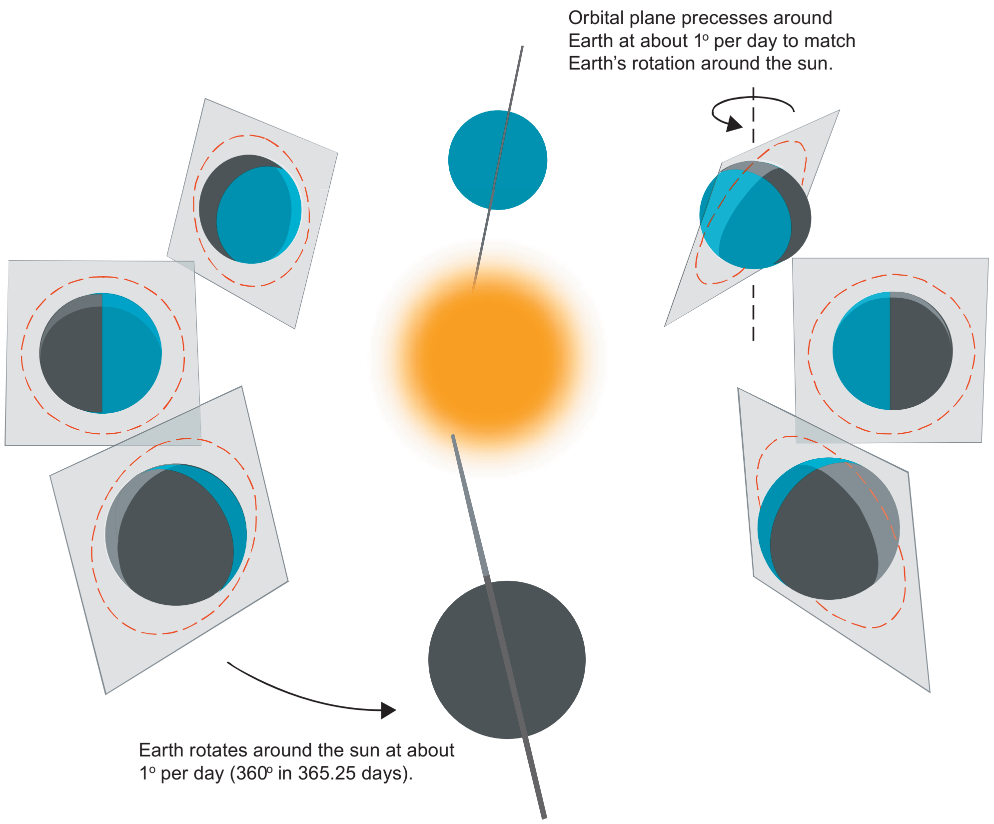


Illustration 1.25b – A Sun-synchronous orbit. In this arrangement, the satellite is inclined in a slightly retrograde orbit. The result of this inclination and the Earth's oblateness is that each day the orbital plane precesses by 1/365th of a revolution (i.e. just less than 1° per day). This has the effect of preserving the orientation of the satellite orbit throughout the Earth year. In the example shown, the orientation maintained is the midday/midnight one we used above, but any other orientation could be maintained so long as the satellite was placed in it with the correct inclination. The exact inclination required varies with altitude, but for a typical LEO, inclinations of about 98° (the value used in this illustration) give the required rate of precession.

d. Sun-synchronous orbits can be used for also have advantages for power generation, which may be important for some missions. If, for example, they are aligned to follow the day/night terminator on the Earth's surface, they never pass behind the Earth as seen from the Sun. This means they stay permanently illuminated by the Sun, and consequently that their solar panels produce power continuously. This can be very important for example if they carry a radar payload.

e. The IRS 1A satellite used as an example in the vignette on Page 1-40 is in fact sun-synchronous; its 99° inclination is the correct value for its altitude to make it so (remember that the exact inclination to achieve the required rate of precession – about 1° per day – varies slightly with the satellite's altitude).

SECTION IV – PRACTICAL SPACE VEHICLES

In this Section you will find a description of:

- How rockets enable access to space.
- The various kinds of rocket motors and engines, and how they are combined into useful systems.
- The implications of the geographical position of the launch site for a satellite.
- Some comments on expendable and re-useable launchers.
- How and why satellites manoeuvre in orbit.
- How to effect deliberate changes of orbits.
- Some of the constraints and factors that apply to the design of practical space platforms.

Launch Strategies

153. **Getting into Orbit**. We have now seen how a satellite can be made to follow a variety of orbital paths around the Earth, and potentially how a spacecraft might navigate interplanetary space. We have not looked, however, at some of the practical implications of gaining access to space. This section addresses issues relating to launch technology, control of spacecraft, and some of the engineering issues faced by designers.

154. **Rockets**. Despite considerable ingenuity and some eye-catching artists' impressions of space planes, a rocket has so far been the only demonstrated method of launching a payload into orbit. It is the only system simultaneously capable of delivering the large amount of thrust required for acceleration and lift in a manner that the payload can survive. High-speed cannon (so called 'super guns') have been considered, and some spacecraft have had a passing resemblance to conventional aircraft – most notably the US Space Shuttle and its tentative Russian equivalent the 'Buran'. Even those, however, are launched using rockets. We will therefore confine ourselves to a brief examination of rocket technology.

155. **Rocket Motors**. Rockets come in 2 principal varieties: solid- and liquid-fuelled. By convention, solid-fuel rockets have a 'motor'; the term 'rocket engine' is not fuel-specific, but is usually taken to mean liquid-fuelled. There are technical advantages and disadvantages for each type, and modern launchers frequently use a combination of both types to achieve complex launch profiles. Both types contain a fuel, which is burned to produce thrust, using an internal source of oxygen (an oxidiser). They can thus operate independently of the atmosphere, generating thrust by accelerating the combustion products in one direction through the exhaust, and harnessing the reaction in the opposite direction.[77] We now examine briefly some of the design factors affecting each type of engine.[78]

156. **Solid Rocket Motors**. A solid rocket combines a fuel and an oxidiser in a stable and safe mixture.[79] For their weight and/or volume, solid-fuel rockets usually produce significantly more thrust than liquid fuel rockets. Also, since the oxidiser and fuel are pre-combined, there is no need to

---

[77] This is Newton's third law, Paragraph 120c.

[78] Examples of applications of each of the major design variants can be found at Annex E, Illustration E.6.

[79] The oxidiser will therefore be a solid chemical that breaks down during combustion within the motor to release free oxygen.

incorporate complicated pumps and valves to control the flow of volatile liquid fuels. They are thus much simpler and inherently more reliable than liquid-fuel models. However, once lit, they cannot be extinguished or controlled.[80] Nor can they be shutdown and restarted in flight. Finally, their total power output cannot be predicted exactly. Designers will try to achieve consistency, but a system using only solid rockets will have some uncertainty in its final position and/or speed on orbit.

157. **Liquid-fuel Rocket Engines**. Liquid-fuel rocket engines use pumps and valves to mix a liquid fuel and oxidiser in a combustion chamber where they are burnt to generate thrust. Compared to a solid rocket, a liquid-fuel rocket is a complex device. The fuel and oxidiser may well be inherently hazardous materials, which are frequently stored at very low temperatures in the interests of maximum efficiency.[81] They have the crucial advantages, however, that their thrust can be controlled, by regulating the flow of fuel and oxidiser, and they can be shut down and re-started during a mission. There are various design options available, and these have differing implications for the missions involved. Key factors include:

a. **Hypergolic Fuels**. Where it is necessary to restart the engine repeatedly during a mission, there are advantages in using **hypergolic** fuels; these are fuel and oxidiser combinations that ignite spontaneously in contact with each other. Hypergolic fuels are almost invariably unpleasant and dangerous chemicals to store and handle, however. Hydrazine (fuel) and red fuming nitric acid (oxidiser) are a popular combination. The benefits include reliability, and saving the weight and complexity of an ignition system.

b. **Turbo-pumps**. The pumps for a large liquid-fuel engine are potentially the most complex and technically challenging part of the engine. They have to move liquids in large quantities from the fuel tank to the combustion chamber at high pressures, and often in difficult conditions (e.g. if the liquids are at cryogenic temperatures). They are almost invariably **turbo-pumps**, i.e. pumps using turbine impellers to move the pumped liquids. Most of the engineering decisions revolve around how to power the turbo-pumps, which then determines how the engine burns its fuel.

c. **Chamber and Nozzle Design – Rocket Science and Rocket Art**. The design challenges at the combustion stage are considerable. The combustion chamber material must withstand very high temperatures and pressures, and its shape must promote smooth burning and efficient flow of the combustion products to the nozzle. The nozzle shape is critical to the production of thrust, as it is the reaction to the acceleration of the exhaust products that is harnessed. The optimum shape for the nozzle will depend on the range of flow rates it must cope with, and whether it is operating against atmospheric pressure or in the vacuum of space. Common pitfalls afflicting early launchers included combustion chambers burning through if cooling flows were not optimised, high-frequency oscillations building up in the combustion process and difficulties containing the initial explosion when combustion began. Many of them were solved empirically, i.e. by trial and error, giving rise to the observation that ‘rocket science’ was something of a misnomer, and that ‘rocket art’ was closer to the truth.

158. **Pressure Fed Cycles**. Having just stated that high-capacity pumps are an almost inevitable part of a high-thrust engine, it is worth noting that low-thrust engines, such as are used for attitude

---

[80] This does not mean they produce constant thrust. Modern solid rocket motors have a varying composition of the rocket fuel physically arranged within the motor to give whatever thrust profile is required during the duration of the burn. This is built in at the design stage, however; it does not allow any controllable variation during launch.

[81] Particular propellants are chosen depending on the thrust required and other characteristics of the motor, but liquid hydrogen as a fuel and liquid oxygen as an oxidiser, stored in insulated tanks at down to-250°C, are an efficient option.

control or manoeuvre in orbit, may avoid the need for fuel pumps altogether. This is achieved by storing fuel and oxidiser under pressure so that they flow to the combustion chamber spontaneously when a valve opens. Helium is often used to pressurise such systems, so-called pressure-fed cycles.

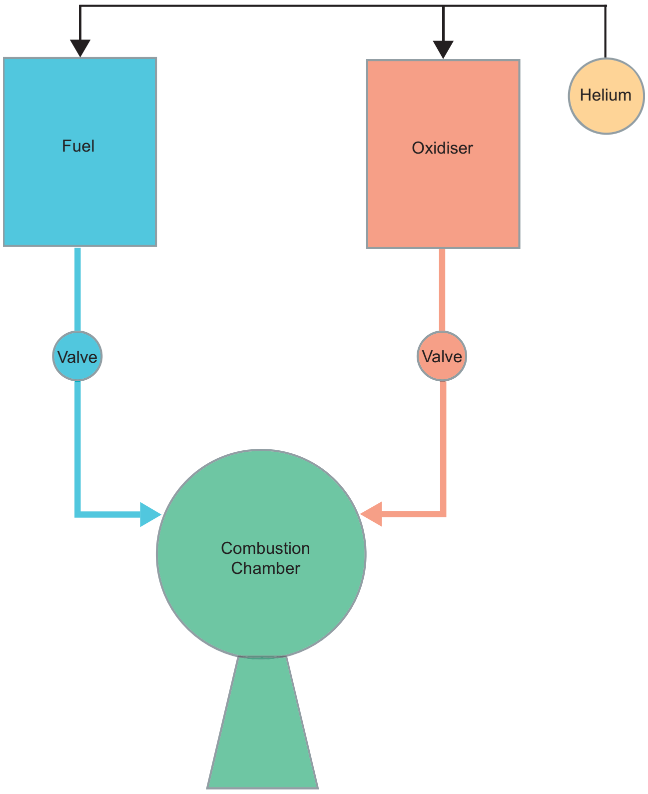


Illustration 1.26a – A Pressure-Fed Rocket Engine. The simplest way to force fuel and oxidiser to the combustion chamber in a rocket engine is to store it under pressure. Helium is a popular choice of pressurising agent, since it remains gaseous at very low temperatures. Valves control the flow rates.

159. **Pumped Systems**. As mentioned earlier, where higher thrust is needed, and particularly where the engine must run both in the atmosphere and in the vacuum of space, pumps are required. The pump impellers are normally mounted on a common rotating shaft, driven by a power turbine wheel. There are several possible motive power sources for the power turbine. Most make use of some of the rocket fuel, but the exact way they do this varies. Details of the various configurations can be found at Annex E.

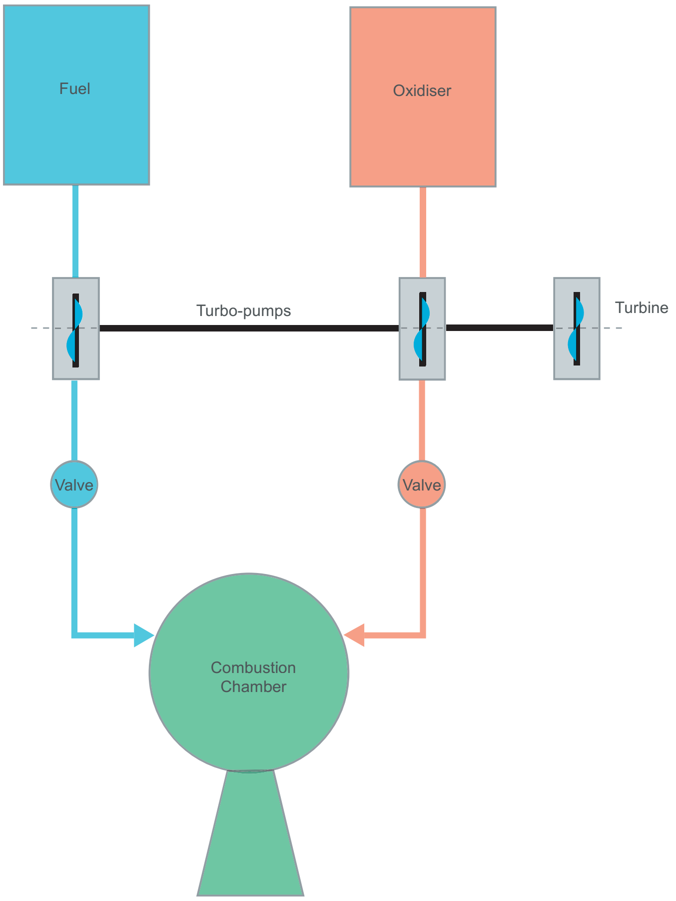


Illustration 1.26b – A Generic Turbo-pump Rocket Engine. This diagram shows a generic turbo-pump engine. The two turbo-pumps usually run on a common shaft, driven by a single turbine. The engineering decisions revolve around how to power the turbine. See Annex E for explanations of some of the available options.

160. **Hybrid Rocket Motors**. Recent experiments suggest that it might be possible to combine some of the best features of solid and liquid propellants in a hybrid motor. One type uses a solid fuel, forming the core of the rocket as normal, but without an integral oxidiser. A liquid oxidiser is then introduced via a pump and valve. The flow of oxidiser controls the combustion and thus the thrust. Such a motor can be extinguished and then restarted. For example, the Virgin Galactic Voyager Space Plane uses such a motor, illustrated at Illustration 1.27 below, once released by its launch aircraft. The fuel is a solid rubber compound, with a liquid/gaseous oxidiser used to control combustion.

161. **Impulse and Specific Impulse**. When used with respect to rockets, impulse refers to the effect of the motor or engine on the momentum of the vehicle.[82] Impulse is the product of the force generated by the motor and the time it is applied for.[83] Thus a small force applied for a long time applies the same impulse as a larger force applied for a correspondingly shorter time. In itself, it is not a particularly meaningful quantity, but for rocket designers, specific impulse is a much more useful measure. Specific impulse is the '*impulse per unit mass or weight of propellant*'.[84] It can

---

[82] Momentum is the product of mass and velocity. Conventionally, it is abbreviated 'P' (since 'm' is already taken for 'mass') i.e. P = m x v.

[83] Strictly, impulse is the *integral* of force with respect to time, but for constant thrust, this simplifies to the product of force and time. (or as a mathematician would put it $\Delta P = \int F\,dt$)

[84] Whether mass or weight is used affects the *units* that specific impulse (I_{sp}) is calculated in, since mass is a property of matter (typically measured in kgs), while weight is a force (typically measured in Newtons). If mass is used, I_{sp} turns out to be a speed; if

usefully be thought of as how powerful (or explosive) a fuel combination you are using. Thus for example a ton of TNT would have more *impulse* than half a ton, since it would create a bigger explosion, but the *specific impulse* would be the same for both, yielding so much 'bang' per ton, whatever amount you exploded. Conversely, high-grade explosive like TNT would have greater specific impulse than low-grade alternatives. Specific impulse must not be confused with total thrust, however; ion-engines (see Paragraph 168c) have very high specific impulse despite not producing much thrust. They produce their small thrust levels using miniscule amounts of propellant, and consequently achieve high specific impulse.

Illustration 1.27 – Hybrid Rocket Propulsion. This picture shows (foreground) the hybrid rocket motor used in SpaceShipOne built by Scaled Composites LLC in Mojave, California which won the Ansari X prize for private space-flight in 2004. The motor is mounted here in its test stand; the solid fuel is the cylinder and nozzle protruding at the rear. The large metal tank in the background is the ground storage and delivery system for the oxidizer (nitrous oxide in this application). As fitted to SpaceShipOne, the engine is not designed to provide thrust control (i.e. a throttle) to the pilot. It is, however, intended to support shut-down and re-start during a flight. (Image © 2004 Mojave Aerospace Ventures, LLC. SpaceShipOne®, A Paul G. Allen Project).

162. **Multi-stage Rockets**. One of the limitations of a simple rocket for reaching space is that as the fuel is consumed, the empty tanks represent an increasing proportion of the total mass of the rocket. Expending fuel to accelerate this empty mass as well as the payload is inefficient, to the point where orbit is practically unachievable with a single rocket.[85] Additionally, the satellite operator is unlikely to want the payload to be accompanied by a large empty rocket in orbit,[86] so at some stage the payload and launcher must be separated. Both these concerns, but particularly the efficiency issue, are addressed by a multi-stage launcher.[87] The concept is relatively simple – rockets, usually of reducing size, are stacked one on top of another; the first stage is used at initial launch to accelerate the whole system clear of the ground. When its fuel is expended, the stage

---

weight is chosen, I_{sp} is actually a time. Both conventions are in use, the 'weight' convention is slightly more common, with I_{sp} quoted in seconds.

[85] This may not be true for a very small satellite, or possibly for an air-launched system where the launch aircraft performs the function of the first-stage. The US 'Mercury' manned capsules in 1961-2 used single-stage rockets for their initial sub-orbital flights, but needed a two-stage system to achieve orbit.

[86] See Chapter 3, Paragraph 318b for a counter-example from the early days of military spaceflight – Project SCORE. This was, however, a very early system, driven by political imperatives as well as technical ones, and it did not represent a trend.

[87] A further gain in efficiency comes from being able to optimize the rocket motors in each stage for their intended environment, so the first stage can be optimised to work in the atmosphere, while the second and third stages can be optimised for use in a vacuum. See Paragraph 157c for more details of why this may be important.

separates, and the next stage ignites and accelerates the much smaller (less massive) system on the next stage of its journey. Since total mass is much reduced, greater acceleration is easier to achieve. Use of up to three stages is normal; where necessary, the first stage can be augmented by solid-rocket boosters. Using them in conjunction with the first-stage gives maximum scope for any correction required due to their imprecise thrust output, and uses their high thrust level when the mass is greatest. Calculations show that for a given total mass at launch, a two-stage launcher could typically deliver six times as big a payload into orbit as a single-stage system. Subsequent stages do not gain quite so much, and diminishing returns become apparent. Thus there has been little need to date to go beyond three stages. Additional drawbacks to multiple stages include cost (e.g. extra rocket motors in the launcher) and reliability. Each additional stage introduces complexity and the possibility of failure from separation problems or the overall reliability of the rocket motors.

163. **Launch Vehicles and Payloads**. Atop the chosen rocket motor, the designer will place the payload or payloads for a mission. A critical design requirement is to keep all extraneous weight and volume to a minimum, so the launcher will contain the bare minimum of additional structure and functionality. However, this will have to include systems to ignite motors in the correct sequence, to separate the various stages of the rocket during its flight, and to steer it into the correct orbit. Steering may initially depend on aerodynamic fins like an aircraft's, but as the rocket leaves the atmosphere, it will steer by small variations in thrust angle while under power. Much of the mission will be controlled and monitored from a ground station, so room must also be found for a communications package. Finally, to guard against failure during the ascent, launchers will usually include a launch-termination system to destroy the entire package at a safe altitude if control is lost.

164. **Launch Sites**. Practical launch sites have to satisfy several competing requirements. We will not look in detail now at launch profiles and flight paths, but note for the moment that while a rocket initially flies straight up to clear its launch site, it must accelerate along its required flight path soon after launch.[88] This is the direction in which it needs to achieve orbital velocity. Thus our launcher needs a clear flight path out of the atmosphere for dynamic reasons, with a clear ground track below to guard against mishaps *en route*.

a. Different nations have different views on what constitutes an acceptably safe ground track. The USA is able to launch over water from its two principal launch sites at Cape Canaveral in Florida and Vandenberg AFB in California.[89] Russia (and previously the USSR) uses land-locked launch sites in remote areas of Asia, while the European Space Agency has access to an east-facing track out into the Atlantic for Ariane rockets from its launch site in Kourou in French Guiana. Israel has a launch site that only permits westward launches, with significant payload implications (see sub-paragraph g below for the reason). They have achieved safe launch of heavier payloads only through international cooperation.

b. Completely separately from any consideration of launch safety, the latitude of the launch site has a critical effect on the orbits that can be achieved directly from it. The latitude of the launch site is equal to the minimum achievable initial inclination. So, for example, from Kennedy Space Centre at Cape Canaveral (latitude 28.5°N), all the orbits that can be reached directly have an inclination of at least 28.5°. This is because the launch site, or at least a point very near it, must lie on the orbital path, and both it and the centre of

---

[88] We circumvented this in our earlier discussion by putting Newton's cannon at the top of the mountain (i.e. already at orbital altitude) and inclining the barrel horizontally (i.e. all the cannon acceleration was in the final flight direction).

[89] Since the Space Shuttle launch profile includes the impact of the re-useable Solid Rocket Boosters into water shortly after launch, an over-water track is in this case essential. One of the implications of the USA's decision not to develop the planned second Shuttle launch site at Vandenberg AFB was that the Shuttle could not then be used to launch payloads directly into polar orbit, as only Vandenberg has an over-water southerly launch track.

the Earth must lie in the orbital plane.[90] The lowest inclination that you can achieve is therefore to launch such that your launch point is the northernmost or southernmost extremity of the initial orbit. But that, as we saw earlier, defines the inclination of the orbit.[91]

Illustration 1.28a – Launch Site and Orbital Inclination. Because a satellite launcher needs to clear the atmosphere as quickly as possible, it is still relatively close to the launch site when it reaches initial orbital altitude and speed. It also needs to accelerate along its intended orbital path. This means that a point quite near the launch site will lie on the ground track. We have already seen that the inclination of an orbit constrains the ground track of the satellite; here the reverse applies, it is the ground path that constrains the inclination. The example above shows a launch from Cape Canaveral in Florida. The satellite achieves orbit relatively soon after launch, so a point near Florida must be on the orbital track. Kepler's First Law then says that the centre of the Earth is in the orbital plane, and by implication, equal portions of the orbit will lie in the Northern and Southern Hemispheres. Thus any initial orbit must have an inclination at least equal to the latitude of the launch site – 28.5º in this case.

c. Another consequence of the launch site being in the orbital plane together with the centre of the Earth is that unless the satellite manoeuvres extensively during its initial ascent (which would be very inefficient in fuel terms), any craft will pass over the antipodal point of the launch site – the point exactly on the opposite side of the Earth – half an orbit after launch. So for example a satellite launched from a hypothetical launch site at 10°N, 90°W will overfly somewhere close to 10°S, 90°E half an orbit after launch. This is regardless of altitude and the inclination of the orbit. Plainly this might have operational implications for somebody wishing to monitor launch activity from a particular site.

---

[90] Due to Kepler's First Law, Paragraph 126.

[91] See Paragraph 130b. For inclinations greater than 90º (i.e. retrograde orbits in the opposite direction to the Earth's rotation), the limit is calculated as (180-inclination)º.

d. We will see below that changing the plane of an orbit once it is established is very expensive in terms of fuel – far more limiting than it might first appear. Nonetheless, this is required to achieve low inclination orbits from launch sites away from the equator. Either the flightpath is managed during the initial rocket burn to fly the satellite onto the desired orbit, or (more commonly) an initial inclined orbit is accepted then modified by manoeuvre. Either approach consumes energy however, leading to a lower maximum payload that can be inserted into orbit.

e. Latitude only imposes a lower limit on inclination. From any site, for a higher inclination orbit you can easily aim the rocket to intercept the required path, though you may still be constrained by ground track considerations.

f. Although a given orbit might be entirely accessible from a launch site, it is unlikely that the availability will be continuous. This is because we need to achieve correct values for all 6 of the orbital elements (Paragraph 130). So far we have really only considered the plane of the orbit and its inclination. We also want the orbit oriented correctly in space, with the satellite at a given point on it. This effectively means waiting until the launch site rotates through the required orbital plane. This gives rise to the launch window for a particular mission. Other engineering factors, such as the need to launch into a sunlit portion of the orbit to provide solar power early in the satellite's deployment may further constrain launch. Potentially this could have military implications for response times when launching a satellite for operational reasons.

g. The final launch site consideration we must account for is the contribution made by the Earth's rotation. We have already stated (Paragraph 132) that orbital dynamics are not affected by the Earth's rotation; satellites would orbit a stationary Earth in the same way as a rotating one (and the ground tracks would be easier to envisage!). However, to achieve orbit, they need to accelerate to speeds of the order of 18 000mph, and some of this speed can come from the Earth's rotation. A point on the equator travels a distance equal to the Earth's circumference every 24 hours. This confers an eastwards motion of about 1000 mph. This is plainly not nearly enough for orbit, but is still a significant gain, and an advantage favouring launch sites near the equator for many orbits. The gain is correspondingly less for launch sites displaced from the equator, where the contribution from Earth rotation is less, and for orbits in a high inclination, where eastwards velocity is not as helpful.

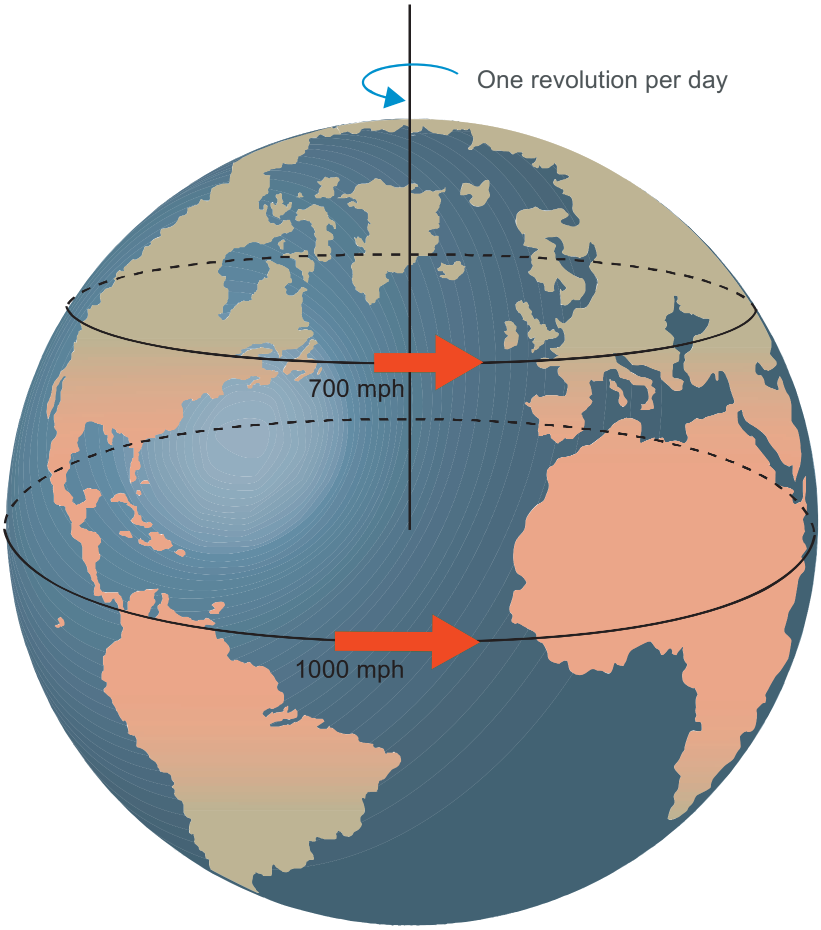


Illustration 1.28b – The Contribution of Launch Site Position to Orbital Velocity. For a launch site on the Equator, the Earth's rotation contributes about 1000 mph to the eastwards component of orbital velocity. For launch into a low-inclination (e.g. equatorial) orbit, this is a very useful bonus. The effect diminishes with increasing latitude north or south, and is less beneficial where a higher inclination orbit is required; for a polar orbit, it is essentially irrelevant. The increasing need for heavy-lift launch into (equatorial) GEO explains the attraction of launch sites such as Kourou in French Guiana (latitude 3° S) for this activity.

165. **Other Launch Strategies – Sea-launch**.

a. **Sea-launch Practicalities**. Since the position of the launch site constrains a launch through safety, achievable orbit, timing and payload considerations, a moveable launch platform is obviously attractive. At least 2 systems are currently operational. A multi-national consortium led by Boeing uses a modified offshore exploration rig to offer commercial launches into geo-stationary orbit. The platform and its supporting vessels operate from the USA and launch from the equatorial Pacific Ocean. Additionally, a Russian system currently offers launches from a modified ballistic-missile submarine, using a launcher derived from the original missile. There were also a series of launches by NASA from a platform moored off the coast of Kenya between 1964 and 1988.

b. **Sea-launch Potential**. In addition to those outlined above, sea-launch might have additional attractions to a military user; it confers an element of discretion to launch activity by making it relatively inaccessible. However, we will see that satellites can be detected with reasonable ease once launched, and at least in the case of the US-based system, the launch convoy is substantial. Sea-launch also exposes the launch system and the payload to the relatively hostile oceanic environment. Its commercial advantage is to allow launch from the equator, and thus to maximise payload. Because of this, its military potential cannot be completely ignored.

Illustration 1.29 – Sea Launch. The Sea Launch Company's Odyssey launch platform illustrates one advantage of sea-launch. This launch is of the DirecTV11 broadcast satellite, using a Zenit-3L launcher to place a Boeing-built commercial broadcast satellite in geo-stationary orbit. The launch took place from the equator to maximise the contribution to launch velocity made by the Earth's rotation. (Image courtesy of SLC)

166. **Other Launch Strategies – Air-launch**:

a. **Air-launch Practicalities**. At launch, a rocket is at its maximum weight, which rapidly reduces during flight as fuel is burned. Thus the initial climb from launch incurs the greatest specific fuel consumption. Additionally, the rocket and payload are exposed to significant aerodynamic stress in the early stages of launch while they accelerate in the thick lower levels of the atmosphere. Using a conventional aircraft to lift the rocket as high as practical prior to its release and ignition can mitigate both these drawbacks.[92] There is currently an operational launcher system in the USA which uses a converted TriStar airliner to launch payloads into LEO.[93] Other systems have been proposed, including one using an air-dropped launcher carried in the cabin of a C-17 transport aircraft and deployed via the ramp. The Virgin Galactic 'Space Ship One' space tourism programme also uses air-launch for sub-orbital flight, conducted from their 'White Knight' dedicated launch platform.

b. **Air-launch Potential**. Air-launch definitely confers military advantages in terms of responsiveness and flexibility, as deployment times are much shorter than for sea-launch, and an aircraft has more potential for covert launch, or for rapid deployment to an optimum

---

[92] The USA built up significant experience of launches of this type during the development of experimental research aircraft in the 1940s and 50s. The Bell X-1, which was the first aircraft to exceed Mach 1, was launched from a B-29 bomber, and the X-15 research aircraft mentioned when we discussed near-space and shown in Illustration 1.1 was carried aloft by a B-52.

[93] This is the 'Pegasus' launcher, operated by the Orbital Sciences Corporation and shown at Illustration 1.30.

position for a given orbit. However, there are severe size and weight constraints imposed on the launcher by the need to integrate it with the launch aircraft, so its potential for heavy payloads or high orbits is limited. There may also be safety considerations relating to carriage of relatively hazardous payloads overland during transit to the launch area. The growing utility of small satellites, however, should ensure continuing interest in the concept.

Illustration 1.30 – Air Launch. Orbital Science Corporation's (OSC) 'Stargazer' converted L1011 TriStar takes-off, carrying a 'Pegasus' launch vehicle capable of placing satellites in LEO. The pace of Pegasus launch activity has reduced in recent years as OSC have developed other conventional launchers, but the Stargazer/Pegasus combination proved the concept of commercial air-launch. (Image courtesy of Orbital Sciences Corporation). See also Illustration 4.3 in Chapter 4.

167. **Expendable and Re-useable Vehicles**. The vast majority of space launches to date have been made by expendable launchers. We will see shortly that the majority of payloads are also expendable. Nevertheless, the lure of a re-useable space plane is persistent. The US Space Shuttle is perhaps the most obvious instance, but there have been other attempts and examples. For the majority of applications, however, where payload weight atop the launcher is paramount, re-usability would incur a structural penalty in more robust construction to survive re-entry. So far, designing the satellite to achieve maximum lifetime in orbit, using the weight saving to launch a bigger payload or provide more fuel for use on station, has proved more attractive. The Space Shuttle has filled a niche with its ability to lift large components to LEO and to perform some manned activities in orbit, but it has faced severe challenges in support and refurbishment between missions and has never met its initial design goals for turn-round times or reliability.

Manoeuvre in Orbit

168. **Propulsion Systems for Manoeuvre and Attitude Control**. Once a satellite achieves orbit, it only needs gentle nudges produced by a low-thrust system to maintain orbit. In practice, placing a satellite in its desired orbit will be a combination of deliberate orbit changes and some final

manoeuvring. For substantial pre-planned orbit changes, the propulsion system will usually be similar to the launch system, i.e. some kind of relatively large rocket. For fine adjustments, and for attitude control, there are a variety of possible systems:

a. **A small conventional rocket can be used**. To provide control, it must be liquid fuelled (see Paragraph 157). To provide reliability, the fuel will probably have to be hypergolic (see Paragraph 157a). An alternative is to use a mono-propellant – a fuel that can be made to decompose in isolation.[94] The amount of fuel carried will determine the operational life of the satellite.

b. **Cold-thrusters** simply expel a pressurised propellant, stored as a liquid or compressed gas.[95] They are very simple and reliable, but again the propellant capacity determines satellite operational life.

c. **Electric thrusters** use electricity to accelerate a charged vapour or vaporised propellant out of the thruster. The ion-drive is a specific example of this, and successful prototypes have been flown. They produce very low levels of thrust, but can do so almost continuously,[96] so to date they have been used for long-duration and deep space missions, and for attitude control. Since the energy comes from the electrical supply, they can turn solar power directly into thrust. Consequently, they are less limited by fuel supply, though they do consume propellant which will eventually be exhausted.

---

[94] A common example is hydrazine, a very reactive chemical that decomposes in the presence of a catalyst such as platinum, yielding heat to expand its own decomposition products as exhaust gas, just like a conventional rocket.

[95] This is in principle exactly the same kind of rocket propulsion as blowing up a toy balloon and releasing it into a room.

[96] They are also very efficient in terms of the ratio of thrust produced to propellant mass, even though the total thrust levels are low. See Paragraph 161 for a discussion of this factor, known as *specific impulse*.

Illustration 1.31 – Orbital Manoeuvering. Orbital manoeuvres by satellites usually takes place out of sight of any human observers, but these photographs capture what happens. The top picture shows the instant of the Space Shuttle *Atlantis* firing its Orbital Manoeuvering System (OMS) engines during flight STS-84 in May 1997. The OMS engines are at the rear of the Shuttle, separate from the main engines, and use hypergolic fuel (see Paragraph 157a). They provide the final push into orbit on each Shuttle launch after the main engines shut down, and are also used to boost orbital altitude when coupled to the ISS. In the upper image, the exhaust highlights the vertical fin, as seen from the flight deck looking aft. The lower image shows a range of Space Shuttle engine exhausts: the Main Engines (3 largest exhausts), the OMS engines (two smaller adjacent exhausts) and the rear portion of the Reaction Control System for attitude control (small nozzles facing aft and upwards outboard of each OMS engine). The windows through which the upper image was taken can also be seen at the forward end of the payload bay. (NASA Shuttle-Mir History images)

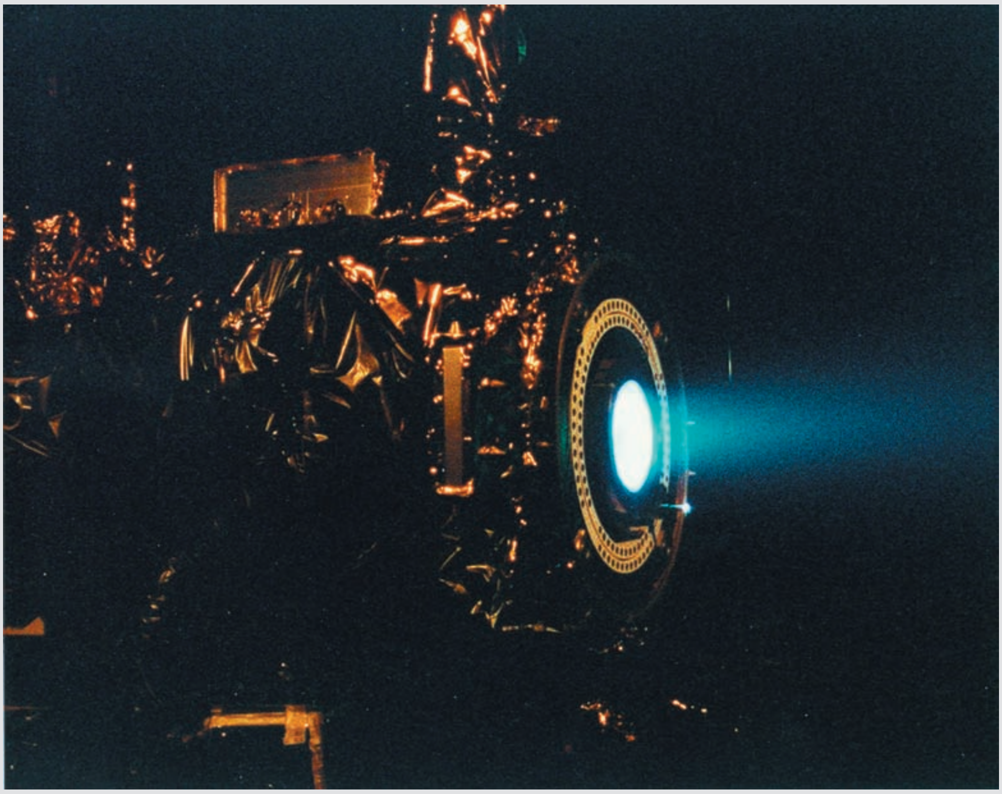


Illustration 1.32 – Ion Propulsion. This picture shows an ion engine under test in a NASA vacuum chamber. The blue glow emerging from the engine is a stream of ionized (neutral atoms given an electrical charge by the addition or removal of electrons) xenon gas, expelled electrically by the engine. The thrust is the reaction to the acceleration of the individual ions; typically, they are accelerated to speeds of about 30 km/sec as they leave the engine. Total thrust produced is low, and such engines would never be able to lift a payload from the Earth's surface, but they have advantages for station-keeping or manoeuvre in orbit, and for missions where low levels of thrust applied for very long periods could be useful. NASA used an engine similar to this in the Deep Space 1 interplanetary probe during 1998-2001. (NASA Image)

169. **Station Keeping**. Once a satellite achieves its required orbit, it is still affected by various external and internal factors that tend to move it away from its ideal location. Among these, LEO satellites are prone to drag, precession affects most satellites in inclined orbits, and over extended missions, it may be necessary to account for perturbations by other planetary bodies. In extremis, it might also be necessary to manoeuvre to avoid debris. In addition, using thrusters for attitude control (see Paragraphs 171-2) will also impart some change to the orbit of the satellite, which may need to be corrected. All these, and other, effects are countered by use of a thruster system. The challenge for the satellite operator is usually to maximise the life of the satellite by doing this in the most efficient manner possible, using the minimum amount of propellant. This will often mean allowing errors to build up towards a tolerance limit, then resolving them in one operation, rather than continuous small corrections.

170. **Orbit Changes**. We have seen that satellite launch is often critically limited by available thrust balanced against the desired payload, and allowing enough fuel for manoeuvre and control during the required mission duration. One of the biggest potential demands on fuel supply during a mission would be an intentional orbit change. To study these sensibly, we must distinguish between in-plane and out-of-plane manoeuvres.

a. **In-plane Manoeuvre**. An in-plane manoeuvre is one where the plane of the orbit remains unchanged. In terms of orbital elements, this means changing the eccentricity and/or semi-major axis of the orbit, without altering the inclination or the other elements that define the plane.[97] Thus, boosting the altitude of a LEO satellite that has been subjected to drag is an in-plane manoeuvre. So too is the method by which geo-stationary satellites are placed in orbit. The initial launch places the satellite in a HEO, with the apogee out at geo-stationary altitude. The satellite includes a substantial boost rocket. Once the plane of the orbit has been checked, and initial deployment of the payload established, the boost rocket is fired to circularise the orbit (i.e. reduce the initial high eccentricity to zero).[98] Although circularising a HEO requires substantial thrust, such an in-plane manoeuvre is inherently more efficient than an out-of-plane one, because it exploits all the momentum the satellite already possesses. In-plane manoeuvres can also be used to adjust ground track for a satellite in LEO or MEO. If the period has been set at an exact fraction of an Earth day, the ground track will be frozen, repeating daily. To change the ground-track, it is possible to change the altitude slightly, thus altering the period (Kepler's third law). The ground-track will then drift, as it is no longer an exact fraction of a day. When the desired point on the surface is under the satellite, the orbit can then be restored to its frozen state by returning the altitude to its original value.[99]

b. **Out-of-plane Manoeuvres**. An out-of-plane manoeuvre changes the orbital plane of the satellite.[100] It therefore needs to overcome at least some of the satellite's existing momentum, and is thus enormously greedy in fuel.[101] Out-of-plane manoeuvres are sometimes unavoidable. For example, orbit inclination may need to be reduced after launch from a site at high latitudes. This is the only way to reach an equatorial orbit (such as a geo-stationary orbit) from an off-equator launch site. In cases such as these, the requirement is predictable and is budgeted for in fuel/payload terms before launch. This may not be the case, however, if an operational requirement arises to change a satellite's ground track substantially – for example to achieve a rapid look at, or revisit to, a specific ground target. Whether such a manoeuvre is possible depends principally on the available fuel. Large manoeuvres that deplete the fuel supply significantly will shorten the life of the satellite appreciably.

171. **Attitude Control – Requirement**. Spacecraft attitude control may be needed for a number of reasons:

a. A payload element may have to be pointed at a target – for instance a reconnaissance sensor or an antenna.

b. Solar panels must be kept pointing at the Sun to maintain power levels.

c. A high degree of pointing accuracy will be required if thrusters are to be used to maintain or change an orbit.

d. Radiators for cooling may need to be pointed into deep space.

---

[97] Orbital Elements are explained at Paragraph 130, and further expanded in Annex C.

[98] Just as in station keeping, the operator aims for maximum efficiency, in this case to ensure that the boost stage (which is otherwise non-effective payload) is no bigger than required. The single circularising burn is the most efficient way of doing this. Note that when a large rocket is used to place multiple satellites in initial orbit, each may then need to be placed in its desired final orbit by its own boost stage.

[99] The period-change from a small change of altitude will also be small, so the urgency of the required change might dictate how great an adjustment is made.

[100] Any manoeuvre that changes the inclination of an orbit is thus 'out-of-plane'.

[101] For example, if the Space Shuttle expended its entire on-board manoeuvring fuel supply (from a typical LEO), it could reportedly change its inclination by about 3°.

e. The satellite must also be manoeuvred to negate small random perturbations caused by debris impact, solar radiation pressure, atmospheric drag and other sources.

172. **Attitude Control – Engineering**. Attitude control must be achieved without imparting significant changes to the spacecraft's orbit. There are several engineering solutions to this problem:

a. Individual components of the spacecraft may be pointed in isolation, using conventional actuators

b. Satellites may be attitude-stabilised, often by gyroscopes or the use of weights on long booms to orient the satellite. The boom will tend to point towards the Earth under the influence of gravity, stabilising the body of the satellite, though this does not stop rotation around the axis of the boom. Deliberate spinning of the satellite can also impart stability and may be used on a temporary or permanent basis. Some sensors rely on the spin to scan target areas (see Paragraph 358a).

c. Some satellites use reaction wheels for attitude control. These consist of small heavy wheels which can be rotated electrically. Rotating the wheel tends to make the satellite rotate in the opposite direction, and this reaction is actually the desired movement. A set of mutually perpendicular wheels allow manoeuvre in any direction. Since the drive is electric, power for manoeuvre can come from solar panels, there are no emissions from a thruster or propellant to exhaust. There is, however a weight penalty from including flywheels of useful size.

d. Another method of attitude control is the magnetic torquer. These devices exploit the interaction between the Earth's magnetic field and that generated by a simple electrical coil fitted to the satellite. A satellite would again typically have 3 of these devices pointing in mutually perpendicular directions to provide complete manoeuvreability. Their advantages include relatively low mass, the absence of stored propellant since they use electricity to generate torques and once again no exhaust issues to manage. Against that, they are relatively weak devices due to the tenuous nature of the Earth's magnetic field, they are relatively imprecise, since the Earth's magnetic field also fluctuates, and they need to be located both separately from each other, to avoid the magnetic fields they generate interacting, and also remotely from any sensor on the satellite that they might affect adversely.

e. Remaining stabilisation and attitude control often comes from a series of thrusters dispersed around the satellite, and possibly shared with the manoeuvre system. Once again, fuel capacity may ultimately determine operational life.

Out-of-plane Manoeuvres – the Hubble Space Telescope (HST) and the Space Shuttle

- The HST sits in LEO, and was designed to be serviced periodically in orbit. It was launched from the payload bay of the Space Shuttle in 1990, and servicing missions were subsequently carried out using the Space Shuttle (in 1993, 1997, 1999, 2002 and 2009).
- Following the loss of the Space Shuttle '*Columbia*' in 2003, a policy decision was made that all future Space Shuttle missions should be able to reach the International Space Station (ISS). This was to provide a refuge for the crew if the Space Shuttle was damaged at launch.
- The HST orbits at about 500 km altitude in a 28° inclined orbit, the ISS at about 350 km altitude in a 55° inclined orbit. Both orbits are within the range of the Space Shuttle launched from Cape Canaveral.
- There is, however, no way that the Space Shuttle could manoeuvre from the HST to the ISS – the change in energy required to move between the orbits far exceeds its available fuel for manoeuvring on-orbit. The final HST servicing mission, which could not comply with the policy described, thus relied on the availability of a second Space Shuttle on the ground ready for launch as a 'lifeboat'.

Illustration 1.33 – The Implications of Out-of-Plane Manoeuvres. This 2009 image shows the unusual (though not unique) scene of 2 Space Shuttles simultaneously on Launch Pads 39A and 39B at Kennedy Space Center in Florida. In the foreground, Shuttle *Atlantis* is prepared for the final servicing mission to the Hubble Space Telescope. In the background, Shuttle *Endeavour* stands ready to act as a lifeboat in the event of launch damage occurring to *Atlantis*. For reasons explained above, if this happened, the crew on *Atlantis* would be unable to manoeuvre to safety at the International Space Station, and a second Shuttle mission would be required to rescue the crew. (NASA Image)

Satellite End-of-life Issues

173. **De-orbit and Re-entry**. There are various reasons why it may be necessary to de-orbit a satellite:

a. The satellite may be occupying an orbit required for re-use by a replacement.

b. The satellite may pose a hazard if it re-enters in an uncontrolled fashion and debris falls on land. Intentional de-orbit into a safe area avoids this.

c. If there is any likelihood of the satellite breaking up either through collision or structural failure due to age, intentional de-orbit mitigates against the resulting debris hazard in orbit.

Illustration 1.34 – Uncontrolled Re-entry. The dangers of uncontrolled re-entry of spacecraft components are illustrated here in a photograph of a Delta 2 third-stage motor casing which re-entered the atmosphere in 2001. Weighing about 70kgs, the titanium casing ended up in the desert about 240 km from Riyadh, Saudi Arabia. The attractions of intentional re-entry into a safe area (typically over-water) are apparent. (NASA Image)

174. **End-of-Life at GEO – Graveyard Orbits**. For a satellite in a high orbit, such as GEO, de-orbit and re-entry would require a significant expenditure of propellant, thus shortening the operational mission. The alternative usually adopted is to save a smaller amount of fuel and to use it to boost the satellite beyond GEO. This frees the slot occupied by it in the highly-valued geo-stationary belt, at the cost of creating a ‘space graveyard’ beyond it. Orbital decay from a graveyard orbit beyond GEO is genuinely negligible; modelling yields decay time predictions of millions of years. It is important, however, to ensure that the graveyard orbit is truly circular. If it has

significant eccentricity, there is a danger that over time the eccentricity will increase[102] and eventually the orbit might intrude back into the GEO belt, posing a collision risk. Circularising the graveyard orbit is thus good housekeeping for end-of-life satellites.

175. **Intentional De-orbit from Lower Orbits**. Once a de-orbit manoeuvre is initiated, the process is irreversible. Assuming the satellite is not designed to survive re-entry, much of its structure will burn up or break into small pieces in the upper atmosphere. For some satellites, there is still the real possibility of substantial pieces surviving to the surface, so operators must consider planning safe impact areas for the debris. A controlled re-entry is thus attractive, despite the resulting fuel penalty.

176. **Returning Small Payloads to Earth**. We explained above that designing a satellite to survive re-entry for re-use has penalties that have so far made it an un-attractive proposition.[103] However, some satellites have returned film and similar products to the Earth for exploitation. Designing a small capsule for such a payload is an easier proposition, and the propellant requirement for such a small mass is manageable. Various methods have been exploited to recover the payload on the surface.

Illustration 1.35 – Returning Small Payloads. This picture shows a USAF C-119 transport aircraft capturing a film capsule returned from Earth orbit by an early reconnaissance satellite. After the ‘trapeze’ mechanism snagged the parachute, the capsule was winched back into the aircraft via the ramp and the undeveloped film retrieved for processing. C-130 aircraft were also used for the purpose. Despite its unwieldy appearance, the system proved rapid and reliable enough for regular employment between 1960 and 1972, although a similar system had a very public failure in 2004, when the NASA *Genesis* mission, due to return particles sampled from the solar wind to Earth, suffered a parachute failure and eluded the helicopters waiting to capture it near Salt Lake City, Utah. (USAF Picture)

---

[102] This does not contradict the observation in Paragraph 141 about atmospheric drag tending to circularise an elliptical orbit; both GEO and the graveyard region are at altitudes beyond measureable atmospheric drag. Other effects tend to increase eccentricity in this region. See also Paragraph 142 concerning gravitational anomalies at GEO.

[103] See Paragraph 167.

Characteristics of Spacecraft

177. **Size, Shape and Mass**. Size constraints apply principally to launch, and are governed by the need to integrate the spacecraft onto the launch vehicle. It is common for parts, such as antennae and solar arrays, to be extensively folded at launch to fit inside the protective fairing at the top of the rocket. After launch, when the spacecraft is clear of the atmosphere, the fairing is discarded and the craft unfolds. While this maximises the potential payload, failures in deployment of elements are not unknown, with potentially terminal effects on the mission. The spacecraft mass incorporates the payload, whatever support infrastructure it requires and fuel for manoeuvre and attitude control. Mass is at a premium for the designer as it drives launcher size and cost, though there is no physical constraint on the size or mass of a satellite from orbital considerations.[104] Conversely, small, lightweight satellites increase the options possible at launch, both in terms of orbits achievable with a given launcher, and the possibility of using small launchers (such as air-launch systems) if mass and size can be reduced enough.

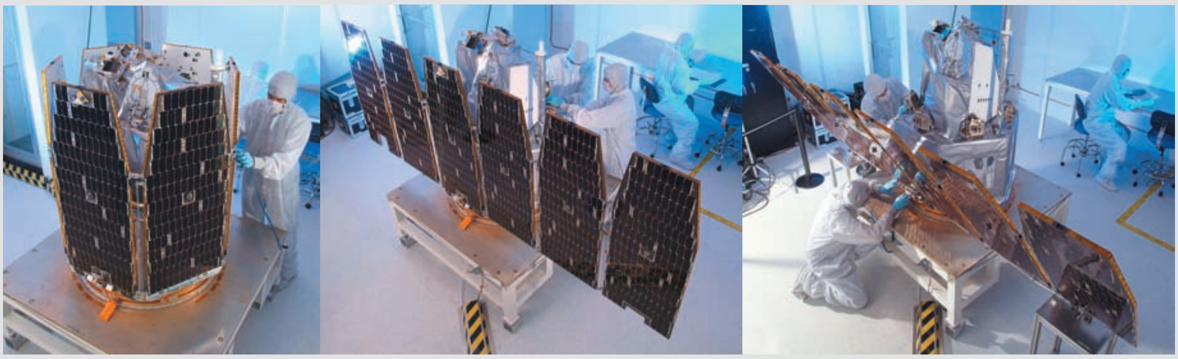


Illustration 1.36 – Deploying Spacecraft Components after Launch. This sequence of three images shows (left to right) a ground test of the solar panel array deployment sequence of the NASA AIM scientific research satellite. The first shows the array folded in a form that will fit into the fairing of the launch vehicle. After launch, the panels unfold then pivot to the correct angle to capture sunlight. (Image courtesy of Orbital Sciences Corporation)

178. **Common Designs and Multiple Payloads**. Multiple payloads may share a common core space vehicle.[105] This may be driven by the nature of the mission, for example adding an additional experiment to a research package. In other cases, it may be the characteristic of the orbit that is attractive to multiple users.[106] For instance, early search-and-rescue surveillance packages were mounted 'piggy-back' style on meteorological satellites;[107] there was no overlap of mission at all, but both required an orbit with a regular revisit of the Earth's surface, from a co-incidentally similar altitude. For some applications there are commercial advantages to hosting secondary payloads onboard a satellite and sharing launch costs, and the term 'hosted payload' is commonly applied in this context. Other launchers simply lift multiple satellites for separate deployment, possibly into disparate orbits by smaller booster rockets once initial orbit is achieved.[108]

---

[104] The International Space Station has grown continuously through the addition of modules since its 'core' was launched, and will not reach its final size until late-2010. Even then, there is probably still 'growth potential', though retirement of the US Space Shuttle fleet will bring construction to an end. The only obvious physical constraint is the need to boost its LEO periodically. Increasing size is probably increasing drag and speeding orbital decay, and at the same time increasing the demands on the thruster used for periodic re-boosting.

[105] Note this is not the same as using a single launcher to place multiple satellites in orbit. We are discussing combining multiple payloads in one satellite.

[106] Where there is no coincidence of mission, the term 'hosted payload' is sometimes used.

[107] See Paragraph 366e.

[108] Illustration 4.4 on page 4-10 shows two satellites being mated together for combined launch – in this particular case into two similar, but widely spaced orbits.

179. **Onboard Power Generation – Solar Panels**. We do not have space or scope to consider all the engineering demands that payloads place on a spacecraft as a whole, but power generation can have direct impact on potential missions so we will look at it a little closer. The need for electrical power is ubiquitous, and it is generated in most spacecraft by solar panels. These use semi-conductor materials to generate electricity directly from sunlight. Depending on its orbit, a satellite may or may not pass through the Earth's shadow once per orbit.[109] If it is in shadow regularly, it will almost certainly require a substantial battery for power while in shadow, though this creates an immediate mass penalty. Missions that involve continuous transmission of some kind – typically radar or communications broadcast – may be constrained significantly, while others are less affected. Solar panels need to be kept oriented directly at the Sun for maximum efficiency, and their output deteriorates over time due to age and debris impact.

Illustration 1.37 – Power Generation in Space. The International Space Station (ISS) depends entirely on solar panels to generate electrical power. The extensive array of panels (at least 14 are visible in this image) is one reason that the ISS is so bright when seen in the night sky. Two Soyuz spacecraft are docked with the ISS in this view (top right corner and extreme right as seen here), each of which also has its own solar panel array extended. (NASA Image)

180. **Onboard Power Generation – Other Approaches**. Other power sources are possible. Manned spacecraft have used fuel cells, which produce electricity directly from combining oxygen and hydrogen,[110] while deep-space probes that will operate too far from the Sun for reliable solar

---

[109] See Paragraph 323c(2).

[110] Fuel cells are attractive on manned missions because hydrogen and oxygen may already be carried as rocket fuels, and the cells generate (clean) water as a by-product.

power may use radio-active power sources.[111] Some Earth-orbiting satellites have also used radio-active power generation, despite environmental concerns relating both to launch failures and to eventual re-entry of orbiting systems.[112]

181. **Payload Integration and Potential Conflicts**. We will explore individual payloads in detail in Chapter 3, but note for now the need to reconcile the characteristics of the payload or payloads with the rest of the spacecraft. There are numerous possible conflicts, which may be mitigated by careful design. Apart from orbit constraints inherent in the payload, such as required altitude, ground track or footprint, the spacecraft must not interfere with the payload mission. Thus an ELINT payload may constrain communications, one using an infrared sensor may place specific cooling demands on the rest of the vehicle, and payloads raise their own pointing and steering demands. Where multiple payloads share a spacecraft, apart from obvious conflicts already mentioned, issues like shared communications may need to be addressed

SECTION V – AN INTEGRATED SPACE CAPABILITY

In this section, you will find a discussion of:

- The various components that make up an integrated national space capability.
- Alternatives to developing an indigenous national space capability.
- The ground infrastructure required to support a satellite in orbit.
- Geographical constraints on the ground component.
- Possible incremental steps by which national capability can be developed.
- Resilience and redundancy in an integrated capability.
- The growing capability of small satellites, which may offer an alternative approach to responsive space capability.

182. **Launch Capability**.

a. **Organic Launch Capability**. In the earliest days of space exploration, launch capability was the preserve of governments, although in the USA at least, commercial use of government launchers quickly followed. As exploitation of space capability has spread to scientific and commercial users as well as to the military, the situation has become more complex. Distinctions between government-owned, commercial and 'hybrid' facilities vary between countries, and critical national security facilities are now routinely provided through public-private partnerships around the world. We have also seen that geography

---

[111] The Pioneer 10 and 11 deep-space probes were both powered by such sources and achieved operating lives of about 30 years before contact with them was lost. Their successors, Voyager 1 and 2, using similar devices, have both exceeded the 30-year point and are expected to maintain power at least until the 40-year point. See also Paragraph 420b regarding deep-space power generation issues.

[112] During the Cold-War, Soviet 'RORSAT' radar reconnaissance satellites employed nuclear power generation on a routine basis in LEO. At the conclusion of the mission, the radioactive component was ejected into a higher 'graveyard' orbit and the remainder allowed to re-enter the atmosphere. Accidents were not unknown; in 1978, the Soviet Cosmos 954 RORSAT mission re-entered the atmosphere in an uncontrolled fashion, polluting of a tract of Arctic Canada, after its reactor core had failed to achieve a safe storage orbit. Legal proceedings for compensation followed (see Chapter 2, Illustration 2.1 for details). The USA also developed an experimental reactor core, test flown as the 'SNAP 10' mission, but did not proceed with development. SNAP 10 was a more ambitious system than its Soviet equivalent, with a higher power output. It too concluded with the core in a high orbit, where it remains to this day.

imposes significant constraints on what can be achieved from a given launch site.[113] Any country that does not have the capability it needs within its boundaries, but which requires assured access to space, must strike an acceptable agreement, either on a government-to-government basis, or with a commercial partner. Novel capabilities, such as air-launched satellites, may also have particular appeal to some nations.

Illustration 1.38 Real-estate Demands for Launch. The operational demands of a launch facility should not be underestimated. This picture shows Kennedy Space Centre (KSC) at Cape Canaveral, Florida on 8th April 1964. The launch in the foreground is an unmanned orbital test of the Titan II/Gemini spacecraft combination, but almost all the land in the background is part of KSC too. Note also the advantage of over-water access for increased safety, and that Cape Canaveral is about as far south as is practical for launch in the continental USA. This minimises constraints of achievable inclination, and maximises the potential contribution of Earth rotation to orbital velocity. (Image courtesy of NASA)

b. **Access via Commercial Launch**. Aside from private use of government facilities, there are now launch enterprises that operate on a wholly or mainly commercial basis. One of the largest is the Arianespace organisation, which has established a track-record for the launch of large payloads from French Guiana. It would seem reasonable that such facilities will continue to proliferate. The extent to which a government can rely on a commercial enterprise to deliver a critical national capability, and the mitigation they require against failure is not a new problem, and is not unique to space operations. There is also growing availability of space *products* on a commercial basis, where the risk of supplying capability lies solely with the operator; commercial telecommunications companies provide services to governments worldwide, and there is growing commercial availability of surveillance products of various kinds.

---

[113] See Paragraph 164.

c. **Access to Allied Space Assets**. If a country requires access to space capability but lacks some component of the system, plainly it can negotiate with an ally to fill the gap. This may involve launch of a national payload by the ally, or direct or indirect access to the ally's space capability. The degree of sensitivity associated with such support depends entirely on what is being offered. Access to surveillance assets may be exceedingly sensitive and constrained, whereas promulgation of threat warnings derived from space assets, such as ballistic missile launch, may be widely disseminated by all available means.

183. **Control of Space Platforms – the Ground Segment**. A satellite may require complex facilities for launch, but once established in orbit, it will also require facilities for its management and exploitation. These are generally separate capabilities, though they might be physically co-located. Equally, where the payload is nationally sensitive, they might be deliberately distant. We will look separately at each. The 2 categories cannot be completely separated however, as the role of the space vehicle is to support the payload mission.

a. **Managing Spaceflight**. This series of capabilities provides management and monitoring of the space vehicle:

(1) The satellite's position in space must be monitored to ensure it is in the required orbit. If it deviates, or its mission requirements change, commands may need to be issued to manoeuvre the satellite.

(2) Onboard systems such as cooling, communications, power generation and attitude control must be monitored and if necessary adjusted or re-programmed.

b. **Managing the Payload**. The second series addresses the requirements of the payload:

(1) Sensors may need to be aimed and operated, antennae steered and communication channels allocated.

(2) Broadcast signals may be sent up to the satellite for re-transmission.

(3) Satellites watching or listening for specific events may need to promulgate results or alerts.

(4) Payload elements may be ejected for re-entry and recovery.

184. **Location of Ground Elements**. There may be constraints on where ground facilities are located, depending on the orbit and the nature of the mission. These must account for communication with the satellite at the required intervals, balancing the communication needed, the height of the orbit (which defines frequency of communications opportunities and duration of visibility of the satellite), and ground track (which dictates from where the satellite can be seen). Security concerns may also need to be addressed.

Illustration 1.39 – Ground Support Elements. Two aspects of the ground segment illustrated. The picture on the left shows an early satellite ground antenna installed at NASA/JPL's Goldstone ground station in the Mojave Desert in California. This particular antenna was installed to support the Project ECHO passive communications experiment described in Chapter 3. The picture on the right shows Schriever AFB in Colorado. The Master Control Station for the Global Positioning System (GPS) constellation operates from here, alongside other military space capability. (Images courtesy of NASA/JPL (left) and USAF (right))

185. **Relay Satellites and Inter-satellite Communications**. Where a satellite needs continuous communications, but where the ground track or orbit precludes this being achieved directly, it may need to communicate via relay satellites in higher orbits. In other applications, such as within the Iridium constellation,[114] communication is direct between the system satellites. Inter-satellite communications can be achieved by RF (radio) or by laser-based systems. Some systems, such as NASA's Tracking and Data Relay Satellite System (TDRSS), exist solely to enable and enhance communications between on-orbit systems and the ground segment.

186. **Design and Construction – Platforms and Payloads**. Countries that do not possess a launch capability and which do not aspire to one can nonetheless develop their own payloads for launch by others. Projects of this nature vary from building a single item for incorporation into a multi-national payload, (for example a scientific experiment or sensor), through to developing a complete vehicle for launch by others, followed by national control and management. There is, plainly, a possible progression for a country along a track such as this, and the precise route taken will depend on national policy and aspiration. Developing a launch system requires niche capabilities, though it may be a logical development for missile-operating states. Payload development may, however, come from, and draw on, a variety of national capabilities in various engineering and science disciplines.

187. **Reaction Times**. Military users of space can be expected to take a particular interest in time constraints and the overall responsiveness of the capability. The issues raised straddle the tactical, operational and strategic domains.

a. **Instant Response**. Some aspects of space capability demand near-instantaneous levels of response. Since the era of cold-war nuclear deterrence, space-based sensors have watched continuously for signs of missile launch, forming part of a firing-chain that required detection, analysis, dissemination and determination of response, all within the short flight time of a ballistic missile.[115] The development of credible Ballistic Missile Defence (BMD) systems has tightened the timeline even further, and has introduced the

---

[114] See Paragraph 327a.

[115] For argument, typically 20-30 minutes.

capability to respond to a missile in flight. BMD capability is not solely space based, but is certainly space-enabled, with consequent demands for continuous availability.[116] The capability is not only technically challenging but also expensive.

b. **Redundant Systems**. Many aspects of current military capability depend either directly, or indirectly, on space. Where these dependencies can be established and quantified, the extent of national security risk implied can be measured. Since, in many cases, there is no prompt mitigation for acute failures, the only practical alternative is to build redundancy into the system. Thus the demand for satellite communications channels may deliberately be over-matched, providing both scope for growth (or requirement creep) over time, and a measure of system redundancy. Where a capability such as GPS is delivered by a constellation of satellites, a small number of extra satellites can be launched as on-orbit spares. If a single failure occurs, capability may be degraded until the spare can be manoeuvred to fill the gap, but the timelines implied are much shorter than those associated with replacing a satellite by procuring and launching a spare.

c. **Operationally Responsive Space**. Some space capability cannot be supplanted or protected by over-provision. An example of this might be a requirement for high-resolution optical surveillance, or a temporary enhancement or provision of satellite communications in a specific area of operations. Growing interest in the potential of small satellites is leading to exploration of the possibilities of launching responsive missions to provide focussed capability, constrained by location, duration or both. The concept is in its early stages, but interest is high and there is a clear willingness to explore relatively unconventional approaches. The USA's implementation of ORS is examined in Chapter 4.[117]

---

[116] See Paragraph 353 for discussion of missile warning, and Paragraphs 386 onwards for BMD practicalities.

[117] See Paragraph 425d.

(INTENTIONALLY BLANK)

CHAPTER 2 – SPACE AND THE LAW

This Chapter considers the legal constraints and privileges resulting from or applying to military uses of space.

201. **Sources of International Law**. Public international law is classically derived from 3 main areas: treaties between states; customary law[1] (rules developed from the practice of states which are binding on states generally); and finally, as a subsidiary source, precedents and the teachings of prominent scholars. As space law is a relatively young subject, substantial customary practice has yet to emerge, and scholars have had little time to teach. You will not, therefore, be surprised to learn that much space law is based on treaties. The origins of space exploration were firmly rooted in national activity, so space as a subject of international law was a logical development. Increasing commercial and civil use of space now complicates the situation. It is not within the scope of this publication to explore these developments in detail; instead we will concentrate on the law as it affects national applications.

202. **Overflight**. Given the sensitivity that states generally display about overflight of territory, it is interesting to construct an argument that allows satellites to overfly territory on an intrusive basis, from orbit.[2]

a. **Borders**. The origins of the law in space relating to overflight lie in limits on the Earth's surface. A country can regulate its land borders, and with very few constraints decide who and what can or cannot be admitted to it or pass through it.

b. **Territorial Waters and Innocent Passage**. When this law was extended to the sea, the principle of **territorial waters** emerged. A country's sovereignty extends to its territorial sea (also known as territorial waters) over which it exercises a full jurisdiction to prescribe and enforce laws. Foreign ships, including warships may exercise innocent passage through the territorial waters, and innocence is defined negatively as being 'not prejudicial to the peace, good order or security of the coastal state'. Passage must be continuous and expeditious. The coastal state should not hamper innocent passage or apply laws in a discriminatory way, and may only temporarily suspend innocent passage if essential for the protection of its security, for example to permit weapon exercises. Where an international strait is completely overlapped by the territorial waters of coastal states,[3] a special regime applies. International Straits Transit Passage allows the exercise of freedom of navigation solely for the purpose of continuous and expeditious transit. By contrast with innocent passage, transit passage may never be hampered, impeded or suspended by the coastal state, and vessels may transit in their normal modes.[4] Thus submarines can transit submerged, and aircraft can be launched and recovered, both of which are impermissible in innocent passage.

---

[1] International law scholars differentiate within customary law between 'custom' and 'peremptory norms', but this distinction is not pursued here.

[2] Some authorities would maintain that the argument here is a *post hoc* justification, and that overflight is sanctioned by custom alone. In this view, once the former USSR and the USA had launched satellites that overflew each others' territories, and as other entrants to the space age did the same, intrusive overflight became the norm, and thus acceptable. The precedents of maritime and air law were conveniently amenable to this, but essentially irrelevant.

[3] For example in the Straits of Dover, where British and French territorial waters abut.

[4] If a vessel was to act in a hostile manner, then in doing so, it would not be exercising its right of straits transit passage, and the state, of course, would have the right of self-defence.

c. **Aviation**. When aviation developed, the law adopted was a hybrid of land and maritime principles. A state controls the airspace above its land territory and territorial waters, and no right of innocent passage exists for a 'state aircraft'.[5] Thus overflight of territory by state aircraft in peacetime is controlled by diplomatic clearance, where permission is agreed between specific states, either on a standing basis or individually. However, the right of 'straits transit passage' also permits transit through national airspace between blocks of international airspace, subject to the over-riding right of a state to protect itself from attack from the air.[6] Within international airspace, aircraft routes are actually unconstrained, though states may agree to limit them by consent, for example to enhance safety by creating airways. Failure to comply with those routes by an aircraft may be a breach of the membership rules for International Civil Aviation Organisation (ICAO),[7] for example, but would probably not be a breach of international law in any other way.

d. **Overflight in Space**. Countries can limit overflight by air because generally, alternative routes exist. Straits overflight specifically covers a situation where an alternative does not exist. The principle of overflight in space is arguably analogous to straits overflight. As demonstrated in Chapter 1, a satellite would not be able to alter its orbit on a continuous basis to avoid overflight of a particular landmass, so essentially the principle of straits overflight is applicable everywhere.[8] A stable orbit is thus seen as necessary to effective spaceflight, akin to a strait, and a country cannot, therefore, restrict overflight of its territory by a 'state satellite'. The provisions of the Outer Space Treaty, which will be examined in more detail, strengthen the presumption of right of overflight, and also protect activity in orbit.

203. **The Legal Definition of Space**. Neither the Outer Space Treaty, nor any other international agreement, expressly *defines* outer space. The informal definitions given at Paragraph 104 suffice for now. There have been periodic attempts to establish a legal definition, which plainly would have implications for overflight by creating a minimum altitude at which it could take place. Opposition to such a definition is based on the belief that no pressing need exists, and that the creation of a definition may inhibit exploitation of emerging technologies. Demonstration of a practical application of capability in near-space, however, may force resolution of this issue.

204. **The Outer-Space Treaty**. The Outer Space Treaty of 1967[9] emerged from the activities of the United Nations (UN) Committee on the Peaceful Uses of Outer Space (COPUOS), which had been established in 1958. It remains the over-arching treaty covering space activity, though its provisions have been extended through several additional conventions described below. Except where specifically required by context, it makes no distinction between military and civilian activity in space, so applies equally to both. As of 1 January 2008, 98 nations, including all the major space-

---

[5] States have complete and exclusive sovereignty over their airspace, and no right of innocent passage exists for either civil or state aircraft. However, the Chicago Convention on International Civil Aviation provides a framework to regulate civil aviation overflight.

[6] The classic example of this was Operation ELDORADO CANYON, the USA's bombing of Libya from the UK in 1986. France and Spain refused overflight for the Operation. The aircraft involved asserted 'Straits Overflight' to pass over the Straits of Gibraltar en-route from international airspace over the Atlantic to international airspace over the Mediterranean, without seeking consent from Spain or Morocco. The aircraft plainly did not harbour hostile intent with respect to those countries, which could not therefore prevent the passage. As detailed in footnote 4 above, a state retains its right to act in self defence.

[7] The International Civil Aviation Organisation (ICAO) was established as a result of the Chicago Convention of 1944, an example of treaty law, applicable to those nations who have agreed to its provisions.

[8] A space vehicle cannot, however, infringe national airspace without consent during launch or re-entry.

[9] Full title: *The Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and other Celestial Bodies*.

faring states, are parties to the Treaty, while a further 27 have signed the treaty but not yet ratified it.[10] The key provisions of the Treaty are as follows:

a. **Weapons of Mass Destruction**. The Treaty prohibits the placing of any Weapons of Mass Destruction (WMD), including nuclear warheads, in orbit around the Earth. It also prohibits placing them on the Moon (explicitly), on any other celestial body or otherwise stationing them in outer space. Note that it does not prohibit stationing conventional weapons in orbit, nor does it regulate nuclear weapons passing through space without achieving orbit, so it does not inhibit possession or use of ballistic missiles. Other terrestrial provisions plainly continue to apply to WMD, and the Treaty says nothing about the nature of operations using space, which are still bound by the Laws of Armed Conflict in the normal way.

b. **Other Weapons and Military Activity**. The Treaty specifically prohibits establishing military bases, testing weapons of any kind or conducting military manoeuvres on the Moon or any other celestial body. It does not, however limit such activity in Earth orbit.

c. **Sovereignty**. The Treaty forbids any country from claiming territory on a celestial body,[11] maintaining that the exploration and use of outer space is 'the province of all mankind'. While this provision is of little practical importance now, it may have implications if exploitation of space resources such as mineral-rich asteroids ever became practical. The launching nation would also probably maintain sovereignty and jurisdiction over any activity occurring in a manned spacecraft, and a satellite remains the property of its owner. See Paragraph 205a(2) for the implications for liability, and Paragraph 207 for more on territorial claims.

d. **Peaceful Use**. An overarching aim of the treaty, set out in its preamble, is to promote the peaceful use of space for the benefit of all mankind. However, with the exception of the specific prohibitions applied to WMD in Space, and to projected activity on the surface of space bodies, there is no express prohibition on military activity in orbit. Thus, for example, reconnaissance activity, missile warning, space surveillance and the use of terrestrial weapons that rely on space capability (such as Global Positioning System (GPS) guided munitions) are not regarded as breaches of the Outer Space Treaty.

205. **Other Regulations Relating to Space**.

a. **Conventions Expanding Provisions of the Outer Space Treaty**.

(1) The **Rescue Treaty** of 1968 commits signatories to assist in the rescue of spacecraft personnel where able, the retrieval of space objects outside their territory of origin and covers the safe return of people and property to their original owners.

(2) The **Liability Convention** renders states liable for damage to people or property caused by their space activities, whether caused in space or on Earth, and sets out the rules governing such liability. The Outer Space Treaty and the Convention accordingly make launching states responsible for damage caused by

---

[10] The newest signatory at the time of writing is North Korea, which signed the treaty in March 2009, prior to their attempted satellite launch in April 2009.

[11] This is another close parallel to maritime law, which prohibits claims of 'ownership' of international waters, in the absence of any surface territory in the claimed area. In this respect, the action of the Apollo astronauts placing an American flag on the Moon's surface was viewed by some as controversial.

certain non-governmental activities in space, and thus also implies launching state responsibility for commercial activity and its consequences in space.

Illustration 2.1 – Liability for Activity in Space. The 3 pictures above show the consequences of the uncontrolled re-entry of the Cosmos 954 satellite, including its nuclear power source, during January 1978. The wreckage that survived re-entry impacted in Northern Canada. The top picture shows a very small piece of contaminated debris being secured in a container for safe disposal. Bottom left is a picture of the first piece of debris found, and on the right 3 of the search party explore a possible further find. Canada claimed against the USSR for the costs incurred in the clean-up operation. The claim was finally settled by a bilateral protocol between Canada and the USSR, which invoked the principles of the Liability Convention explicitly. A comprehensive account of the cleanup can be found at http://cgc.rncan.gc.ca/gamma/ml_e.php , and the text of the protocol is available at http://www.jaxa.jp/library/space_law/chapter_3/3-2-2-1_e.html (Images courtesy of Natural Resources Canada)

(3) The **Registration Convention** establishes a UN register of space objects. Launching nations must create and maintain a public register of orbital elements for objects placed in orbit, and separately notify the UN of such actions. Different nations interpret the provisions of this convention in different ways, and in the absence of an effective enforcement mechanism, compliance is not consistent. The detail required for compliance provides little information on the mission payload, so those details must be inferred in other ways.

(4) The **Moon Treaty** was an attempt to extend and update the Outer Space Treaty. Although the text was agreed through COPUOS action in 1979, only 12 nations, none of them space-faring, have ratified the treaty, and consequently it is widely regarded as having failed.

b. **Other Aspects of Law Affecting Space**:

(1) The **Limited Test-Ban Treaty** specifically banned nuclear explosions in space, at least in peacetime.[12]

Illustration 2.2 – High-Altitude Nuclear Test. An aerial photograph (the wingtip of the KC-135 aircraft from which it was taken can be made out at the bottom of the picture) of the USA's STARFISH PRIME high-altitude nuclear test in 1962. One of a series of high-altitude tests, conducted at the same time as similar experiments by the USSR, their severe consequences for orbiting satellites (including TELSTAR 1 as described in Chapter 1) led to their explicit banning under the Limited Test Ban Treaty. (Image courtesy of US Government Defense Threat Reduction Agency).

(2) The **Anti-Ballistic Missile Treaty** regulated and limited the deployment of anti-ballistic missile systems. It was drafted during the Cold War, with the intention of stabilising deterrence between the USA and the USSR. The USA no longer regards itself bound by the treaty.

---

[12] As of July 2008, France and China are not signatories to this treaty.

(3) Various **Disarmament Agreements** implicitly touch on space capability. The **Anti-Ballistic Missile Treaty** and the **Strategic Arms Reduction Treaty (START 1)** relied on verification of disarmament actions by surveillance from space. De-commissioned bombers for example had to be dismantled and left in the open for verification of such action by other parties from space. Interference with the verification means in any way, for example by obstructing space operations or affecting communications with a surveillance satellite potentially would have been a *de facto* breach of the disarmament agreement.

Illustration 2.3 – Verification of Treaty Compliance from Space. This picture shows the remains of B-52 airframes dismantled at Davis Monthan Air Force Base (AFB) in compliance with the US/USSR START 1 agreement. The treaty required that after dismantling, the airframes had to be left for inspection from space for a period of at least 3 months to verify compliance. Similar constraints applied to USSR activity. Although this picture is an aerial photograph taken at close range, sufficient detail could be seen from space to satisfy both parties. The treaty explicitly forbade interference with any space system carrying out verification activity. (USAF Photo courtesy of 309th AMARG Business Affairs Office, Davis-Monthan AFB).

(4) Where the context permits, space activity may be bound by terrestrial law in addition to space treaties and other agreements. Thus, the principles of the Laws of Armed Conflict such as proportionality and discrimination apply equally to activity reliant on space capability and to alternative terrestrial courses of action. Similarly,

a government agency conducting surveillance of its own nationals or territory from space might find that national privacy regulations still applied.

206. **PAROS – The Prevention of an Arms Race in Outer Space**. The relatively minimal provisions relating to restrictions on arms in space enshrined in the Outer Space Treaty have caused some nations to consider bolstering its provisions. One attempt at this is the **Prevention of an Arms Race in Outer Space (PAROS)** proposal. This is a draft treaty that has been presented to the UN via the Conference on Disarmament at Geneva. The current draft is sponsored by Russia, with support from China and other nations. There are no immediate prospects of its adoption or ratification, but it is indicative of continued interest in the topic.

207. **Territorial Claims**. Notwithstanding the provisions of the Outer Space Treaty, there have been attempts to establish territorial claims in space. In 1976, 8 equatorial nations drafted the **Bogota Declaration,** claiming sovereignty over the portions of geo-stationary orbit that lay over their territory. Since then, other equatorial nations have periodically revived the issue with additional similar claims. The issue has been discussed periodically at the UN, but has received little international support.

208. **Allocation and Use of Geo-stationary Orbits**. Despite opposition to territorial claims, there is plainly a need for de-confliction of space activities in peacetime between states. This is nowhere more obvious than in the location and use of geo-stationary satellites. In Chapter 1 we demonstrated that a geo-stationary orbit is only possible at a specific altitude and inclination. Additionally, a geo-stationary satellite has a fixed view of the Earth, so the available orbits above populous areas of the Earth are particularly valued. Since the vast majority of geo-stationary activity relates to communications and broadcast activity, the allocation of orbital ‘slots’ is coordinated (by mutual agreement of the states concerned) by the International Telecommunication Union (ITU), a UN agency for information and communication technology issues. There is a well-established process for nations to take requests to the ITU to launch a satellite into geo-stationary orbit. Once an allocation has been given, the nation or entity concerned has a fixed time to occupy the allocated position and make use of it. Failure to do so means the location can be re-allocated to another user. This prevents the established space-faring nations from restricting access by ‘blocking’ useful locations with pre-emptive bids.

209. **RF Spectrum – Allocation and De-confliction**. Although it is felt most acutely in the allocation of geo-stationary orbits, a requirement also exists to de-conflict other space uses of radio-frequencies (RF). The consequences of interference between satellites could be serious from safety considerations, as well as potentially constituting unlawful interference with another party’s Outer Space Treaty rights to peaceful use of space. Depending on mission, a satellite may well require protected communications with its control authority, as well as de-confliction of any signal it transmits in the course of its mission. This de-confliction is conducted under the same ITU processes as other RF allocation. The ITU Constitution notes that military activity is not restricted by the ITU, but requests that military users observe de-confliction measures ‘so far as possible’. Allocation and use of frequencies by deployed forces may also be regulated by the host nation as part of the Status of Forces Agreement covering their deployment.

(INTENTIONALLY BLANK)

CHAPTER 3 – MILITARY USES OF SPACE

301. **Categories of Operation**. The UK MOD Space Strategy provides a framework for analysing Space Operations. Such a framework is useful given the multiplicity of roles executed from space and the wide variations in spacecraft design. The Strategy divides Space capability in to 4 main areas: Data Networks in Space, which further divides into Satellite Communications and Position, Navigation and Timing Applications; Surveillance from Space; Surveillance of Space and Space Control. We will follow this breakdown in this part of the Primer, though for completeness, we will begin with a brief examination of ballistic missiles and potential space weapons, since their enabling technology is inextricably linked with that of true Space capability, and many of the mechanisms countering relevant to them are also Space based.

302. **Ballistic Missiles and Space Weapons**. The origins of military rockets can be traced back to medieval China. They came to Europe via India and British colonial military experiences. The Second World War saw the emergence of practical modern rocket motors and missiles, with reach beyond the close battle. Within one generation of military experience,[1] missile technology matured from relatively crude systems such as the German V2, to modern nuclear-armed missiles with global reach. During the same period, speculation mounted that weapons in space could pose a continuous threat to Earth's surface, or alternatively lead to the migration of terrestrial combat to Outer Space. We will look shortly at the development and categorisation of ballistic missiles, and the technical possibilities of Space-based surface weapons.

303. **Data Networks in Space**. Data Networks in Space have become pervasive, and in many cases transparent to the user. In fact they were the first true Space capability to be exploited by military users. As the term is understood today, Data Networks include directed systems – where messages are transmitted on a point-to-point basis; and broadcast systems – where multiple users access a signal from Space. The most obvious example of the first category is satellite communications ('SATCOM'), while the second encompasses such systems as the Global Positioning System (GPS).

304. **Surveillance from Space**. Space platforms enable a number of surveillance tasks including both conventional passive imagery (gathered at various wavelengths and resolutions), Electronic Intelligence (ELINT), and the use of active illumination systems such as Space-based radar. An obvious example of the utility of surveillance from Space is the huge leap, in modern times, of the accuracy of weather forecasts – facilitated by a wide range of Space-derived data.

305. **Surveillance of Space**. Surveillance of Space is the means by which any country, whether it possesses a national Space capability of its own or not, derives information about other operators' Space activity. Some of this information may best be derived from Space, but usually, it is based on information from terrestrial sensors.

306. **Space Control**. Awareness of Space activity by others is of limited use if no action can then be taken to counter its effects. Space control encompasses that activity.[2] Space Control can be thought of as analogous to air defence, and just like air defence, it may consist of active and passive elements.

---

[1] For example, General Bernard Schriever USAF (1910-2005) entered the US Army as an artillery officer in 1933, flew as a B-17 pilot during World War 2, contributed to exploiting captured German technology immediately afterwards, transferred to the USAF on its formation in 1947, and by the time of his retirement in 1966 was known as the 'father of the ballistic missile' within the US Military.

[2] There are conflicting definitions in use in this area. See Paragraph 378 for details.

SECTION I – FIRES THROUGH SPACE – BALLISTIC MISSILES AND SPACE WEAPONS

In this Section you will find a description of:

- Legal Aspects of Space Weapons.
- The Phases of Flight of a Ballistic Missile.
- Range and Other Classifications Applied to Ballistic Missiles.
- A Short Description of Other Kinds of Space Weapons.

307. **Weaponisation and the Law**. We looked in Chapter 2 at the legal constraints on the use of Space. It is worth summarising those constraints as they apply to weapons in Space. Firstly, the normal Laws of Armed Conflict continue to apply. Thus any weapon residing in or passing through Space must for example conform to the principles of military necessity, humanity, proportionality and discrimination. Secondly, there is a blanket prohibition, resulting from the Outer Space Treaty (OST), on the *stationing* of Weapons of Mass Destruction (WMD) in Space.[3] The customary reading of the OST is that this prohibition does not apply to nuclear weapons passing through Space atop a ballistic missile, but that it would apply to any system that placed a WMD (whether nuclear or otherwise) in orbit. See also the description of 'fractional orbital' weapons in Paragraph 313. Finally, the OST is not generally understood to place a restriction on stationing conventional munitions in orbit.

Missiles

308. **Ballistic Missiles – Phases of Flight**. The dictionary definition of a 'ballistic missile' is '*a missile which is initially powered and guided but falls under gravity on to its target*'.[4] Although rockets have been used in warfare since medieval times, it was only with the development of the German V2 missile in the closing days of WW2 that ballistic missiles as we would recognise them today became a reality. The definition gives rise to the customary description of the phases of flight of a ballistic missile:

a. **Boost Phase**. This is the phase from launch until the engine cuts out. The missile may climb clear of the atmosphere, and will accelerate continually while the engine is running, but it does not reach orbital velocity.

b. **Mid-Course Phase**. This is the phase where the missile coasts in Space. Its trajectory is a sub-orbital ellipse, with the major axis aligned vertically.[5] The ground track of the missile will approximate to a great circle, slightly modified to account for the rotation of the Earth (and consequent movement of the target in Space) while the missile is in flight. Speed reduces until apogee then increases again to a value near the original burn-out speed as the missile approaches the surface.

---

[3] See Chapter 2, Paragraph 204 for further details.

[4] Compact Oxford English Dictionary definition at www.askoxford.com.

[5] Bearing in mind the study of conic sections in Chapter 1, this may seem slightly counter-intuitive; over short ranges, a ballistic trajectory is roughly parabolic (though within the atmosphere, air resistance distorts it away from a true parabola). Over the range of a typical ballistic missile, however, the direction of gravity (always towards the centre of the Earth) changes significantly, and this changes the shape of the mid-course phase from parabolic to elliptical. During the boost phase, the missile is accelerating due to thrust and is being guided actively. During re-entry, air resistance exerts a very large effect that shapes the path significantly.

c. **Re-entry/Terminal Phase**. This is the final phase of flight, from first encounter with the atmosphere on re-entry until impact. For a short-range system, which may never truly leave the atmosphere, the distinction between mid-course and terminal phases is arbitrary.

As a rough rule of thumb, the ground range of a ballistic missile is about twice the maximum altitude achieved (or alternatively, the maximum altitude reached in the mid-course phase will be about half the range.)

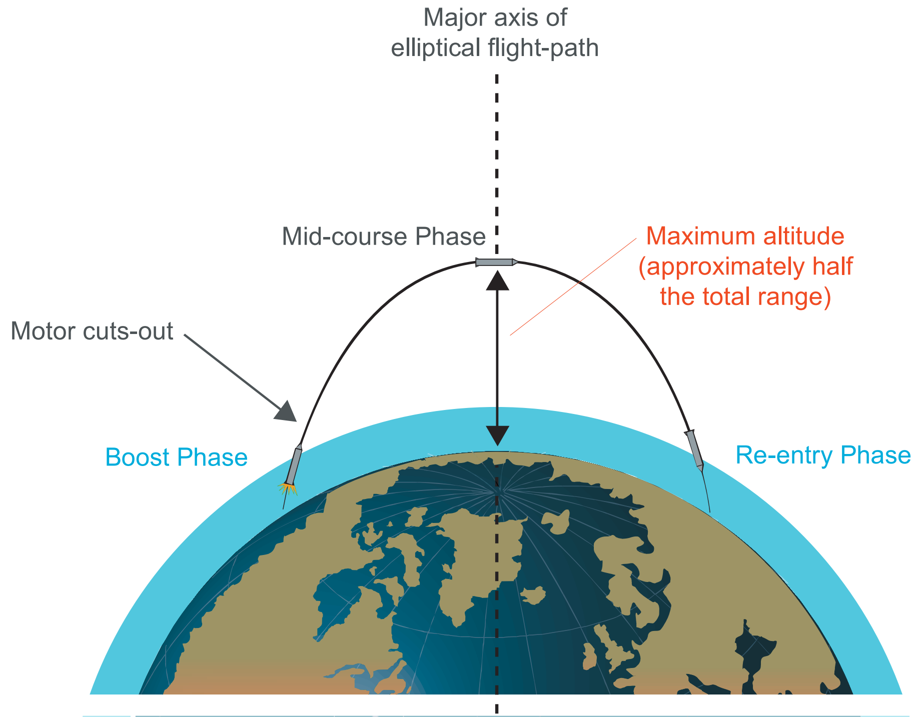


Illustration 3.1 – The Phases of Flight of a Ballistic Missile. The diagram shows the stages of flight of a ballistic missile. During the boost phase, the rocket motor is producing thrust. In the mid-course phase, a modern system may deploy decoy warheads and disperse multiple operational warheads. In the re-entry phase, the operational warheads descend on their targets. See also Paragraph 311.

309. **The V2**. Because of its relative simplicity, the German V2 missile[6] used during 1944-5 in Europe, is a good system to illustrate the concepts of a ballistic missile. It was a single-stage missile, weighing about 12 500 kgs at launch, powered by a liquid-fuel rocket engine using an alcohol/water mixture as fuel, and liquid oxygen as oxidiser. The payload was 980 kg of HE with an impact fuse. On launch, the missile initially climbed vertically (under the guidance of a gyroscope in the nosecone). Manoeuvre was achieved by a combination of moveable aerodynamic surfaces, and deflection of the rocket thrust by graphite vanes inserted into the exhaust.[7] After a short vertical climb, the missile flight-path was inclined onto a pre-set compass heading.[8] The missile continued to accelerate on that heading until the engine cut out (controlled by an accelerometer, which effectively ran the engine until a pre-set speed was reached). It was then in free flight until impact.

---

[6] The initial German designation of the missile was ‘A4’. Dr Goebbels, the Nazi propaganda minister, substituted the ‘V2’ title, denoting ‘vengeance weapon’ when the operational missiles were deployed. A4 and V2 were thus the same system.

[7] These vanes were simple and reliable, and avoided the complexity of having to make the rocket thrust-line moveable for steering purposes. They incurred a significant penalty in the total thrust produced, however, by obstructing the exhaust gases, this limiting the range of the missile. Post-war US experiments with captured V2s included replacing these vanes with alternative steering mechanisms, which yielded significant improvements in range and/or altitude achieved.

[8] Later versions of the V2 used a simple radio beam transmitted from the launcher for initial guidance.

continued to climb, decelerating, and ultimately reached about 88 km altitude (beyond the Karman line). It then fell under gravity to the target. There was no attempt to guide the missile in its descent. At the top of its trajectory, it simply tipped over and fell at the target. The whole missile fell as one unit, with the fins stabilising it as it re-entered the atmosphere. Because it had not achieved anything approaching orbital speed, there was no insuperable problem with aerodynamic heating on re-entry. There were, however, major issues relating to aerodynamic stability at supersonic speeds throughout flight (the missile was supersonic soon after launch and impacted at about Mach 4), complicated by the centre of gravity of the missile moving radically as fuel was consumed. Maximum range was about 320 km.

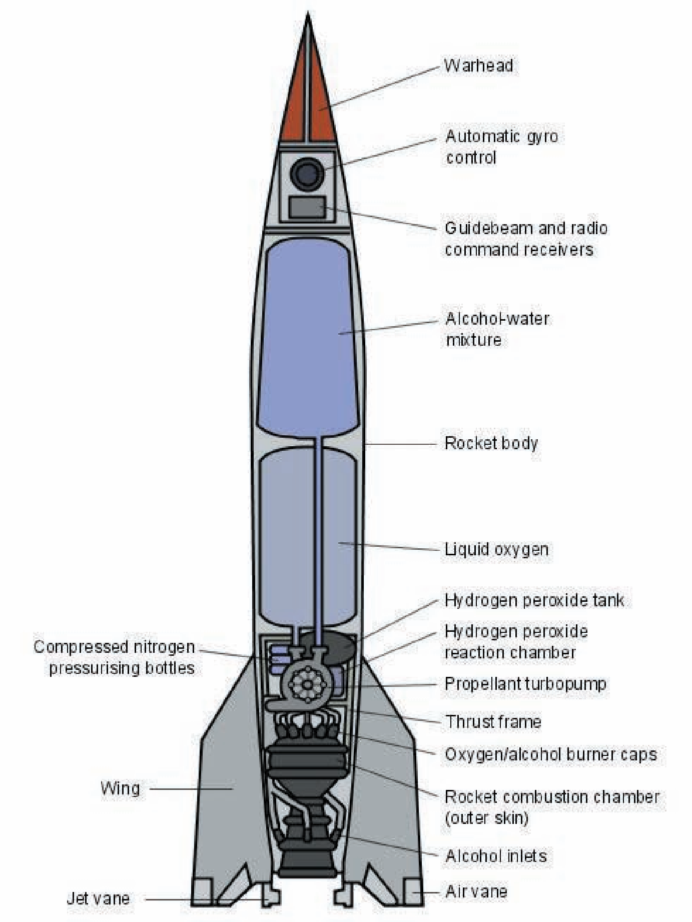


Illustration 3.2 – V2 Missile Components. This diagram shows the major components of a V2 missile to support the description in Paragraph 309. See also the description of the propulsion system at Paragraph E3 and Illustration E.2.

310. **Missile Classifications**. Ballistic missiles are produced in a variety of sizes, which broadly determine their range. Most categorisations sub-divide by range, although for some purposes, launch characteristics (typically submarine, and fixed or mobile surface-launched) are more meaningful. The scheme used by the US DoD is included below in Illustration 3.3. This classification is widely accepted, but note the following additional factors:

a. **Intermediate-Range Nuclear Forces Treaty Classification**. The (US/USSR) Intermediate-Range Nuclear Forces (INF) treaty covered all surface-to-surface missiles with ranges between 500 and 5000 km. This straddles several classes in Illustration 3.3.

b. **Missile Technology Control Regime Limits**. The Missile Technology Control Regime (MTCR) is an informal (i.e. not based on a treaty) agreement to restrict the export of critical missile technology to certain countries. MTCR limits apply to missiles with a range of at least 300 km and a payload of 500 kg or more.

c. **Submarine Launched Missiles**. Submarine launched ballistic missiles (but not submarine launched cruise missiles such as 'Tomahawk') are generically referred to as '**SLBMs**', regardless of their range.

Classification	Abbreviation	Range (km)	Range (nautical miles)
Short Range Ballistic Missile	SRBM	< 1100	< 600
Medium Range Ballistic Missile	MRBM	1100-2750	600-1500
Intermediate Range Ballistic Missile	IRBM	2750-5550	1500-3000
Intercontinental Ballistic Missile	ICBM	5550-14800	3000-8000

Illustration 3.3 – US DoD Missile Range Categorisation

311. **Modern Ballistic Missiles**. Shortcomings in early missiles such as the V2 were addressed to yield the long-range systems now in use. Multiple stages provide increased range and solid-fuel rockets[9] are widespread. Errors in terminal velocity due to variations in solid-rocket thrust characteristics are corrected by complex control systems, utilising highly accurate sensors and active guidance during the latter stages of flight. To this extent, such weapons are therefore not pure ballistic missiles. Nonetheless the key characteristics and phases of flight are still recognisable:

a. **Boost Phase**. This will now include the sequential ignition of multiple stages. It ends when the last stage is expended. For an ICBM, this phase will typically last 3-5 minutes and will reach 150 to 400 km altitude.

b. **Mid-Course Phase**. A modern ICBM may use this phase of flight to deploy decoys of various kinds to ensure survivability of the main payload in the face of missile defence systems. Depending on the range that is required, this stage may last up to 25 minutes, and reach an altitude determined by the required range.

c. **Re-entry Phase**. This stage is usually deemed to commence at about 100 km altitude, and will only last 1-2 minutes. Decoy payloads are not usually designed to survive this phase and consequently are destroyed as they encounter the atmosphere. Because range is ultimately dependent on the flightpath and velocity achieved at the end of the boost phase, modern long-range systems reach higher velocities than earlier systems, and re-entry heating has become a greater issue for the designer.

312. **Ballistic Missile Payloads**:

a. **Nuclear Warheads**. The capabilities offered by nuclear-tipped ballistic missiles have sustained deterrence for many years, and we have noted that they are not regarded as contravening the OST. Their possession and some of their capabilities have been constrained by the various (mainly bilateral, USA-USSR) Strategic Arms Limitation Treaties.

---

[9] Solid-fuel rockets are more compact, due to their greater thrust, and safer and more responsive since they do not require fuelling before launch. They can remain ready for use for months or years

b. **Other WMD Options**. A ballistic missile tipped with any other WMD, such as a biological or chemical payload, would be regarded as a breach of the treaties prohibiting their possession. There would, however, be few practical difficulties with their construction – Saddam Hussein's threat during the Gulf War to deploy 'Scud' missiles with chemical or biological payloads was judged to be credible. The greatest practical difficulty might be delivery of the payload in an effective form given the impact velocity of a missile.

c. **Conventional Warheads**. Improvements in guidance accuracy are such that there are now also discussions on the potential of conventional (HE) warheads mated to intercontinental ballistic missiles.[10]

313. **Fractional Orbital Systems**. During the Cold War, the Soviet Union developed an orbital weapon system described as a **fractional orbital bombardment system**. Essentially, this was an ICBM that achieved low-Earth orbit (LEO) for a short period, prior to de-orbit and re-entry of a (nuclear) warhead. The system offered several perceived advantages. It would allow attack on the USA from any direction (e.g. approaching from over the South Pole), as opposed to the more limited trajectories imposed on sub-orbital missiles as a function of range and launch positions. It could also threaten any point under the orbit's ground track, thus concealing the intended target until late in flight. The Soviets avoided OST stipulations by only conducting trials with unarmed missiles. Fractional Orbital systems were specifically banned by the SALT II disarmament agreement of 1979.

Space Weapons

314. **Categorisation**. 'Space Weapons' in their broadest sense encompass weapons threatening ballistic missiles as part of **ballistic missile defence**, weapons countering orbiting satellites as **anti-satellite systems**, and (hypothetically at present) Space-based weapons threatening targets on the surface of the Earth. In this Section of the Primer, we will look at those systems that might threaten surface targets. We will look at anti-missile and anti-satellite systems in Section VI.

315. **Practicalities of Anti-Surface Orbiting Systems**:

a. As far as is known, stationing a conventional weapon in orbit, with the intention of posing an enduring threat to surface targets, has not been attempted by any Space-power. There would undoubtedly be significant political issues posed by development of such a capability. Setting these to one side, there would also be practical difficulties to be overcome.

b. If the intention were to pose an enduring threat, a stable orbit at an altitude sufficient to make decay negligible would be desirable. Higher orbits would also have the advantage of bringing a greater surface area under threat; the achievable re-entry paths for payloads would be constrained by orbital manoeuvre considerations. On the other hand, increasing altitude also imposes energy constraints on controlled re-entry for the payloads. Re-entry energy management might thus be a critical limit for practical payloads.

---

[10] V2 accuracy (of the order of kms) was sufficient for psychological effect on a city-sized target, but not for any recognisable degree of precision – though in general, this was true for the conventional bomber aircraft of the time too. Accuracy (perhaps to hundreds of metres) quickly improved enough to make nuclear warheads credible as counter-force weapons. The accuracy required to achieve precise effect with a conventional payload (tens of metres or better) is still at the very bounds of the possible; short-range battlefield rockets such as MLRS have for a long time, of course, made wide use of conventional payloads.

c. Re-entry to a desired target area would also be constrained by over-flight intervals. To pose a quasi-continuous threat, even to a restricted area, would require a relatively elaborate constellation of satellites.

d. Since the payload would initially be travelling at orbital velocity, it would need to be constructed to withstand the heating caused during re-entry. It might be possible to design a robust kinetic payload that simply re-entered and impacted on the target. A means of safe disposal at the end of the satellite's life would have to be a key design consideration.

e. Directed-energy systems might offer an alternative route for Space-to-surface effects, but they have at least two serious design issues to address. They would be required to operate through the Earth's atmosphere (and potentially through cloud), and would therefore be subject to issues of turbulence and absorption. They would also require a source of energy – possibly a chemical reaction of some kind, or maybe even a nuclear reactor. There are potential issues of limited firepower or endurance for a practical system which might make the overall concept expensive relative to the potential effect.

SECTION II (A) – DATA NETWORKS IN SPACE – COMMUNICATIONS

In this Section you will find a description of:

- The drivers behind the development of satellite communications.
- Implications for use of the RF spectrum by communications satellites.
- The practical constraints and characteristics of SATCOM systems, including selection of orbits, and the implications of different frequency bands.
- Some practical examples of military and civil SATCOM systems.
- The possible impact of internet applications in space.
- Satellite broadcasting.

Introduction

316. **Long Range Communications**. Communication beyond visual range has been a military goal since ancient times; the limitations of line-of-sight communications persisted into the 19th century, when semaphore towers dotted the European countryside. The advent of radio communication probably seemed like the answer to many prayers, but its limitations quickly became apparent when military wireless sets were deployed. Much terrestrial radio communications is a compromise between range and portability, particularly when two-way communications, as opposed to broadcast reception, is required. Range is limited either by power, specifically the ability to generate a strong enough transmitted signal or to receive a weak one, or by terrain.[11] The higher the radio frequency used, the more information a signal can carry. Unfortunately, the physics of transmission mean that high frequencies are confined to line of sight use. Low-frequency radios, which offered the first practical over-the-horizon option, were (and are) plagued by poor voice quality due to interference from atmospheric effects, bulk (due to the required antenna size), and by slow data-rates when used for digital signals.

---

[11] Transparency and propagation at various wavelengths are described in slightly more detail in Annex F.

317. **Relay through Space**. SATCOM is an attempt to overcome the limitations of long-range radio by allowing line-of-sight signals (with all their inherent advantages of quality and capacity) to be relayed via Space. This, and the implication that they are ultimately still just a form of radio communications, should be remembered.

Satellite Communications – SATCOM Theory and History

318. **Active and Passive Satellites**. The fact that communications signals can be reflected has been recognised for a long time, with some terrestrial long-range systems depend on reflection of the signal off layers in the atmosphere.[12] SATCOM simply moves the relaying or reflecting to orbital altitudes. When artificial satellites were developed, both reflection and re-transmission were explored.

a. **Passive Satellites**. The USA experimented with the 'Echo' series of satellites in the early-1960s. Essentially they were orbiting aluminised balloons, off which signals could be bounced. They proved the feasibility of the concept but were not taken further. Other early experiments proved the feasibility of bouncing signals off the surface of the Moon. A major limitation of such devices is that the signal strength after being sent up to the satellite, reflected (at less than 100% efficiency) then re-directed back to the surface is marginal for most purposes. Such a system has, however, the advantage of only using one frequency for the link up to and down from the satellite. Passive communications satellites are no longer in use.

Illustration 3.4 – A Passive Communications Satellite. The picture shows a test inflation of the NASA 'ECHO' satellite in a hangar. The operational satellite was launched into orbit then inflated, where it acted as a passive reflector for ground transmissions. Note the onlookers and their saloon car to gain a sense of scale. The associated ground installation is illustrated in Chapter 1. Although the principle proved successful, active satellites quickly supplanted such systems. (Image courtesy of NASA).

---

[12] Both the **troposphere** and the **ionosphere** are used for this, though to achieve differing ranges and using different frequencies. The troposphere is the lowest layer of the atmosphere, reaching to about 12 km altitude, and reflection to achieve short ranges can be achieved at its upper boundary. Ionospheric effects take place at much higher altitudes, typically hundreds of kilometres, and the resulting propagation can span continental distances.

b. **Early Active Satellites**. In 1957, Sputnik 1, the first artificial Earth satellite, transmitted a radio signal that was received over much of the populated surface of the Earth. Sputnik 1 was not, however a true communications satellite; rather it was transmitting a tracking and data signal back to its ground station and, perhaps not accidentally, achieving political effect by broadcasting its presence in orbit.[13] The first dedicated communications satellite was the experimental SCORE[14] satellite launched by the US in December 1958.[15] SCORE could receive and re-broadcast a signal. It could also store a message on a tape-recorder for later transmission,[16] an important breakthrough in the days before geostationary (GEO) satellites.[17] Communications applications were some of the earliest drivers of satellite technology, and by 1960 the US Army Signal Corps had demonstrated a practical relay satellite (the COURIER project) that could re-broadcast teletype data over long ranges. COURIER was also, incidentally, the first satellite fitted with solar panels to recharge its batteries. By 1962 the TELSTAR satellite, first described in Chapter 1, could relay telephone and television signals between ground stations.

Illustration 3.5a – The Project SCORE Launcher. Seen just before launch, the SCORE payload mounted on an Atlas booster. (Picture courtesy of USAF)

---

[13] The characteristics of the received tracking signal were explored to determine Sputnik's orbital elements, based on the known locations of the receiving stations. The idea occurred to several scientists that if the signal characteristics were known exactly beforehand, the process could be reversed and the location of the receiving station deduced. See Paragraph 331 for more detail.

[14] 'SCORE' stood for 'Signal Communications (by) Orbital Relay Equipment'.

[15] Just like Sputnik, SCORE had an ulterior political imperative – to demonstrate the orbital potential of the Atlas missile used as a launcher. The entire Atlas missile body (less two solid boosters jettisoned during launch) entered orbit, with the SCORE payload a rather lonely 150lb package incorporated in the 9000lb missile body.

[16] To guard against failure of the uplink to the recorder, SCORE was launched with a pre-recorded message already aboard. The first message it broadcast was thus a pre-recorded Christmas greeting from President Dwight D Eisenhower.

[17] SCORE sat in an elliptical orbit with a relatively low perigee of 114 miles. Thus users would typically have to wait for the satellite to come into view over the horizon before they could send or receive messages.

Illustration 3.5b – Project SCORE. A contemporary publicity diagram explains the principles. A smiling President Eisenhower (who recorded the inaugural message) has his words relayed to the ground. The depiction of virtually the entire rocket fuselage entering orbit is essentially accurate, although the depiction of multiple formed beams being projected to Earth is probably more speculative. (Picture courtesy of NASA).

319. **Modern Communications Satellites**. In the same way as you can see the progression from the V2 to the modern ICBM, there is a clear line of development from SCORE, COURIER and TELSTAR to modern capability such as SKYNET V. We now examine each stage of the process in turn.

a. **Uplinking**. A ground station, which may be the source of a broadcast signal, or a hub for users intent on point-to-point communication, sends a signal up to a satellite in orbit. The signal may, of course, have originated elsewhere, and reached the ground station via a terrestrial link such as a telephone system. It is transmitted to the satellite on the **uplink frequency**. The signal need only be strong enough to reach the satellite in a useable form, though note that as it still travels through the atmosphere on its way up it is thus subject to atmospheric attenuation and other interference sources.

b. **At the Satellite**. At the satellite, the received signal is processed. This may just be 'cleaning-up' to remove noise, but may also involve analysis to extract routing information. The satellite then rebroadcasts the signal on a (different) **downlink frequency**. This frequency shift allows the uplink receiver and the downlink transmitter to run simultaneously without mutual interference with each other. Each receiver/transmitter pair on the satellite is commonly referred to as a **transponder**, and a modern system will often contain multiple transponders, to provide redundancy and increase capacity.

c. **Modern Satellite Antennae**. Some satellites broadcast a signal down to a large area of the Earth's surface; commercial TV broadcast satellites, for example, may serve a relatively large area, while military broadcast systems may span an operational theatre. For many other applications, however, there are great advantages to be had by using narrow beams to communicate:

(1) There are power advantages in not sending wasted signal to illuminate areas of the Earth's surface where they are not required. Concentrating the transmitted signal allows the receiver to use a smaller, less sensitive antenna e.g. for mobile reception.

(2) There is growing pressure on allocations of RF channels for communications, and there is a need to avoid interference between multiple users of the same frequency. If the satellite transmissions can be directed to specific areas of the Earth's surface, the same frequency can be re-used many times in different areas without the signals becoming confused. Similarly, national policies on frequency allocation vary, and it may be necessary to avoid radiating a particular signal to a particular geographical area.

(3) There are communication security advantages in only transmitting where required. Not only is eavesdropping made more difficult, but the satellite is also made more resistant to jamming.

Modern communications satellites thus deploy sophisticated antennae, typically capable of forming multiple discrete beams on specific ground targets and steering and switching those beams as required. The limits on this capability are driven by the frequency used – forming narrow beams gets harder at lower frequencies– and by antenna size. Larger antennae can form narrower beams, but the antenna must be of a size and weight that can fit into the launcher and deploy reliably once the satellite is in orbit.[18]

d. **Downlinking**. The downlink transmitter sends the signal back down to Earth, either to a discrete receiver, or by broadcasting to multiple receivers. As the downlink transmitter has to rely on the satellite power supply, there are greater constraints on downlink signal power than for fixed ground stations. Thus efficiency at the satellite end may be at a premium. Where either end of the ground chain is a mobile installation, however, that too may face power, efficiency and size constraints.

---

[18] This is another example of 'resolution' as described in Annex F, governed by the same parameters. The constraints of aperture and wavelength work in an analogous way for radio as they do for optical systems.

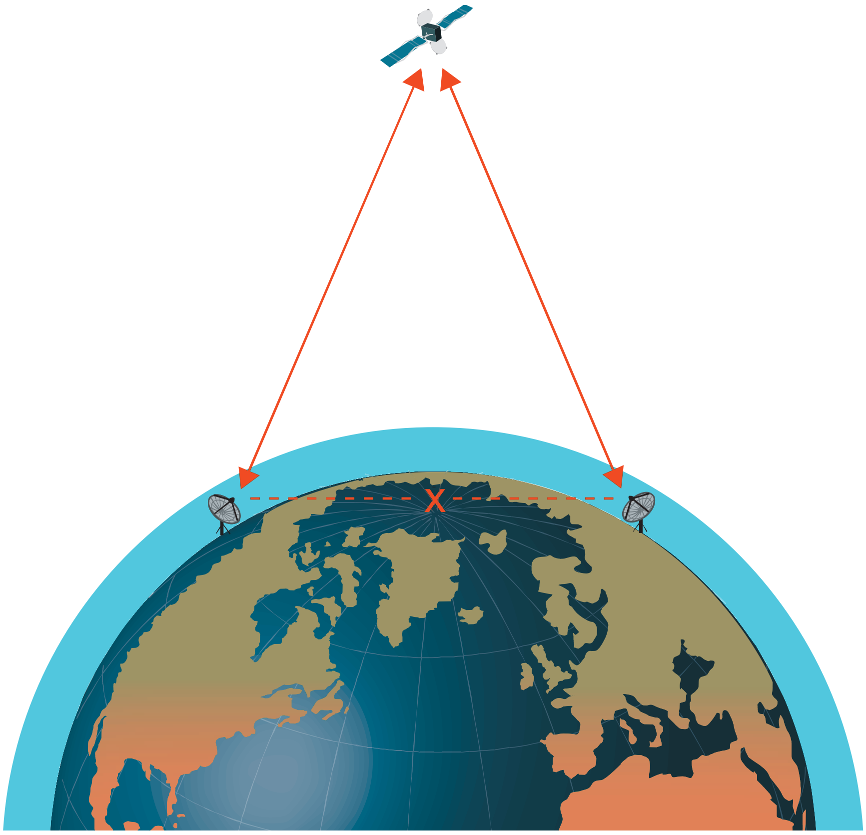


Illustration 3.6 – Simple SATCOM. So long as two ground stations share visibility of an appropriate orbiting satellite, they can use it for communications, even if they do not share a terrestrial line of sight. If the satellite appears to move across the sky, however, the area on the Earth's surface covered by the satellite will also move, and thus coverage will not be continuous. In the interests of ensuring a strong signal, the ground station antennae may also have to track the satellite's movement, at a considerable cost in complexity. See Paragraph 323b-d for how and why GEO satellites offer one solution to this problem.

Communications and the RF Spectrum

320. **Communications Frequency Bands**. In the explanation of the Electromagnetic (EM) Spectrum contained in Annex G, we use the informal categorisations of 'Radio' and 'Radar' waves, within which we included microwaves. This is actually a very large portion of the EM spectrum, and to understand communications, and particularly SATCOM, we need greater discrimination. It is impossible to discuss them in any depth without encountering apparently ambiguous and overlapping definitions and naming conventions, so we now attempt to simplify matters.

a. **International Telecommunication Union Definitions**. The International Telecommunication Union (ITU) has produced a comprehensive and relatively simple classification of communication bands based on frequency. There are no gaps or overlaps between bands. The full listing is reproduced in Annex G; this is where you find definitions of terms such as 'Very Low Frequency', 'High Frequency' and suchlike.

b. **Military Frequency Band Conventions**. While military usage generally follows the ITU convention, there are minor conflicts at the 'edges' of the bands. In particular, note that at least for aviation applications, the VHF/UHF boundary would normally be defined as 225 MHz rather than 300 MHz. Although this could have implications for SATCOM, in practice, frequency bands are usually superseded by letter designations described below.

c. **Wavebands**. In everyday conversation, the use of waveband names was once common, and has still not entirely died out. In particular, **short-wave (SW), medium-wave (MW)** and **long-wave (LW)** descriptions were often used in connection with broadcasting. Note that there is no direct correlation between these and the ITU bands. LW signals lie in the LF band, MW signals straddle LF and MF, and SW signals straddle MF and HF.

d. **Modulation**. In this plethora of abbreviations, it is also important to separate out references to modulation schemes, which describe how the information is encoded within the transmitted signal. **Amplitude Modulation (AM)** and **Frequency Modulation (FM)** are the most common analogue systems in use, and in the commercial arena are often associated with specific frequency bands (AM on MW, FM on VHF), but note that they refer to different characteristics of the overall communications system. Digital signals employ other modulation systems again; space does not permit detailed explanation of how these work.

321. **Frequency Band Letter Designation**. Even the ITU Frequency Band system is too imprecise for accurate SATCOM designation. The American Institute of Electrical and Electronic Engineers (IEEE), which plays an active role in defining many international technical standards, maintains a convention, known as **IEEE 521**,[19] which sub-divides the UHF to EHF bands, denoting the sub-divisions by letters. This is the most common system used to describe SATCOM; it is used by MOD. A full listing can be found in Annex G. This is where you find descriptors such as ‘X-band’ and ‘K-band’. As UK users, we will mostly be concerned in this Primer with systems using X-band, which lies in the SHF range, and S-band, which lies mainly in the UHF range. US Military SATCOM use occupies a large portion of the IEEE 521 range.

322. **Competing Letter Designations**. There is plenty of scope for confusion with letter designations. IEEE 521 is updated periodically, so the latest version may not match its predecessors exactly. The ITU has its own radar-band nomenclature, which is similar to, but not the same as IEEE 521. NATO and US military authorities have a letter-based frequency band system used in ECM which does not match IEEE 521, though it covers a similar range. Finally, radar engineers have an agreed standard for waveguide design used in radar construction, which is similar to the ITU standard, though again it does not match it exactly. For our purposes, we do not need the precise level of detail defined in these systems, but the reader should at least be aware of the pitfalls.

---

[19] The latest iteration of ‘521’ at the time of writing is IEEE Std 521-2002, which superseded 521-1984.

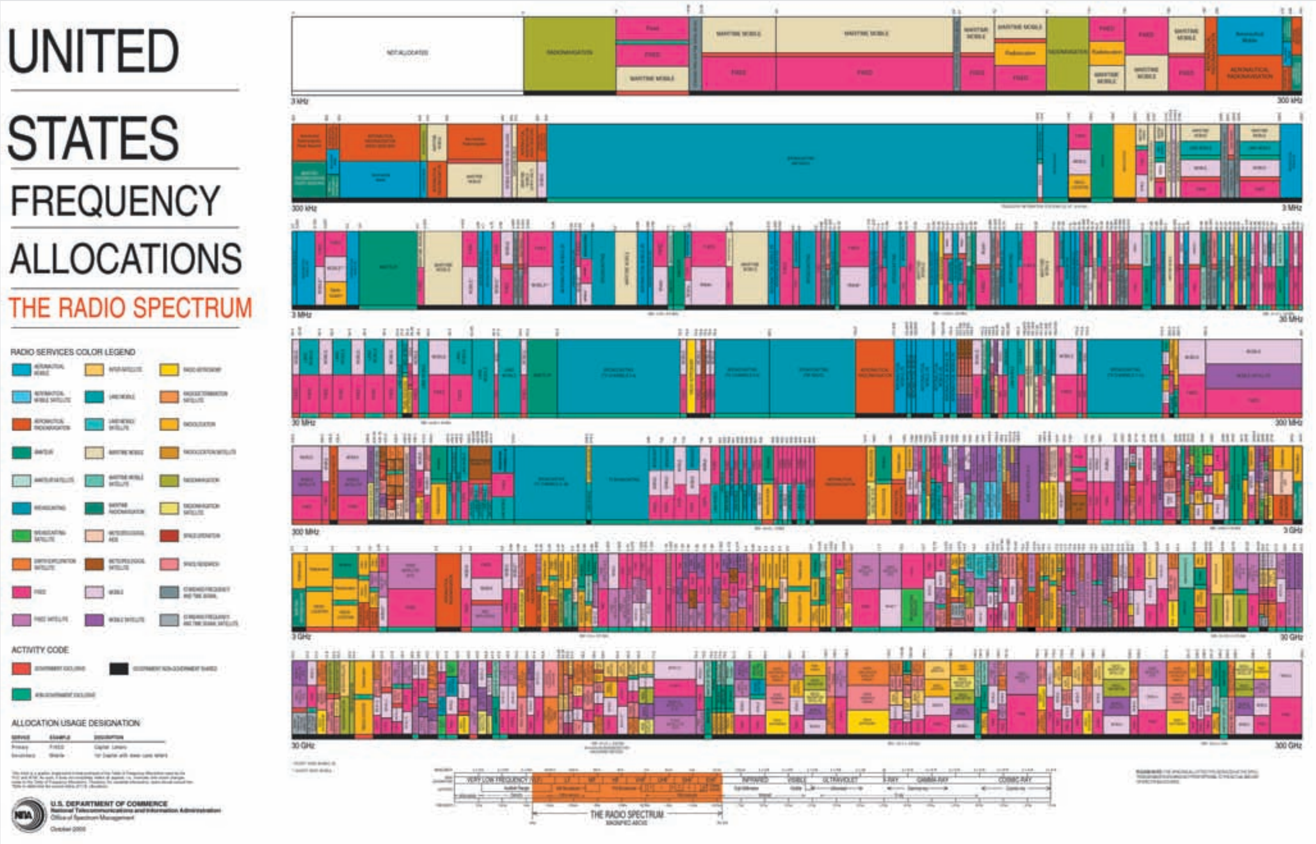


Illustration 3.7– The Complexity of the RF Spectrum. This diagram is not intended to be legible at this scale. It is included simply to illustrate the complexity of RF Spectrum management. It shows the US Department of Commerce's understanding of RF Frequency band allocation in 2003. The RF Spectrum's place in the overall EM Spectrum is illustrated by the orange portion of the small diagram bottom centre. The original of the diagram can be consulted at a legible scale at http://www.ntia.doc.gov/osmhome/allochrt.pdf, although readers should note that it is probably already obsolete. (US Department of Commerce graphic)

Practical SATCOM

323. **Design Constraints on a Communications Satellite**:

a. **Bandwidth**. A common constraint on satellite communications is bandwidth, which is explained in more detail in Annex F (Paragraph F25). Bandwidth describes the overall capacity of a communications channel. It may be helpful to consider it as being like the capacity of a ship's hold or of a railway wagon, which might be able to carry a variety of loads, possibly simultaneously, but which will still have an overall capacity limit, governed by constraints of weight or volume. It is a real constraint on effective satellite communications. The bandwidth capacity provided by a transponder can be sub-divided according to the needs of the users, but trade-offs need to be made between the number of signals carried and the nature of each signal. Typically, a satellite transponder will be flexible (within limits) on the kind of signal it can re-broadcast (data, voice, video etc), but at any given instant, the operator has to allocate its capacity to either a few high-usage signals (e.g. video) or multiple lower-grade signals (e.g. voice). It cannot do both at once. Practical examples are generally complex, and lie beyond the scope of this Primer, but the reader should understand what the trade-offs are and why they have to be made. Encryption tends to increase the bandwidth requirements of a specific signal.

b. **Orbit**. Many communications satellites occupy geostationary orbits, for reasons that were explored in Part 1[20] and are recapped and expanded below. There are both advantages and drawbacks to GEO, though their popularity indicates that the advantages outweigh the drawbacks for many applications.

c. **GEO Advantages**:

(1) The satellite appears stationary in the sky to a ground user. This obviates the need for complex tracking mechanisms on the ground to keep antennae pointing at the satellite. Commercial Satellite TV broadcasters, for example, can therefore utilise simple dish antennae, which need to be aimed carefully on first installation, but can then be left unattended for years.

(2) The sun more or less permanently illuminates a satellite in GEO.[21] This is important for 'power-hungry' downlink transmitters.

(3) At GEO, the satellite can see nearly half the Earth, though with limitations of low elevation at high latitudes that we have already discussed, and by analogy also near the edge of its coverage in longitude (see Illustration 3.8a). A single satellite can therefore service widely separated users. For global coverage at sub-polar latitudes, a relatively small number of satellites spaced around the world are required. In fact, 3 GEO satellites spaced evenly in longitude would constitute a practical minimum.

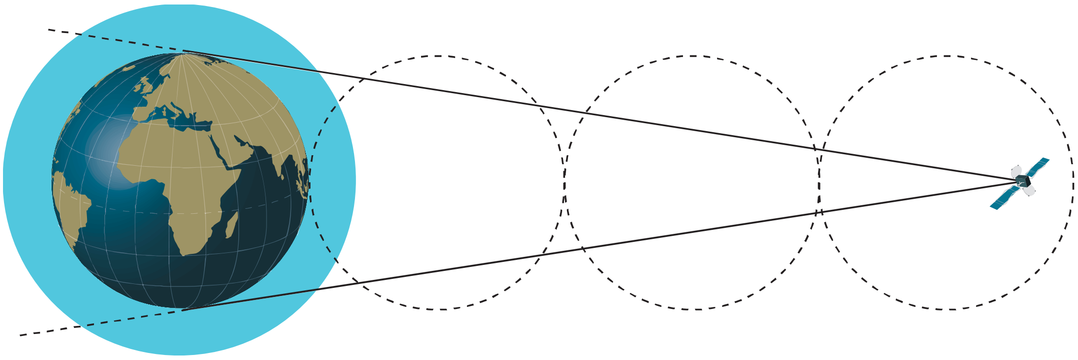


Illustration 3.8a – GEO to Scale – Lines of Sight. The altitude of GEO is about 2.8 times the diameter of Earth, so this illustration shows the altitude (though not the size of the satellite!) roughly to scale. Note that with the extreme lines of sight drawn in, a satellite at GEO is invisible at or beyond about latitudes 81° N or S, and even at those extremes, the satellite will be very low in the sky and the line of sight is through a thick portion of the atmosphere.

---

[20] See Paragraphs 148-9.

[21] There are in fact short seasonal periods where the Earth will eclipse a GEO satellite, blanking the solar panels, so for continuous availability there is still the need to provide battery power. These periods are easily predictable for given orbital elements.

d. **GEO Disadvantages**:

(1) GEO is difficult to reach, due to its altitude. The wish to place heavy payloads in GEO has been one of the main drivers in the development of large launchers, but such launchers are still at the limits of technology, and (expensive and dramatic) failures are not unknown.

(2) The time-of-travel of a radio signal from the surface to the satellite and back is not negligible. While it does not impact on normal voice or data transmission, it may pose a limit to applications such as remote operation of machinery (e.g. remotely-piloted air vehicles). For relay using multiple satellites and ground stations, the delay may become more noticeable (see Illustration 3.8b).

(3) Because of the distance from the surface to the satellite (and back again), received signal strengths are relatively low, with performance implications, particularly for users of mobile equipment.

(4) Because GEO lie over the equator, signal coverage at higher latitudes is relatively poor. We saw in Chapter 1 that Molinya orbits offer an alternative in some areas, at the cost of periodic unavailability and the need for users to track a moving satellite.[22]

(5) Commercial pressure on available GEO slots continues to grow as satellite technology matures. The allocation system for these slots is described in Chapter 2. The implication for military users is that access may become harder and more expensive with time.

(6) GEO is difficult to reach, particularly in a responsive manner, due to the size of launcher required, the need to secure an operating location in the congested GEO belt and the time taken after launch to achieve a stable orbit. There is interest in the potential of small satellites in GEO, as well as in the uses of geo-synchronous[23] orbits that may circumvent some of the congestion in the 'true' GEO band, but these applications are not yet mature.

---

[22] We will see another approach to this problem in Paragraph 327A.

[23] Recall from Chapter 1 that a geo-synchronous orbit is one at geo-stationary altitude, but inclined to the equator. The satellite thus appears to hold position (roughly) in longitude, but to move north-south over a 24-hour period. The path of the satellite over the ground is actually a long, thin, 'figure-8' aligned north-south. Although this imposes some complexity on the ground antenna required to track the satellite, and leads to degraded coverage at high latitudes, such orbits are still useful.

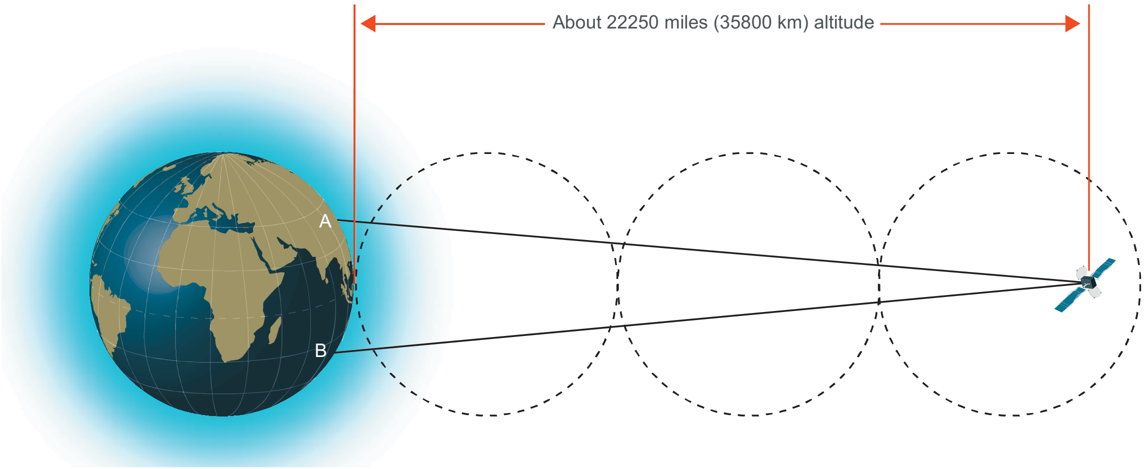


Illustration 3.8b – GEO to Scale – Time of Flight. A potential drawback of communications via a satellite at GEO is the latency, or time taken for the signal to propagate. For users at A and B in the diagram, let us say the distance to the satellite is roughly 38000 km, so the total distance travelled by the signal is 76000 km for the SATCOM link alone. At 300 000 km/sec (the speed of light, and hence of propagation of the radio signal), that will take at least ¼ of a second. This may not seem much, but could prove critical for applications such as controlling equipment remotely. Loss of signal strength over the relatively long distances can also become limiting.

e. **Visibility and Interference**. 'Traditional' SATCOM uses the field of regard of a satellite at GEO to span potentially vast distances; if both ends of the chain can see the same satellite simultaneously, they can communicate, however far apart they are on the ground. The theoretical limit is thus just under half the circumference of the Earth, though practical considerations may limit this further. SATCOM is still bounded by line of sight and transparency; local visibility of the satellite can be constrained by terrain and other obstructions such as urban environments or overhead foliage. Weather and other environmental factors can also affect the RF signals to and from the satellite. Prediction of some of these effects is possible.

f. **Frequency**. Practical SATCOM, whether military or civilian, has to date been concentrated into the UHF and SHF bands. There is, however, increasing interest in the development of EHF (K_A Band) capability. The drivers are practical.

(1) Successively higher frequency bands are defined by the frequency increasing by a factor of 10. Thus each band has a potential for greater numbers of channels, with less interference between adjacent channels, or greater capacity for high-bandwidth/high data-rate signals such as video.

(2) Extra bandwidth available at higher frequencies has advantages for employment of anti-jam techniques.

(3) At higher frequencies (shorter wavelengths), narrower beams can be formed.

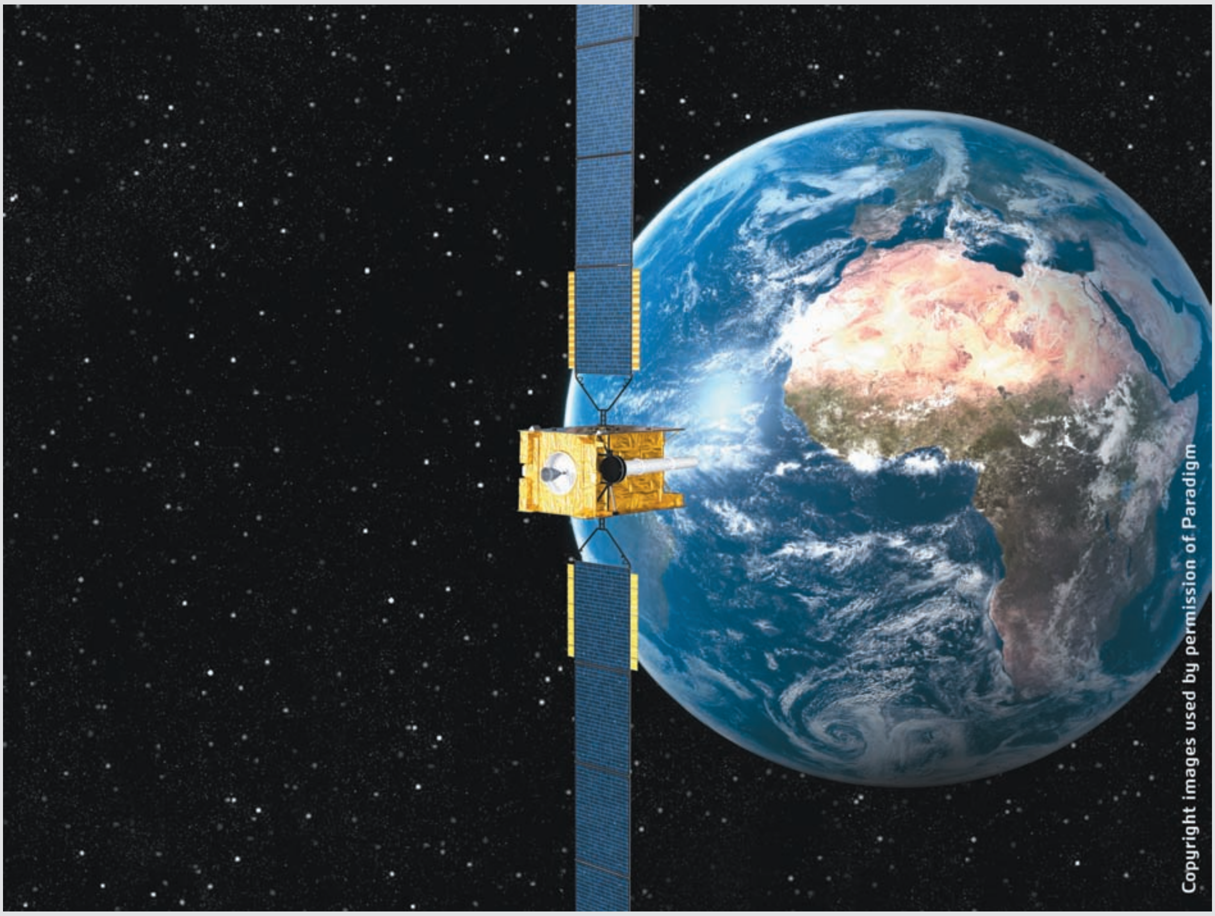


Illustration 3.9 – Modern Satellite Communications. Another illustration of the limitations of GEO; this artists' impression of a SKYNET satellite at GEO shows both that a SATCOM link can easily straddle Africa, or connect the Mediterranean region to the South Atlantic, and that despite this, the polar regions remain stubbornly out of sight. It also shows the limitations in longitude; this satellite could link the Horn of Africa with the East coast of the Americas, but not much beyond that. (Image courtesy of Paradigm Services.)

324. **Military SATCOM Frequency Bands – Practical Implications**. There are potential advantages and disadvantages for the military user in each of the possible frequency bands.

a. UHF SATCOM lends itself to small ground terminals with modest antenna sizes.

b. UHF antennae can be relatively simple, achieving adequate performance without complex design or accurate pointing.

c. UHF SATCOM performs relatively well in poor weather.

d. The UHF spectrum is relatively congested. Available bandwidth will support voice communications, but video or similar will probably be impractical, and there is limited scope for anti-jamming measures.

e. SHF requires large antennae when used for high-bandwidth applications.

f. SHF antennae have to be accurately pointed at the satellite with which they are communicating.

g. SHF performance in weather is poorer than UHF.

h. EHF would provide greater capacity still, but suffers the most from atmospheric attenuation and interference. This can, however, be turned to military advantage, as the relative opacity of the atmosphere at the higher end of the EHF band (around 60 GHz) provides a secure frequency for inter-satellite links, which cannot be intercepted from the ground.

Illustration 3.10 – SATCOM Antennae vary greatly in size. (Clockwise from top left) Tactical use of an Iridium handset, REACHER X-Band ground equipment installation, SKYNET Ground Station at Oakhanger, SKYNET UHF antenna installation at Colerne. (Copyright images used by permission of Paradigm Services).

325. **UK MOD SATCOM**. The bulk of UK military SATCOM capability is currently delivered via the SKYNET constellation of satellites, which provide UHF and SHF services. No further specific details are provided in this Primer as they are subject to change and update, but details are widely available via MOD sources.[24] The MOD also makes limited use of other commercial satellite communications systems, including INMARSAT and Iridium.[25] INMARSAT operates 'conventional' communications satellites in GEO just as we have described here. Iridium uses an alternative approach, which is described shortly. Contractors supporting MOD may also buy commercial SATCOM capacity.

---

[24] SKYNET is operated for the MOD by a commercial partner – 'Paradigm'. The Paradigm website provides basic information about the system.

[25] There are constraints on procurement of commercial services within MOD due to the relationship with Paradigm. The relevant IPT should be consulted for details and implications.

326. **Maritime Satellite Communications.** Ships have made extensive use of satellite communication since the earliest days of its availability. The services provided have expanded with time to encompass contributions to maritime safety as well as traditional communications. Safety requirements for commercial shipping are captured in the Global Maritime Distress and Safety System (GMDSS) regime.[26] GMDSS is a complex regime with differing requirements in various locations and for various categories of vessels, but for large merchant and passenger carrying vessels, it requires automatic alerting of rescue services in the event of emergency, location of an incident via floating buoys following a sinking, and alerting of adjacent vessels in either case. Commercial SATCOM providers offer a variety of installations that comply with GMDSS requirements, but note that those based solely on GEO constellations may not meet the requirements at high latitudes.

327. **Satellite-enabled Mobile Communications**. There are several companies offering mobile personal and business communications via satellites. This Primer does not aim to provide any comment on the commercial or performance advantages of any particular system, though readers would probably recognise some of the relevant brand names.[27] Currently, mobile SATCOM systems can be divided into LEO and GEO based systems.

a. **LEO Systems.** The only LEO system currently in widespread use is the Iridium system. This employs an extensive constellation (of satellites (at present numbering 65 operational spacecraft) in LEO. The constellation is dispersed so that from any point on the Earth's surface, a satellite is continuously visible, though each moves relatively rapidly across the sky while in view. The handset establishes a connection and communicates directly with one of the satellites, and routes voice traffic in the same way as a mobile phone. Handsets interface with terrestrial telephone systems via dedicated ground stations. From the Space-engineering point of view, the interesting feature is how the constellation manages connection with an individual handset, and can transfer a connection from one satellite to another during a conversation when the first satellite sets as seen by the user. This is controlled automatically by the system; Iridium satellites communicate extensively amongst themselves to arrange these handovers and to allow conversations to reach beyond the coverage of the local satellite.[28]

b. **GEO Systems.** The major commercial SATCOM providers have diversified into mobile communications as miniaturization of receiver sets has become possible. They rely on the service provider securing a GEO slot for the satellite and can then exploit its relatively large footprint to provide a

c. **Technical Choices and Comparisons**. Note the implications of the technical choices made by mobile SATCOM providers. LEO systems can use polar orbits, which enables true global coverage with no drop in service at high latitudes. This, however, requires an extensive constellation due to the limited communications footprint of LEO

---

[26] GMDSS is mandated by the International Maritime Organisation, an organisation to which most sea-faring and costal states subscribe. The GMDSS regime began to be implemented during the 1970s and since 1999 has essentially been global in reach. It should not be confused with the **Automatic Identification System (AIS)**; AIS is a terrestrial system for ships, using VHF broadcast from ships and shore stations. AIS may use satellite derived position information, but the linkages between the ships and the shore stations do not rely on satellite communications.

[27] Most of the relevant company websites provide basic details of their satellite architecture and specifically of the geographic coverage they offer. Many of the commercially available handsets are dual-mode, capable of connecting with terrestrial mobile networks as well as with satellites, in order to extend coverage outside the satellite footprint, or where there are cost or service benefits for the user.

[28] See http://www.iridium.com/about/howitworks.php for a slightly more detailed explanation of the Iridium constellation, including the rationale for the orbits used and details of the satellite cross-talk arrangements. There are also details of various proposed enhancements, which may include secondary imaging payloads.

satellites. Because of the short ranges involved, there is very little latency in a LEO relay. LEO may additionally offer advantages for hosted/secondary payloads in forthcoming systems, such as comprehensive imaging coverage from LEO ranges. GEO systems require fewer (but much bigger) satellites, require less management of connections since the satellite does not move relative to the user during a conversation, but the time taken for signals to travel to and from the satellite can cause latency. They depend on allocation of geostationary orbital slots to enable global reach, and have degraded or non-existent coverage in high-latitudes. .

328. **Internet Applications in Space**. Whether based in GEO or LEO, SATCOM as we have described it here is based on **circuit-switched** routing; there is a direct connection from sender to recipient, via some ground path, an uplink and a downlink. Both commercial and military users are now pondering the potential of communications based on Internet Protocol, or **packet-switched** technology. Essentially, this means flying an internet router on a satellite, and allowing a signal to be broken up into packets which can be sent discretely across multiple routes prior to re-assembly at the receiving end. It is important to realise that this is not constrained to e-mail or similar text messages. When two users communicate using voice or video signals over the internet (commonly referred to as **voice over internet protocol** or **VOIP**), their message (e.g. the voice signal) is similarly broken into packets and rebuilt.[29] Whatever the practical limits now, the spread of Internet applications into Space will yield efficiency savings in the use of communication channels, and probably open up applications enabled via satellite that we can only guess at now.

329. **Broadcast Communications**. For many civilian users, the most pervasive application of satellite communications is *Broadcast Communications*, specifically reception of satellite TV signals in the home.[30] From the technical standpoint, this is essentially a one-way version of long-haul SATCOM. The broadcaster assembles his programme material, uplinks it to a satellite in GEO where it is re-broadcast to consumers. True GEO is almost essential,[31] so that users can employ fixed antennae that do not need to track a moving satellite. Commercial digital encryption techniques, decoded via information on a card distributed to subscribers, are used to control access to the system. Most systems only use the satellite link in one direction, from broadcaster to user, and require a terrestrial link back to the operator for interactive services such as pay-to-view programming. International standards do exist, however, for very low data-rate communications from the consumer back to the supplier via the user's dish antenna and the satellite.[32] From the space engineer's point of view, broadcast TV has been a main driver for the development of heavy-lift launchers to place large systems in GEO and has encouraged development of complex beam-forming antennae to concentrate signals in the required user area. Military users do not yet make extensive use of broadcast communications, probably because of the need to make two-way communications at least potentially available to the end-users, but there are theatre-specific exceptions to this.

---

[29] This is one reason why low-grade video-conferencing signals occasionally become 'jerky' – the Internet packets are not arriving quickly enough for seamless re-assembly.

[30] Satellite TV is not confined solely to direct broadcast to the consumer. In areas with cable TV coverage, the signals to be distributed to consumers are usually sent by satellite link to a local hub, prior to distribution across the cable network.

[31] Molniya HEO orbits (see Paragraph 151a) are used to serve specific geographic locations where the line of sight of a GEO satellite is too close to the local horizon for reliable reception.

[32] This standard, known as 'SATMODE' is currently used for a demonstration/prototype satellite-enabled internet service. The signal format would support a variety of interactive broadcast applications.

SECTION II (B) – DATA NETWORKS IN SPACE – POSITION, NAVIGATION AND TIMING

In this section you will find a description of:

- The evolution of the GPS constellation.
- Combining GPS with other navigation systems.
- Alternatives to GPS – Galileo, GLONASS, Beidou/COMPASS and other possible systems.
- Applications of PNT information.

330. **Introduction**. As we noted above, some communications satellites enable point-to-point communication, but others broadcast, sending a continuous message to multiple receivers. The most obvious example of this is broadcasting for entertainment or information purposes, but the same principle – broadcasting a signal containing useful information to a dispersed user population – also describes how satellite-based navigation systems (SATNAV) work.[33] In both cases, the satellites transmit without any direct method of monitoring who receives the signals, or how, or why. The important difference in principle is, however, that broadcast satellites usually control the geographic area they transmit to, while navigation systems usually aim to serve as wide a population as possible.

331. **Early SATNAV Ideas**. Ideas about space-based navigation go back to the dawn of the space-age. In 1957, when the Soviet Sputnik 1 satellite was launched, there was considerable interest in determining its orbital elements, which would in turn yield much useful information on the performance of the launcher, orbital decay and suchlike. One technique employed in the USA was simultaneous reception of Sputnik's broadcast signal by two widely separated receivers on the ground. Both ground stations received the signal modified by the Doppler effect, due to the satellite's orbital velocity. For widely separated ground stations, the Doppler effect was significantly different at each location. By comparing these differences against consistent timing information, orbital data could be deduced. Accuracy depended, however, on the positions of the ground stations being known. This was not a significant challenge in this instance – their positions were well established. It occurred to scientists however, that although the technique relied on using known ground positions to determine an unknown position in space, the same process would work backwards – you could use a known space position to determine an unknown ground position. This was the origin of space-based navigation. Once reliable launch systems were perfected, the US Navy was the first user of the new technology. The first practical SATNAV system in orbit was the US TRANSIT system, whose main purpose was to provide position information system to SLBM systems such as Polaris, fielded by the 'Lafayette' class and other early SSBNs. The first TRANSIT satellite was placed in orbit in April 1960, and the system supplied navigation information until 1996.

---

[33] Because of the way we have subdivided space capability in this Chapter, we will not look here at passive tracking and location applications such as ARGOS and SARSAT, since while these systems provide position information, they do so to a remote operator rather than to the user. Such systems are regarded as instances of surveillance from Space, and details can therefore be found in Section III.

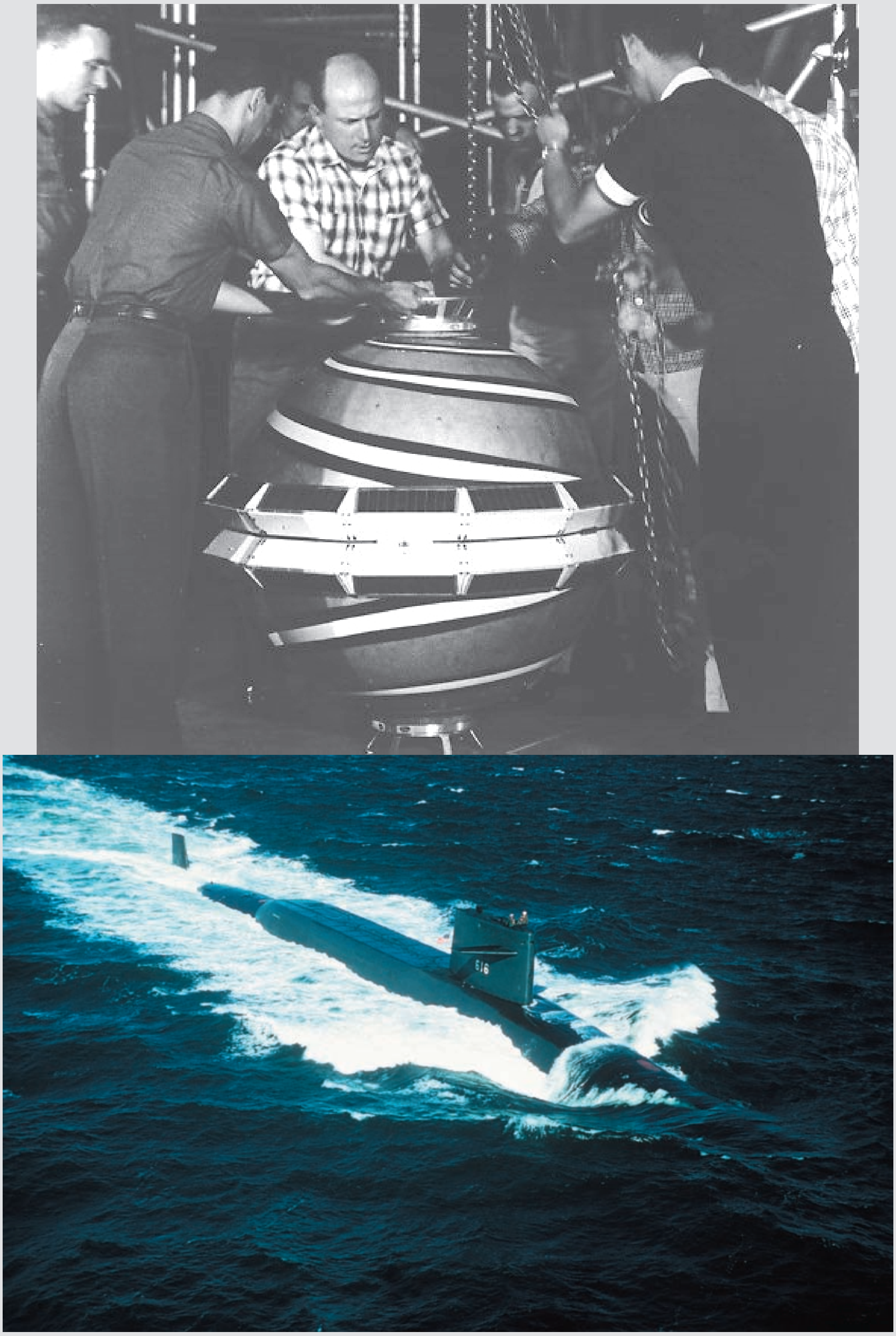


Illustration 3.11 – The TRANSIT Satnav System. Drawing on ideas developed from tracking the SPUTNIK 1 satellite, the USN sponsored the first operational SATNAV system. The first picture shows the TRANSIT 1 satellite being readied for launch at Cape Canaveral in 1960. The aim of the system was to provide positional information to early SSBNs such as the Lafayette class (USS LAFAYETTE shown). Initially, coverage was not continuous or global, but was designed to support the SSBN fleet at sea, without system availability giving away too much about the location of patrol areas. (USAF Patrick AFB Heritage Collection and USN images).

332. **The Global Positioning System**. The US DoD developed a more capable satellite-based navigation system during the 1970s and 80s. The **Global Positioning System** has become the most ubiquitous of several competing systems. Because of its widespread use, a reasonably detailed understanding of how GPS derives position and time information may be of use to the reader; such an explanation is provided at Annex H. Other systems, which are described below, use similar, though not necessarily identical techniques. In summary, GPS exploits the extreme accuracy of timing information available from an atomic clock. Although such devices are expensive and cumbersome, clocks of subsidiary accuracy can be placed in orbiting satellites. A constellation of such satellites is arranged so that multiple satellites are continuously visible to a user. Each satellite transmits radio signals containing this very accurate time information. The GPS receiver compares the time signals from multiple satellites in the constellation, and from the differences in reception time of the signals is able to compute its position. The system is arranged so that the accuracy of the satellites' positions and of their on-board clocks can be maintained despite the usual perturbations of their orbits.

333. **Military and Civil GPS codes – Rejecting Spoofing**. The GPS signal at the receiver is actually very weak. The GPS satellites are in relatively high orbits (i.e. at long range from the receivers), and must radiate their signal in all directions to ensure that they can be received everywhere. The format of the signal is unique, however, and this assists the receiver to separate the signal from background noise, though jamming a GPS signal remains a possibility. Since the signal format is public knowledge, it would also be relatively easy to generate an incorrect signal to 'spoof' receivers. At present, this is overcome to an extent by the existence of two separate but integrated signals, with one portion encoded. Mimicking the encoded signal, which is known as the '**P(Y) code**', would be a much harder proposition than mimicking the public code, which is known as the '**C/A component**'.[34] The fact that all current SATNAV systems occupy similar frequency-bands also makes them vulnerable to common jamming sources, though again, aspects of ongoing signal enhancements may address some of these issues.

334. **Selective Availability**. When the GPS system was introduced, the US DoD was concerned not to make its full potential accuracy available to non-military users. They thus reserved the right to introduce a deliberate error into the system to degrade it. This feature, known as *selective availability* was disabled during the year 2000. The legacy GPS satellites in orbit could implement it again in the future, but the US Government has indicated that future replacement GPS satellites will not support it.

335. **Integrated Navigation Systems**. GPS is commonly used in combination with other navigation systems. We look very briefly at the integration of GPS with inertial navigation systems.

a. **Inertial Navigation**. Prior to the introduction of GPS, the only widely available systems of comparable accuracy were based on **inertial platforms**. These used a physical platform, stabilised by gyroscopes and capable of measuring very small accelerations to calculate position. Inertial navigation (IN) suffers, however, from accumulating errors over time, which need correction by periodically updating the platform to a known position.

b. **GPS Integration with Inertial Systems**. Modern precision navigation systems, particularly those used on weapons and vehicles such as aircraft, typically combine GPS with an inertial (IN) element. The combination of the GPS system's highly accurate position, but with vulnerability to occasional loss of signal, and an IN platform's ability to

---

[34] See Annex H, Paragraph H6, for greater detail of the signal format, frequency allocation scheme and implications on anti-jam capability.

derive position and velocity without a continuous external signal, but tending to drift over time, gives a highly accurate source of hybrid navigation information. Various methods of integration exist. So-called **loose integration** comprises the two elements (GPS and IN) running as independent systems with a simple computer providing a 'best-guess' navigation solution. Improved accuracy, at the cost of processing complexity, can be achieved by **tight integration**. In this arrangement, the GPS and IN systems are closely linked, with, for example, the individual satellite ranges and their rate of change being used to correct errors in the IN portion. So called **ultra-tight integration** systems are now being developed, which, for example, use the fused navigation solution to allow for the small frequency variations in the satellite signals caused by the Doppler effect. All these refinements improve the overall accuracy of the navigation system.

336. **Single-Channel Receivers and Other Legacy Issues**. When GPS receivers were first introduced into the 'consumer' market, a common limitation was that the device contained a fairly basic radio receiver to detect the satellite signals. This meant that in practice, it would detect the signal from one satellite, calculate a range from it, then detect the next satellite, calculate range from that and move on again. Such systems were therefore intended for use on slow-moving platforms. The danger arose where a system designed for hand carried use (e.g. hill-walking) was used for example in a vehicle or aircraft. Older receivers were also sometimes limited by the need to acquire a full set of almanac[35] data via one of the satellites before being able to resolve the satellite signals, or by storing their position when switched off, then using that stored position as an assumed position at next switch-on. These problems have largely been addressed by manufacturers, but may still constrain older systems.

337. **The Practical GPS Constellation**. The constraints inherent in GPS go some way to explaining the orbits chosen for GPS satellites.[36] They operate in MEO (stable orbits with little drag), at an inclination that provides signal coverage at higher latitudes. The orbits are chosen so that the satellites can pass over a ground station that checks and updates their position regularly. Sufficient satellites are flown to ensure that from any position on the Earth's surface, there are at least 4 satellites above the horizon at any time.

---

[35] An *almanac* is a catalogue of information for multiple items – in the context of GPS, it is a broadcast message providing details of the overall health of the constellation and any information required to acquire signals from the other satellites. GPS satellites broadcast an almanac message for the constellation, and additional *ephemeris data* for the satellite in question. The ephemeris portion of the message is a highly accurate set of orbital elements, which the receiver needs to calculate its position precisely. See Annex H, Paragraph H6b.

[36] For greater details of why these orbits work, see Chapter 1, Paragraph 147.

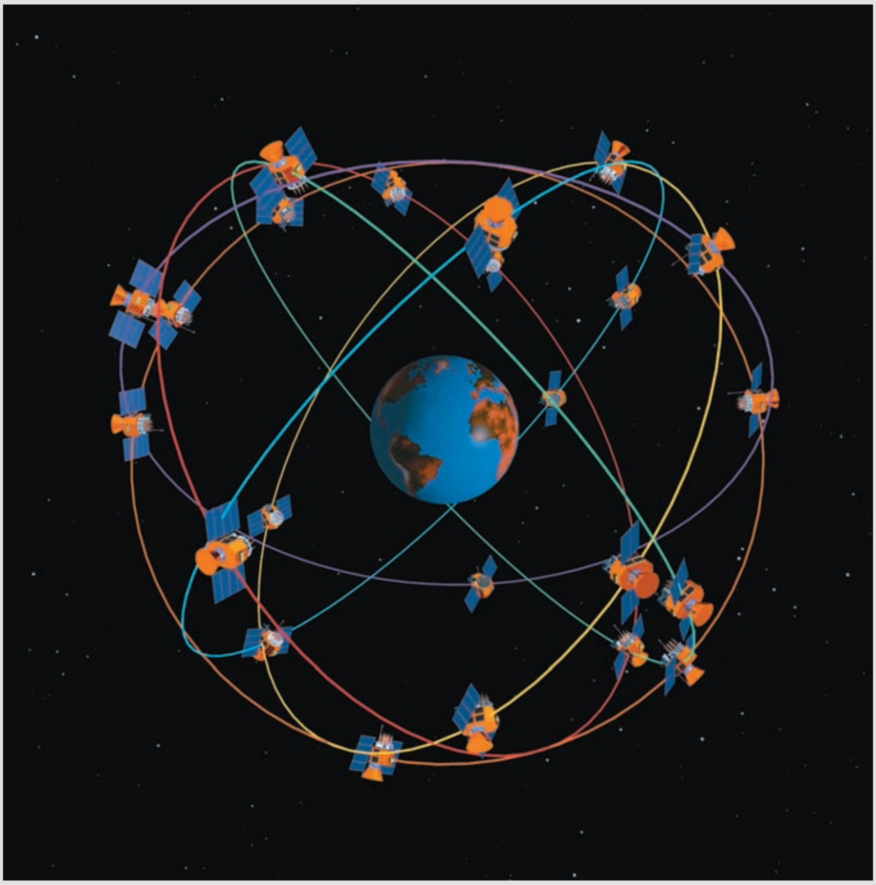


Illustration 3.12 – The GPS Constellation. This Illustration is a schematic view of the 24-satellite GPS constellation. Although the diagram looks cluttered, it illustrates that there are actually 6 underlying orbits (orange, yellow, pink, purple, green and blue in this diagram), each with 4 satellites evenly spaced around it. The 6 orbits have orbital planes separated by 60° in longitude of the ascending node, and the 4 satellites in each are separated by 90° in true anomaly (see Paragraph 130 and Annex C if you need to recap these terms). (Copyright illustration from The Aerospace Corporation, used with permission).

338. **Enhanced Accuracy**:

a. **More Than Four Satellites**? The description of GPS operation in Annex G establishes that a minimum of four satellites must be in view of the receiver for successful operation. The working constellation of 24 satellites, with 6 satellites in each of 4 evenly spaced orbits ensures that this will be the case. In practice, however, there will often be more than four satellites visible to a receiver at any given time. Additionally, there are typically more than 24 satellites in orbit at any given time, to provide a measure of redundancy in the event of satellite failure. These extra satellites transmit a useable signal too. The result is that a GPS receiver can improve the accuracy of its position estimate by processing the extra signals and using statistical techniques (coupled with knowledge of the relative positions of the satellites) to combat the residual errors that remain in a four-satellite solution. The system operators can also distribute the extra satellites around the constellation, and manipulate the regular updates of clock and position carried out on the satellites to optimise accuracy in specific ways. This could be a very significant factor for military planners, for example in relation to GPS-guided munitions, though further details of how this is achieved cannot be given here.

b. **Differential GPS and Augmented Systems**. Errors in GPS position include a geographic element that derives from the properties of the atmosphere through which the local GPS signal is passing.[37] Typically, this is the biggest residual error. It is essentially independent of the receiver used (though there will also be an error component due to receiver accuracy, but little can be done about that portion). **Differential GPS** corrects for local errors by using reference GPS receivers placed at surveyed (known) locations, which then generate correction information. This can be broadcast to GPS receivers to allow them to correct for these errors.[38] Resulting accuracies can be better than 1m. Aircraft precision approach guided by GPS is one obvious application.

339. **Alternative SatNav systems – Galileo, GLONASS and Others**. Although GPS was the first widely available PNT system, it is not the only one.

a. The EU and other international partners are developing a system called *Galileo*. The concept of operation is similar to GPS, and it uses the same frequency-band to transmit the satellite signals, although the signal coding is different. The system is in the early stages of deployment, so is not, at the time of writing, offering a full capability. The intention is that it will eventually become a true global system. The similarity in signal would make a dual-mode receiver capable of decoding GPS and Galileo a relatively simple development. There are, however, important differences in the service delivered by Galileo, based on the differing business models that support a system developed for military use (GPS) and one designed as a commercial proposition from the outset (Galileo). Among these are varying levels of service provided at cost, and guarantees of level of service provision and/or warning of degradation which are built into Galileo. These may become essential for commercial users building safety-critical systems reliant on satellite capability.

b. **GLONASS** was developed by the Soviet Union, and is now being refurbished by Russia and India. GLONASS is a global system and is also similar in concept and frequency-band to GPS and Galileo.

c. China has developed a regional system called **Beidou**. They are proposing to extend it to provide global coverage, under the new name **COMPASS**.

d. India and Japan are both developing systems providing regional coverage.

340. **Applications of PNT**. The output from a satellite navigation system can be solved to derive position, timing or velocity information. The means by which position and time solutions are derived is described in Annex G. Velocity information (i.e. speed and direction) is principally of importance in GPS/IN integration detailed above.

a. **Precision Navigation**. The most common use of the GPS signal is for navigation. The very high accuracy of GPS enables such capabilities as conducting precision instrument approaches in an aircraft without relying on ground transmitters. When installed in a weapon system, (missile or bomb), it can also guide the weapon to a target location. Hand-held and automotive applications have also become very common.

---

[37] Most of this error derives from the properties of the ionosphere; the speed of light (and thus of the GPS signal) varies with the instantaneous properties of the atmosphere. While these properties can be measured and modelled, this does not fully correct for the error. See Annex H, Paragraphs H2f and H6f for more details.

[38] See Annex H, Paragraphs H4-5 for details.

b. **Tracking Objects**. Low-cost GPS receivers will not offer the accuracy and adaptability of high-precision models, but where basic awareness of location of an item or individual is all that is required, and where a relatively slow update rate is acceptable, cheap GPS circuits can be embedded in many common electronic systems. Applications are limited only by the imagination of the designer, and can include, for example, cameras that record positions on digital images, theft-preventing devices embedded in valuable items that can detect and transmit their position, asset management systems and in the military sphere, survival radios that can report locations automatically.[39]

c. **Timing**. Accurate timing data is fundamental to GPS operation. You can see in Annex H how a GPS receiver uses signals from multiple satellites to correct its own internal (relatively low-accuracy) clock. A by-product of this is that the receiver has ‘corrected time’ available, which is very closely tied to the precise system time generated by an atomic clock. This means even a fairly simple GPS receiver can provide timing information to an exceptional degree of accuracy. This has several practical applications. Civil systems make extensive use of GPS timing information, in areas from timing financial transactions to controlling vital infrastructure such as electrical power distribution grids. Military applications are equally diverse:

(1) **Cryptography and Anti-Jam Techniques**. Very accurate timing information has utility for encoding information or synchronising ‘frequency-hopping’ systems.

(2) **Data-Links**. Modern data-link systems may share information among users by synchronising its transmission. Each user is allocated a time slot to broadcast to other users. The more closely these slots can be controlled, the more efficient the network becomes and the more information it can carry. GPS Time is accurate enough for this task and is thus widely used.

---

[39] See Paragraph 367 for a description of the ARGOS data relay system, which can incorporate GPS information in data logging applications.

SECTION III – SURVEILLANCE FROM SPACE

In this Section you will find a description of:

- Factors driving the development of surveillance from Space.
- The different kinds of resolution that can be achieved from Space.
- Imaging in visual and Infrared (IR) wavebands, including descriptions of pan-spectral and hyper-spectral imaging systems.
- Non-optical systems for missile warning, nuclear detonation detection and other applications.
- Principles and practicalities of synthetic aperture radar techniques.
- Electronic surveillance, illustrated by descriptions of the COSPAS/SARSAT and ARGOS systems.
- The implications of orbits and ground tracks on surveillance from Space.

Reconnaissance Principles

341. **Reconnaissance from Above**. The observational advantages of high ground are well understood. Attempts to exploit aviation for this purpose go back as far as 18th and 19th century use of manned balloons and kites, through faltering steps with artillery spotting from aircraft in WW1 to the development of dedicated photo-reconnaissance platforms during WW2. The pinnacle in terms of sensor/platform capability was probably reached with Cold War reconnaissance assets such as the U-2, SR-71 and Canberra PR family; contemporary UAVs continue to exploit similar technology. Although these platforms have next to nothing in common with Spacecraft, the sensors employed by them have direct read-across to those now deployed in Space. The capabilities of specific platform/sensor combinations are often highly classified, so we confine ourselves in this Chapter to examining general principles. The interested reader with a justifiable need to know will have to research particular systems elsewhere.

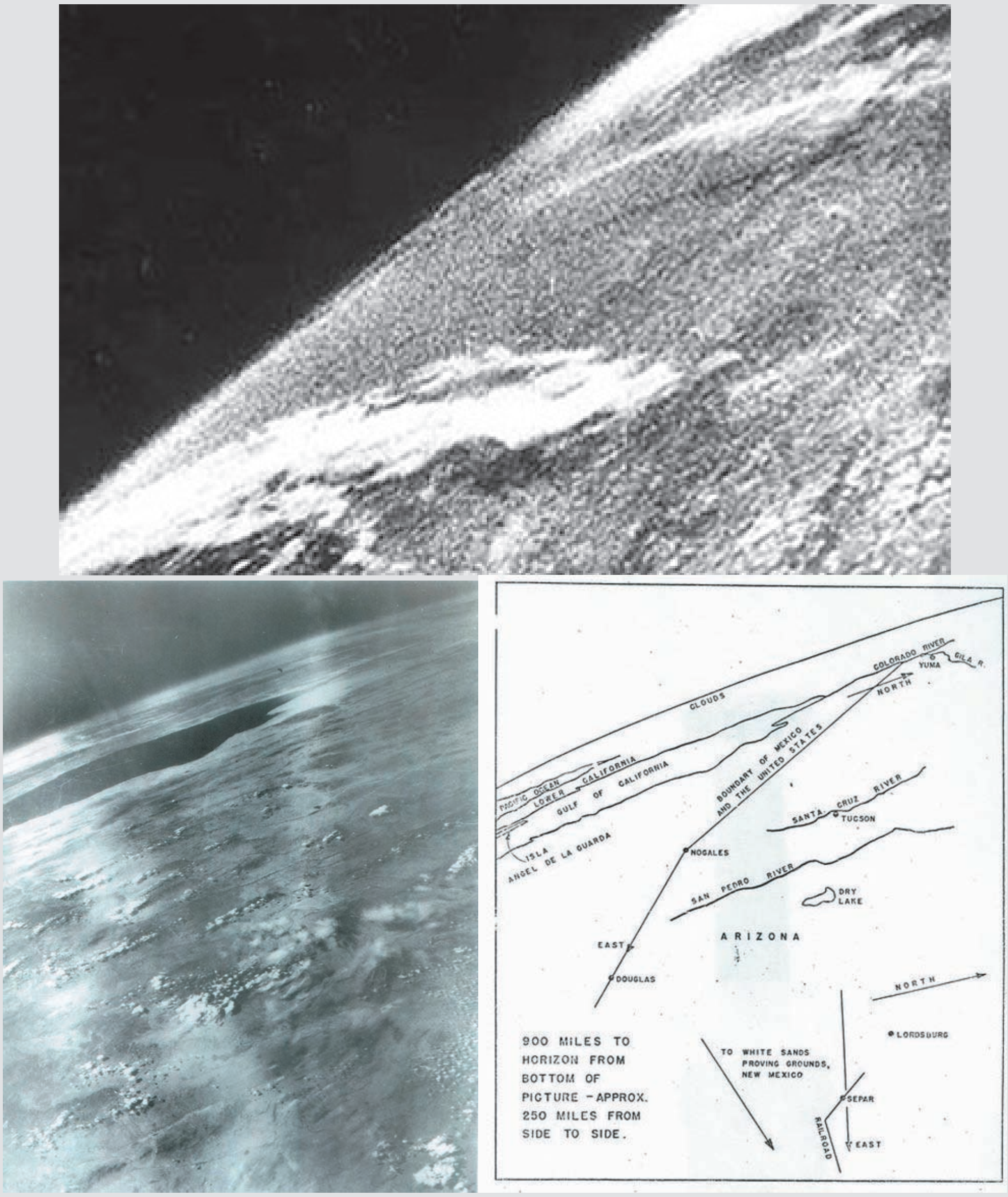


Illustration 3.13 – Early Progress in Space-based Surveillance. The picture top centre is the first image ever taken of the Earth's surface from space, actually from the test-firing of a captured German V2 rocket at White Sands Proving Grounds, New Mexico on 24 October 1946, Flight 13 of the test series. Technique quickly improved, and the bottom picture and associated key show the results from Flight 21 on 7 March 1947. See Paragraph 418 for further details. (Photos and diagram courtesy of White Sands Missile Range).

342. **Freedom of Manoeuvre**. We discussed the legal implications of overflight in abstract terms in Chapter 2. The potential for Space reconnaissance was appreciated early in the Space age, and the advantages of it emphasised by the loss of the US's U-2 aircraft piloted by Gary Powers in 1961 during an intrusive overflight of the USSR. This highlighted the fragility and inherent risks (to say nothing of the political implications) of reliance on air-breathing platforms – even the most capable.

343. **Military and Civil Systems**. This section of the Primer attempts to outline some of the fundamental principles that govern surveillance from Space, and does so dwelling on capabilities that

have military utility. It is important, however, to remember that even though much of the development work grew out of military and intelligence requirements, they have now been taken up by civil and commercial communities. In the same way as satellite communications and navigation systems have been adopted by civil users, the field of surveillance has proved equally useful. Some pointers to civilian requirements are included here and in Chapter 4. For the military reader two points are worth emphasis. First, civil capability is growing continuously both in quantity and quality, and for military users may provide (at a price) an alternative capacity in some areas,[40] for example if military systems are degraded or over-stretched. Second, since the capability is offered on a commercial basis, the results may be available to nations with no indigenous Space capacity, and possibly to irregular actors too.

344. **Resolution**. Anyone conducting reconnaissance from Space is, albeit perhaps unconsciously, seeking **resolution**, that is, the ability to discriminate between objects or events. There are, however, different kinds of resolution, and over-concentration on one may lead to poor choices regarding the others. It is important to understand their relationship and where trade-offs are made.

a. **Spatial Resolution**. This is the first kind of resolution that a user will probably think of. We examine some of the factors that limit it in Annex E. Spatial resolution is the ability to resolve physical detail, probably defined for the customer by physical size (Illustration 3.10 shows a very quick and clear improvement in spatial resolution during trials). In Annex E there is an explanation why a sensor is usually constrained by ability to distinguish an angle, rather than a distance or size. The link between angular resolution and the size of detectable target is range. For a satellite, the ultimate near-limit of range comes from orbital altitude, which is probably fixed for a given platform. This is then modified by slant-range considerations, which derive from the distance between the target and the satellite's ground track. Ultimately the far-limit is governed by distance to the horizon, another implication of altitude, as discussed in Chapter 1 (see Paragraph 144a). Depending on role and design, imagers may or may not have the ability to steer their sensors significantly. Where spatial resolution is quoted for optical systems, it will often refer to that achieved **at nadir** (looking straight down).[41] Oblique resolution will not be as good. We will examine the factors that limit a particular sensor's resolution as an angle in more detail shortly.

b. **Temporal Resolution**. A recurring factor in reconnaissance is change detection. Analysts spend long hours comparing current and historical data, looking for variations. Temporal resolution is governed principally by the duration and frequency of overflight, which derives directly from the orbital elements explored in Chapter 1. For the purposes of change detection, repeated overflight at the same local time of day on successive days may have distinct advantages. This is due to consistent and predictable illumination of the target,[42] assuming that we are interested in visual or near-visual wavelengths. At other wavelengths, different factors may be dominant, but these may still favour particular times

---

[40] Civilian capability concentrates on optical (including IR) imaging, and to an extent radar (including SAR). Currently, there are no civil equivalents of systems such as missile warning.

[41] The term is derived from ground observation, where the point in the sky directly above an observer is referred to as the **'zenith'**, and its opposite point (which is unobservable, being through the Earth below the observer) is the 'nadir'. A satellite can in theory point both at its zenith and its nadir locations, but if it is observing the Earth, it will be looking at or around its nadir.

[42] Illumination of the target will of course be affected by seasonal factors, but over short periods these may be negligible, and over longer periods they are at least predictable.

of day.[43] All these again are governed by the orbital elements of the observing platform. Illustration 4.2 shows an imperfect example of temporal resolution; the two images of New Orleans are in fact temporally separated by 5 years, but the real temporal parameter was the ability to generate data within days of the incident being monitored (Hurricane Katrina).

c. **Spectral Resolution**. A third kind of resolution is achieved by observation at multiple wavelengths and comparing and contrasting the results achieved at each. This is principally a function of sensor performance, rather than of orbital elements, though this may again be influenced by illumination. Additionally, spectral resolution may be achieved by comparison of data from multiple sensors, which may or may not share a platform. Thus the intersection of multiple orbital ground tracks near a target may also be a factor. Illustration 3.17 shows an image of a city (Munich) where the different response of ground features to varying radar frequencies and polarisations allows characterisation of the target area.

345. **Spatial Resolution and Sensors**. Fundamental factors affect the theoretical and practical resolution of a particular sensor – completely independently of range to the target. We look briefly at each.

a. **Wavelength**. The wavelength at which you observe puts an upper limit on sensor resolution. Short wavelengths (high frequencies) equate to greater resolution, longer wavelengths (lower frequencies) give less resolution. Thus, at least in principle, visual wavelengths give you greater resolution than IR, and IR greater than microwaves (radar).[44]

b. **Aperture**. Increasing aperture yields increased resolution. Thus, all other things being equal (a challenging assumption, but the principle is true), big apertures (e.g. big lenses or reflectors) see more detail than small ones. This is due to diffraction effects, which are described in more detail in Annex F. (It is the same factor that says that big antennae can form narrower communications beams than small ones). Plainly, this has implications for satellite design; increasing aperture probably implies increased weight, and greater problems integrating the sensor with the platform and launcher.

c. **Magnification**. This is a factor principally affecting visual and near-visual systems – there is no real equivalent, for example, in a radar system.[45] Magnification trades field of view for detail, and needs to be related to the capability of the detector system. Any imaging system will need some kind of optical system to focus the image, yielding some level of magnification. Satellite systems will use high-quality optics, though the design may be relatively simple. Typically, fixed focal-lengths are used (i.e. no 'zoom' lenses). If an off-track spot-look capability was required, one option would be to include a steerable mirror in the system. This would have the advantage of keeping the primary sensor and its optics fixed, leaving the lightest part possible to move, but at the cost of significant complexity. Another option is to manoeuvre the sensor and platform as one. Magnification will enhance not only the detail of the target, but also any atmospheric turbulence between

---

[43] There could of course be other temporal factors of interest for specific targets, such as surveying a coastal target at a consistent or specific phase of the tide.

[44] We are ignoring transparency here, i.e. assuming that the wavelength can reach the detector. This is not true in the atmosphere, where there are significant transparency issues in both the UV and IR bands.

[45] Civil-sector scientific satellites may need to form images at exotic wavelengths, typically at high frequencies/short wavelengths such as UV and X-rays, but very few military applications have emerged from this. (See the discussion of NUDET and gamma rays later in this Chapter for a counter-example, though note that NUDET systems do not need to produce detailed images at these wavelengths, just to detect them).

the sensor and the target. It will also magnify relative motion caused by the orbit of the satellite, so it cannot be increased without limit.

d. **Detector Performance**. Detector performance describes the ability of the sensor (which could be a receiver, camera or other detector) to capture the information presented by the rest of the payload. Practical implications vary with wavelength.

(1) Even the most exquisite optical system imaginable ultimately focuses an image onto some kind of imaging device, which then tries to capture it in as much detail as possible in a short period of time. Factors governing performance here include the size and sensitivity of the individual elements in the detector, (now probably an element in an electronic system, but previously the composition of photographic emulsion). More detail is provided below. Matching the characteristics of the detector with that of the imaging component is a problem for the payload designer.

(2) At IR wavelengths, sensitivity of imaging chips is again a factor, complicated by the effects of temperature on IR signals (see the discussion of black-body radiation in Annex F). This means that the satellite structure may interfere with the 'wanted' signal.

(3) At micro-wave and radio wavelengths, detector performance is concerned mainly with the ability to receive, amplify and process faint signals. As far as spatial resolution is concerned, the dominant factor may in fact be the antenna design, as few systems form 'images' (but see the discussion of synthetic-aperture radar below).

Imaging – Visual and IR Surveillance

346. **Optical Sensors**. Turning a focused image from an optical system into a picture used to require photographic film but is nowadays almost certain to employ a digital sensor. A brief understanding of the processes involved in both may be useful.

a. **Film**. The first reconnaissance satellites utilised photographic film. The satellite payload thus resembled a conventional camera, not unlike systems in use in aircraft of the time. Different types of film had differing sensitivity and resolution characteristics (which were often traded off against each other), which also varied with the illuminating wavelength. Most film used was **monochrome** (black and white), sometimes exposed through a coloured filter to capture specific information. The system designer had to choose between returning the exposed film to Earth to be developed and exploited,[46] or using onboard development to process the film, and subsequently transmitting the information revealed back to Earth via some kind of scanner or TV camera system. Both methods were used – again each had specific advantages in differing circumstances. Film-based systems are now certainly obsolescent, and possibly extinct.

b. **Digital Sensors**. Digital sensors are now ubiquitous in the consumer market and widely used militarily. They work on the principle that certain semiconductors[47] can be

---

[46] See Paragraph 176 and Illustration 1.35.

[47] Generically, a semiconductor is a material, often based on silicon combined with traces of other elements, which because of its controllable resistance to electric current can be used to make electronic devices. We will see when we look at IR sensors that much more exotic semiconductor materials than silicon may be required for specific applications.

manufactured which are sensitive to EM radiation – principally visible light. They can convert incoming radiation into electrical charge, which can then be stored and measured. Although the charge-quantities could in theory be manipulated directly in what would be an 'analogue' mode, practical devices convert the charges to numerical values that can be stored and manipulated as 'digital' images. Key parameters include the size of the individual elements and their sensitivity, which is often wavelength-dependent.[48] Because they constitute a tiny sub-element of the finished picture, rather like one 'dot' in a newsprint photograph, the individual cells are referred to as 'picture-elements', commonly abbreviated to '**pixels**'. Just as with film crystals, small pixels equate to fine resolution, but because of their smaller surface area, also to lower sensitivity. Because of the detail of the mechanism used, some digital imaging chips are referred to as '**CCDs**' and others in common use as '**CMOS**' devices, but the features and characteristics distinguishing them are immaterial here.[49]

347. **Cooled Detectors**. A digital imaging device is directly affected by temperature, because the semiconductor material it is made of generates small electrical charges spontaneously, at a rate that is governed by its temperature. Normally this can be tolerated, particularly for the kind of device used for visual imaging. Where it would be unacceptable, this **thermal noise** can be minimised by cooling the chip. This is a separate process to cooling the Spacecraft or camera structure to avoid emitting IR radiation. There are various methods of cooling detectors, depending on the environment and the required temperature. Where significant cooling is required, stored coolants such as liquid helium may be used. The exhaustion of the coolant – which cannot be regenerated onboard – may prove to be a limit on mission duration and satellite life.

348. **Optical Payloads**. Optical payloads are essentially cameras. In the early days of Space-based reconnaissance, these were a logical development of conventional airborne reconnaissance cameras. These were already sophisticated packages, designed to operate in the relatively hostile environment of high-altitude aviation, though air platforms at least implied shorter ranges than an orbiting satellite. When digital cameras first became practical, they could not match the sensitivity of film, so both systems co-existed, but digital detectors are now widespread. Modern satellite payloads are too complex and sensitive to discuss in detail, but we can outline a few of the design issues and resulting impact on performance in general terms.

a. **Focal length and Folded Systems**. Focal length is the physical distance between the optical centre of an imaging system and the plane where the image is produced. In practical terms, long focal lengths equate to high magnification, so typically satellite systems will require longer focal lengths than airborne system for a similar field of view. This sits uneasily with the need to compress the payload physically to fit inside the launcher. Whereas it is relatively easy to deploy an antenna after launch, the need to align in-use optical elements precisely limits what can be done with an optical system. One approach to get round this is use of folded optical systems, where combinations of lenses and mirrors are configured to give an effective focal length greater than the physical size of the system. The trade-off is the increased complexity of multiple optical elements, which need to be held in rigid alignment and increasing losses from less than perfect transparency and reflectance within the system.

---

[48] For both film and digital detectors, wavelength-sensitivity can be traced back to the characteristic energy of individual wavelength/frequency elements, as discussed in Annex F.

[49] 'CCDs' are 'Charge-Coupled Devices', 'CMOS' chips use 'Complimentary Metal-Oxide Semiconductor' technology. The underlying semi-conductor principle is the same in each, but the information is read-out from the differing designs in different ways.

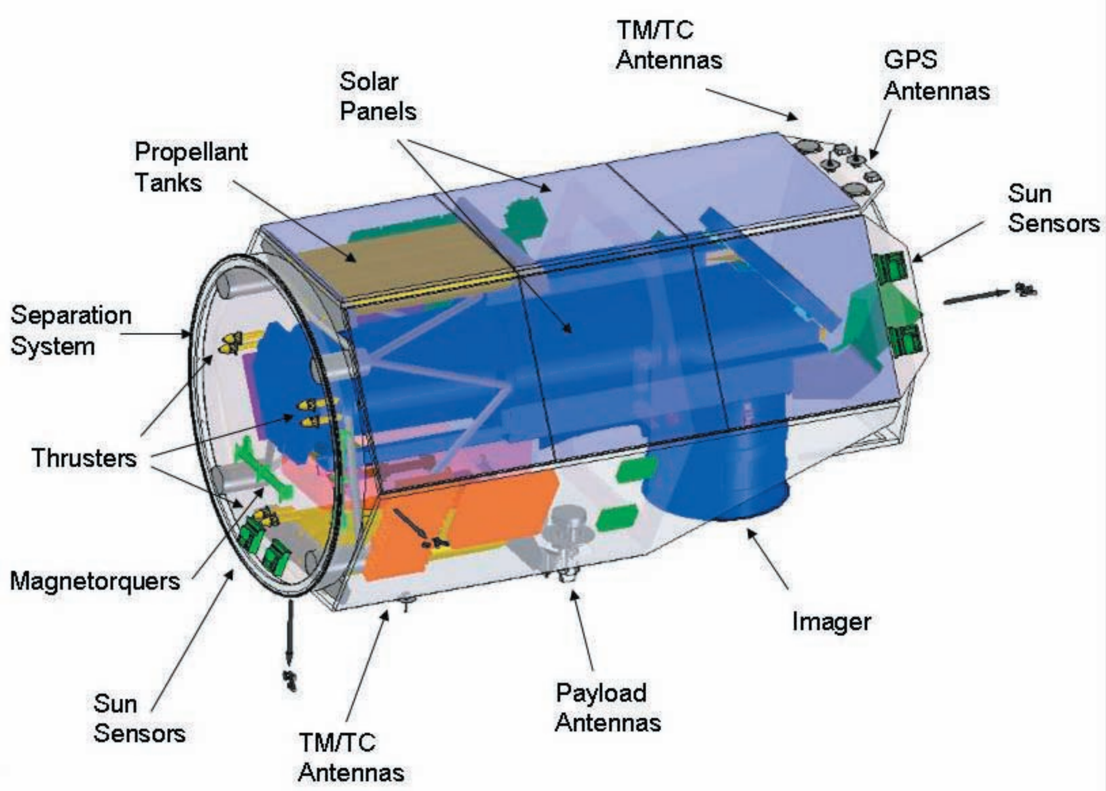


Illustration 3.14 – A Folded Optical System. This sectional view of the SSTL 'Topsat' imaging satellite shows an instance of a folded optical system. As drawn, the imaging system (the blue L-shaped component) points 'down' at the target area, but the image is then reflected 'along' the satellite structure to provide a more favourable overall shape. Typically, the constraints on the shape may be driven by the overall focal length necessary, or by thermal characteristics of the satellite or other engineering requirements. The other components shown are of passing interest only for this Section, but note the combination of Magnetorquers (Paragraph 172d) and thrusters for attitude control and station keeping, and the plethora of antennae: the TM/TC antennas support telemetry to control and monitor the satellite, the GPS antennae allow the satellite to determine its position, and the payload has its own antenna system to transmit results from the imager. (Graphic courtesy of SSTL)

b. **'Pushbroom' Imaging**. Having described imaging systems using terms based on cameras, the reader may be surprised to learn that practical satellite imaging devices often owe more to photocopier technology! In practice, rather than build large two-dimensional chips to capture a snapshot of a scene, there can be advantages in using the satellite motion over the target area to provide one dimension of the image. The camera projects the scene from the target onto a flat area, where the image moves as the satellite proceeds along its path. A long, thin array (many pixels wide, but only one 'deep') of light-gathering cells is arranged, nominally at right angles to the direction of motion, and the scene from the target area moves across the detector array, driven by the motion of the satellite. The detector is then 'read-out' at high speed, to construct the scene one row at a time; in just the same way as a photocopier or fax machine typically scans a page of paper fed into it. The analogy of the detector sweeping across the scene to be imaged gives rise to the 'pushbroom' title applied to the technique. On some systems it is possible to manoeuvre the satellite so that the array is at an oblique angle to the forward motion, rather than being oriented directly across it. This results in a degree of oversampling. The imaging element sees a smaller part of the scene below than would otherwise be the case, but it is possible to trade the lower total swept width of the target area for higher resolution.

c. **Other Options**. Not all imaging systems use pushbroom imaging, though it is a common choice. Some systems use conventional two-dimensional chips to capture a full frame image. The satellite then moves on and captures its next image after a suitable interval. Alternatively, it is also possible to use a single detector cell (a one-pixel capture device) for imaging. In this instance, the satellite motion provides one dimension to the image, just as in the push-broom technique, and a nodding mirror in the optical system scans the across-track image onto the single cell, which is read out rapidly to yield cross-track resolution. This is sometimes referred to as 'whiskbroom' imaging, to suggest the rapid scanning of the system in one direction, while making slower progress in another.

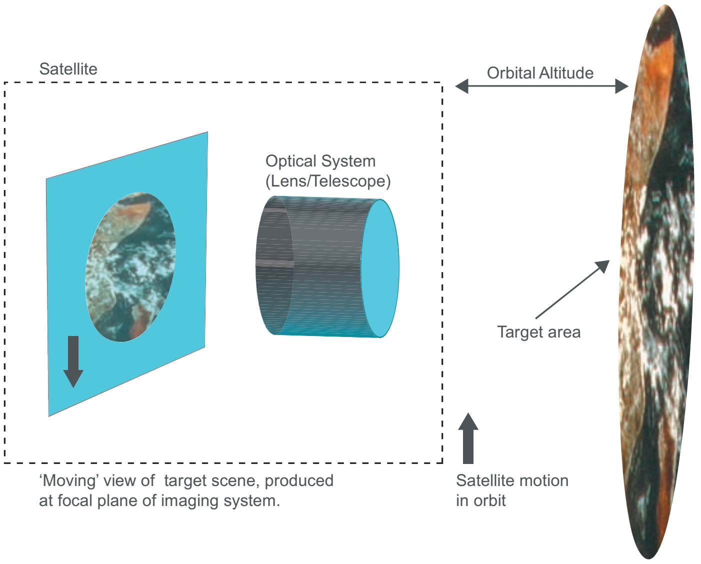


Illustration 3.15a – Pushbroom Imaging 1. Imagine a surveillance satellite orbiting the Earth. An optical system of some kind – a variation on a camera lens or a telescope – produces an image of the scene below, and projects this onto a focal plane. The focal plane need not exist physically as a flat plate in the equipment, but there will be a place in the system where the image exists. As the satellite moves in orbit, the projected scene also moves. For clarity, we will omit the optical system from future illustrations.

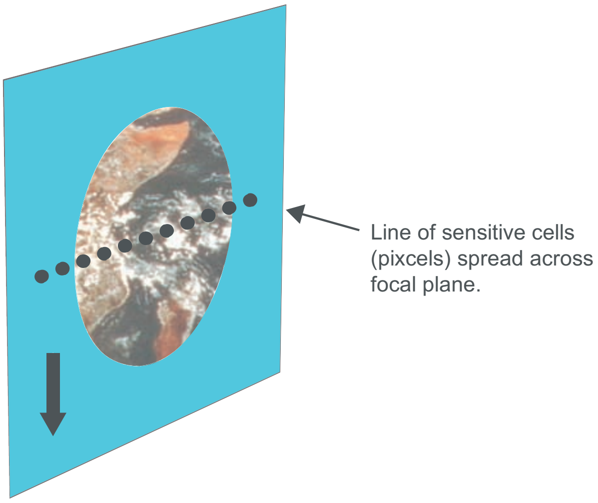


Illustration 3.15b – Pushbroom Imaging 2. As the image moves across the focal plane, it is possible to arrange a row of sensitive cells across the scene to detect the image. The drawing grossly under-represents the number elements in a real imaging system. In that case, the size and number of pixels would probably be matched to the theoretical resolution limit of the optical system.

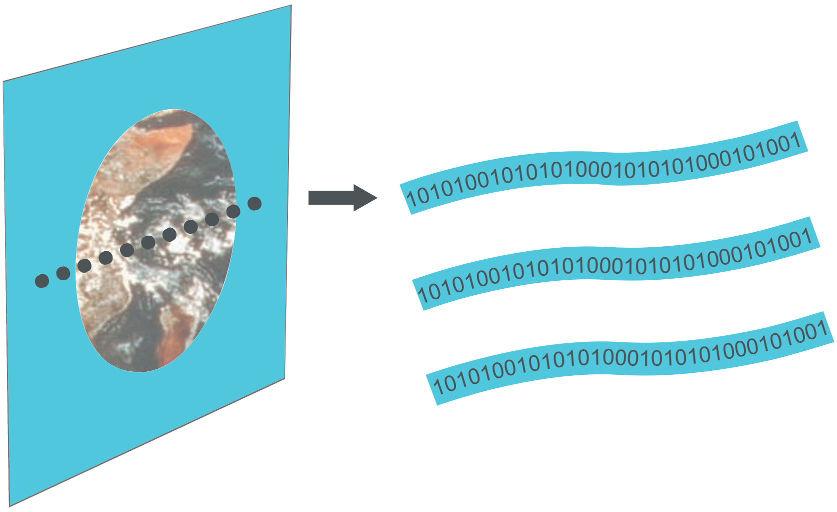


Illustration 3.15c – Pushbroom Imaging 3. As the scene moves across the pixel array, the digital data is read out repeatedly (and rapidly). The output then forms one row of a composite picture of the scene. The individual rows are re-assembled to create a (monochrome) image of the scene. The mechanism is similar to a fax machine or photocopier digitising a document which is fed into it and then passed under a similar pixel array.

349. **Imaging Specific Wavelengths – Terminology**. For reasons of sensitivity and resolution, Space surveillance uses black and white (monochromatic) imaging, whether film based or using digital sensors. To make sense of the various imager types, it is necessary to understand the various descriptive terms used.

a. **Pan-spectral Imaging and Panchromatic Film**. Black and white digital output or film, containing or sensitive to all wavelengths of visible light.[50] It typically produces a realistic black-and-white view. It gains sensitivity because any photon of light of any colour can stimulate the system, but discrimination is then lost – the pixel in the sensor or crystal in the film cannot 'tell' you what colour stimulated it – it just 'knows' that it was stimulated.

b. **Orthochromatic Film**. An obsolescent term applied to film that lacked sensitivity in specific visible colours, but which conferred advantages in other areas. Not strictly relevant to modern satellite imagery, it was nonetheless common in air reconnaissance, and the term may still occasionally be encountered. The output was a black and white image that appeared unnatural in tone.

c. **Multi-Spectral Imaging**. Multi-spectral images are made by combining monochrome images taken in different wavelength bands. Thus, a 'true-colour' image is made by combining mono (black and white) images taken through, typically, red, green and blue filters into a multi-spectral image. In a multi-spectral system, the EM spectrum used may also include near, mid and far IR bands, and in principle could also include ultraviolet information at one end and radar information at the other. It is important to separate in your mind how the image is gathered and what it contains from how it is presented. See below for a discussion of presenting results.

---

[50] Note it is not necessarily *equally* sensitive at all wavelengths; the system will have a 'sensitivity curve' showing how sensitive it is at varying wavelengths (colours).

350. **Practical Multi-Spectral Imagers**. Bearing in mind the description of a 'pushbroom' imager in the paragraph above, how do satellites image at more than one wavelength? Depending on the resolution required, there are two principal methods that can be employed:

a. **Multiple Detectors**. There is usually room at the focal plane of the imaging system to accommodate more than one detector array. Each then sees the scene passing below in turn. Each can also be filtered to observe at a specific wavelength. The resulting image from an individual detector is still a monochrome image, but it contains (for example) the red or blue or green information from the scene. These images can then be combined as required to produce natural or 'false' colour images. The analogy would be pushing three or more different brooms across the same scene, each sweeping up a different kind of debris. In practice the arrangement of detectors at a focal plane can be complex, with, for example, pairs of identical detectors aligned to give overlapping coverage of a scene to improve resolution.

b. **Multiple Cameras**. Another tactic available to the designer is simply to place multiple cameras on the same satellite, each with a filtered detector fitted. These cameras each then yield a separate picture which can be combined as described above.

c. **Design Trade-offs**. Each design approach has different advantages. For a given size of satellite, the choice will lie between a single large-aperture device and multiple smaller single-waveband systems. Since aperture limits resolution (see Paragraph 345b above), the 'multiple small cameras' approach has inherent resolution limitations. Conversely, however, the multiple detectors in a single large aperture system each see the scene below at slightly different times, which limits ultimate resolution when fitting the different frames together.

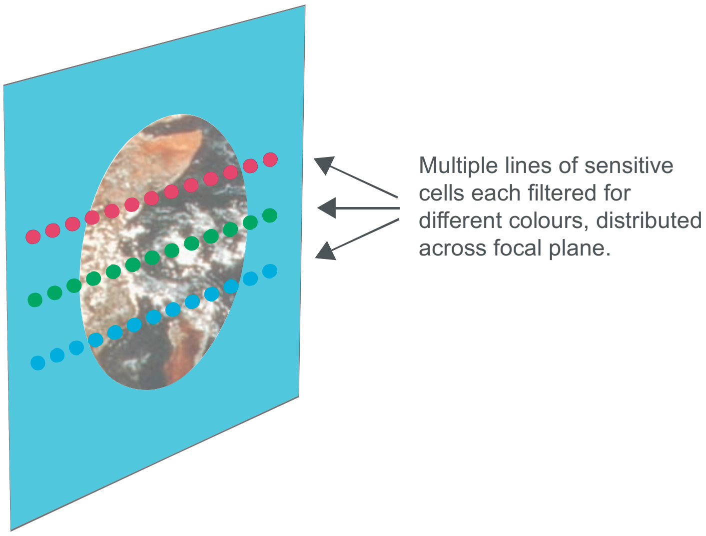


Illustration 3.16a – Multi-spectral Imaging 1. We can develop the simple imager in Illustration 3.12 to give colour results. The pixels in that imager were unfiltered (they may in fact have been filtered to exclude unwanted wavelengths, but in principle they were sensitive to all visible light). Now we insert multiple pixel arrays, physically adjacent to each other, and each filtered for a specific waveband. In the example above, we imagine red, blue and green-filtered arrays, which would combine to give something close to natural colour, but this is not essential. As the scene moves across the focal plane, it is sampled by each array in turn.

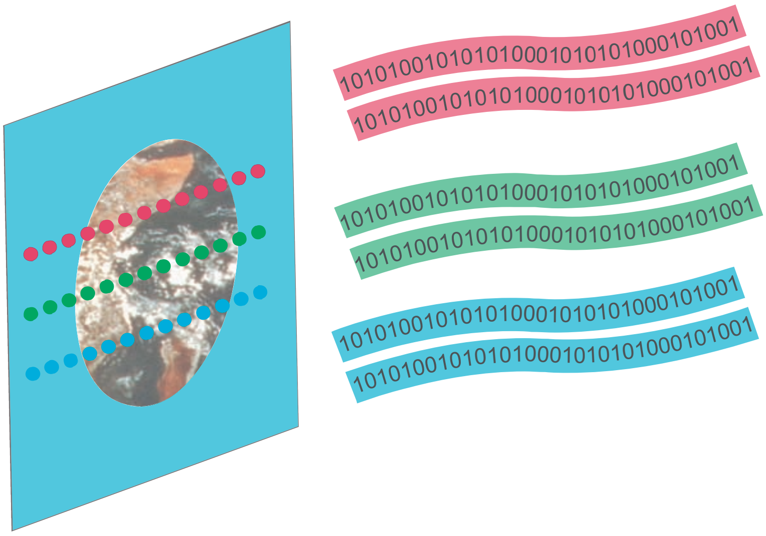


Illustration 3.16b – Multi-spectral Imaging 2. Each filtered array is read out repeatedly, and reassembled to give red-filtered, blue-filtered and green-filtered views of the target scene. These can then be combined to yield a colour (multi-spectral) image. For specific systems, the colours chosen for filters may vary (they may, for example include an IR sensing array), there may be more than 3 of them (though there will be a physical limit on how many can be fitted onto the focal plane), and their layout may be more complex than shown, but we have captured the principles of such a system.

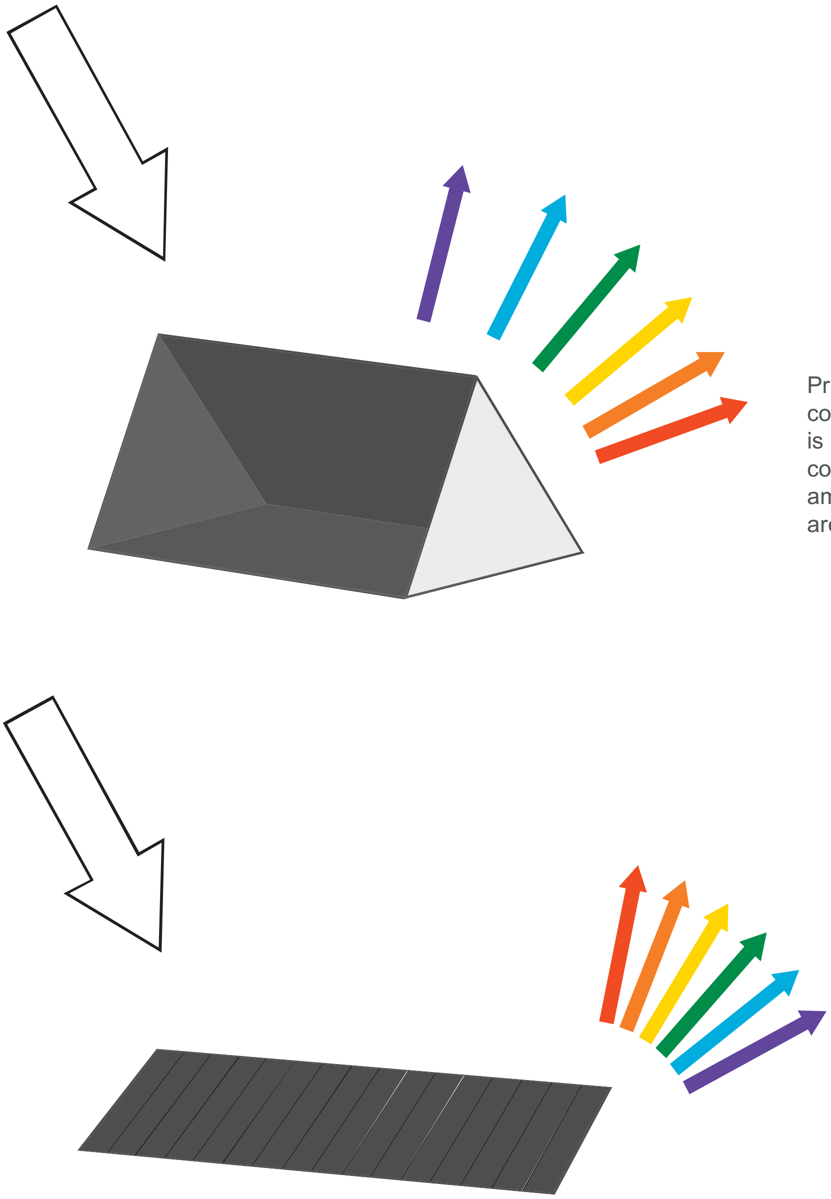
351. **Hyper-Spectral Imaging Theory**. Hyper-spectral imaging is a development of multi-spectral imaging. In a multi-spectral system, the spectral bands are relatively broad, and typically there will not be many of them – possibly tens of bands, but not more. Also, in a multi-spectral system, the images may be gathered by separate sensors operating through various filtering mechanisms. A hyper-spectral system refines this process by utilising hundreds of very selective (narrow) wavelength bands, and by processing them coherently. The output from a hyper-spectral sensor can be referred to as a **cube**. A cube is a stack of multiple images of the same scene, each taken in a slightly different band. This means, as we mentioned above, that resolution as a term becomes ambiguous. It may mean **spatial resolution** – how small or fine a physical detail can the system resolve, but it may also refer to **spectral resolution** – how subtle a distinction in output in different bands can the system differentiate between and/or how narrow are the bands. The resulting output cube is also numerically complex, so effective hyper-spectral processing is a computationally intensive task, although well within the capacity of modern computer systems.

352. **Practical Hyper-Spectral Imaging**. If you wanted to sample a scene in a very large number of discrete wavelengths to generate a hyper-spectral image, it would in theory be possible to arrange a very large number of linear detectors in the focal plane of the imager to do so. You would, however, encounter insuperable practical difficulties. It would be very difficult to locate the arrays physically so that they were evenly illuminated in the focal plane, and it would also be practically impossible to construct filters for them that each passed only a narrow band of wavelengths, but which in total sampled each wavelength consistently. A hyper-spectral imager avoids these problems by using a two-dimensional imaging chip – a so-called **focal-plane array**, in conjunction (usually) with a **diffraction grating**. This is a device similar in principle to a prism,[51] but using very narrow lines ruled parallel and very close together on an optical surface. The grating splits incoming

---

[51] The key advantage of a diffraction grating over a prism is **linearity**. Prisms diffract different colours of light by varying amounts, depending on wavelength in a complex way. Diffraction gratings diffract in a much more predictable way, with a simple relationship between wavelength/frequency and the angle of diffraction. Despite that, both prisms and diffraction gratings are in use. Experiments have also been carried out with wedge filters, using interference effects to pass different wavelengths across the varying thickness of a wedge onto an imaging system behind it.

light into its component colours, reflecting or transmitting each colour in a slightly different direction[52]. The scene to be imaged is reflected off a diffraction grating and projected onto the focal-plane array, which is similar to the imaging chip in a conventional terrestrial digital camera. Instead of each pixel representing a different physical location in the imaged scene, however, the chip is arranged so that rows of pixels lie across the satellite motion, just like a simple pushbroom imager, and adjacent rows are each illuminated by a slightly different wavelength. Thus along columns of pixels, you see spectral rather than spatial resolution. The chip is downloaded rapidly and regularly, so it acts like a pushbroom imager, but with each row of the chip seeing a different wavelength to its neighbour.



Prisms split incident (white) light into component colours, but the process is relatively inefficient and the colours are refracted by varying amounts. High quality optical prisms are relatively difficult to make.

Diffraction gratings are optical surfaces with numerous parallel lines ruled on them (many more than shown here). They split incident light into its component colours too, albeit through a radically different process. Some gratings reflect light (as shown) while others are transparent (so called *transmission gratings*) and transmit the light. The angles the different colours are diffracted through are much more regular, the process can be made to be be more efficient and the gratings are simpler to manufacture.

Illustration 3.17a – Hyperspectral Imaging 1 – Prisms and Diffraction Gratings. For hyperspectral imaging, we need to be able to separate out the component wavelengths of incoming light with high precision – more accurately than can be done with simple filters. Prisms and diffraction gratings can both be used for this purpose, though diffraction gratings are common.

---

[52] This is the same phenomena that gives domestic CDs their characteristic 'rainbow sheen' when viewed in daylight. The spacing between the lines of digital signal embedded on the CD is comparable to the wavelength of light, and the multiple closely-packed tracks act as a diffraction grating.

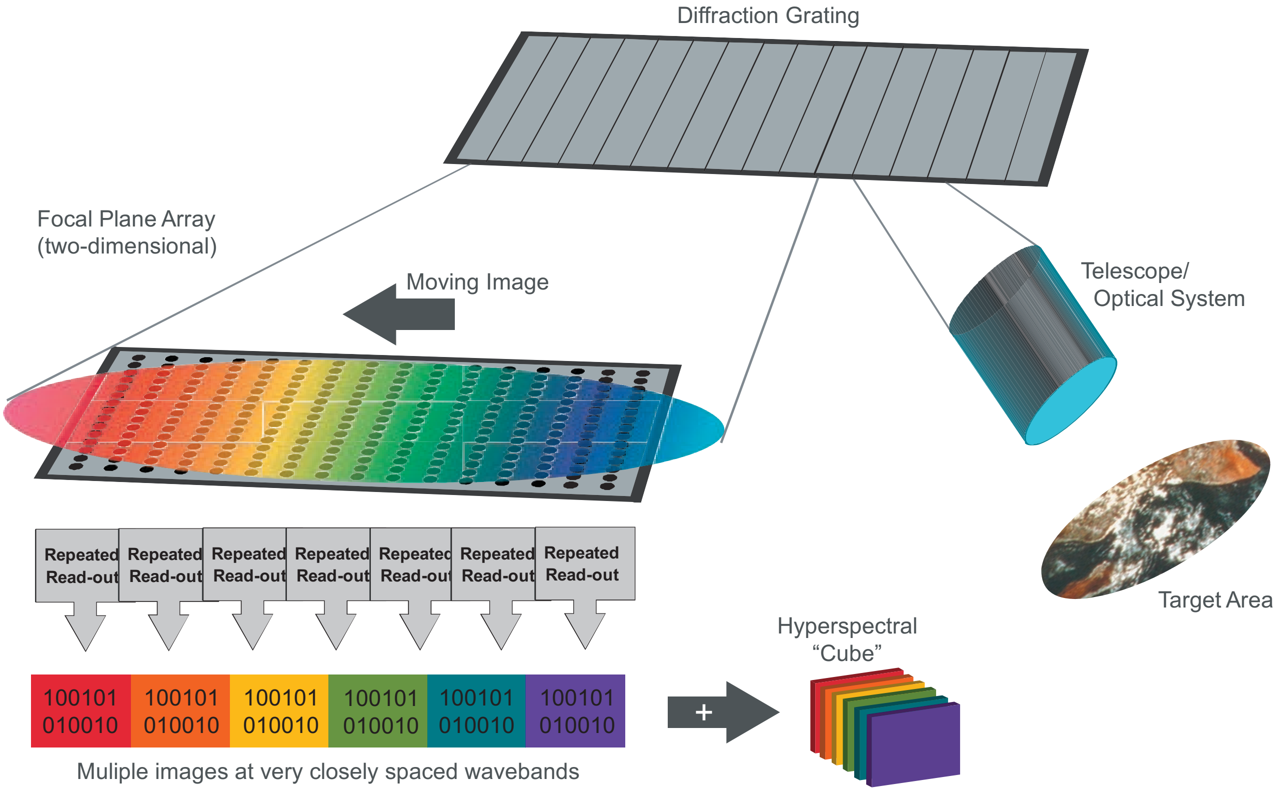


Illustration 3.17b – Hyper-spectral Imaging 2 – Focal Plane Arrays. A hyper-spectral imager uses a 2-D array of pixels, not as a means of capturing a 2-D scene as in a conventional camera, but rather as a means of capturing very closely spaced wavelengths. A diffraction grating is inserted in the imaging train; we show it above as a reflection grating between the optical assembly and the imaging plane, but the physical arrangement is unimportant. What matters is that the different wavelengths in the incoming light are each diffracted by slightly different angles at the grating. The imaging chip then becomes an array of unfiltered pixels, where each row sees the same scene but only at a specific wavelength. The rows are then read-out as if they were each a push-broom imager looking at a very narrow waveband. The resulting 'cube' can be interrogated in multiple ways to yield target information.

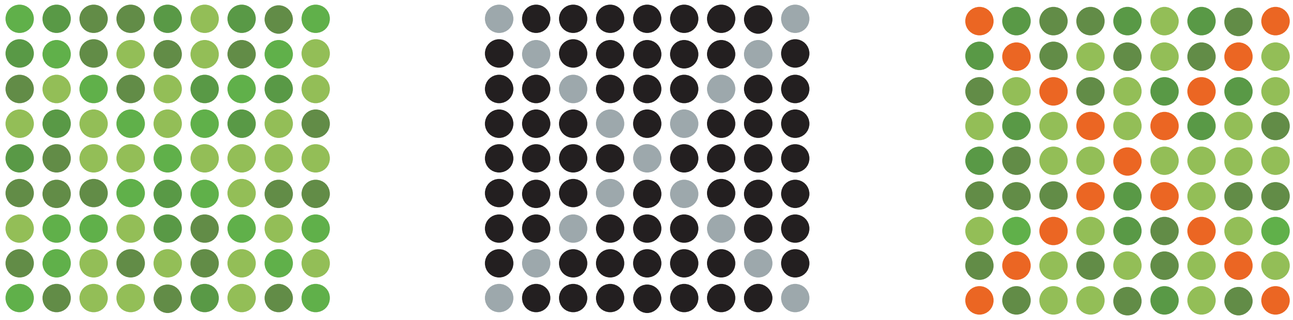


Illustration 3.17c – Hyper-spectral Imaging 3 – The Power of Narrow-band filtering. There are many ways of interpreting hyper-spectral images, but we illustrate one possibility. The pattern of dots on the left is in fact coloured in only four shades of green, but they are relatively similar shades and the pattern of colouring is apparently random. If you could filter a view of the pattern to exclude all but one of the shades (easily done if the spectral resolution of the imager is good enough), you would see (as in the middle) that there is actually a clear pattern in that shade. If that shade was characteristic of some activity or property of the scene, you would be able to detect it clearly and un-ambiguously. To display it clearly, you could then use the information to re-colour the original scene, as shown on the right. This last step is *false-colour presentation*, discussed further in Paragraph 354.

353. **IR Payloads**. One of the earliest drivers for satellite surveillance was to provide prompt warning of ICBM launch during the Cold War. In this case, the aim was to detect the thermal effects

of the launch of a missile, and most of these took place in the IR region of the spectrum.[53] Almost everything here could be sacrificed for prompt alert; certainly, there was little need to explore high-resolution imagery. We will cover how IR detection of missile launch works in Part IV, but in parallel with this early driver, there was military awareness of the potential of IR imaging.

a. **IR Imaging**. A major problem with early IR surveillance was developing effective detectors. While film can have limited sensitivity in the near-IR band, for true IR imaging, electronic detectors have always been required. Ordinary silicon-based semiconductor detectors are not sensitive at the appropriate waveband[54] and a variety of alternatives, based on materials such as mercury, cadmium, arsenic, gallium and tellurium are employed. These are much harder to fabricate than visual detectors, and for many years, small one-dimensional rows of sensitive cells were all that could be produced reliably. Since many IR systems were used for missile detection, where surveillance rather than high-resolution imaging was required, they were scanned mechanically in the field of view of the optical system (see Paragraph 358a for an example application).[55] They were known as '**scanned arrays**', especially more recently, to contrast them with '**staring arrays**'. A staring array is a relatively recent technological advance, which can be thought of as the IR equivalent of a visual-spectrum CCD or CMOS two-dimensional chip. Compared to visual devices, they are much harder to fabricate, and suffer much greater thermal noise and characteristic defects such as varying sensitivity. Nonetheless, they are now at least practical devices. They will almost certainly require much more significant cooling to provide useful results.

b. **IR Targets**. An obvious implication of visual surveillance is the need for daylight illumination of the target. IR detectors are usually designed to detect thermal radiation, which is emitted by objects all the time, and so can be used for surveillance at night as well as during the day. IR radiation can provide information on the temperature of the body radiating it (see the discussion of black body radiation in Annex F), so IR sensors can differentiate temperature. The resulting target set quickly becomes apparent. Obvious examples include factories and power-plants, the status of major pieces of equipment such as ships or submarines, and natural events such as flows of water and forest fires. Ballistic missile launch is also a valid target.[56] Weather forecasting also relies on the IR characteristics of water-vapour in the atmosphere (see Paragraph 356).

---

[53] This may appear counter-intuitive, since many aircraft missile warning systems today use the UV portion of the rocket plume signal as the warning cue. The IR portion is relatively much stronger, however, and consequently easier to detect from orbital ranges; additionally, there are problems with IR 'clutter' on aircraft-mounted warning systems which make UV detection more attractive at the relatively shorter range required. From space, the IR advantages are overwhelming.

[54] Yet another impact of the frequency-energy relationship outlined in Annex F.

[55] Airborne 'IR-Linescan' devices were also developed and employed for airborne tactical reconnaissance in the late 1960s and 1970s. The TSR2 was intended to carry an IR-linescan payload, and the RAF first employed them in the EMI reconnaissance pod fitted to the F-4 Phantom. Similar systems were deployed on the F-111 and various Mirage variants.

[56] See Paragraphs 357-8 and 364 for discussion of missile launch detection and the related topic of nuclear detonation detection.

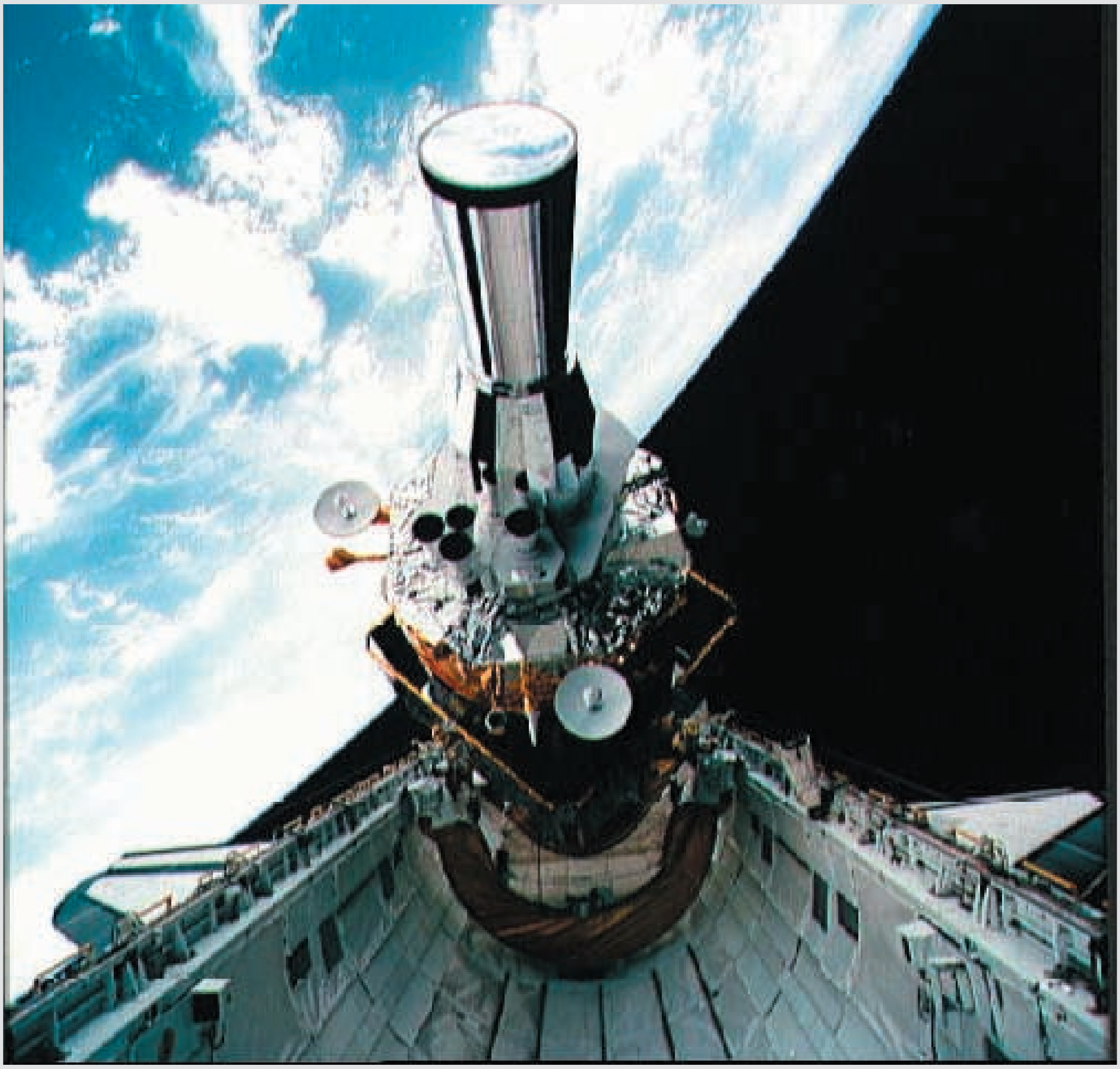


Illustration 3.18 – A Large IR Surveillance System. This picture shows an unusual scene – a US Defense Support Program (DSP) satellite being deployed from the payload bay of the Space Shuttle *Atlantis* in November 1991 (Mission STS-44). DSP satellites operate at GEO to detect missile launch through the IR signature of the launch event. The DSP satellite pictured is mated to an *Inertial Upper Stage* booster, which will lift it from the Space Shuttle LEO to the required GEO. Most DSP satellites were launched by conventional boosters. (NASA Johnson Space Centre image)

354. **Presenting Results – False Colour Output**. Since multi-spectral and hyper-spectral images can produce results that could not be seen with the naked eye (by definition the eye is blind outside the visual range!), a problem arises in how to display them. Invariably the solution is some kind of false-colour image; a colour picture where the colours used are a deliberate distortion of the natural colour of the scene. Some false-colour displays are subtle; the composite image may be a mainly visual representation of a scene, employing largely natural colour, with a particular characteristic emphasised using one shade that contrasts well.[57] Others use completely unrealistic colours to emphasise contrast. For particular applications, there are standard colour schemes assigning a colour a consistent 'meaning', but image analysts may also create custom schemes for an ad-hoc task. See Illustration 3.17c for a hypothetical example.

---

[57] An analogy is using a highlighter pen to emphasise one phrase in a paragraph of text.

355. **Applications of Multi-Spectral and Hyper-Spectral Images**. Some multi-spectral images are just slight adjuncts to the visual picture – for example to overlay temperature information on a scene. Hyper-spectral imaging offers many other additional applications, which cannot be discussed in detail here. Most of the ‘interesting’ applications rely on activities that can be distinguished by very subtle differences of colour. In the public domain, many Earth-sensing applications rely on these, for example to monitor crop growth. Individual varieties of crop, applications of fertiliser or pesticides and the spread or control of diseases or pests can all be detected. There are similar applications in mineralogy, chemical processes and small changes in water temperature. The military potential must be left in this publication to the reader’s imagination.

356. **Meteorological Satellites**. Conceptually, meteorological (weather) satellites are just a special case of surveillance, with the added nuance that what is clutter or obstruction for reconnaissance may be the desired target for the weather forecaster. Modern weather forecasting (whatever the public think) is vastly more accurate than it used to be, particularly in the short term, and like much else in reconnaissance and surveillance, the gains have come about through increased computing power as much as from any revolution in sensor technology.

a. Weather is forecast by producing a very large mathematical model of the Earth’s atmosphere, populating it with weather measurements from the present and recent past, then running the model to extrapolate them into the future. The keys to accurate and useful solutions are the processing power to run the model faster than real time, allowing prediction, and harvesting data to feed into the model in the first place.

b. Space assets greatly enable data capture. A variety of sensor types, working at different wavelengths provide useful data. Visual imaging can help to identify cloud locations and weather fronts, though only by day; IR imagery provides evidence of cloud and weather systems by day and night. The precise temperature of a cloud alters its IR characteristics, and since temperature and altitude are related, the IR image can also indicate cloud heights. Yet other varieties of IR imagery of the Earth’s surface can provide temperature information clear of cloud. Other weather surveillance techniques approach cross-over with Earth-resource monitoring.

c. Weather satellites operate in a variety of orbits. Some are in GEO, providing wide-area surveillance at relatively coarse resolution. Others sit in a Polar LEO to provide regular close-range survey of smaller areas.[58]

d. Much additional weather forecasting information is derived from radar data, specifically synthetic aperture radar returns,[59] but also including radar altimeters, which can measure ocean surface features, microwave radiometers, which can measure the temperatures of sea surfaces and of different levels in the atmosphere, and scatterometers, which can derive wind speed and direction at the surface from their effect on radar returns over water. Synthetic aperture radar is explained further below.

---

[58] A coincidental advantage of this orbit can be found in the discussion on COSPAS/SARSAT at Paragraph 366.

[59] See Paragraph 363.

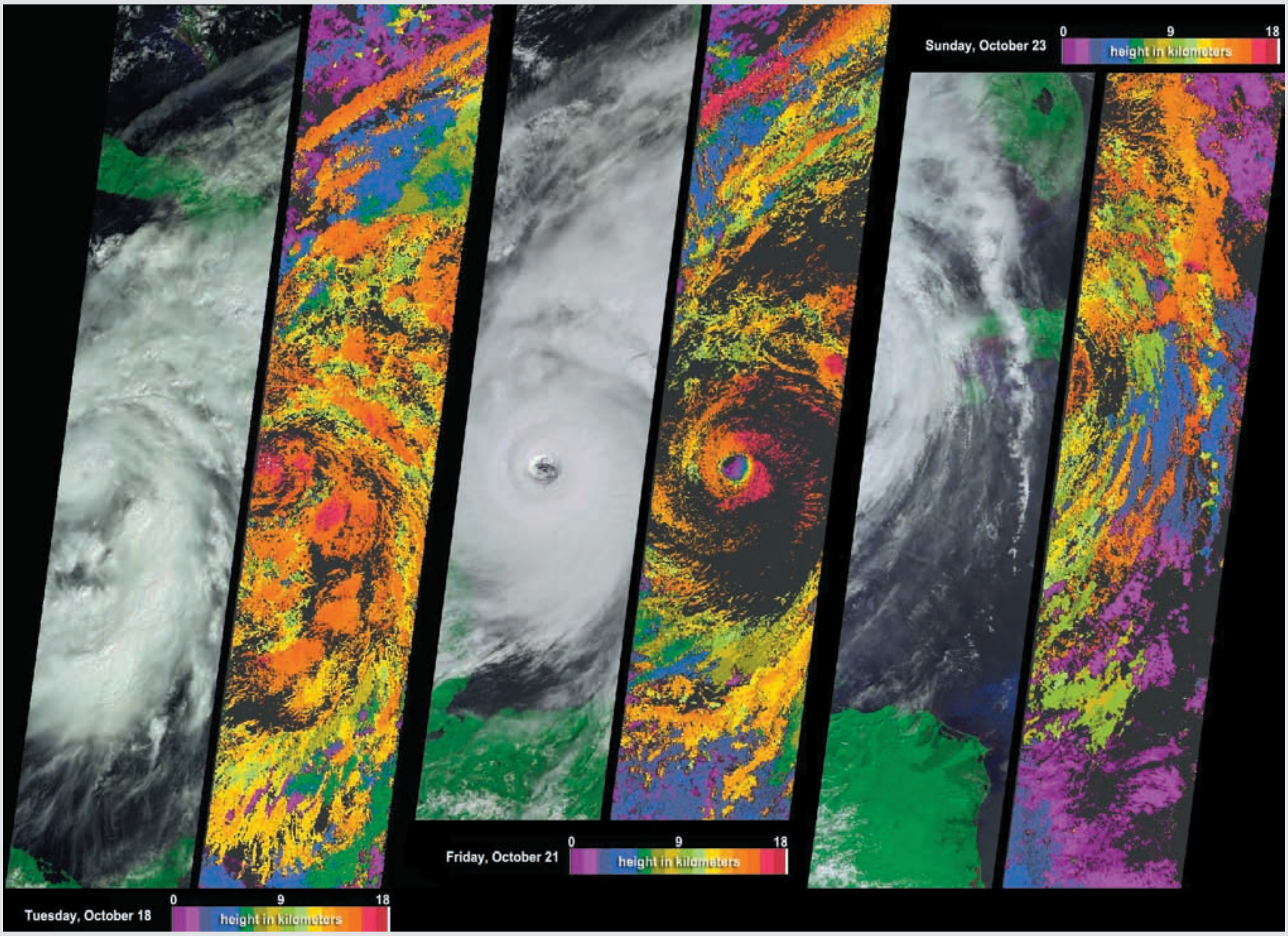


Illustration 3.19 – Meteorological Information from Space – a 3-D model of a hurricane. The photograph shows matched pairs of images of the progress of Hurricane Wilma across the Caribbean region in 2005. The third pair of images (Sunday 23 October - on right) indicates scale – Florida is at the top, Cuba is central and the South American coast is at the bottom. The left of each pair shows a visual representation of the region, the right shows cloud-top height, using a false colour presentation to give contour information. In this instance, cloud top heights are determined by perspective from orbit. The sensor (NASA's MISR sensor on the *Terra* satellite) has 9 cameras arranged to give overlapping fields of view ahead, under and behind its path as it overflies the ground. Computer processing then looks for consistent change in cloud features from one image to the next, and uses the perspective shift as the satellite passes over the region (at a known height) to calculate the height of the cloud tops. Monitoring of the vertical structure of a hurricane contributes greatly to validation of forecasting models. (NASA JPL Image).

357. **Missile Warning – DSP and SBIRS**. Even before the development of ICBMs, Cold-War deterrence depended on prompt detection of enemy action that might have required a strategic response. While nuclear weapons were solely aircraft-delivered, this was provided by chains of radar stations located along likely approach routes to provide the required warning time of attack. With the advent of nuclear missiles, launched from remote land-locked sites and with flight times of the order of 30 mins,[60] this became impractical. Although radar still plays a part in missile warning, by the time the warheads are within radar range, it might be too late to respond. Earlier cueing was plainly necessary. Research into possible Space-based surveillance to detect missile launch centred on the characteristic IR signature of a missile during the boost phase. Achieving a working system took time and encompassed many failures, but for the USA and allies, warning is currently provided

---

[60] See Paragraph 311.

by the US Defense Support Program (DSP) System.[61] The technical approach adopted is to have a number of surveillance satellites in GEO scanning large areas, looking for the characteristic IR plume emitted by the boost phase of a rocket in flight, and providing basic location and trajectory information. DSP is due to be replaced by a system called the Space-Based Infrared System (SBIRS). SBIRS will have elements in GEO and elements in HEO to provide comprehensive coverage.

358. **Missile Warning – Techniques**:

a. **Detector Technology**. We described in Paragraph 353 the technical challenges of fabricating electronic IR detectors. DSP and its precursors had to cope with these challenges, and resorted to scanning large areas onto a single row of detector chips. The sensor was an IR telescope pointing Earthwards, and the scanning mechanism was to spin the whole system around the axis pointed towards the Earth, so that the field of view of the telescope swept across the row of cells. The rotation rate was chosen to ensure that the surveillance area was scanned often enough to be sure of catching a missile in flight. The detector cell outputs were sampled regularly, many times per rotation; a valid target would be detected as a coherent series of IR responses as the satellite rotated and the missile flew across the field of view.

b. **Target Discrimination**. A missile warning system needs to detect missiles against the IR background radiation from the Earth. Invalid targets likely to be detected include volcanic eruptions, forest fires and other large thermal sources, mixed with the Earth's black-body radiation and the IR component of reflected sunlight.[62] If a missile is launched below cloud cover, it is unlikely to be detected until it breaks cloud. Several features of a missile's signature aid discrimination, however. The IR characteristics of missile exhaust are distinctive, and concentrated around specific IR wavelengths; this also discriminates between solid and liquid fuelled devices. Target motion is also characteristic; as well as 'filtering out' static sources such as volcanoes, motion will also differentiate between ballistic missiles and other rockets such as satellite launches, as well as giving an indication of flight-path, likely target and time to impact.

c. **Expanded Missions**. Steady performance improvements in the detector satellites, and elements of military necessity, have led to some expansions in the Missile Warning mission. Detection of static thermal sources, such as fires, can be supplied to civilian agencies and aid scientific environmental monitoring. More usefully for military operations, the warning satellites proved to have utility against tactical as well as strategic targets. Thus during the first Gulf War in 1991, SCUD missile launches were detected in Iraq, and this information was used to provide tactical warning and cueing. We discuss the anti-SCUD campaign in greater detail in Paragraph 391d. Among the challenges faced were the short flight-time of a tactical missile, as well as the C2 challenges inherent in disseminating information from a system designed and integrated with strategic-level organisations at a tactical level.

---

[61] The evolution of DSP, including some technical explanation and an analysis of overall capability can be found in the book by Jeffrey Richelson referenced in the Bibliography.

[62] See Annex F for an explanation of black body radiation.

Other Forms of Surveillance

359. **Radar Payloads**. The devices we have looked at so far have generally depended on external illumination. Usually this has been from daylight, or the inherent emission of the target at IR wavelengths. Radar payloads differ by providing their own illumination. This has implications for power generation and consumption at the satellite, but can also overcome a significant limitation of imaging systems, in that cloud is generally transparent at radar wavelengths. They must, however, be able to detect targets amongst ground returns and other clutter, from a platform that is moving at very high speed relative to the target. Accordingly, they have in the past tended to be used against large targets such as ships. Such radars provide little information about the nature or identity of the target, just its position and motion. These limitations are addressed in some respects by synthetic aperture systems, which are described below.

360. **Synthetic Aperture Radar**. Synthetic Aperture Radar (SAR) arose originally as an aircraft-mounted system, and is still used in air-breathing applications.[63] Nowadays, SAR can be aircraft or satellite mounted, or even (Inverse SAR) ground-based against Air or Space targets. SAR has revolutionised Space-based radar surveillance, combining the merits of radar and optical systems. Because it is particularly well suited to Space use, and is therefore becoming widely available in both civil and military applications, we will describe it in some detail. SAR retains use of radar wavelengths, and thus transparency through cloud, and independence from daylight illumination, while simultaneously compensating for the resolution shortfall of a conventional radar antenna. We have noted earlier, and in Annex F, that resolution varies directly with aperture (big aperture = good resolution and *vice versa*), and inversely with wavelength (long wavelength/low frequency, such as radar, gives poor resolution compared to short wavelength/high frequency visual systems). To build a simple radar system with the same angular resolution as a typical camera lens would be impractical;[64] roughly speaking, radar wavelengths are at least tens of thousands, and potentially millions of times longer than optical wavelengths, so the aperture (diameter) of the antenna would have to be tens of thousands to millions of times bigger than the lens to compensate. Since a lens might typically be inches to a few feet across, the impracticality of the equivalent radar antenna becomes apparent. SAR addresses this by creating a virtual aperture (the 'synthetic' aperture implied in the title) in software.[65]

361. **How SAR Works – More Details**:

a. **SAR Targets**. The first point to make clear is that SAR is not intended to detect small discrete targets in motion; moving targets are not normally plotted correctly by a SAR system. It is thus not a 'surveillance' system like an air or surface search radar. Rather it is a ground-mapping system that produces a pseudo-photographic plan view of the ground and of the features on it. Search radars usually try to reject radar returns from the surface, in order to see their intended targets. SAR instead exploits the ground return to build a picture of terrain and surface features. SAR still depends on reflection, so basic considerations of reflectivity etc still apply.

---

[63] SAR should not be confused with Sideways Looking Airborne Radar (SLAR). SLAR, like SAR, uses a fixed radar transmitter looking obliquely from an aircraft, but is based on different processing techniques. The two similar abbreviations should not, therefore, be confused. SLAR is invariably aircraft mounted; we will not describe it further.

[64] This does not imply that SAR and visual systems will in practice achieve the same resolution – we will not provide real resolution data in this Primer. SAR still represents an improvement of several orders of magnitude in radar resolution, however.

[65] There are in fact two common and complementary explanations of SAR principles; the explanation here based on the apparent size of the synthetic aperture, and an alternative explanation that relies on considering the changing Doppler effects seen by the radar as it moves relative to a stationary target. Both are explaining the same phenomenon. A reader wishing to see the 'Doppler' explanation should consult one of the radar references in the Bibliography.

b. **Basic Principles**. SAR improves basic radar resolution (in one dimension) by employing the motion of the vehicle to synthesise the effect of a large antenna aperture. What this means in practice is that the radar system transmits and receives looking ‘down and out’ obliquely from the satellite, at right angles to the ground track. Each radar return contains range information (how long the pulse took to go out and back from the transmitter) and bearing, or azimuth, information (the intensity of the return in different directions). By storing successive pulse returns and comparing adjacent ones, the SAR processor is able to deduce what would have been resolved by an antenna as big as the distance travelled between pulses (remember we are relying on the radar being in motion during the process). An effective SAR system will store many pulses and pulse returns, so that it has travelled a greater distance between the first and last pulses of a particular frame.

c. **Limits on Resolution**. SAR resolution is not limitless, but some of the constraints on it are complex. There are memory limits on the number of pulse returns that can be stored and processed, an area of ground will only remain in the radar beam for so long, and the electronic stability of the radar transmitter is critical to the system (specifically the processor needs *phase* information to relate adjacent pulses to each other). Also note that the radar is moving in one direction only, relative to the beam. It is thus creating a ‘long, skinny’ virtual antenna rather than an enormous circular one. Consequently the resolution of the radar is improved in one direction only; in principle the ‘along track’ resolution is being improved rather than the ‘across track’ resolution. Note, however, that the resolution of ‘non-SAR’ radar may not be equal in range and in azimuth in the first place, so we are not comparing like with like.

d. **SAR Modes**. We have described basic **swath SAR** above. It produces a ‘strip map’ of terrain swept by the radar beam. It is also possible in some systems to operate in a **squint SAR** mode where the beam is processed looking ahead or behind the perpendicular line to the platform motion. This can have subtle advantages in some circumstances, but a more common refinement is to process while moving the beam in azimuth. This effectively means that the SAR beam lingers on a target, giving improved resolution at a cost of lost coverage in adjacent areas. Accordingly, it is usually referred to as **spot SAR**, or **spotlight mode**. The radar beam is steered electronically, rather than by physically moving the antenna.

e. **Moving Targets**. We mentioned above that SAR does not resolve moving targets correctly. Anything moving relative to the ground return still sends its reflection back to the receiver, but the processor cannot resolve the fact that some of the shift is not due to the underlying platform motion. Accordingly, the target would still appear in the SAR output, but in the wrong position. In some applications, it is possible to combine SAR with separate *moving target indication* to correct for this.

f. **Inverse SAR**. We described SAR above as using platform motion to improve resolution against stationary targets. It is possible to reverse this process where the target has smooth, predictable motion and the radar is fixed, yielding ‘inverse SAR’ (ISAR). The obvious Space application for this is radar surveillance of orbiting objects, and this technique is used successfully to determine the size, shape and orientation of satellites in orbit.[66] See Section IV of this Chapter for more details of Surveillance of Space.

---

[66] Additionally, just as there are airborne (i.e. aircraft-mounted) SAR systems, there are also airborne ISAR systems, though the relative velocity of the target from the sensor is much less in these cases.

362. **Polarisation**. Polarisation has an inherent connection to reflection; specifically, a reflecting process often polarises the returning signal.[67] This is true for microwaves as well as optical wavelengths. Since radar depends exclusively on reflection, it is possible to exploit this within SAR systems. If the receiver can distinguish the state of polarisation of the incoming signal, this can be used to discriminate between a target and the background clutter, particularly on a smooth surface that may display consistent polarisation characteristics.

Illustration 3.20 – SAR Imagery, Polarisation and False Colour Presentation of Results. This is a SAR image of Munich, Germany taken by the NASA SIR-C/X-SAR system mounted on the Space Shuttle *Endeavour* on 18 April 1994. The image illustrates several features of SAR capability and processing. The radar was a SAR system operating on two frequencies and using two different polarisation modes. Note the following:

- The pseudo-visual appearance of the overall result; despite appearances, this picture could in theory (and may in fact have been) taken in total darkness.
- The false-colour presentation is used to separate out 3 of the 4 possible combinations of frequency and polarisation. These were assigned to red, green and blue channels in the output, allowing approximate characterisation of the target area. White output (i.e. responding to all frequency/polarisation combinations) is urban, pink identifies forested areas, while green distinguishes agricultural and other cleared open ground.

(Image courtesy of NASA/JPL)

---

[67] See Annex F for further details.

363. **SAR Meteorological Data**. As well as allowing pseudo-visual imaging, the level of detail captured by SAR sensors has applications in weather forecasting. A SAR picture of the sea surface can yield details of oceanographic features including fronts, wind and wave heights, strengths and directions. This, and other similar detail, can be useful for maritime operations, as well as for more general meteorological purposes.

364. **Nuclear Detonation Detection**. Separately from detecting the launch of missiles, there was also great interest during the Cold War in detecting nuclear detonation (NUDET). As a technical challenge, the problems were not insuperable; we discuss detection strategies below. As a strategic concept, detection was also at the time controversial.

a. **Technical Challenges**. As you might imagine, detecting an atmospheric nuclear detonation from Space is technically possible. The IR profile of such an explosion is very distinctive. It is shaped by the initial detonation, and by the thermal characteristics of the resulting fireball. Additionally, the detonation is accompanied by an equally characteristic emission of gamma rays and x-rays, which can also be detected from Space. Finally, an exo-atmospheric explosion is accompanied by a strong **electromagnetic pulse** (EMP).[68] The effect of such an explosion on satellites can be catastrophic, as we will see in Section V.[69] For a suitably hardened payload, however, especially if it is distant from the event, EMP sensing provides another detonation detection mechanism.

b. **US NUDET Systems**. Early US NUDET payloads were carried on dedicated satellites – the 'VELA' series of satellites. These comprised a suite of IR wide-field cameras and gamma ray, x-ray and EMP detectors, mounted together in the satellite. Although classified at the time, considerable detail of the VELA system performance is now available, largely because the gamma ray and x-ray detectors fortuitously yielded valuable scientific data. A VELA payload made the first detection of a gamma ray burst from a star, an event of great theoretical interest to astrophysicists. The decision to release this data into the public domain raised the visibility of the VELA constellation performance. VELA has been replaced by NUDET payloads hosted on a variety of other satellites, including, at present, the GPS constellation. Astrophysicists meanwhile have designed and flown customised scientific x-ray and gamma ray orbiting observatories to continue the work prompted by VELA.[70]

c. **Strategic Implications**. Some commentators maintained that detection of detonations was a confidence building process which verified the various agreements and restrictions on nuclear tests – particularly atmospheric tests. Additionally, it provided critical information to police nuclear warhead counter-proliferation. Others maintained that it potentially enabled fighting a nuclear campaign over an extended period – essentially providing nuclear battle damage assessment –and that by making such a campaign possible, it was inherently destabilising.

365. **SIGINT**. This Primer can say very little about monitoring of any other frequency bands from space. Suffice it to say that a little thought about the applications that we have already studied will reveal possibilities. A variety of techniques can be used on the ground to locate transmitters, and some of these would transfer easily to orbiting platforms. Considerations of overflight,

---

[68] Endo-atmospheric (.i.e. within the atmosphere) nuclear detonations also produce an EMP, but over much shorter effective ranges due to the density of the atmosphere.

[69] See Paragraph 392d.

[70] See Chapter 4, Section IV for a brief description of scientific payloads and missions.

visibility, transparency and range have been covered elsewhere in the Primer. To provide some illustration of joining these together, in the next two paragraphs we look at the operation of the COSPAS/SARSAT rescue systems and the ARGOS data retrieval system.

366. **COSPAS/SARSAT – Electronic Surveillance of a Cooperating Target**. COSPAS/ SARSAT was (and is) an application of Space surveillance to aid location of casualties for Search and Rescue purposes. Although the advent of dedicated survival radios with embedded GPS may have made it less relevant to military operations, it is still in use, and it illustrates several interesting design features of a satellite system. Accordingly, a brief description and explanation follows:

a. By international agreement, the radio frequencies of 121.5 MHz, 243.0 MHz and 406 MHz (among others) have been dedicated for distress and rescue purposes. Many survival systems incorporate distress transmitters that broadcast a continuous signal on one or more of these frequencies. Locating survivors involved traditional directional receivers, mounted on the ground or in surface or airborne systems, which could provide a bearing from the receiver to the signal source (hopefully to the survivor).

b. In the 1980s, it became apparent that these signals could be received in Space by orbiting platforms. A dedicated receiver could thus provide all the advantages of persistence and range/coverage that we have discussed elsewhere. A geostationary receiver would maximise coverage and persistence, but direction finding from geostationary altitude was impractical. An alternative was chosen to provide good coverage and fixing ability in a less complex package.

c. If a satellite in LEO receives a signal from a fixed ground transmitter, the satellite's speed over the Earth's surface imparts a Doppler shift on the received frequency.[71] If the satellite over-flies the transmitter exactly, it will see the signal at a steady, relatively higher frequency as it approaches the transmitter, and a steady lower one afterwards as it flies away. It will see a very sudden shift from one to the other as it passes directly overhead.[72] If its path is such that it passes to one side or other of the transmitter, it will still see a Doppler shift, but the total shift will be smaller, and the rate of shift in frequency at closest approach will be less abrupt.

d. This effect is exploited in the COSPAS/SARSAT systems. COSPAS was a Soviet system, SARSAT its western equivalent. The systems are interoperable, and shared information even during the Cold War. Working on the assumption that the satellite 'knows' the transmitted frequency, the receiver can analyse the Doppler shift of the received signal, compare it with its own motion, and derive location information for the transmitter. The satellite can assume that a received signal is within its horizon (this is a rare example where the relatively close horizon from LEO is a positive benefit). The point where the received signal frequency is at its 'middle' value gives the closest approach of the satellite to the transmitter. At this point, the satellite is either directly overhead the transmitter, or the transmitter is directly 'abeam' its path. The rate that the frequency changes at (abrupt or gradual) gives an indication of range. A single satellite pass may not provide a very accurate location; typically, the 'fix' may be an ellipse of some tens or hundreds of kms extent. This can still be a life-saver if used for initial localisation, and if there is confidence that only a single transmitter is involved, two or more passes can provide accurate fixing.

---

[71] This Doppler shift must also be allowed for in any ground-segment that communicates routinely with the satellite. A short description of Doppler effects can be found in Annex F. We also discuss its relationship with deriving orbital data in Paragraph 331.

[72] The audible analogy is the sound of a police or ambulance siren changing when the vehicle drives past the listener.

e. The next problem was ensuring regular, global coverage for the receivers. The chosen solution was to 'piggy-back' them on weather satellites, specifically those in polar LEO. These constellations are designed to provide detailed weather imagery from LEO on a recurring basis. The combination of polar LEO and the Earth's rotation ensure that the whole of the Earth's surface is surveyed regularly. An additional factor favouring them was the overt nature of their mission, and the atmosphere of international co-operation that surrounds their work. There was therefore no sensitivity about flying simple receivers of deliberately limited capability on a global basis.

f. Since the system was deployed in the 1980s, it has been expanded and augmented. It now includes a GEO alerting capability (no fixing) to complement the doppler-based location carried out from LEO.

g. The final technical hurdle of note related to the ground transmitters. Doppler fixing as described relies on knowing that the transmitted frequency is constant, and assuming that it is somewhere close to its intended value. The widely-used 121.5 MHz and 243.0 MHz transmitters lacked the stability to provide optimum fixing from the satellite, particularly in varying temperatures. The solution was the introduction of 406 MHz transmitters manufactured to higher standards, to aid fixing accuracy. COSPAS/SARSAT fixing of 121.5 MHz and 243.0MHz beacons was discontinued in February 2009, as 406 MHz transmitters are in widespread use.

367. **ARGOS**. Many environmental monitoring projects have been enabled via the ARGOS system, a joint venture between the French Space Agency CNES, and in the USA NASA and the National Oceanic and Atmospheric Administration (NOAA). ARGOS uses dedicated beacons which can be incorporated within data gathering experiments such as oceanographic buoys, or incorporated into tags that can then be attached to experimental subjects such as migratory animals. The ARGOS space segment is hosted, like COSPAS/SARSAT, on meteorological satellites to ensure comprehensive surface coverage. The message format allows an ARGOS beacon to encode information from its experiment and to broadcast it in messages at pre-determined times. The satellite package retrieves the messages and returns them to a ground processing centre. Here, location data can be extracted from Doppler information in the received signal and attached to the experimental results. The whole message is then sent to the customer. The message format is deliberately flexible, for example allowing the encoding of GPS position if the experiment can support it, which improves accuracy and obviates the need to perform Doppler fixing. As well as supporting historical track data, for example to monitor migration, and applications like remote climate monitoring by purpose built stations, it has been used for one-way communications, for example to enable medical alerting for public health purposes in remote locations. No direct military applications have been reported, although military users benefit from the meteorological information retrieved via ARGOS, and in the security sector, ARGOS is used for maritime security and intruder alerting on merchant ships to comply with International Maritime Organisation standards.

Surveillance and Reconnaissance from Space – Special Factors

368. **Context**. All that we have said so far in this section about surveillance and reconnaissance from Space needs to be seen through the filter of the constraints outlined in Chapter 1 relating to orbits, ground tracks, the Space environment and satellite engineering. Some details have been alluded to throughout the Chapter, but it is probably useful to gather recurring themes together. We summarise these themes below, but before doing so, it is also worth briefly listing the advantages of ISTAR from Space:

a. Potentially global reach, including concurrent support to widely separated operations, and to deep, close and rear aspects of a single theatre.

b. Access to otherwise denied areas (through legal overflight).

c. Near invulnerability, compared to air-breathing systems.

d. Wide area coverage.

e. High look-down angles to avoid terrestrial shadowing.

f. Minimal in-theatre logistical tail, compared with that required to derive similar information conventionally.

g. Potentially autonomous operations, with low head-count of personnel overall for results gained.

369. **Orbits, sampling and revisit calculations**. The orbit of a surveillance satellite may in many cases offer clues as to its purpose. For the user, the orbit needs to be appropriate to the task in hand. We can consider this in terms of range and ground track.

a. Range between target and system is generally seen as the enemy of resolution, which is in itself true, but for satellite systems, this hides other factors. Range is determined by a combination of orbital altitude and satellite ground track. Low altitude thus aids resolution, but it does it at a price:

(1) LEO has the most limiting field of view in terms of distance to the horizon; the lower the orbit, the shorter the range to the target, but the less distance on the ground you can see before curvature of the Earth intrudes.[73] Optical sensors may well look straight down to the nadir point, though they may also have the capability to image off-nadir at reduced resolution. When imaging at nadir, distance to the horizon is irrelevant to them. Radar systems usually look obliquely sideways so need to account for distance to the horizon. LEO thus favours reconnaissance, if the ground track can be adjusted to suit, but hinders wider-area surveillance.

(2) LEO has the shortest orbital period.[74] This leads to fastest overflight of the target area, and hence to the most fleeting detection opportunities for ISTAR, though it may also facilitate more rapid revisit.

(3) Looking at the longer term, LEO is the least stable orbit, suffering the greatest rate of frictional decay.[75] This either implies the ability to re-boost the satellite periodically, leading to increased complexity and bulk, or shorter overall life. For sophisticated (= expensive) sensor systems, this latter approach may be unacceptable, though for smaller, simpler packages it may be a perfectly acceptable price to pay for shorter range.

(4) Because of the relatively fleeting ground passes, LEO poses the greatest problems in returning results to the ground. This may be acceptable for

---

[73] See Paragraph 144.

[74] See Paragraph 146.

[75] See Paragraphs 111 and 146d.

reconnaissance when coupled to the relatively short orbital period, but for surveillance, where warning time may be a critical factor, this may be a greater constraint. Communication via other satellites may offer a solution.[76]

b. Ground track is driven by a combination of orbital inclination and period, when compared to the Earth's rotation rate. Remember also from Chapter 1 (Paragraph 164b) that launch location directly constrains achievable inclination.

(1) Ground track disappears in GEO (Paragraphs 148-9) – the satellite has a fixed field of view of the Earth's surface, there is no overflight or revisit calculation, and the communications solution is fixed. The satellite is also, however, constrained to orbit over the equator, with consequent limitations at high latitudes, or alternatively to use a geo-synchronous, rather than a GEO, with consequent north-south excursions from a fixed position.

(2) Where the advantages of lower altitudes are adopted, ground track becomes a complex issue that will require modelling. Orbital period shortens, but the ground track may either favour wide-area surveillance, with eventual over-flight of a large area but long intervals between individual passes, or else favour more regular over-flight of a more restricted set of locations, with other areas permanently out of view.

(3) Polar orbit (Paragraph 150) provides the quickest method of ensuring global coverage. Note, though, that from polar orbit, both poles are over flown regularly, probably with significant overlap from one orbit to the next. Assuming a paucity of targets in polar regions, this is, to an extent, 'dead time' in the satellite's orbit, but it also makes high latitudes a potentially useful location for the ground segment stations, as all the assets in a constellation in polar orbit will regularly come within range. This will probably outweigh the environmental disadvantages of operating in such locations. See also the vignette in Chapter 1 following Paragraph 134 relating to orbital periods and polar orbits.

(4) Sun-synchronous orbits (Paragraph 152), which are very nearly polar, provide consistent illumination on repeated overflights of a given target.

(5) Any requirement to change the orbital plane of the satellite (by definition an 'out-of-plane' manoeuvre – see Paragraph 170b), for example to alter inclination, comes at a great price in manoeuvre/station-keeping fuel. The balance between short-term, possibly tactical, gain against long-term cost must be established. The relative merits of these factors may also vary during the planned life of a system, if, for example, coolant is being depleted or other items are approaching the end of their design life. For present-day systems, such decisions are unlikely to be devolved to tactical level.

(6) Where comprehensive ground coverage of a specific area is required, one possibility is to 'freeze' the ground track, by making the period a simple fraction of a day. So for example, if the satellite orbited the Earth in 90 minutes, it would make exactly 16 orbits per day. This would give a predictable return to a given point on the Earth's surface, at the cost of never overflying other points. The advantage of this regime, which has some characteristics of **persistence**, is that to move the path

---

[76] See Paragraph 185.

to overfly a *different* point on the ground, it is only necessary to raise or lower the orbit slightly, which will (must) change the period (Kepler's Third Law, see Paragraph 128). The orbital period is then out of step with the Earth day, and the track-pattern will move slowly over the Earth's surface. When the desired point is in coverage, the orbit can then be adjusted back to freeze the ground track again. To raise or lower an orbit, and thus to adjust the period and ground track is an in-plane manoeuvre, relatively much less costly in fuel than an out-of-plane manoeuvre. The bigger the change in altitude and period, the quicker the ground track will move, so there is still a fuel-cost/tempo decision to be made, bearing in mind the urgency of the required change.

(7) True persistence, with all points on the Earth's surface being covered continuously, or even nearly-so, would only be achieved from LEO by a large constellation of satellites. The size of the Iridium constellation (66 intended satellites) required to provide continuous coverage from LEO for communications,[77] probably illustrates why so far, no-one has constructed one, although proposals are beginning to emerge for such systems.[78]

c. The underlying calculations that support modelling of ground track and range, or of field of view of a system, are not enormously complex, but they would be time-consuming to carry out manually. Happily, there are many modern orbit and payload software modelling packages that can make them quickly and present the results in easily understood formats. This both aids planning for friendly assets, and may allow rapid assessment of the potential of hostile systems.

370. **Spots and Swathes – Choosing Resolutions**. Modern imaging systems may have adaptable sensors that can vary their field of view (this is only really the equivalent of a zoom lens on a terrestrial camera, although other means may be used in orbit in preference to complex optical systems). The trade-off is between the area covered and the achieved resolution. The implication for users is not to request resolution beyond that required for the task in hand. Requesting lower-resolution data than the 'best' available eases several constraints on the intelligence chain. It may increase the number of potential collection systems, it may allow re-exploitation of already collected data and by collecting a larger field of view, it may allow a single image to satisfy several requests. Where wide-area coverage is required, with adjacent images being 'stitched' together, accepting lower resolution allows the ground to be covered in fewer swathes, greatly speeding the production time. This may be particularly important where the satellite ground track is such that there will be a long delay between adjacent passes over the target area.

371. **Returning Surveillance Results to the Ground**. We have noted in several parts of the Primer the implications of keeping a satellite in communications coverage. This was initially principally a limit on satellite control, very little of which was autonomous. Nowadays, where satellites carry much more complex payloads, returning results to the ground may pose equal or greater problems. A LEO reconnaissance satellite over its target in hostile territory may well be out of direct contact with a friendly ground station. There are several potential solutions to this problem:

---

[77] See Paragraph 327a.

[78] In 2008, CNES, the French Space Agency, proposed a constellation of high-resolution imaging satellites to provide a consistently updated world picture. The proposal is referred to as e-CORCE, 'A webcam for the Earth'. With suitable investment (not yet secured), CNES believe IOC could be achieved in 2014. See also the discussion of enhancements to the Iridium constellation at Paragraph 327 (and associated footnote).

a. We saw that some early reconnaissance packages exposed film then returned it to Earth for processing via re-entry capsules. On one hand this was a simple system, but on the other, the potential for failure through flawed re-entry, missing the target area or hitting the ground too hard was obvious.

b. Satellites can themselves employ satellite communications to communicate from LEO to a distant ground station. Some dedicated communications satellites exist to support this capability.

c. Much of the need to communicate with dedicated ground stations arose from the fact that ‘raw’ satellite product was unintelligible to the end-user. This is now being addressed through increased on-board processing power and more capable portable ground installations. Some platforms may therefore broadcast direct to a user in the field, offering tactical capability in near real-time. There will still be issues of tasking and allocation to be addressed, but many communications burdens are eased by such systems.

SECTION IV – SURVEILLANCE OF SPACE

In this Section you will find a description of:

- Methods of detecting satellite launch.
- Methods of detecting objects in orbit using radar and optical systems.
- How to gather information about objects in orbit.

372. **Introduction**. Most of the Space activity we are reviewing in Chapter 3 depends on access to a Space-borne platform of some kind. Surveillance of Space is an exception to this; it is an activity that can be conducted from the ground, though as we will see, Space-based sensors provide valuable information too. For military planners, the aim is to develop **Space situational awareness (SpSA)**. We will see that this parallels the same sort of activity in the terrestrial domains.

373. **Detecting Launch Activity**. There are a variety of overt indicators of potential Space activity. A Space-faring nation may publicly announce its intentions to launch a satellite to achieve prestige or influence. There may be open-source announcements relating to fabrication of the payload or launcher. Traditional terrestrial or Space-based reconnaissance may reveal activity at a launch site. Bearing in mind the constraints on achievable orbits from various launch sites outlined in Chapter 1 (Paragraph 164), there may also be activity when the payload is deployed to a distant launch site. However, most of these cues could at least in principle be circumvented. The first unambiguous indicator of Space activity is launch.

a. **IR and the Launch Plume**. We described the missile-warning function of IR sensors in some detail in Paragraphs 357-8. Plainly these sensors will also detect a launch to orbit if the site is within range. The tracks and flight paths of a missile and a satellite launch will quickly diverge. We cannot discuss here exactly how much information an IR detector can derive about the nature of the launch activity, but the cueing potential is obvious.

b. **Radar Search**. Radar is also an integral part of missile detection, and in the same way as radar can detect a ballistic missile in flight, it can detect items en-route to orbit as soon as they are above the horizon at the radar site or in range of an airborne platform.

Most radar systems are strictly line-of-sight, though missile warning has been one of the main drivers for development of over-the-horizon radar systems. Given their size, few launcher/payload combinations should pose problems for radar detection.

c. **Other Cues to Launch Activity**. The role of other traditional collection methods to detect launch activity should not be ignored. Dependent on weather, launches can be observed visually over hundreds of miles. Peacetime launch activity may be notified via NOTAM for aviation safety and via Notices to Mariners for seamen to enforce range-safety considerations, though these notices are likely to be for extended periods rather than identifying the instant of launch. During the 1960s and early-1970s, an enthusiastic schoolmaster at Kettering Boys School in the UK achieved fame for the group he led, which regularly detected and analysed Soviet Space activity. They had no direct Space observing capability, but instead, coming from an amateur radio enthusiast background, exploited the telemetry and other signals broadcast during a launch.

374. **Detecting Objects in Space**. Once an object achieves orbit, the two main methods of tracking it are via radar and via optical systems. The USA and Russia maintain catalogues of on-orbit items, updated through a combination of radar and optical observations, predominantly from the ground.

a. **Radar**.

(1) Large radar systems sweep Space regularly to maintain a catalogue of orbiting objects. These include known objects – satellites of confirmed identity in established orbits – as well as unknown items. The latter include debris from launch activity and from collisions or failure of established satellites. The usual constraints on radar performance apply – small objects are usually more difficult to detect than large ones, distant ones give weaker signal returns than closer ones (GEO may be very challenging in this respect), and the shape and composition of the target also affect its likelihood of detection. Finally, the implications of elliptical orbits must be considered. These can place a satellite close to the surface over one area, but as distant as GEO at another, with implications for radar detection depending on where the sensor is.

(2) The USA maintains a network of Space-surveillance radars, of which RAF Fylingdales forms a part. France and Germany have some additional capability, and there are extant proposals to extend European Space Surveillance capability through a combination of optical and radar sensors. Further details are given in Chapter 4.

(3) We noted above (Paragraph 361f) that complex radar signal processing (inverse SAR), could yield detail of the size and shape of moving targets from fixed ground locations. This technique has been used to characterise orbiting bodies. The best known of these systems is the 'Haystack' radar operated by the Lincoln Laboratory of the Massachusetts Institute of Technology.[79]

---

[79] See www.haystack.mit.edu/obs/haystack/LincolnUpgrade.pdf for a description of the Haystack installation and an outline of its management as a joint research capability of MIT and an element of the US space surveillance network. The details of planned upgrades also verify once again the relationship between wavelength/frequency and resolution, including illustrative radar-imagery of a satellite.

b. **Optical Detection**:

(1) **Visual Detection**. Some satellites are big enough to be detected visually from the Earth's surface; on any clear evening in the UK, if you look long enough in the sky you are likely to see a satellite pass overhead. Most are relatively dim objects moving steadily across the sky, and you can tell little from what you see, other than that the time and direction of rising, transit and setting, coupled with your own position may serve to identify the satellite. Transient events in orbit can occasionally be seen; the Iridium communication satellites in LEO carry highly reflective solar arrays for power generation. When these reflect sunlight in the direction of a ground observer, the resulting 'flash' is dramatic. These events, known as 'Iridium flares', typically last up to 20 seconds, and can be bright enough for daylight observation. They have been mistaken for air accidents, re-entry events and suchlike.[80] Note, however, that compared to the thousands of tracked objects in orbit, very few reach the brightness necessary for visual detection, so even ignoring the impact of weather, it is not a reliable or comprehensive search or tracking method.

(2) **Searching GEO**. Satellites at GEO are sufficiently distant that unassisted visual detection is impossible. Dedicated telescopes are, however, the most common means of searching for objects in GEO. For the ground observer, orbits in GEO are stationary, so all that is necessary is to scan the desired portion of the GEO belt and look for objects that are stationary against the background of stars apparently moving due to the Earth's rotation (See Illustration 3.21a). There are several such systems engaged in coordinated monitoring of the GEO band. Note that any given ground location can only ever see a fixed portion of the GEO band. Comprehensive coverage thus requires a network dispersed in longitude around the world. Coverage in latitude is less important, and can be optimised to maximise favourable weather opportunities.

(3) **Gathering Information in LEO**. Relatively modest optical systems may show some size and shape for an orbit in LEO, and a dedicated system would yield more information. Amateur observers frequently publish such images (See Illustration 3.18b below), particularly of large targets such as the ISS, and exploit other details such as variations in apparent magnitude (i.e. apparent brightness) to speculate on possible shapes of objects or manoeuvres such as rotation. Where they relate to sensitive payloads, the accuracy of such speculations is understandably completely unconfirmed. As noted above, radar imagery of satellites in LEO is also productive (See Illustration 3.21c).

---

[80] In the Bibliography, you will find a reference to the 'heavens-above' website, which provides predictions of visual satellites passes for specific locations. The orientation of the Iridium solar panels is public information, and the flares are thus also predictable – details are available on the same site.

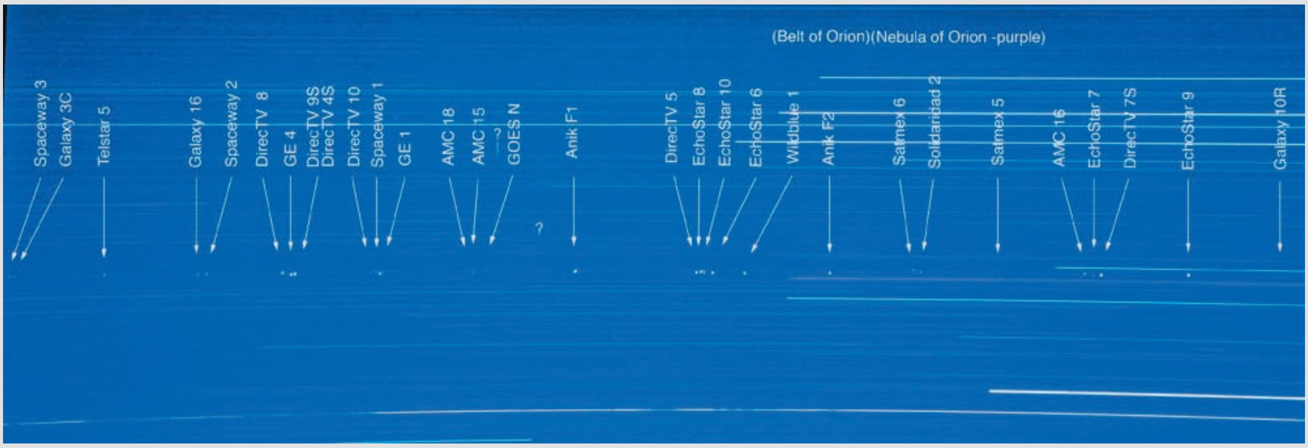


Illustration 3.21a – Visual Surveillance of GEO. This picture shows how objects at GEO betray their presence. This is an 8-hour exposure of the night sky using a fixed camera. Over the long exposure time, stars appear to move due to the Earth's rotation, and consequently appear as streaks. The GEO satellites keep station with the Earth as it rotates, and appear as fixed dots. Good weather, a favourable location and pointing the camera at the correct part of the sky were all that was required to produce this picture; the camera was a professional-grade Hasselblad camera with an 80mm lens, but no complicated optics or camera mount were required. The photographer's identification of individual satellites came from unclassified sources. (Image courtesy of Dr Bill Livingston at Kitts Peak National Solar Observatory, AZ.)

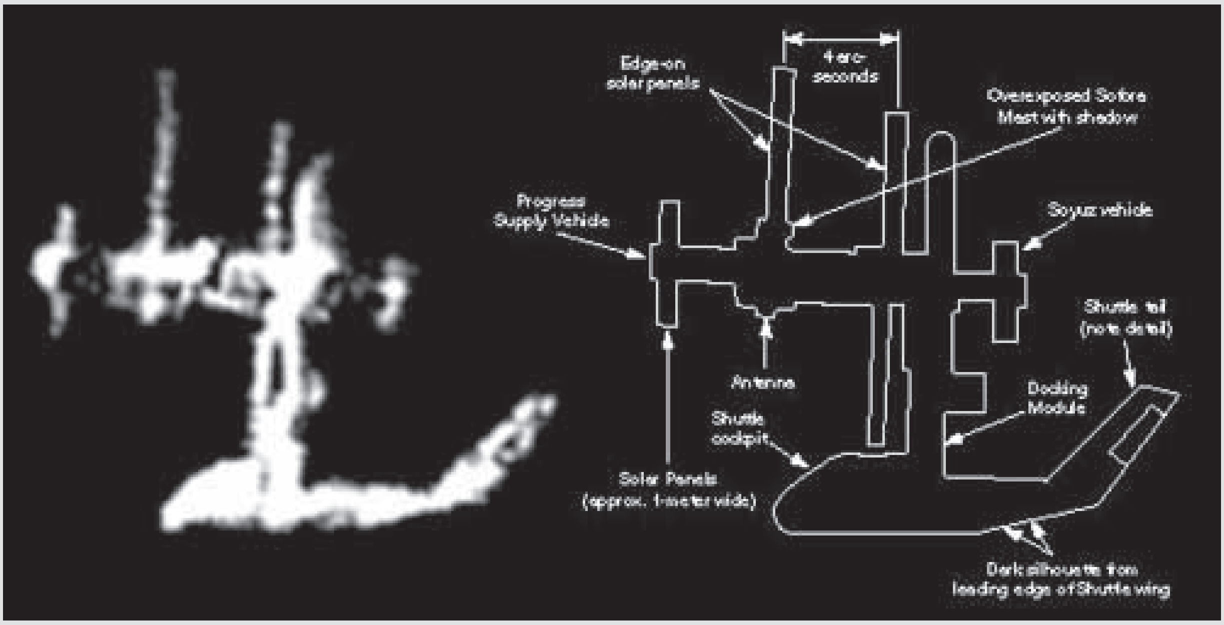


Illustration 3.21b – Imaging an Object in LEO. The picture shows the Space Shuttle *Atlantis* docked with the Russian MIR space station in November 1995. The key on the right identifies individual components down to about 1m resolution, taken from the ground (just outside Boston, MA) with the Mir/Shuttle combination at 350 miles altitude. This picture was taken using a high-quality software-controlled amateur telescope (12" aperture) and a consumer grade video-camera, but nevertheless clearly demonstrates the potential of ground-based observations. (Copyright: Museum of Science, Boston, Massachusetts, USA. Image by Ron Dantowitz. Reproduced with permission).

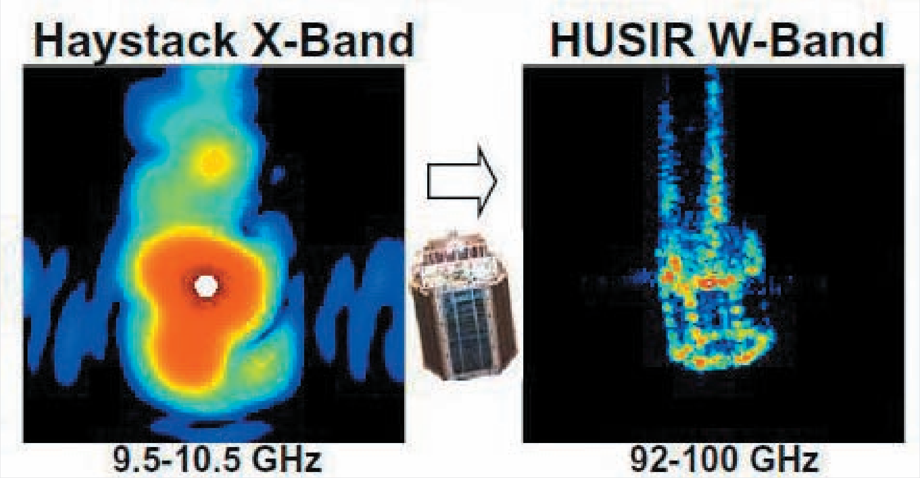


Illustration 3.21c – Imaging with Radar and the Frequency/Resolution trade-off. This picture serves two purposes. It illustrates the ability to process radar returns to generate pseudo-visual images, and demonstrates the relationship between frequency/wavelength and resolution. It shows the current, and potential future resolution of the 'Haystack' radar installation at the Lincoln Laboratory of the Massachusetts Institute of Technology against an orbiting satellite. Currently operating in X-band (left image), the Haystack radar is being upgraded to operate in W-band (right image). The W-band sensor operates at about 10 times higher frequency than the X-band version, which implies 10 times shorter wavelength and consequently the resolution improvement (Paragraphs 344, 361f and F24 refer). The images shown are actually of a laboratory test of the radar upgrade against a model target, but the potential is obvious. (Images used with permission from the Lincoln Laboratory of the Massachusetts Institute of Technology).

375. **Track Maintenance**. We will not look in detail at how the results are derived, but a series of coherent observations of an object from Earth allow its orbital elements to be deduced. Subsequent observations, which can be predicted from these either confirm and refine the extant elements, or alert observers to an accidental or deliberate perturbation of an orbit. Ultimately, the disappearance of an object from tracking may indicate that its orbit has decayed and it has re-entered the atmosphere. Notice from this what we discussed earlier in Chapter 1; particularly in LEO, orbits decay at varying rates and orbital elements will alter over relatively short periods of time. Thus the operator of a space platform may have as much interest in tracking his own assets as in maintaining awareness of the activities of others. A sudden change in orbital elements may well indicate an intentional manoeuvre, for example re-boosting a satellite in LEO to a higher altitude to counter orbital decay, or an intentional change of orbit related to its task. In order to maintain a coherent Space picture, regular surveillance of known items is thus critically important.

376. **Analysing Objects in Space**. Once an object has been detected and tracked, and its orbit established, it is possible to make some deductions about what it is, or at least might be. Throughout this Chapter, we have noted how specific orbits favour particular tasks. Thus an observer may be able to work backwards from the observed orbit to the task. The orbit also leads to the ground track and field of view for the satellite – at least for as long as the extrapolated orbital elements are meaningful. If any physical characteristics of the satellite can be deduced from radar or optical observations, these too may yield useful information. Examples include the dimensions of antennae revealing the frequency band they may be optimised for, the dimensions of solar panels allowing deductions about power capacity, and features such as the dimensions of the payload at launch giving constraints on overall sensor size. All these clues, when integrated, allow an observer to build a conceptual model of a satellite and make valid deductions about its purpose and capability.

377. **Surveillance of Space from Space**. IR detection of launch is inherently Space-based, but everything else we have talked about in this section can be (and usually is) done from the ground.

Nonetheless, some of the obstacles to effective surveillance can be overcome from orbit. In particular, more stable and predictable observation, without the obstructions of weather and atmosphere would have obvious benefits. Surveillance and observation could be carried out at several wavelengths. Capability (and counter-capability) in this area are relatively immature, but the potential is clear. Applications might include close-up inspection of other satellites, as well as self-surveillance of the immediate vicinity of a satellite as a defensive measure.

SECTION V – SPACE CONTROL

In this Section you will find a description of:

- Definitions relating to Space Control.
- Space Situational Awareness
- Defensive and Offensive Space Control capabilities.
- Ballistic Missile Defence systems.
- Anti-satellite technologies.

Space Control and Space Situational Awareness

378. **Introduction and Definitions**. 'Space Control' is still a relatively young discipline, and definitions of it are maturing slowly. The most comprehensive definitions originate in the USA.[81] They describe the field in some detail, but readers should note that they are liable to change, and that there are both USAF and US Joint Staff definitions in circulation.[82] They are not entirely congruent.

a. **US Joint Definition (from US JP 1-02 Definitions Glossary)**. '*Operations to ensure freedom of action in space for the US and its allies and, when directed, deny an adversary freedom of action in space. The space control mission area includes: operations conducted to protect friendly space capabilities from attack, interference, or unintentional hazards (defensive space control); operations to deny an adversary's use of space capabilities (offensive space control); supported by the requisite current and predictive knowledge of the space environment and the operational environment upon which space operations depend (space situational awareness).*'

b. **USAF Definition (US AFDD 2-2)**.[83] The 2006 re-issue of AFDD 2-2 acknowledges the Joint definition cited above, but then offers the following as a USAF-only 'clarification': '*Operations to attain and maintain a desired degree of Space superiority by allowing friendly forces to exploit Space capabilities while denying an adversary the ability to do the same (e.g. protection, prevention and negation). Space Control is achieved through offensive counterspace and defensive counterspace operations. Note: the Air Force uses counterspace as an equivalent definition of the Space Control mission.*'

---

[81] See Paragraph 426 for more information on US Space Doctrine.

[82] In the UK, leadership for Space Control Capability is vested in TIO, and their staff should be consulted for the latest national description of capability and policy. There is also a UK definition of Space Control in Chapter 3 of AP3000 (4th Edition), which was published in 2009.

[83] US AFDD 2-2 is the USAF manual of Space Operations in the round. Space Control operations are further elaborated in AFDD 2-2.1 'Counterspace Operations'. Sources of space doctrine are discussed in greater depth in Chapter 4, Sections I and V.

Note the USAF introduction of the 'counterspace' term. *We will use the term 'Space Control' for the remainder of this section, but take it to refer equally to the USAF 'counterspace' concept.*

379. **Access to US Publications**. Both the USAF documents (AFDD 2-2 and AFDD 2-2.1) and US JP 3-14, the manual of Joint Doctrine for Space Operations, are at the time of writing, open-source public access documents, available on the Internet. The latest versions of each, which can be found via common search-engines, should be consulted to determine any developments in US thinking.

380. **Space Situational Awareness**. SpSA[84] underpins offensive and defensive space control. We touched briefly on some aspects of it when we looked at surveillance of space above, but as we noted then, SpSA is more than a catalogue of orbital elements.

a. Orbital elements enable basic tracking and allow projection of ground track or footprint.

b. Friendly space assets will communicate with their ground segment, confirming their serviceability and status.

c. Tracking a hostile or unknown space target over time may provide correlation between any declared purpose and its actual capability. It may also give warning of any unexpected manoeuvre in orbit.

d. Any measureable characteristics of the satellite (size or shape for example), may also indicate purpose.

e. Space weather data contributes directly to SpSA.

f. Information from other sources (some open, others covert) may also provide correlation of SpSA. Commercial operators may, for example, advertise their capability openly.

The implication of this analysis is that technical surveillance of Space is a necessary component of SpSA, but is not enough in isolation. True SpSA requires integration of information from many sources, using analysis and correlation techniques common to many other intelligence problems. The particular features of the Space environment introduce their own certainties and uncertainties. Ideas of 'track integrity' common in terrestrial air traffic control may offer useful templates for SpSA.

Defensive Space Control

381. **Breakdown of Defensive Space Control**. This description of defensive Space control largely follows the outline of AFDD 2-2.1, though without its emphasis on the detail of US Space C2. The USA recognises deterrence, defence and recovery as aspects of defensive Space control. We will examine each in turn.

382. **Deterrence**. Plainly, elements of space capability can be regarded as critical national infrastructure, and the owners of such would regard any hostile action very seriously indeed. This Primer will not explore the policy aspects of this any further, but readers should note three factors

---

[84] Beware the ambiguous use of 'SSA' as an abbreviation for Space Situational Awareness in a joint context, where SSA commonly refers to '**Shared** Situational Awareness' – potentially in any domain, and not necessarily related to space capability at all.

directly relevant to space assets. Firstly, a reaction to a space attack could be enforced in any domain. Secondly, a country would probably grade space assets in order of importance or national sensitivity when considering how to deter attack. Finally, given the current state of SpSA, any response might well be complicated by questions of attribution of the incident. While some kinds of hostile action might easily be tracked to source, space weather and undetected debris could yield similar effects. The sudden failure of a satellite without obvious physical impact could thus be very hard to analyse, and the penalty for false attribution might be severe. 'Plausible deniability' might be within reach of the instigator, and if nothing else this underscores the importance of developing SpSA capability.

383. **Defence**. Defence can be analysed as a combination of passive and active measures.

a. **Passive Measures**. Some passive measures are direct extensions of terrestrial passive defence. These include **camouflage**, **concealment** and **deception**, and apply principally to the ground segment. Some elements of the ground segment may be impossible to conceal; large communications antennae, radar installations and launch facilities spring to mind. Others, however, such as tactical hubs and nodes, headquarters, fabrication facilities and other essentially administrative facilities may pose fewer problems. **Hardening** can apply both on the ground and in Space. It includes physical hardening to resist kinetic and electronic attacks, and irradiation by laser or nuclear radiation, for example from the detonation of a nuclear weapon, as well as anti-jamming techniques to protect uplink and downlink control circuits. Physical hardening in Space may well be constrained by launch mass, though the hostile nature of the Space environment dictates a base level of protection as a consequence of exposure to Space weather (see Paragraphs 106-114 in Chapter 1 for details of Space environmental issues). Another factor to bear in mind regarding any sort of radiating weapon in Space is the absence of any protection offered by atmospheric absorption. Finally, **dispersal** may be possible on the ground, but in Space really overlaps with redundancy, which we will examine shortly.

b. **Active Defence**.[85] Active measures to protect Space assets include interfering with an adversary's ability to attack them in the first place, as well as responding in space when an attack is imminent or occurs. Little need be said about the first aspect once capability is evident. Roughly, such action is analogous to Offensive Counter-Air activity in the air domain. Active defence in Space is a more involved subject. If a threat can be identified in time, it may be possible to manoeuvre a satellite to avoid it. Note, however, the stipulations in Chapter 1 about the implications for fuel consumption of satellite manoeuvre, and consequent impact on anticipated ground track in the short term and ultimate mission duration in the longer term. Where a threat is to a sensor rather than the satellite overall, it may be possible to protect it by shielding or attitude control. Finally, where the threat is to satellite control, e.g. by attempted jamming or interference with control signals, there are active measures that may protect the system or at least mitigate effects. Note, however, that there are 'soft-kill' implications in most of these actions. The choice between risking destruction on the one hand or accepting mission degradation or suspension on the other is a decision for the operational commander to take.

384. **Recovery**. Where defence has been unsuccessful, recovery of Space capability is achieved either via redundancy or reconstitution. Complex systems may already incorporate elements of redundancy; recall that the GPS constellation requires 24 satellites for full capability, but has

---

[85] This section of text solely considers capability and technical feasibility. The legality of any active measure must of course, also be considered.

contained up to 32. This does not mean that **any** satellite could be moved to replace **any** other; constraints of manoeuvre and fuel supply would still apply. Equally, however, even if availability dipped below 24 satellites, the degradation of capability would initially be intermittent, local and essentially predictable. Similar factors would come in to play for other Space capabilities, and mitigation might include the extent to which the commercial Space sector could fill gaps in availability. Reconstitution might at first sight seem a long-term project, though in some capability areas there may be a regular programme of platform replacement in place. The need to mitigate acute localised gaps in coverage may also be a justification for the development of operationally responsive capability.

Offensive Space Control, Ballistic Missile Defence and Anti-Satellite Systems

385. **Policy, Legal and Technical Aspects of Offensive Space Control**. US Doctrine sub-divides Offensive Space Control into deception, disruption, denial, degradation and destruction options. Essentially, however, these are simply different (increasing) amounts of temporary or permanent impairment of an adversary's capability. Plainly there are significant policy, legal and strategic issues posed by offensive Space control options, though few of these can be examined further in this Primer. Some of the technical issues exposed by Offensive Space Control can, however be discussed openly, at least to indicate what is possible. In this section we look at aspects of Ballistic Missile Defence (BMD), including a historical review of capability to date, differences between the demands of Ballistic Missile and Anti-satellite (ASAT) Operations and finally at damage mechanisms that might provide credible ASAT options. This grouping is strictly for ease of reference, and does not imply any UK Government policy position on any aspect of these capabilities.

386. **BMD Categorisation**. In the same way as we can regard ballistic missiles as an element of Space capability, even though their reliance on Space is transitory, there are aspects of BMD which overlap aspects of Space Control, particular where the enabling technology is similar. Doctrinally, BMD and Space Control are considered to be separate disciplines, but the technical similarities are unchanged.

387. **BMD and ASAT Compared**. For a variety of historical, technical and doctrinal reasons, discussions of BMD and ASAT weapons and techniques also often overlap. There are again similarities between both fields that make such comparisons useful, but equally there are fundamental differences that should be kept clearly in mind.

a. **The Target**. There are substantial differences between the targets. Ballistic objects can be assumed to be warheads, or related decoys. Relatively speaking, they are robust objects, containing either conventional or Chemical, Biological, Radiological and Nuclear (CBRN) munitions. Size varies; warhead dimensions are constrained by the launcher and the size of the munition needed for the desired effect. Satellites, as we have seen elsewhere, are inherently fragile.

b. **Timelines and Tempo**. The total flight-time for an ICBM from launch to impact might range up to 30 minutes or so, but not much longer. Shorter-range missiles have even shorter flight times. Thus the BMD process must operate entirely inside that cycle, from detection, to decision, to execution. This will always be a challenging constraint. In comparison, a decision to interfere with a satellite can follow a relatively slower pace, including refinement of the targeting solution, consideration of various attack options and possible re-attack. This is not to say that tempo is not a constraint. An adversary might manoeuvre the target, there may be pressing operational reasons to deny him his capability, and the development of Operationally Responsive Space capability may force a re-appraisal of what tempo is acceptable.

c. **Cueing and Tracking**. At the instant of launch, ballistic missiles and orbital payloads may be indistinguishable, though the identity of the launch site may give an initial indication of which is more likely. Space-based IR surveillance is almost certainly the quickest and most reliable method of detection. The flight paths of a missile and a satellite will quickly discriminate between them and both bodies will probably need to be tracked on radar either to derive orbital elements for a satellite or to calculate intercept geometry to counter a ballistic missile.

d. **Debris**. Any kinetic attack on an orbiting satellite will almost certainly leave most of the satellite and at least some of the weapon in orbit.[86] In all probability, subsequent collisions between portions of debris will multiply the total number of pieces of debris. Whether this debris disperses will depend on the impact geometry.[87] The altitude at intercept will control how quickly the orbits of these debris pieces will decay and re-enter the atmosphere. For a ballistic missile, the debris will be sub-orbital but the implications vary, depending on where along its flight-path it is intercepted. We will look at this in more detail shortly.

e. **Discrimination**. Identifying a satellite target is a function of SpSA as outlined above. Decoys are unlikely, though dispersal through redundant systems, and concealment of a satellite's true or multiple functions pose similar problems and may introduce issues of collateral damage. Decoys intended to confuse ABM systems, however, may well accompany ballistic missile warheads. These are deployed after the boost phase, so they follow the same path as the true payload during cruise. Some publications divide the cruise stage of a ballistic missile flight path into stages before and after deployment of decoys. Decoys are not usually designed to survive re-entry, so their effectiveness diminishes rapidly once that phase begins.

388. **Damage Mechanisms**. A rocket under thrust, whether it is acting as a missile or a launcher, is inherently vulnerable if anything can be done to rupture the rocket casing. A ballistic missile warhead is relatively more robust, though depending on its range and velocity, it will be subjected to varying forces on re-entry to the atmosphere, where any damage may lead to its destruction. By comparison, a satellite is very fragile indeed. We have already looked implicitly at soft-kill or mission-kill options, but even where a hard kill is desired, relatively 'soft' interaction can be catastrophic. EO sensors on observation platforms can be dazzled either temporarily or permanently, communications links can be jammed and the potential of interfering with telemetry and control signals should not be overlooked. For kinetic impact, the fact that the satellite is moving at orbital velocity means that a relatively small object impacting the satellite can do extensive damage.[88] There is very limited scope to physically harden a satellite against substantial impacts due to payload-weight restrictions.

---

[86] We look at the different implications of direct-ascent and co-orbital ASAT weapons below at Paragraph 392.

[87] Space Debris is also discussed in Paragraph 113.

[88] This is just an extension of the debris hazard described in Chapter 1.

Ballistic Missile Defence

389. **BMD at Various Phases of Flight – Debris Implications**.

a. **Boost-phase Intercept**. If a missile is intercepted while its motor is still burning, by definition it will not have attained the speed required to reach its intended target. Also, the motor breaking up under power may well scatter the debris. How long the motor had been running will dictate where the debris ends up. If the intercept is at or immediately after launch, it may fall back straight onto the launch site.[89] If on the other hand intercept was just before the motor would have cut out anyway, it is probably outside the atmosphere, and the debris may well follow most of the intended flight path and hit the ground near the intended target. For any intermediate stage, the debris will probably fall somewhere along the intended flight path.

b. **Cruise-phase Intercept**. During the cruise-phase, any intercept will make some change to the missile's velocity, and may disable any guidance mechanism. Thus accuracy will degrade, but the degree of deflection will depend on exactly how the intercept took place.

c. **Re-entry Phase Intercept**. If a warhead is intercepted during re-entry, the aim must be to break it up and disrupt its intended function. Even if this is achieved, however, some debris (possibly substantial portions of it) will almost certainly impact in the target area.

390. **BMD at Various Phases of Flight – Practical Possibilities**.

a. **Boost Phase Interception**. There are clear tactical advantages to boost-phase intercept. One 'kill' takes out all the warheads carried on the missile, decoys are not a factor and the debris problem is minimised. The target is also at its most vulnerable and is clearly identifiable due to its exhaust. There are, however, major issues with access to that phase of flight – any air-breathing system may have to penetrate significant defences to gain access, and the timeline constraints outlined above are maximised; detection to execution must take place within the burn-time of the rocket motor.

b. **Cruise Phase Intercept**. During cruise, the target is outside the atmosphere, and is inherently more accessible due to its flight path. Decoys may well be a significant problem for the weapon system, however, debris will probably reach the intended target area and if missile-based interceptors are used, their flight time, including any ground reaction time prior to their launch, must be factored into the calculation.

c. **Re-Entry**. Intercepting a re-entering missile gives easiest access to it – by definition, intercept takes place in or adjacent to 'friendly' airspace. Warning and reaction time taken from initial launch detection is maximised, and decoys cease to be a factor as atmospheric re-entry proceeds. However, the target is travelling at close to its maximum speed and debris is an issue. In the case of theatre and short-range missiles, intercept at re-entry may be all that is possible.

---

[89] The USA provided a graphic illustration of this kind of failure on 6 December 1957. The first attempted launch of the four-stage Vanguard launcher (Vanguard TV-3) failed two seconds after ignition. The entire launcher exploded, consuming the launch pad too. When the dust settled (literally), the satellite payload was found at the edge of the launch site, still in its protective nosecone, successfully transmitting its radio signal, though unfortunately out of range of the planned receivers.

391. **Practical BMD – a Short History**. Countering ballistic missiles has been a military concern since the V2 campaign in WW2.

a. **Early efforts**. There was only one recorded instance of a V2 being intercepted in flight, when a USAAF B-24 bomber returning from a raid in Europe happened to overfly a launch site as a V2 was launched. One of the B-24 gunners engaged the missile during its initial climb with a 0.5” machine gun and destroyed it. There was research in the UK into the practicality of a large AAA barrage capability (essentially firing with minimal radar cueing following a launch detection) to counter incoming V2s. Operational analysis suggested that the collateral damage from shrapnel and unexploded shells would exceed that from the V2 and the project was abandoned. Unlike the V1, which was usually launched from a fixed site, the V2 mobile launcher proved an elusive target and there were no recorded instances of an operational launcher being interdicted in action. German tactical skill verified the utility of the mobile Transporter Erector Launcher combination. Among the effective V2 counters[90] were the bombing of factories and other production facilities, a dis-information campaign feeding erroneous impact reports back to the Germans to induce them to aim the missiles incorrectly, and most importantly, pushing the land front forward to render the intended V2 targets out of range. Acoustic launch detection using sensors along the UK South Coast was also used to provide early warning of likely impact in London.

b. **Cold War BMD**. In the earliest days of nuclear deterrence, both the USA and USSR explored potential counters to incoming missiles. Both faced enormous technical challenges in intercepting a high-speed, potentially exo-atmospheric target, and both sides resorted to nuclear tipped missiles as a result.[91] The USA developed a system called **Sentinel**, consisting of nuclear-armed **Spartan** and **Sprint** missiles to explode in the vicinity of incoming warheads, as a counter to fractional-orbital systems. Spartan, which was itself a development of the Nike[92] SAM system, was the primary interceptor, backed up by Sprint. The USSR developed a missile system called A-35, NATO code-name **GALOSH,** to counter western ICBM systems. This too was a large nuclear-tipped missile, similar in concept to Spartan. Both systems were pursued to completion, but were then constrained by arms-control negotiations. Both systems also faced considerable technical challenges posed by the development of decoy warheads and multiple warhead missiles as ICBMs matured. Ultimately, Sentinel technology was developed into a system known as **Safeguard**, but that was only briefly operational, and was then withdrawn. An updated GALOSH remains operational to this day to protect Moscow. The ABM Treaty of 1972 (as amended) constrained the USA and USSR to one operational protected site, but the USA publicly abrogated its ABM commitment in 2001 to allow development of its NMD concept.

---

[90] The integrated campaign to counter V1 and V2 weapons in Europe was known as Operation CROSSBOW.

[91] Certainly the US (the Douglas GENIE AAM) and probably the USSR had already developed nuclear-tipped AAMs to counter incoming manned nuclear bombers, so this was not the conceptual leap it might at first appear.

[92] The lineage and nomenclature of the Nike family of missiles is complex, and sources vary on precise details. Nike began development for the US Army in 1944 as a SAM to counter jet aircraft, but was ultimately used as a component of several systems.

Illustration 3.22 – Early US BMD Systems. SPRINT and SPARTAN missiles were developed during the Cold War. Both missiles utilised nuclear warheads, but had very different flight profiles. SPARTAN was a relatively large, 3-stage missile designed to counter incoming ICBMs with a high-altitude exo-atmospheric explosion. SPRINT was smaller and fielded a less powerful warhead, designed for atmospheric intercept. Because of the speed of the incoming targets, SPRINT had a sustained acceleration of over 100g during launch – the missile name was certainly apposite. The pictures show a Spartan missile mock-up being examined by technicians, a SPRINT missile being loaded into a silo, and a SPRINT test-firing. (Photos courtesy of US Army).

c. **The SDI or 'Star Wars'**. During the 1980s, more sophisticated ABM systems were contemplated in the USA. President Reagan's 1983 announcement of the Strategic Defense Initiative (abbreviated 'SDI' – but more commonly referred to as 'Star Wars') was a multi-faceted research programme exploring many potential technical solutions, incorporating elements of ASAT and BMD, and including ground-based, air-breathing and Space-based sub-systems. Ultimately, it was not implemented as intended, though elements of its technology survive.

d. **The Gulf War – SCUDs and Patriots**. The SCUD[93] was an early USSR attempt to develop V2 technology for post-WW2 use. Although operational derivatives have little physical similarity with a V2, there were, and are, shared technical details. The SCUD family was widely developed both within the USSR and indigenously by subsequent customers, and includes conventionally armed, nuclear and chemical variants. Iraq had significant experience of SCUD operations dating from the Iran-Iraq War, and during the first Gulf War deployed SCUDs principally to threaten Israel. The USA deployed Patriot SAMs to theatre as a counter, with software modifications to enable them to engage ballistic missile targets. Successful engagements occurred, although public controversy surrounds the overall success rate. Once again, mobile launchers proved elusive, with a very small number of successful attacks claimed, compared to the resources deployed.

e. **National Missile Defence and the Missile Defense Agency**. With the end of the Cold War, attention turned to missile and WMD proliferation and the USA withdrew from the ABM treaty to allow development of a limited but geographically extensive system to protect against 'rogue states'. The system is still being developed and deployed, but at the time of writing has achieved IOC. The overall concept is an enhanced network of search radars, coupled with extant launch-detection systems, and networks of engagement radars and conventionally-armed, dedicated missile interceptors. These are designed for cruise-phase intercept. The fixed radar installations can be supplanted by naval air-defence radars, and potentially by ship-launched enhanced surface to air missiles (SAMs) to give flexibility in deployment and coverage. An Air-borne Laser (ABL) capable of destroying a missile is also under development to increase the number of attack options within the system.[94] The entire system is managed from within the US by the Missile Defense Agency, which was established specifically for this purpose.

Anti-Satellite Systems

392. **ASAT Technology**. Interdiction of missile warheads is almost by definition aimed at a 'hard' kill – a system that left an intact, unexploded warhead somewhere near its intended target is not going to impress many users. At least partly because of orbital debris considerations, however, ASAT systems may well offer soft-kill options.

a. **Electronic Attack**. An ASAT system may simply jam the target satellite, either interfering with its control uplink or downlink, or in the case of a communications satellite its traffic. It might also be possible to disrupt or damage a satellite by use of directed-energy systems.

---

[93] Soviet names for the SCUD family included R-1, R-11, R-17 and R-300. There were also Soviet navalised versions (D-1 and SS-N-1), and local derivatives manufactured under numerous names.

[94] The ABL utilises a chemical laser employing iodine and oxygen as fuel, and producing a high-intensity IR laser beam. The beam is designed to rupture the casing of a rocket in flight. The complete system occupies all of a Boeing 747-400 airframe. As of April 2009, it appears possible that the program will be curtailed, continuing only as a research and development effort.

b. **Manoeuvre**. Any system that poses a credible threat to a satellite may force the operator to manoeuvre the intended target as a defensive countermeasure. This may frustrate its mission, constituting a 'soft-kill'.

c. **Damaging Sensors – Dazzle**. We have noted that EO sensors are inherently fragile and can be damaged either temporarily or permanently if a strong enough illuminating beam can be injected into the optical system. The fact that the optics are designed to focus incoming light/radiation onto the imaging chip or plate enhances this effect. If the satellite operator is aware of the threat, protection may be possible by closing the shutter on the imaging system or by manoeuvring the satellite attitude or track, but again this would constitute a soft kill.

d. **Destruction**.

(1) **Nuclear ASAT Weapons**. We noted above that nuclear-tipped weapons were initially attractive as a BMD mechanism due to their relatively large lethal radius and problems at the time with aiming at incoming missiles. A similar argument and the results observed with high-altitude nuclear burst experiments noted in Chapter 1, led to proposals to use exo-atmospheric nuclear weapons as ASAT systems. Aside from the obvious proliferation issues posed, additional issues of discrimination and collateral ground effects made them undesirable. The resulting systems were short-lived in the case of the USA (see Paragraph 391b). While there are no technical bars to their use, the other disadvantages are certainly serious, and possibly overwhelming. Their most destructive effect was through the energising of particles trapped in the Van Allen Belts, which in turn damaged electronic components on the victim satellites. They were thus also rather slow-acting and unpredictable in their effect, and were limited to those satellites transiting the energised regions, which were already recognised as hostile and unfavourable locations for satellite orbits.

(2) **Co-orbital weapons**. One possible kinetic kill platform is a satellite in a similar orbit to the target that manoeuvres close to it and then either hits it, scatters debris in its path or explodes.[95] Other kill mechanisms may also be possible. The USSR developed and tested an interceptor system[96] using a conventional fragmentation warhead during the 1960s and 1970s. It is believed that it was operational during 1973-83. It was intended to achieve co-orbit with a LEO target within two revolutions of the Earth, giving a total mission duration of the order of 2-4 hours. Within the geographic constraints of the launch site used (to intercept a target, the interceptor needs a launch site that will provide a launch-window into the target's orbit, as discussed in Paragraph 164f), the system was tested successfully. More recently, there has been theoretical discussion of the possibility of **parasitic satellites**, which might manoeuvre slowly and stealthily into the vicinity of a target satellite and explode or otherwise interfere with the target when commanded. The term **space mine** is also used in this context, although the name is slightly counter-intuitive, since conventional mines are static, waiting for a target to encounter them, rather than dynamically keeping station on their intended goal.

---

[95] The remark that 'any ship can be a minesweeper, once' has thus a corollary that 'any satellite can be an ASAT, once'.

[96] The system was initially known as the 'Istrebitel Sputnikov', literally 'destroyer of satellites' and was referred to by various 'IS' designations in open press. This was complicated by the multiple different Soviet launch vehicles used, although 'IS-2' appears to be the most common designation for the operational system.

(3) **Direct-ascent Weapons**. A direct ascent ASAT weapon is essentially a SAM extended into Space. Current examples overlap technically with extant SAMs, in the same way that the Patriot SAM system has been employed for BMD (see above). Design issues to be overcome include range and terminal homing. The potential target will by definition be at orbital altitude, so even if the flight path is minimised by careful intercept geometry, the total distance for the missile to travel is certainly over 100 miles. This places the likely class or size of SAM in context. The missile must also somehow sense and close with the target, either to impact, or at least to the lethal radius of its warhead/damage mechanism. This must be done between a target at orbital speed and a sub-orbital missile, so the challenges are considerable, although the Chinese ASAT test of 2007 and the US interception of their errant satellite in 2008, both described in Chapter 1, indicate that they are soluble.

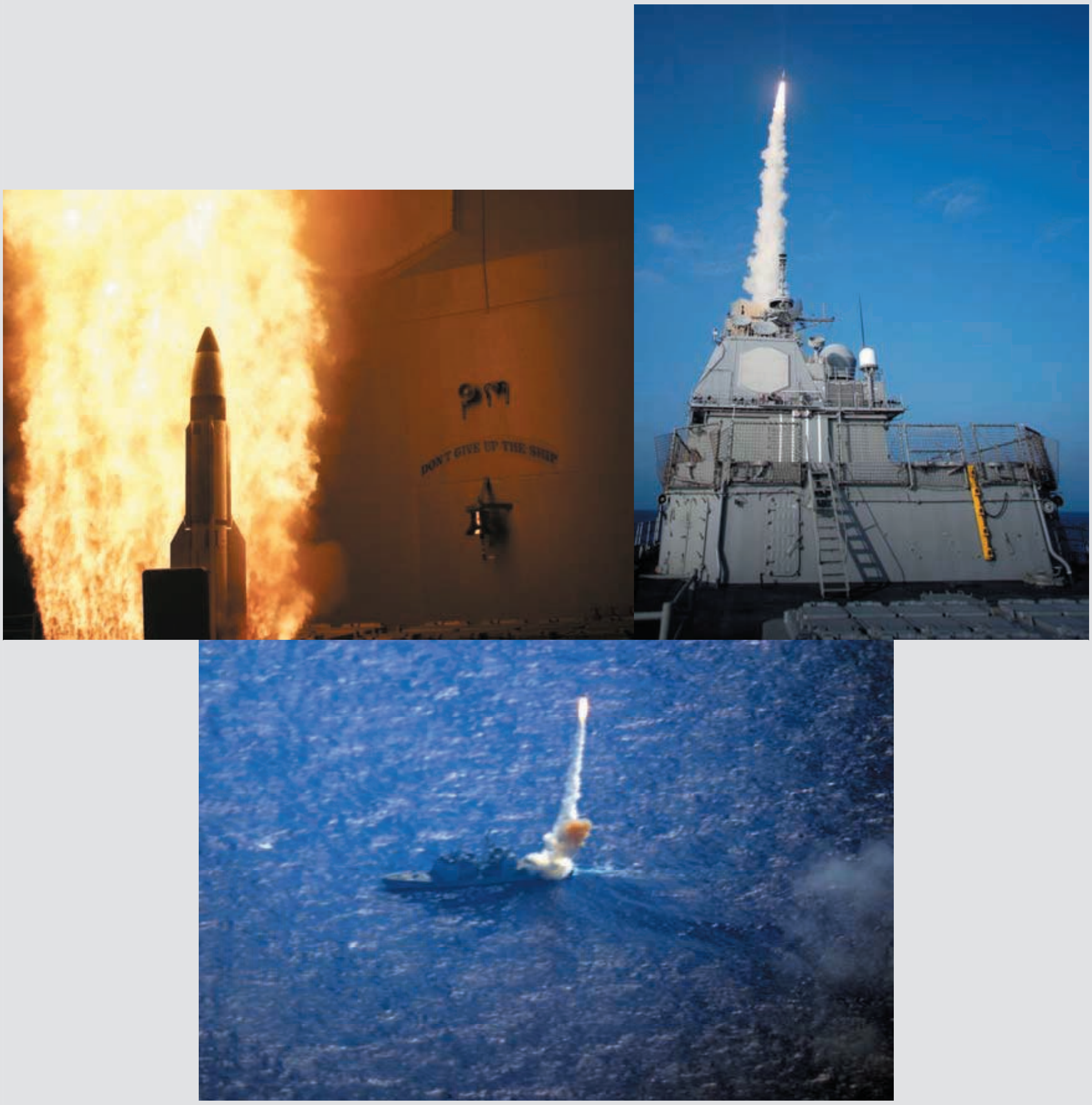


Illustration 3.23 – An ASAT System in Action. The pictures show the actual firing of the SM-3 missile from the USS LAKE ERIE that destroyed the errant US-193 payload on 20 February 2008. This, of course, was only one part of a much larger capability that provided space situational awareness and targeting information to enable the successful shot. (US Navy images)

(INTENTIONALLY BLANK)

CHAPTER 4 – SPACE IN MODERN SOCIETY

401. **Co-existence of Military and Other Space Users**. Thus far this Primer has looked at the theoretical, legal and practical aspects of military space operations. Since the dawn of the Space Age, however, military space has co-existed with the scientific and commercial space community. In this final Chapter we look at some of the consequences of this co-existence.

402. **An Outline of Chapter 4**. This Chapter first looks at how military space has been analysed and situated by strategic thinkers. Some have compared space to other domains, while others have attempted to analyse it as a new environment altogether. Both approaches will be examined briefly. This is a fast-growing area where little consensus has yet been reached. Nonetheless, we will try to establish at least some of the more common categorisations applied to space capability. Military space users are also becoming aware of their *dependence* on space, but precise quantification of the degree of dependency, which would greatly aid planning of mitigation strategies, is frustratingly difficult. The Primer will try, however, to provide some pointers to the implications of these failures. Just as was the case in the pioneer days of aviation, the military were early-adopters of the emerging technology, and their specific needs have driven research in particular directions. The size of the civil and commercial space community has grown however, as technology has matured, reflecting the different pressures that the civilian sector operates under. Some applications are purely commercial in intent, while others benefit the wider scientific and academic community. Both of these areas will be examined briefly, including the implications of manned space flight, and flight beyond Earth orbit. In closing, we conduct a brief survey of space doctrine, and of arrangements for governance and organisation of military and civilian space.

403. **Timeless and Perishable Information**. Most of the content of the Primer has been based either on generic or historical information, and is thus, essentially timeless. This Chapter differs; some of its content is likely to date quickly, so it should be read with caution. Some paragraphs are essentially open-ended, highlighting issues without providing resolution. Completeness still justifies their inclusion, but particularly in the final 2 sections, the aim is to provide a roadmap of organisations involved in space policy; they should then be consulted for the current situation in their area. There is a deliberate bias towards British and American organisations; this is not to downplay the varying European perspectives on space, or indeed of other world players such as China, Russia, India, Japan or Brazil. Rather it serves simply to emphasise the importance of the UK/US relationship. There is also only brief mention of the activities of the European Space Agency, which plays a key role in developing civil and commercial activity. The interested reader should, however, have no difficulty in finding reference material for most of these latter aspects.

404. **Terminology and Space Sectors**. US analysts are usually pedantic about the categorisation of space activity into *military*, *civil* and *commercial* sectors. Some sub-divide *military* into true military and intelligence sectors. The clearest distinction is between military/intelligence activity, and the civil and commercial sectors. *Civil* is taken to mean academic or research activity, undertaken on a non-commercial basis, and typified by missions such as the Hubble Space Telescope, the International Space Station (ISS) or planetary exploration. *Commercial* activity is anything profit making, such as commercial SATCOM, broadcasting or marketing imagery. The distinction is not always followed as closely in the UK, but it is probably useful, and is much more likely to be strictly observed in the US. We will follow it in this Chapter of the Primer.

SECTION I – THINKING STRATEGICALLY ABOUT SPACE

In this section you will find a description of:

- Sources for study of strategic space thought.
- Methods of comparing space with other domains.
- Some of the key early publications in the area.

405. **Sources**. Attempting to capture consensus on a developing topic is hazardous at best, and references in this section should be seen in that light; new books are published regularly, and authors come and go. Nonetheless, the reader may find some foundational texts useful. They are presented with the minimum of analysis necessary to situate the writers. Additionally, students at the United States Air Force (USAF) Air University at Maxwell Air Force Base (AFB), and specifically those attending the School of Advanced Air and Space Power Studies, frequently produce space-related theses. In accordance with US Government policy, these are both unclassified and publicly accessible. They are usually placed in the public domain on the Internet for consultation and found via common search-engines. While their quality and accuracy cannot be guaranteed, they invariably contain detailed bibliographies. Those seeking to capture the latest thinking in the field may find that these provide useful onward references.

406. **Comparisons to Other Domains**. Attempts to analyse space strategically have frequently focussed on adapting strategic thought from other domains. Maritime and Air environments have proved the most popular sources, although no school of thought has achieved dominance. In his book *Astropolitik*,[1] Professor Everett Dolman takes classical terrestrial Geopolitical thinking, as typified by Halford McKinder and Nicholas Spyekman, and adapts it to the constraints of space. Air power advocates note the contribution of General Thomas D White, Chief of Staff of the USAF in 1958, for first coining the term *aerospace*.[2] Acceptance of this premise was promoted by the commonality between applications in air and space during the early years of the space-age, specifically in reconnaissance. Others seeking to generalise air power theories to apply to space cite Douhet's observations on the ubiquity of air, supplanted by General 'Billy' Mitchell's similar views. Maritime strategists note the principles outlined in Chapter 2 regarding the legal foundations of overflight, and similarities between useful orbits and sea lines of communication.[3] Finally, it should be noted that there is also a school of thought that rejects analogies as an analytic tool, holding that they distort more than they clarify, and that space should therefore be analysed in much greater isolation. Plainly this debate will run for a long time.

407. **Pure Space Analysis**. Perhaps the most commonly encountered *pure space* discussions are those of Jim Oberg,[4] and of Lieutenant Colonel David Lupton.[5] Lupton's analysis is older, and is more dependent on Cold War paradigms, but still contains much to ponder. He divides space power theories into four families: sanctuary thinking, survivability/vulnerability thinking, high-ground theories and the space control school, and devotes a chapter to analysing each in depth. These

---

[1] Everett C Dolman, *Astropolitik: Classical Geopolitics in the Space Age,* (London: Frank Cass, 2002).

[2] General White did not publish formally on this topic, but references to his remarks may be found in contemporary journals, and he focussed attention on the attempts alluded to in Chapter 1 to define a legal boundary between *air* and *space*.

[3] See for example John J Klein, *Space Warfare: Strategy, Principles and Policy,* (London: Routledge 2006), where the author compares the maritime doctrines of Alfred T Mahan and Sir Julian Corbett and considers their application to space.

[4] James E Oberg, *Space Power Theory,* (USAF Academy: 1999). This work is out of print, but the text is in the public domain on the internet, where it can be downloaded freely. Portions of it form a short primer on orbitology, but Chapters 2, 4 and 6 in particular deal with Space Power theory.

[5] Lieutenant Colonel David E Lupton, *On Space Warfare: A Space Power Doctrine,* (Maxwell AFB: Air University Press 1988). Out of print, but again in the public domain and available free on the internet.

categories are still used. Oberg’s work owes more to contemporary US doctrinal debates and the existence at the time of writing of US Space Command. This remains a topic of considerable interest to strategic theorists, and a steady stream of new publications in the field can safely be predicted.

SECTION II – DEPENDENCIES ON SPACE

In this section you will find a description of:

- The extent of modern dependence on space-based systems.
- Acute and persistent failures of space-based systems.

408. **Space-enabled Systems**. When military space capabilities were first brought into service, any dependencies they created were obvious. In many cases the mitigation strategy was easy to identify too. Reconnaissance could revert to air-breathing platforms, communications to land-lines or High Frequency (HF) radio. More recently, the spread of space-enabled capability, and its increasing transparency to the end user, has 2 main impacts.

a. Firstly, realisation of where the dependencies lie becomes increasingly difficult. High-capacity SATCOM, with large dish antennae outside the headquarters, is easy to recognise as a space-enabled capability, but an Iridium mobile phone looks much like a terrestrial one, and a casual user may not realise that the Iridium system is implicitly reliant on satellites. Even a sophisticated user may struggle to establish whether a terrestrial cellphone network is in fact also space-dependent via Global Positioning System (GPS) timing received at the local network stations (answer: it probably is).

b. Secondly, as dependence has grown, the practicalities of some mitigation strategies need to be considered. Robust replacement capability may be identifiable, but if it is of lower capacity or less responsive, the effect of the loss needs to be quantified. Modern concepts such as *reachback* may be so reliant on high-bandwidth connectivity that they become impractical in its absence. Whether it is therefore essential to train and practise for their absence is a judgement call for the commander.

409. **Impacts of Failures**:

a. **Acute Implications**. Failure of a space-based system can be sudden or gradual, and predicted or unexpected. In Chapter 3 it was implied that it may be impossible to attribute quickly a cause to a failure, for example to differentiate technical failure, debris impact or malicious action. The situation is even more complex than this, however. For example, at switch-on, a GPS receiver calculates which satellites ought to be observable from its estimated position and looks for the relevant signals. It may not, however, receive all of them. Possibly the satellite has failed abruptly, but far more likely is that terrain or some other obstacle is blocking that particular signal at that instant. No analysis is necessary; the receiver simply looks for an alternative signal and displays the derived position. Since the GPS constellation design usually ensures that more than the minimum number of satellites is observable at any time, the transient event passes unnoticed. A more dangerous failure would be an errant navigation satellite broadcasting incorrect information. This would ultimately be detected, either by the receiver being unable to reconcile its signal with others, or by the Ground Control Centre detecting the failure and broadcasting the error via the remaining satellites. The possibility of that sort of failure, and the fact that it could temporarily introduce position errors, provides the basis for decisions about whether to allow aircraft to make instrument approaches based solely on SATNAV information, and to what minima. The important lesson regarding the impact of failure is analysis of the likelihood of the failure and the degree of resilience in the affected system.

b. **Persistent Implications**. A transient satellite failure could have serious implications for individual users, but persistent failures have different characteristics altogether. An individual sending a commercial email for example usually has no way of knowing the route(s) taken between sender and recipient. If a satellite link in the chain failed, the world-wide-web would find another route, and the whole system is constructed to minimise the chance of a critical node even existing. Overall, however, failure of a major trunk system such as a high-capacity satellite or a major undersea cable[6] would slow the system locally or regionally. Military systems, using dedicated hardware, could encounter acute effects if a key component failed. The persistent effects would include an overall loss of capacity, the cost of contracting commercial alternatives and possibly loss of special characteristics of the military hardware, such as anti-jam receivers. Some constellation-based systems may suffer local loss of capacity if the constellation was heavily depleted, but the failures would probably at least be predictable in terms of time and space until repairs or replacements could be arranged. For critical national applications, however, such as missile warning, even transient failures may be totally unacceptable, particularly if an adversary knew details of what had failed. Mitigation of such failures is not discussed in this Primer.

SECTION III – COMMERCIAL USES OF SPACE

In this section you will find a description of:

- The adaption of military space capability to the commercial sector.
- Uniquely civil applications and factors affecting commercial exploitation of space.

410. **Exploitation of Military Applications**. Many of the early military space capabilities proved attractive to civilian users too. The rate of adaptation varied according to cost and classification, amongst other factors, but there are read-across applications to match most military uses. In some cases the military space product has direct application to civilian uses, in others, demand from the civil user community has driven research and development in subtly different directions.

411. **Surveillance – Survey, Earth-Resources, Meteorological Data**. Surveillance of the Earth from space was the earliest military driver for space access; reconnaissance satellites were being planned before rockets were available to launch them. The performance of these satellites was, however, highly classified. Even the existence of the National Reconnaissance Office, which handled satellite imagery in the US, was not publicly acknowledged until 1992. As sensor technology improved, however, the details of legacy systems have migrated into the public domain, and some of the applications have proved particularly useful to civil customers. Examples include civil use of narrow-band imagery for disaster relief, monitoring Earth resources such as agricultural production and mineral extraction, and imagery being used as a basis for mapping, planning and other geographical applications. Civil users have also benefited greatly from the improved accuracy of weather forecasting possible through the use of space-derived data.

---

[6] A series of undersea cable failures in the Mediterranean and Arabian Seas in 2008-9 demonstrated exactly such impacts.

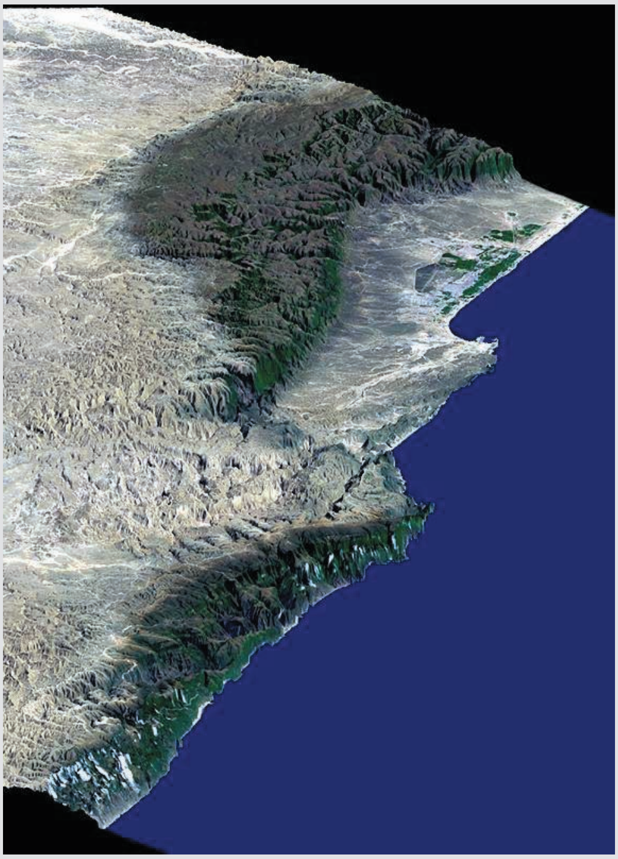


Illustration 4.1 – Earth Resources Monitoring. This image was derived from a combination of 2 space-derived datasets, gathered in 1999 and 2000. The terrain model showing the relief features came from the Space Shuttle Radar Topography mission, which generated a 3-D Earth model using radar measurements (actually the same transmitter as was used for the illustrative SAR image in Illustration 3.17, though on a different mission). The colour data was then supplied by the NASA *Landsat* satellite, which was overlaid on the terrain model. The terrain perspective was deliberately exaggerated by a factor of 3 in this image for clarity, which was used to track habitation, cultivation and the limits of desert. The target area is Salalah and the surrounding mountains and coastline in Oman. (NASA JPL Image)

Illustration 4.2 – Disaster Relief from Space. The images show New Orleans before and after Hurricane Katrina, as seen by NASA's *Terra* satellite. The top image from 2000 (a composite from April and September) provides a reference. The bottom image, taken on 13 September 2005, 17 days after Hurricane Katrina struck, provides clear indications of the extent of the flooding as well as detail of inshore waters and shoreline hazards. Lakefront Airport is visible to the right, New Orleans City Park is centre/left, with Lake Pontchartrain at the top. (NASA JPL Image)

412. **Position, Navigation and Timing**. Although the drivers for the development of satellite navigation were military, the decision to make the NAVSTAR/GPS signal format public allowed the civil sector to exploit it widely. In exactly the same way as the military uses have expanded beyond those originally envisaged, so have the civil ones. Additionally, just like the military, civil use of the precise timing capability delivered by GPS has become widespread. GPS-based navigation has become ubiquitous, with users varying from hill-walkers through truck drivers to airline pilots and super-tanker captains. Civil expertise in using GPS for asset tracking probably exceeds the military's; timing applications include synchronising output and/or delivery on pipeline complexes and power grids, and time-stamping financial transactions.

413. **Broadcasting and Communications**:

a. In terms of turnover, communications in all their formats are the most lucrative commercial space application. We have already noted some of the technical implications of this: development of heavy launchers suitable for placing large payloads in GEO, and of large capacity transponders and complex beam-shaping antennae on satellites.

b. Broadcasting to domestic customers is a uniquely civil application, but again has given rise to interesting developments which may spill back one day into military usage. For example, in the US, a company operating a satellite radio service, offers a range of music, sport, news and other entertainment channels to predominantly ground-based users. Alongside this, however, on another channel it broadcasts a continuous weather radar map of the continental US. Light aircraft owners can specify installation of a receiver not only to provide entertainment to passengers, but also to display the weather radar information in real-time in the cockpit, integrated with light-weight, modern, glass-cockpit style avionics displays. They thus gain some of the advantages of onboard weather-radar without the cost, weight and complexity of installing a traditional weather-radar system.[7] The system also carries weather-related text. A similar service carrying slightly different information operates for mariners.

c. True communications satellites for point-to-point communication, whether carrying voice, data, video or other signals, operate in an area where civil and military systems can be virtually indistinguishable. Military users make extensive use of commercial SATCOM links to supplant dedicated military systems,[8] and as noted in Chapter 3, in the UK the commercial operator of the SKYNET series of satellites has the option of selling spare capacity on the system to other users.

414. **Space Tourism**. Although currently an outlet for the super-wealthy, space tourism is nonetheless a practical possibility. Leisure visitors have made short visits to the International Space Station, and the Virgin Galactic sub-orbital passenger flight project has successfully achieved proof-of-concept flight and is (in July 2009) approaching first flight of their commercial platform.[9] Other competing projects will doubtless emerge. Although strictly a civilian operation for now, such (relatively) affordable space-access opportunities may in time yield military applications.

415. **Space Resources – Mining and Minerals**. As our understanding of the composition of the solar system has improved, some commentators have noted the existence of potentially valuable resources on other bodies. These include minerals that would be marketable back on Earth, as well as sources of fuel and/or water that may be critical for interplanetary travel and exploration.[10] Of these possibilities, the latter seems more likely to be exploited first, pending any revolutionary reduction in the cost of space flight. There would, however, be substantial questions of international law raised in reconciling commercial exploitation of other bodies with the denial of territorial claims enshrined in the Outer Space Treaty.

---

[7] Because the system displays information generated from ground-based radar transmitters, it does not offer complete weather-radar functionality. For example, there is no equivalent of antenna-tilt control to discriminate clouds vertically.

[8] UK policy is to make maximum possible use of SKYNET, and to minimise commercial usage. However, this is not necessarily the case in other nations.

[9] See paragraph 147 for a discussion of the Virgin Galactic hybrid propulsion system.

[10] The advantages in terms of payload saved would be substantial if explorers did not need to carry water supplies for an extended mission or perhaps fuel for the return journey.

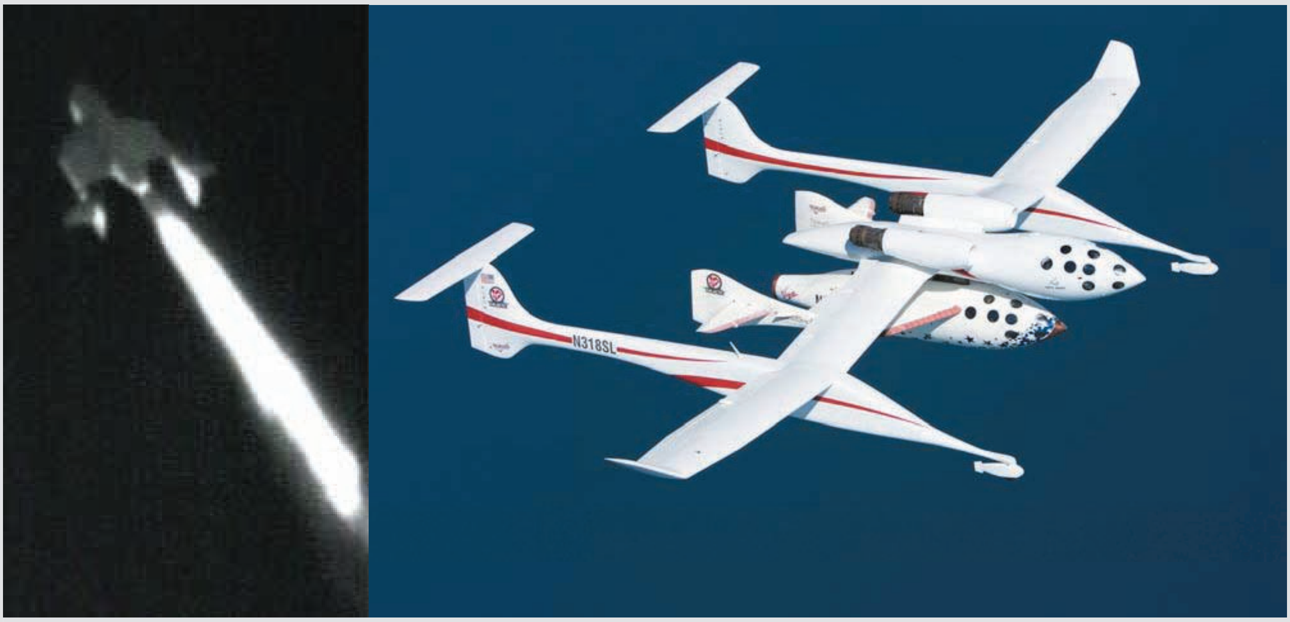


Illustration 4.3 – Space Tourism. The Virgin Galactic company intend to offer private space travel on a commercial basis. The images show (left) *SpaceShipOne* being carried aloft by the *WhiteKnightOne* launch platform and (right) *SpaceShipOne* in sub-orbital flight (travelling at about 2500 mph when this photo was taken). This combination won the Ansari-X prize for private spaceflight in October 2004, and forms the basis for development of the operational commercial vehicle. Other potential uses of such vehicles, such as scientific research and small satellite launch are being considered. (Image courtesy of Virgin Galactic LLC).

416. **It's Life, Jim….** Speculation about extra-terrestrial life pre-dates the Space Age by a long way, featuring in literature from the 19th Century and before. At a more practical level, planetary probes have sought for signs of life within the Solar System, and the Search for Extra-Terrestrial Intelligence (SETI) program looks for communications from intelligent sources beyond it, so far without conclusive results in either case. Any such discovery would of course have wide-ranging impacts; even the presence of bacteria on a planetary surface would confirm that life could exist elsewhere, as well as potentially constraining exploitation and exploration of planetary resources. The reader is free to draw his or her own conclusions on how mankind would react

SECTION IV – SCIENTIFIC USES OF SPACE

In this section you will find a description of:

- Pure and applied science experiments and studies enabled by access to space.
- The implications of spaceflight beyond earth orbit, including potential extension or read-across to military roles where relevant.
- Manned space-flight as it may relate to military activity.

417. **Pure Science**. The USA's exploitation of German rocket equipment and expertise at the end of World War II is relatively well known. Having captured the main V2 production facility in the Harz Mountains, they recovered enough materiel to construct and launch 68 V2s from the White Sands missile range in New Mexico between 1946 and 1952. Less well known is that they simultaneously convened the V2 Upper Atmosphere Research Panel. This body coordinated the

selection of payloads in the top of the test launches in lieu of the warhead.[11] Among the experiments conducted were the return of the first photographs of Earth from space (see Chapter 3, Illustration 3.13 for examples), exploration of the composition, temperature and pressure of the upper atmosphere, measurement of solar and cosmic radiation, including rudimentary X-ray astronomy, and a series of biological experiments. Since then, many fields of the physical sciences have benefited from space-based research.[12] Astronomy, physics, chemistry and other areas have all gained from access to space, which has enabled observations of many kinds free from the shielding effects of the Earth's atmosphere, as well as experiments conducted in weightless conditions.

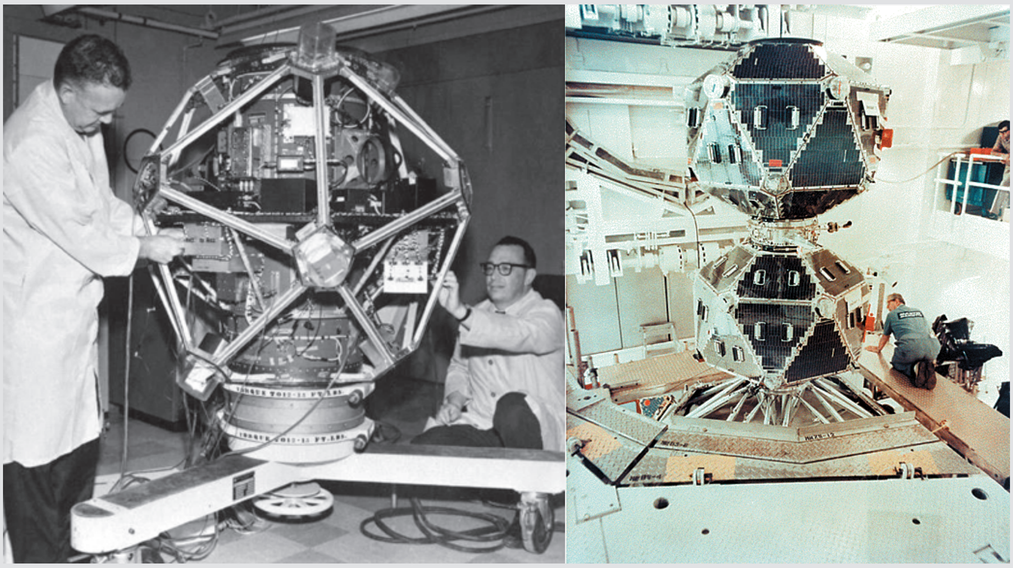


Illustration 4.4 – Fortuitous Cross-over from Military to Pure Science Applications. The pictures show VELA satellites intended for terrestrial nuclear detonation detection under construction (left), and a pair being mated together in a clean room prior to launch (right). VELA payloads were launched in pairs which then separated in orbit and were manoeuvred to their operating locations. VELA payloads detected combinations of x-rays, gamma rays, neutrons and (later versions) visual events. The gamma ray detectors, which yielded the early evidence of gamma ray bursts (see Paragraphs 364b and 418 above) as a by-product of the primary mission, were mounted internally alongside the neutron detectors; the highly-energetic neutrons and gamma rays were able to penetrate the structure. The outer triangular panels are solar arrays for power and the devices at the corners of the triangles are the x-ray detectors. (NASA images).

418. **Applied Science**. Space-based observation has had direct influence on a variety of applied sciences. Studies of possible global warming and climate change have benefited greatly from space-derived data and improved understanding of the links between solar radiation and the Earth's climate. Earth sciences make direct use of many different observations of climate, composition and distribution of surface features and characteristics. In the life sciences and in engineering, claims are also regularly made of the potential benefits of chemical purification of materials including drugs and semi-conductors that could be conducted in space. Verifiable commercial exploitation of these principles remains frustratingly elusive at present. There have also been a variety of claims of the benefits that have accrued from 'spin-off' products of space exploration. The demands of space exploration were undoubtedly a principal driver for the development of computing power in the 1960s.

419. **Interplanetary and Deep Space Exploration**. Almost all of this Primer has considered operations in Earth orbit, but the scientific (and national prestige) value of reaching beyond the

---

[11] The rockets would probably have needed to carry ballast anyway to ensure consistent dynamics if science payloads had not been available.

[12] See Paragraph 364b for description of a fortuitous scientific result gleaned by a military system.

Earth's immediate environs was obvious early in the space age. By 1959, the USSR had obtained pictures of the far side of the Moon from the *Luna 3* probe.[13] Since then, most of the major bodies in the solar system have been explored to some extent and the oldest probes are now leaving the solar system. There are particular technical challenges associated with interplanetary travel for exploratory probes, and although there may seem to be few areas of direct relevance to military space operations, there are some facets that may be useful in the future.

a. **Dynamics and Trajectories**. Deep-space probes move far beyond the local influence of the Earth's gravity, and depending on their exact position, their motion may be influenced by any of the major bodies in the Solar System – often by different bodies at different stages of their mission. In the same way as we needed to accelerate a body to a speed of around 18 000 mph for it to orbit the Earth, there is a need for greater speeds to escape from Earth orbit and to move in predictable ways between the planets. A speed of about 24 000 mph is sufficient to escape from the Earth's gravity, but greater speeds still are required for practical journeys between the planets.[14] A probe intended to enter orbit around another planet, rather than simply fly past it, must also lose speed before it can do so, and a mechanism for this must be provided. One solution would be to use very large rockets for boost and deceleration, but other techniques have been developed. Among these is the *slingshot*. This is a close passage around a planet (usually not the intended destination). The probe is aimed precisely just to miss the planet, and to take advantage of acceleration towards it and the change of direction due to gravity as it passes around it. If the trajectory is aimed precisely enough, the probe ends up travelling faster than before in the desired direction, without expending any of its own fuel. It is not, however, achieving 'something for nothing'. Momentum is conserved during the process, and the probe has traded momentum with the planet. Momentum is the product of mass and velocity, and the probe exploits the advantage of small mass and a large change in velocity, against the planet's enormous mass and an undetectably small change in its velocity. The drawback may be the need to take a circuitous route to the desired destination. If there were ever a military deep space capability, the implied time penalty may have serious implications for the user.

b. **Deep Space Power Supplies**. Probes venturing beyond the Earth towards the outer Solar System are moving away from the Sun, and the power output that can be harvested from solar panels diminishes as this distance increases. Additionally, the efficiency of a solar panel diminishes with time, which over the duration of an interplanetary mission may be a significant factor. Alternatives include **fuel-cells**, which generate electricity directly by reacting hydrogen and oxygen in very controlled conditions, and **radio-thermal generators** (RTGs), which use the heat generated by radio-active decay to produce electricity. Fuel cells are attractive for manned space flight, since they also produce water as a by-product, but they are limited by fuel capacity for long duration missions. RTGs raise concerns about safety and possible pollution following a launch mishap, and for some science applications the heat and residual radiation they generate is undesirable. They are still the longest lasting source of electrical power and this makes them indispensable at present for the longest duration missions. During the Cold War, RTGs and even systems resembling conventional

---

[13] Luna 3 used a very unusual trajectory and some innovative technology. Rather than leave Earth orbit and enter into orbit around the Moon, it was launched into a very highly elliptical orbit around the Earth, which at apogee took it out beyond the distance to the Moon, so that first time around it could look back at the far side of the Moon. It took pictures using wet film, which were processed onboard, scanned, and subsequently transmitted back to Earth by an onboard television camera. One complete orbit took about 14 days, though after the first time around the orbit, the Moon was not in the correct position to be observed. The communication link with the ground station proved to be unreliable, and the UK Radio Telescope at Joderell Bank was used to help the USSR recover data from the probe.

[14] See Paragraph 123.

nuclear reactors were deployed in Earth orbit for critical applications requiring high-capacity power supplies.[15]

c. **Autonomy and Remote Operations**. In Chapter 3 we noted that the time it takes a communications signal to travel up and down to a satellite in GEO may be noticeable in conversation in some circumstances. Where multiple 'bounces' are involved, or for the most critical applications, this latency may also impose limitations on what can be done over a SATCOM link. Once you move beyond Earth orbit, latency in communication is an unavoidable consequence of distance. Probes exploring distant bodies need the ability to store instructions for future actions, and possibly to decide when and how to execute them without reference to the Earth – essentially they need a considerable degree of autonomy. While this does not imply military exploration of distant bodies, the techniques of autonomy, developing goal-seeking systems and related applications may have direct read-across to military automated vehicles and systems.

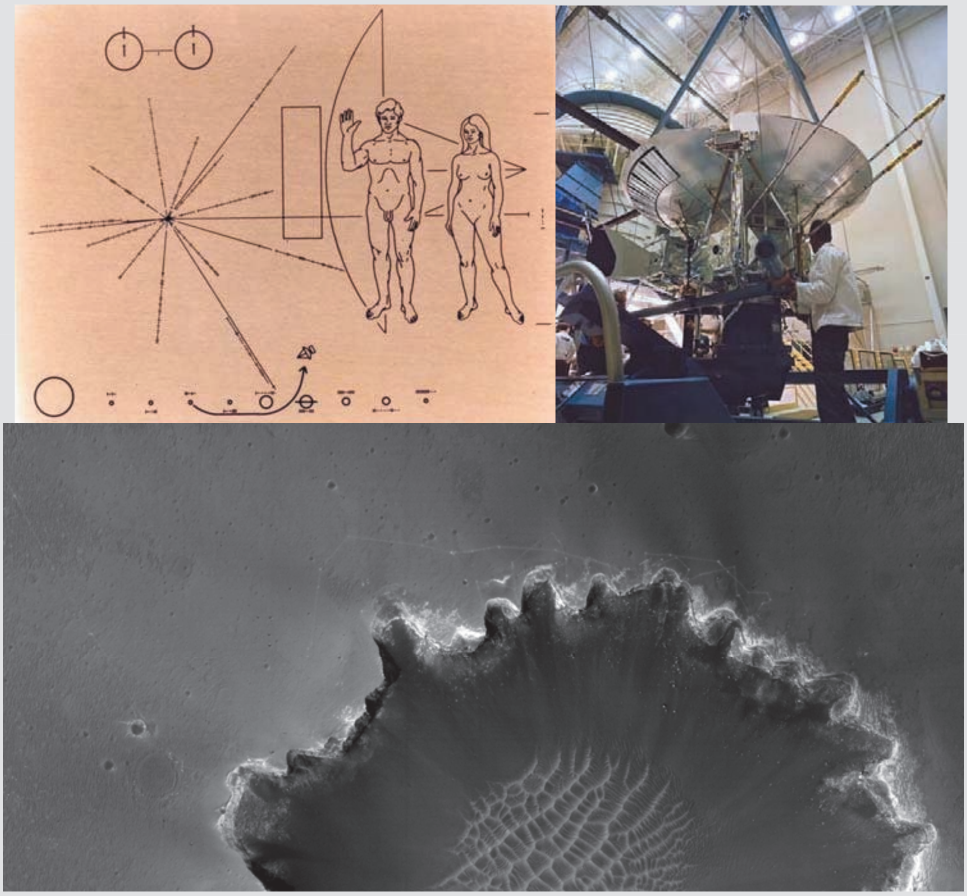


Illustration 4.5 – Spaceflight beyond Earth Orbit. The top right picture shows the Pioneer 10 interplanetary spacecraft under construction. Its flight took it from the Earth to Jupiter, then out of the Solar System – the first spacecraft to do so. Top left is an enlargement of the plaque it carried, against the possibility of it ever being discovered by extra-terrestrial life. The bottom portion of the plaque depicts the flight path from Earth, around Jupiter then out of the Solar System. Contact was maintained with the craft for over 30 years until its onboard power supply depleted. The bottom picture depicts a measure of autonomy in exploration. The image was taken by NASA's Mars Reconnaissance Orbiter mission, and shows a crater on the Martian surface. Clearly visible above the crater are tracks made by NASA's *Opportunity* Mars Exploration Rover in its (semi-autonomous) travels across the surface. (NASA Headquarters and JPL images).

---

[15] See also Paragraph 180. The Cosmos 954 incident alluded to in that paragraph led to an application of the Liability Convention, which is discussed in Paragraph 205.

420. **Manned Spaceflight**. Despite the fact that all the early astronauts, both American and Soviet, were serving military officers with a background in military test-flying, and despite the overt military support for all stages of manned space missions, the programmes they undertook were still implicitly civilian.[16] As space flight matured, however, both the US and the USSR conducted dedicated military manned missions. These included US military manned Space Shuttle flights to launch classified payloads, and perhaps military manned reconnaissance aspects of the Soviet 'Salyut' Space Station series of missions. These missions have been the exception rather than the rule, however. Manned space flight has its own particular constraints as well as advantages, and these can hamper military applications.

a. **Reliability**. Space vehicles rated for manned missions need to be built to higher standards of reliability than unmanned ones, and require launchers with similarly verifiable levels of reliability. For example, on Space Shuttle missions, anything carried onboard as a payload must also be 'man-rated' in terms of reliability and demonstrable safety for all anticipated or contingent stages of the mission, such as safety for an emergency landing with the payload still aboard. This constrains payload design.

b. **Payload Constraints**. There is a payload constraint associated with manned space flight *per se*. The human crew requires life support systems to provide habitable accommodation throughout the flight and safe re-entry at the end. This includes air-conditioning, pressurisation, launch and recovery constraints on allowable acceleration and provision of food and water. All this comes at a price in payload mass for a given size of craft. The International Space Station (ISS), and its various US and Russian predecessors, prove that relatively long duration space missions are possible for a human crew, but only at considerable physiological penalty for the astronauts.

c. **Accessible Orbits**. Apart from the relatively short-duration Apollo missions to the Moon, human space flight has been constrained to low-earth orbit.[17] Radiation shielding constraints would make sustained occupation of medium-earth orbit by a manned craft problematic, and while the problems could probably be solved, we have already noted that there would be a considerable fuel penalty in de-orbiting a geo-stationary satellite, particularly if a prompt return was required, as well as the need to develop a man-rated launcher capable of lifting a heavy manned capsule to a high orbit.[18]

d. **Endurance**. Although the ISS and its precursor Space Stations have proven the possibility of extended human spaceflight, which may have great applicability to future manned exploration of the Moon and planets, they have only done so at great cost and with demonstrable, albeit quantifiable, risk. For military applications, the persistent long-endurance capability typical of unmanned orbiting systems has so far proven more attractive.

---

[16] Statistically, the USAF has provided most US astronauts, but Alan Shepherd, the first US (sub-orbital) astronaut was a US Navy pilot. John Glenn, the first American to orbit the Earth was a US Marine Corps pilot, and Neil Armstrong, the first astronaut to walk on the Moon was another former US Navy pilot. US Army astronauts have flown on Space Shuttle missions. The first non-military astronaut at the time of their flight was probably Valentina Tereshkova from the USSR in 1963, who was also the first female astronaut.

[17] See Paragraphs 106 *et seq*, and Paragraph 146.

[18] See Paragraphs 173-6.

Feature Index

	Page		Page
Abby	14	Editorials	4
Amusements	6	Sports	22
Comics	25	Society	5
Crossword	28	Want Ads	26
Jumble	14	Radio-TV	28

28 PAGES TODAY

The Huntsville Times

Where Progress... Covers The Valley!

VOL. 51, NO. 21 CHICAGO DAILY NEWS SERVICE HUNTSVILLE, ALABAMA, WEDNESDAY, APR. 12, 1961 ASSOCIATED PRESS — WIREPHOTO 45c PER WEEK

Man Enters Space

'So Close, Yet So Far,' Sighs Cape

U.S. Had Hoped For Own Launch

CAPE CANAVERAL, Fla. (AP) — The Redstone rocket which the United States had hoped would boost the first man into space stands on a launching pad here. The Soviet Union beat its firing date by at least two weeks.

"So close, yet so far," commented a technician who is helping groom the Redstone to send one of America's astronauts on a short sub-orbital flight, hopefully late this month or early in May.

"If we hadn't had those troubles last fall and on the chimp and Little Joe shots this year, we might have made it," the technician said.

"But you have to give the Russian scientists credit. They've accomplished a remarkable breakthrough."

Dr. Hugh Dryden, deputy director of the National Aeronautics and Space Administration, told the House Space Committee in Washington Tuesday that the earliest possible date for the manned launching is about April 28.

Project Mercury officials had hoped to achieve a manned Redstone flight last December or January. A series of launch mishaps necessitated additional launchings to qualify the system.

On Nov. 8, a space capsule failed to separate from a Little Joe rocket fired from Wallops Island, Va., in a test of the escape system.

Two weeks later, a Redstone fizzled because of a faulty connection which caused the escape tower to fire, leaving the rocket

Hobbs Admits 1944 Slaying

By BOB WARD
Of The Times Staff

Isham D. Hobbs confessed today to the brutal murder in 1944 of Mrs. Margaret Thornton Fleming, Circuit Solicitor Macon L. Weaver said.

Hobbs, now 43, is held by Air Force authorities at Eglin Air Force Base, Fla. He signed a statement there detailing his knife-slaying of the prominent 33-year-old widow, Weaver said he learned from Air Force officials.

* * *

The suspect, who has undergone psychiatric treatment by military authorities since Feb. 6, has recovered his memory in full, Weaver was told. Psychiatrists said Hobbs' apparent amnesia re-

This is Russian Maj. Yuri Gagarin, history's first man in space. The Russians today rocketed him around the earth in an orbit taking slightly less than 90 minutes and brought him back safely to a prearranged spot in the Soviet Union. (AP Wirephoto via radio from Moscow)

Praise Is Heaped On Major Gagarin

'Worker'

First Man To Enter Space Is 27, Married, Father Of Two

Soviet Officer Orbits Globe In 5-Ton Ship

Maximum Height Reached Reported As 188 Miles

MOSCOW (AP)—A Soviet astronaut has orbited the globe for more than an hour and returned safely to receive the plaudits of scientists and political leaders alike. Soviet announcement of the feat brought praise from President Kennedy and U. S. space experts left behind in the contest to put the first man into successful space flight.

By the Soviet account, Maj. Yuri Alekseyvich Gargarin, rode a five-ton spaceship once around the earth in an orbit taking an hour and 20 minutes. He was in the air a total of an hour and 48 minutes.

The whole sequence of events and the announcements relating to it raised a number of questions. The Soviet announcement said the flight took place today between 9:07 and 10:55 a.m., but some persons in Moscow's Western colony were skeptical that the feat actually came off today.

There was a curious sequence of events leading up to the announcement.

Rumors had been circulating several days that the space coup had been pulled off. Two days ago, Soviet TV technicians moved into the Central Telegraph Office with the evident purpose of getting pictures of correspondents in action as they reported such a story. There were various reports, none verifiable from official sources, that the flight had been made.

* * *

Then Tuesday night the Daily Worker, London Communist newspaper with apparently

VON BRAUN'S REACTION:

'To Keep Up, U. S. A. Must Run Like Hell'

Illustration 4.6 – Faces From Manned Spaceflight. Three pioneers of manned spaceflight. At the top, the *Huntsville Times* (From Huntsville, Alabama, home of NASA's Marshall Space Flight Center) reports Yuri Gagarin's first manned spaceflight on 11 April 1961. Bottom left, John Glenn was the first US astronaut to conduct an orbital flight on 20 February 1962. On 29 October 1998, he became the oldest astronaut to date when he flew in the space shuttle *Discovery* On Flight STS-95 at the age of 77 (Pictured centre with the remainder of the STS-95 crew). Valentina Tereshkova was the first woman in space on 16 June 1963. Her achievement was commemorated widely including by the 6 Kopec stamp shown bottom right. (Gagarin and Glenn images courtesy of NASA-MSFC.)

SECTION V – MILITARY SPACE DOCTRINE AND ORGANISATION

In this section you will find a description of:

- How the UK exercises control and oversight of military and related security space activity.
- How the US military is configured for space operations.
- Recent reviews of us military space capability.

421. **UK Military Space Organisation**. Staff responsibilities within the Ministry of Defence (MOD) change regularly, so this section, which is correct as of January 2010, should be read with caution. Nevertheless, it may provide initial points of contact for specific queries.

a. **MOD Central Staff**. The Assistant Chief of the Air Staff (ACAS) has overall responsibility for Space in the UK MOD. There is a dedicated space desk within the Air Staff. Targeting and Information Operations (TIO) have responsibility for policy relating to Space Control. The senior leadership forum within MOD for space policy is the 2* Space Management Group (SMG), chaired by ACAS. Under this is a Space Working Group (SWG) chaired by Head Joint Capability.

b. **Headquarters Air Command**. Headquarters Air Command (HQ Air) operates the UK's space surveillance radar at RAF Fylingdales on the North York Moors. They are also responsible for the Air Warfare Centre, which maintains technical expertise on the warfare aspects of space at their headquarters at RAF Waddington, and delivers training on space-related topics at its RAF Cranwell site. Finally, the UK Space Operations Coordination Centre (UK SpOCC) is located at Headquarters Air Command at RAF High Wycombe. The UK SpOCC is responsible for execution of the legacy missions of missile warning, satellite warning and space object re-entry advice and warning, provides Space Situational Awareness through compilation of the Recognised Space Picture and functions as a focal point for support to operational users of space.

c. **MOD Scientific Staff**. Defence Science and Technology Laboratory (Dstl) provides technical advice and operational analysis to the MOD on space capability and technology. There is also scientific and technical/engineering expertise within the Defence Intelligence organisation. The MOD has also sponsored research into space capability by third parties from time to time, within the overall MOD research programme.

d. **Other Space Users**:

(1) **JARIC – the National Imagery Exploitation Centre**. JARIC is responsible for supporting defence planning, current operations and the intelligence assessment process. They make extensive use of satellite imagery, culled from a variety of sources, to do this. Their expertise in exploitation of space-derived product allows them to provide support at strategic, operational and tactical levels. Defence Intelligence Staff handle policy aspects of UK space reconnaissance.

(2) **CIOJ6**. As Joint User for Military SATCOM, CIOJ6 is the policy lead for the provision and use of satellite communication services. Responsibilities include the allocation, apportionment and prioritisation of satellite assets. CIOJ6 also provides strategic oversight for military SATCOM, specifically arrangements for the

security of infrastructure, exchange of capacity with NATO and EU nations and the maintenance of UK orbital filings.

Illustration 4.7 – The UK Contribution to Missile Warning and Space Surveillance. RAF Fylingdales on the North Yorkshire Moors houses a Raytheon Solid-state Phased Array radar (SSPAR), which in conjunction with other similar sites elsewhere contributes to Allied missile-warning and space-surveillance roles. The SSPAR replaced a mechanically-scanned radar installation in 2000, housed in formerly well-known 'golfball' radomes. The RAF Fylingdales installation is unique in the Allied radar chain in having a 360º field of view, provided by the 3 faces of the antenna. (Crown copyright photograph).

422. **Other Government Departments**. Other Government departments have legitimate interest in space in pursuit of national security objectives. The details of how and who they cooperate with to achieve this are understandably sensitive, and will not be expanded on here, save to note that as a result of their omission here, there are consequent gaps in the description of oversight arrangements. Effective linkages do exist to prevent conflicts of interest between them and overt users of national Space capability.

423. **UK Space Policy and Doctrine**:

a. **National Security Strategy**. The UK Government published a National Security Strategy in 2008, and its first annual update in June 2009. The implications for Space are discussed in Chapter 7 of the 2009 update at Paragraphs 7.45 to 7.49. Although not formally a National Space Policy, it is the most current statement of government thinking on the security implications of Space for the UK. Below the level of the NSS, there are military and civil space policies. The military elements are examined below and the civil equivalents in paragraphs 428-31.

b. **The UK Space Innovation and Growth Team.** Between June 2009 and February 2010, a government-sponsored Innovation and Growth team considered the wider implications and potential for the UK of a comprehensive Space Policy. The team included representation from the MOD, other government departments, Industry, academia and trade and professional organisations. A formal government response to the team's final report will outline how many of its recommendations are likely to be implemented.[19] This implementation will probably be linked to the formation of the UK Space Agency alluded to at Paragraph 428.

c. **MOD Space Policies and Plans.**

(1) **MOD Space Policy**. A unified MOD Space Policy exists, but is under review following the publication of the NSS. MOD Central Staff can advise on content and status.

(2) **MOD Space Strategy**. The MOD Space Strategy has been managed through the Space Strategy Action Plan (SSAP) to provide a central focus for UK military space activity, analysed across the Defence Lines of Development (DLODs).[20]

(3) **MOD Space Personnel and Training Strategy**. Subsidiary to the Space Strategy, the MOD has developed a Joint Space Personnel and Training strategy. Custody of the strategy is entrusted to a working group at Headquarters Air, overseen by the SMG. Although development is thus led by the RAF, the strategy itself is intended for implementation across all 3 Services.

d. **UK Doctrinal Publications**. As an essential enabler to many operations and capabilities, space is addressed in a similar way to other air power disciplines in the foundational UK publications. Details of UK Concepts and Doctrine relating to space can therefore be found in the Future Air and Space Operational Concept and in Air Publication 3000, the UK Air Power Doctrine manual. It is intended that relevant aspects of space capability will also inform other conceptual and doctrinal work on a routine basis. Development, Concepts and Doctrine Centre retain a desk officer to develop space conceptual thinking.

424. **US Military Space Organisation**. The following description of US Space command and control is believed to be both comprehensive and accurate. It is, however, a slight simplification in that it ignores the impact of 'dual-hatting' of certain key appointments. In particular, the relationship between 14th Air Force and US STRATCOM is more complex than it may appear here. Readers to whom this is relevant should consult the appropriate UK Liaison Officer if necessary.

a. **The US Unified Command Plan – USSPACECOM and USSTRATCOM**. The USA allocates responsibilities for military operations and missions to Combatant Commanders (COCOMs). COCOMs are normally Joint Commanders.[21] COCOMs do not maintain standing forces; rather they employ forces assigned to them by single-Services. They exercise their responsibilities either regionally or functionally. COCOM responsibility is allocated via the Unified Command Plan (UCP), which by US law is reviewed and revised every 2 years. Thus any definitive comment on allocation of responsibility may quickly become outdated. The text of the UCP is classified, so inquiry as to its current status or content is best made via a suitable liaison officer, or in-country personnel. Between 1985 and 2002, responsibility for Space Operations was vested in US Space Command (USSPACECOM) as a functional COCOM. In 2002 a routine reorganisation

---

[19] The Executive Summary of the team's final report can be found via the Intellect web page www.intellectuk.org in its library section.

[20] Production of this Primer was a SSAP deliverable under the 'Training' Line of Development.

[21] Assignment of forces by 2 or more services is mandated to qualify as a 'Unified COCOM'. All COCOMs are currently 'Unified COCOMs'. If a mission were assigned to forces of a single Service, the resulting organisation would be called a 'Specified Combatant Command', though no such construct has existed recently.

dissolved USSPACECOM, and transferred responsibility for space to an expanded US Strategic Command (USSTRATCOM). This remains true in January 2010. USSTRATCOM executes its space responsibilities mainly via the Joint Functional Component Command for Space (JFCC SPACE). (Commander JFCC SPACE is dual-hatted as Commander 14th Air Force.) JFCC SPACE controls the employment of Department of Defense space assets and their integration into global operations via the US Joint Space Operations Center (JSpOC) at Vandeberg AFB, California. Headquarters USSTRATCOM is located at Offut AFB, in Omaha, Nebraska. HQ JFCC SPACE and HQ 14th AF are both located at Vandenberg AFB.

b. **US Air Force Proponency for Space**. Air Force Space Command (AFSPC) is a single-Service (4-star) major command within the USAF, which has responsibility amongst other things for providing operational space capability to COCOMs and managing the space career stream within the USAF. It should not be confused with the now defunct USSPACECOM. HQ AFSPC is located at Peterson AFB, Colorado. The Under Secretary of the Air Force is the Executive Agent for Space within the Department of Defense. This last responsibility has specific meaning relating to the acquisition of space systems through the Space and Missile Systems Centre at Los Angeles AFB in El Segundo, California.

c. **Air Force Space Command Roles and Organisation**. Until December 2009, AFSPC executed its operational missions via 2 numbered Air Forces: 14th AF managed force generation and employment of the principal US orbiting space systems, and 20th AF operated the US Inter-continental ballistic missile (ICBM) capability. In December 2009, a reorganisation took place. 14th AF remains part of AFSPC, with its commander dual-hatted as detailed above. 20th AF retained its ICBM responsibilities but now discharges them jointly on behalf of USSTRATCOM and a new organisation, AF Global Strike Command. Having lost responsibility for 20th AF, AFSPC gained a new numbered AF – 24th AF – to coordinate and develop capability in cyberspace. HQ 24th AF was formed in September 2009, and achieved initial operational capability in January 2010. Details of its ultimate capability, complement and final status have yet to emerge. HQ 14th AF is located at Vandenberg AFB, HQ 24th AF is at Lackland AFB, Texas.

d. **The Joint Operationally Responsive Space Office**. The Operationally Responsive Space (ORS) Office is established at Kirtland AFB, New Mexico to coordinate the development of the ORS Concept. The ORS Office reports direct to Department of Defense Executive Agent for Space (the Secretary of the Air Force) – not up any Service chain. Operationally, they respond to STRATCOM statements of warfighter needs. Currently they envisage ORS being delivered in 3 'tiers' or capability groupings:

(1) **Tier 1 ORS**. Rapidly exploiting existing capability, possibly extending or expanding its original mission, in a timeframe of *minutes to hours*.

(2) **Tier 2 ORS**. Replenishing, augmenting and reconstituting space capability, either using existing assets or 'field-ready' assets deployed on demand, in a timeframe of *days to weeks*.

(3) **Tier 3 ORS**. Replenishing and reconstituting space capability, using rapid development of new technologies or capabilities in a timeframe of *months, but not years*.

The ORS Office will advocate, coordinate and provide resources where possible for Tier 1 ORS, and is currently working to develop enablers for Tier 2 and 3 ORS.

e. **Missile Defense Organisation**. Responsibility for Missile Defence of the US is spread between the Missile Defense Agency (MDA), an independent agency, and the Joint Functional Component Command for Integrated Missile Defense (JFCC-IMD), a component of USSTRATCOM.

(1) **Missile Defense Agency**. The MDA is responsible to the US Department of Defense for developing a layered defence against ballistic missiles. The agency has its origins in the Strategic Defense Initiative which was established in 1983, renamed as the Ballistic Missile Defense Organization in 1993, and then renamed as the Missile Defense Agency in 2002. The MDA's mission is to develop, test and prepare for deployment of a missile defence system. Using complementary interceptors, land-, sea-, air- and space-based sensors, and battle management command and control systems, the planned missile defence system will be able to engage all classes and ranges of ballistic missile threats.

(2) **Joint Functional Component for Integrated Missile Defense**. The JFCC-IMD was established to deter missile attacks against the US, its territories, possessions and bases. JFCC-IMD has the associated mission of planning, integrating and coordinating global missile defence operations and support for missile defence. JFCC-IMD originated in the Implemented Directive issued by the Commander, USSTRATCOM in January 2005. JFCC-IMD is responsible for planning and coordinating global operations and support for integrated missile defence. JFCC-IMD works with other combatant commands, the MDA and joint service components. Headquarters USSTRATCOM retains the responsibility for advocacy of system-level missile defence capabilities, integration of missile defence into strategic level planning and military assessments of missile defence capabilities. Strategic Level Intelligence support is provided by Defense Intelligence Agency's Joint Intelligence Operations Centre.

f. **Other US Services**. US Services are mandated to maintain a cadre of space-qualified personnel, conduct research and, acquire and operate space systems against single-service requirements. Personnel from all services have also flown on manned missions.[22]

425. **US Space Doctrine**. The earliest useful US Space Doctrine emerged in the 1970s. Air Force Manual (AFM) 1-1 (Functions and Basic Doctrine of the USAF) identified space operations as a basic mission of the USAF in 1979. Students of doctrine development seeking more detail of how early US doctrine emerged may find that document and President Carter's Presidential Directive 37 (PD-37) useful starting points.[23] Current US Space Doctrine is contained principally in 2 series of publications.

a. **Joint Doctrine – Joint Publication 3.14**. Joint Publication (JP) 3.14 has been recently revised (January 2009) and is available on the Internet. It is the capstone doctrinal document for US Space. The single Services' publications acknowledge it as their source of authority.

b. **USAF Doctrine – Air Force Doctrine Document (AFDD) 2-2 and AFDD 2-2.1**. AFDD 2-2 *Space Operations* and 2-2.1 *Counterpace Operations* are the authoritative USAF

---

[22] See Paragraph 420 for a description of the astronaut corps.

[23] Robert F Futrell's *Ideas, Concepts, Doctrine: Basic Thinking in the USAF 1961-1984,* Volume 2, Chapter 9 (Maxwell AFB: AU Press, 1989) covers the development of USAF capability and US Space Doctrine in detail and is copiously referenced. Both volumes of Futrell's (lengthy) history are available for free download on the Internet.

Space Doctrine publications. They are relatively current (November 2006 and August 2004 respectively) and are publicly accessible.

c. **Other US Services**. As well as subscribing to JP 3.14, each of the US Services produces its own single-Service space doctrine, similar to the USAF's. The USAF documents have the advantage of being unclassified and easily accessible, but readers with a particular task in hand should bear in mind the possibility of needing to account for other US single-service viewpoints and policies.

426. **US Space Reviews – the 'Rumsfeld' and 'Allard' Panels**. Two relatively recent US policy reviews are responsible for many of the relationships discussed above. In 2000, just before assuming the post of Secretary of State for Defense, Mr Donald Rumsfeld led a presidential review of US National Security Space. Not all its recommendations were implemented, but many were, including the upgrading of AFSPACECOM from 3-star to 4-star status, and the allocation of Executive Agency status for Space to the Department of the Air Force. In 2008, Congressman Allard promoted legislation leading to a Congressional (as opposed to Presidential) review of similar areas. As a Congressional report, it has a less prescriptive status than one initiated by the Executive. It is too early to predict which of its recommendations may be implemented, but as a report, it will undoubtedly be discussed and debated for several years.[24]

SECTION VI – CIVIL SPACE CAPABILITY AND GOVERNANCE

In this section you will find a description of:

- UK space organisation in the civil and commercial sectors.
- The roles and activities of the European Space Agency.
- American civil and commercial space organisations and bodies.
- A very brief summary of other nations' space governance.

427. **Space Agencies**. For many years, it was a common observation that 'the UK does not have a space agency'; the British National Space Centre (BNSC) performed some of the functions of a space agency without maintaining the capacity for independent mission management that may be considered the defining characteristic of a space agency in other nations. In December 2009, HMG announced their intention to form a new Executive Agency for Space. At the time of writing (March 2010), details of its composition and function were still in development, so the following paragraphs describe the *status quo ante* functions of the BNSC

428. **The British National Space Centre**. The BNSC is responsible for coordinating UK civil space activities. It is formed from 11 Government Departments at the time of writing (including MOD and the Meteorological Office) and research councils. It has 3 long-term objectives:

a. To enhance the UK's standing in astronomy, planetary and environmental sciences.

b. To stimulate increased productivity by promoting the use of space in government, science and commerce.

---

[24] The Allard Commission report is sometimes referred to publicly as the 'Young' report; Mr Thomas Young chaired the panel that conducted it. There was, however, an earlier 'Young' report in 2001, produced for NASA and relating to the International Space Station. Context should identify which report is being referred to.

c. To develop innovative space systems, to deliver sustainable improvement in the quality of life.

429. **British National Space Centre Activities and Outputs**.

a. BNSC is responsible for managing UK contributions to the European Space Agency, and all civil space activities including international partnering regarding civil space use.

b. BNSC also licences UK satellite launches, and maintains the statutory register of such activity.

c. BNSC produces a *UK Civil Space Strategy* document. The latest iteration was issued in February 2008.

430. **UKspace**. UKspace is the trade organisation of the UK Space industry. Jointly founded by the Society of British Aerospace Companies and Intellect (the trade body of the UK information technology, telecommunications and electronics sector), its members include the major UK companies employed in the space sector.[25] The role and mission of UKspace is to:

a. Grow the UK's share of the global space market, by promoting the best commercial, political and public environment for the UK space industry.

b. Promote greater awareness in Government, the media, the public and other key stakeholders of the wide-ranging benefits from one of the UK's most innovative, high skilled, value-adding sectors.

c. Provide the focal point for any organisation commercially involved in space systems and related services in the UK.

d. Provide the primary forum for industry to dialogue with Government and with other national and international stakeholders.

431. **The European Space Agency**. The European Space Agency (ESA) is a multinational organisation,[26] established by international convention. Originally founded as a research collaboration exercise (modelled on the *Conseil Européen pour la Recherche Nucléaire* (CERN), the particle physics organisation), it has grown into a true Space Agency capable of conducting independent space missions. ESA's purpose is to provide for, and to promote, for exclusively peaceful purposes, cooperation among European States in space research and technology and their space applications, with a view to their being used for scientific purposes and for operational space applications systems:

a. By elaborating and implementing a long-term European space policy, by recommending space objectives to the Member States, and by concerting the policies of the Member States with respect to other national and international organisations and institutions;

b. By elaborating and implementing activities and programmes in the space field;

---

[25] The *UKspace* website includes a members' page, providing a list of member companies and, where available, links to their own websites.

[26] As of January 2009, membership comprises: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom. Canada takes part in some projects under a Cooperation agreement. Hungary, Romania and Poland are 'European Cooperating States'. Estonia and Slovenia have recently signed cooperation agreements with ESA.

c. By coordinating the European space programme and national programmes, and by integrating the latter progressively and as completely as possible into the European space programme, in particular as regards the development of applications satellites;

d. By elaborating and implementing the industrial policy appropriate to its programme and by recommending a coherent industrial policy to the Member States.

A characteristic of ESA projects is that work is allocated between member nations in proportion to their contribution to costs. ESA members can elect whether to participate in individual projects, according to expertise, affordability and national goals. Recently, ESA has also proposed an integrated civil space surveillance capability, the European Space Surveillance System (ESSS). Details of the proposal were published in the ESA Bulletin during 2008, although timescales for possible implementation have not yet been announced.[27]

Illustration 4.8 – European Space Agency Membership. The flags outside the ESA Headquarters Building in Paris reveal the wide membership of the organisation. (ESA Image).

432. **European Space Policy**. A European Space Policy document promulgated in May 2007 reconciled ESA and EU Ministers' views on European Space Policy. It sets out a basic vision and strategy for the space sector, and tackles issues such as security and defence, access to space and exploration. Full details of its content can be found on the ESA website.

433. **The International Space University**. The International Space University was founded in Massachusetts, US (with the assistance of Massachusetts Institute of Technology) in 1987, and moved to its permanent campus outside Strasbourg, France in 1994. It offers postgraduate higher education in the fields of space science and engineering, and space business studies, through a variety of formal courses, rooted firmly in the ideals of international cooperation. The majority of its teaching activity is based around the Strasbourg campus, although it also organizes an annual external Space Studies Program. Although cognizant of the security and defence aspects of space activity, its main focus is on civil and commercial space. Full details of its activities and of the current faculty can be found on its website at www.isunet.edu.

---

[27] Heiner Klinkrad, Richard Tremayne-Smith, Fernand Alby and Detlef Alwes, *Europe's Eyes on the Sky*, ESA Bulletin 133 (February 2008) 42-48.

434. **American Bodies and Organisations**:

a. **American Institute of Aeronautics and Astronautics**. In the USA, the American Institute of Aeronautics and Astronautics (AIAA) is the professional body of the space industry, akin to the role of the Royal Aeronautical Society in the UK and equivalent bodies in Europe. As such, the AIAA's mission is to advance the state of aerospace science, engineering, and technological leadership within the US.

b. **The National Aeronautics and Space Administration**. NASA is a US Government agency, established by statute in 1958. It grew out of a combination of the extant National Advisory Committee for Aeronautics (NACA) and elements of military research organisations working on aeronautics and space.[28] Today it conducts research into aeronautics and astronautics via a variety of research facilities, as well as managing the US civil space programme, both manned and unmanned.

c. **The Space Foundation**. The Space Foundation is a non-profit advocacy group, founded in 1983 to 'foster, develop and promote, among the citizens of the United States of America and among other people of the world ... a greater understanding and awareness ... of the practical and theoretical utilisation of space ... for the benefit of civilisation and the fostering of peaceful and prosperous world.' Over the years it has expanded its goals; its current mission statement is: 'To advance space-related endeavours to inspire, enable, and propel humanity.' It has expanded the practical implementation of these goals to incorporate education programmes, services for space professionals and a research and analysis capability. It thus (in UK terms) combines some of the functions of a 'think tank', a trade organisation and a professional body.

435. **Other Nations**. Other nations' space programs are developing too quickly to provide useful information about them in this Primer. Aside from general and specialist news stories, the NASA website includes a page of links to other nations' space agency websites. Although it does not claim to be comprehensive, in January 2009 it included links to the Canadian, Russian and Japanese agencies, as well as to some of the European constituent agencies making up ESA.

---

[28] The Defense Advanced Research Programs Agency (DARPA) was formed at the same time as NASA. NASA was an explicitly civilian agency, and the initial negotiations to create both agencies included allocation of extant projects between the agencies.

(INTENTIONALLY BLANK)

ANNEX A – CONIC SECTIONS

A1. **Conic Sections**. The term **conic section** refers to the shape of various cross-sections through a cone. Their study dates from antiquity; they were certainly known to the ancient Greeks. There are four classical conic sections: the **circle**, the **ellipse**, the **parabola** and the **hyperbola**. It can be shown (though we will not do so here), that in any restricted two-body problem, the path of one body will *always* be a conic section. Paths following one of these curves are sometimes referred to as '**Keplerian Trajectories**'. Satellites in Earth orbit follow circular or elliptical paths. Spacecraft that have escaped from Earth orbit follow parabolic or hyperbolic paths. A very brief look at how they are derived will show that they belong to the same family of curves and thus suggest why they recur so often in spaceflight.

A2. **Eccentricity**. We saw in the main text that eccentricity describes the shape of an ellipse. In fact, all conic sections have an eccentricity. For a circle, eccentricity is exactly zero, for an ellipse it is greater than zero and strictly less than one, for a parabola it is exactly one, and for a hyperbola it is strictly greater than one.

A3. **The Cone**. To a pure mathematician, a cone is a pair of identical solid shapes, which look like the common conception of a cone, but touch, point to point. This is because of the form that the equation describing the cone normally takes. For our purposes, we can regard a cone as being one of these shapes and we can ignore the other. A cone is a thus solid shape with two faces; a circular (flat) face, and a single curved face leading to a point. Taking a cross-section through a cone is simply cutting a slice through it – akin to slicing a carrot. If you cut squarely across the carrot, the slices are circular. If you cut at an angle, they are elliptical. Parabolic and hyperbolic slices are harder to visualise (and to eat!), but the same principle applies.

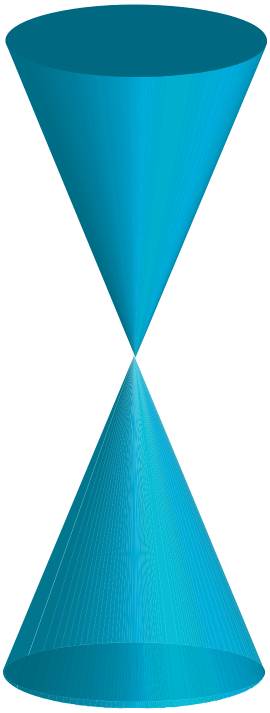


Illustration A.1 – The Pure Mathematician's Cone. This is the 'double cone' shape described in the paragraph above. We show it here because many diagrams of conic sections show this shape, with the conic sections taken through both aspects of the cone. For our purposes, where we are really only interested in the shape of the curves, we will from here on use the simpler 'single cone'.

A4. **The Circle**. A circle is the cross-section through a cone formed by a flat surface, *where the flat surface is parallel with the base of the cone*. The size of the circle changes depending on where on the cone the cross-section is 'cut', but the shape remains unchanged – always a circle.

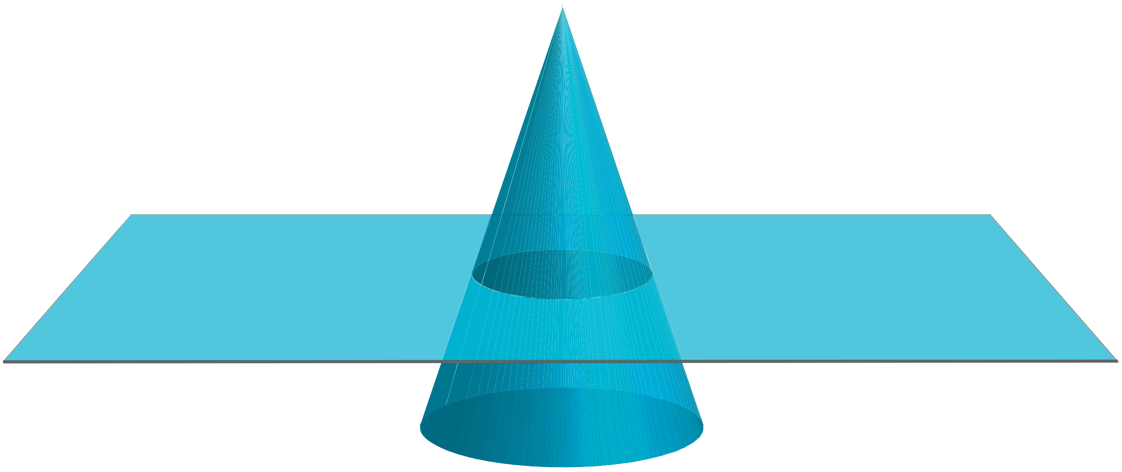


Illustration A.2 – The Circle. The plane of the cross-section is parallel to the base of the cone, yielding a circular cross-section. The size of the circle depends on the place the cross-section is taken, but the shape remains unchanged. (There is also in theory another (degenerate) cross-section which is a point, where the plane only touches the tip of the cone, but that has no meaning for Keplerian paths).

A5. **The Ellipse**. An ellipse is a cross-section through a cone *where the flat surface makes an oblique cut through the cone*. Varying the angle of cut varies the *eccentricity* of the ellipse, moving it up or down the cone varies the *size* of the ellipse. We saw in the main text that closed orbits are either circles or ellipses. Circles are just a special case of an ellipse, as we see here too.

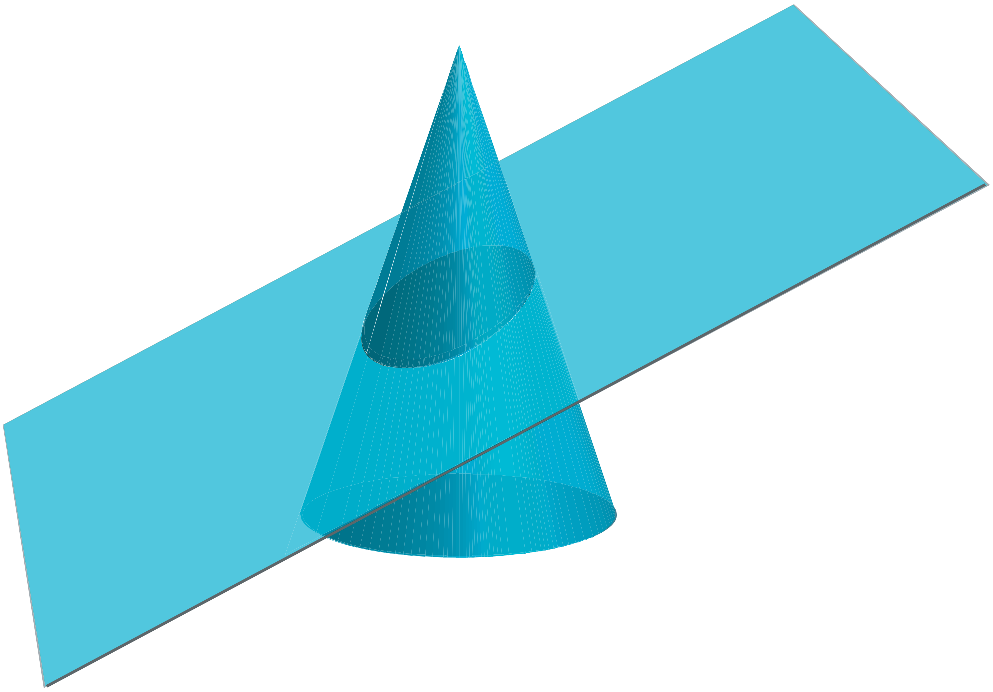


Illustration A.3 – The Ellipse. An ellipse is a cross-section through a cone where the angle between the plane of the base of the cone, and the plane of the cross section is less than the slant angle of the cone. Note that a circle satisfies this condition (the angle is implicitly zero, though in fact the planes are parallel, so they do not intersect), showing that a circle is a special case of an ellipse.

A6. **The Parabola**. A parabola is a cross-section through a cone *where the flat surface cut is parallel with the curved surface of the cone.* For any given cone, all the parabolic sections will be the same *shape*. The *size* depends on where you cut the section. If a spacecraft achieves escape velocity exactly, its path away from the Earth will be parabolic with respect to the Earth. Usually, an interplanetary vessel will exceed escape velocity by a substantial margin, so parabolic trajectories in space are actually unusual. Over short ranges on Earth, however, the parabola is also a good approximation to a ballistic trajectory.

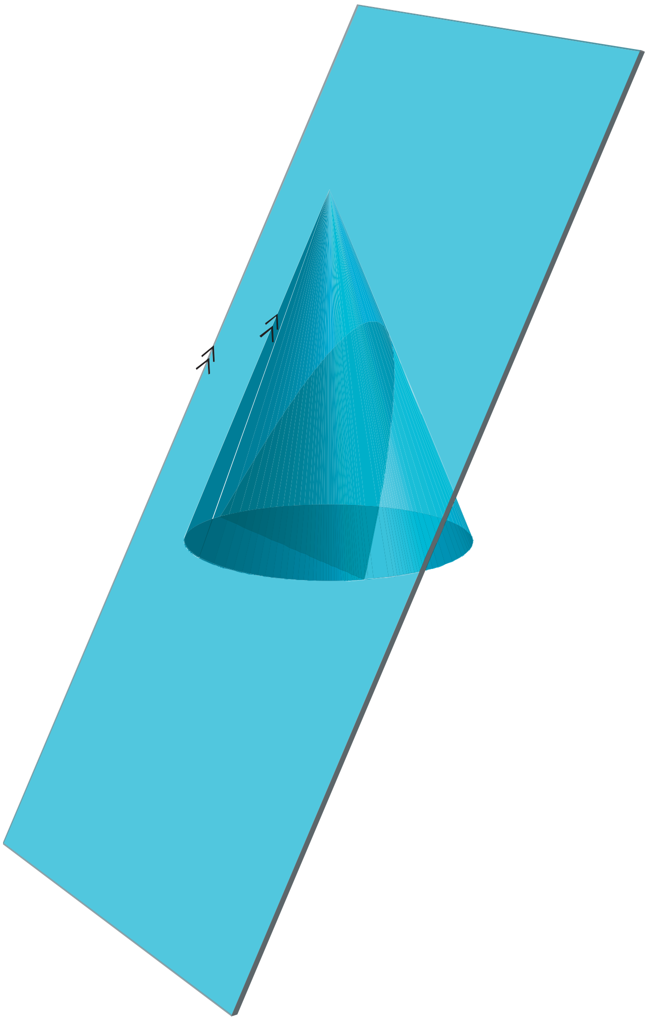


Illustration A.4 – The Parabola. A parabola is a cross-section where the plane of the cross-section is parallel with the slant face of the cone. (Again, there is a (degenerate) cross-section where the plane of cross-section is tangential to the face of the cone, and the cross-section is a straight line; this has no meaning for Keplerian paths).

A7. **The Hyperbola**. A hyperbola is a cross-section through a cone *where the angle of cut of the flat surface is steeper than the angle of the curved face.* Spacecraft that exceed escape velocity follow hyperbolic paths with respect to the Earth.

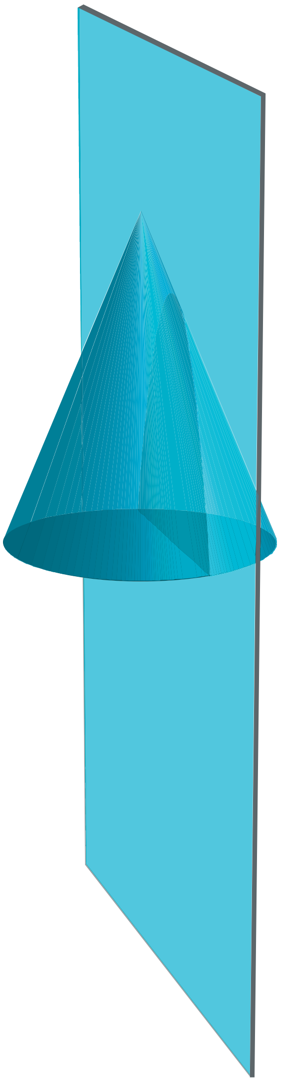


Illustration A.5 – The Hyperbola. A hyperbola is a cross-section where the angle between the plane of the base of the cone and the plane of cross-section is greater than the slant angle of the cone.

A8. **Speed of the Satellite and Keplerian Paths**. We recap and place the conic sections in context. For a two-body problem (with the assumptions outlined in Paragraph 124), the general equation of motion for the satellite can be solved so that the resulting equation is also the generic equation of a conic section. The term that defines the eccentricity of the conic section depends on speed. Thus the speed of the satellite tells us the shape of its path (in an ideal, two-body world). Note that the minimum orbital velocity and the escape velocity are properties of the body being orbited (e.g. they are different for the Earth and the Moon). This gives the following results:

Speed of Satellite	Shape of Orbit or Path	Eccentricity
Less than minimum orbital velocity.	Sub-orbital ballistic path.	-
Minimum orbital velocity.	Circle.	Exactly 0.
More than minimum orbital velocity, less than escape velocity.	Ellipse.	Strictly greater than 0, strictly less than 1.
Exactly escape velocity.	Parabola.	Exactly 1.
Exceeds escape velocity.	Hyperbola.	Strictly greater than 1.

Illustration A.6 – Speed of Satellite, Shape of Orbit or Path, Eccentricity

ANNEX B – BRAHE, KEPLER AND NEWTON

B1. **The Importance of Kepler's Laws**. Kepler's Laws govern orbital motion, and we have looked at how their constraints influence the utility of artificial satellites around the Earth greatly. Partly for background interest, but also to illustrate how they are universal laws, due to fundamental constraints of nature, it may be interesting to set their discovery in a historical context. Kepler's work was actually directed at explaining the motion of the planets around the Sun, but those paths are orbital paths governed by gravity in exactly the same way as satellites orbiting planets, so any theory applicable to one is equally relevant to the other. To understand Kepler, we must take a few steps back in history.

B2. **Geocentric and Heliocentric Theories**. The fact that the ancients believed that the Earth was the centre of the universe (*geo*-centric theory) and that all the other celestial bodies revolved around it, is relatively well known. This idea was formalised by the Greek astronomer Ptolemy, around about 150 CE.[1] He believed that the Earth was the centre of the Universe, that the 'fixed' stars were embedded in some kind of sphere which rotated around the Earth, and that the planets, which could be seen to move against the background of the stars, must be carried by smaller concentric spheres nested inside each other, and lying between the Earth and the stars. Although this became received wisdom, alternative theories that placed the Sun at the centre of the universe (*helio*-centric theory) had predated Ptolemy – probably dating back to the 7th Century BCE. There were no convincing explanations of why one theory was superior to the other, but geo-centric theory held sway until the turn of the 14th /15th Century.

B3. **Copernicus and Galileo**. Nicholas Copernicus (1473-1543) lived principally in what is now modern Poland, and in his youth studied variously medicine, law and astronomy. His observations of the planets' motion led him to question Ptolemy's theories, and while he could not disprove them, he proposed that a much simpler explanation of reality was that the Sun was the centre of local motion, that the planets (including the Earth) revolved around it, that the Moon revolved around the Earth, and that the Earth rotated on its axis. This may not *sound* simple, and controversy followed, complicated by various theological views. Galileo (1564-1642) added fuel to the intellectual fire by demonstrating (by using a telescope) that small satellites plainly revolved around the planet Jupiter, and that the planet Venus displayed varying phases just like the Moon. This strongly suggested that the Earth was not the centre of the whole universe, and made the Ptolemaic view much harder to sustain. Steadily improving accuracy in observation of the planets was already making Ptolemy's theories appear improbable; to account for small deviations in planetary position, Ptolemaic theorists were adding concentric circular motions to planets supposedly fixed in crystal spheres in order to make observed positions fit theory. They had no explanation of why this should be so. Gradually, helio-centric theory triumphed. It was still an entirely empirical theory; if pressed, its supporters would probably have maintained that the orbits of bodies around each other were circular, for aesthetic reasons as much as anything else.

B4. **Tycho Brahe**. Between the lives of Copernicus and Galileo, Tycho Brahe (1546-1601) was born to a wealthy Scandinavian family (he was born in what is now Sweden but was then part of Denmark). He worked as an astronomer and alchemist throughout his life, building his own observatory in Denmark, and working in Prague just before he died. He was determined to resolve the conflict between geo-centric and helio-centric theories, both of which he felt had attractive features. He devoted himself to the most accurate measurements possible of planetary positions,

---

[1] The terms 'common era' (CE) and 'before common era' (BCE) are used in this Primer as alternatives to 'AD' and 'BC', though their practical meaning is the same.

taking numerous measurements for many years. He worked without optical aid (he predated the telescope by a few years), using compasses and quadrants – instruments for measuring angles in the sky using large protractors. Brahe proposed a hybrid model of the universe (known as the 'Tychonic'model), where the Sun revolved around the Earth, carrying with it the planets revolving around the Sun – but he was not really a theorist. His strength was in the accuracy of his measurement of planetary positions, which were typically accurate to about one minute of arc (i.e. 1/60th of a degree). He left serious analysis of his work to others, principally his last assistant, Kepler.

B5. **Kepler**. Johannes Kepler (1571-1630) was a German mathematician and astronomer, who worked in Germany, Austria and (in 1600) in Prague. In Prague he met Tycho Brahe and became his assistant. He believed in helio-centric theory, and wanted to collaborate with Brahe to confirm and develop it. Brahe's unexpected death in 1601 thwarted this plan, but Kepler gained access to Brahe's legacy of detailed observing records and set to work to explain them. Kepler approached the problem principally as a mathematician. He concentrated on Brahe's observations of Mars, and tried to match numerous geometrical theories to the observations and to make valid predictions of the planet's position. After over 40 attempts in 4 years, he tried making the planetary orbits simple ellipses (he had previously discounted this possibility because he assumed it was too elementary for earlier observers not to have discovered it). It rapidly became obvious that the theory fitted the observations, and Kepler quickly developed the first two of his laws of planetary motion ('elliptical orbits' and the 'equal area, equal time' rule for the radius vector), essentially on an empirical basis – *'they fitted the facts'*. It took most of the rest of Kepler's life for him to develop the third law (the relationship between size of orbits and periods of revolution), and he punctuated his work with investigations of telescope optics (he corresponded with Galileo about the satellites of Jupiter) and other observational matters. Additionally, he worked as an astrologer at various European courts, attempted to explain his laws of motion by recourse to a mixture of theology, geometry, numerology and aesthetics, and at one point defended his mother in court against a charge of witchcraft (!) He had some concept of gravitational attraction, based on Galileo's ideas on the subject, but essentially had only discovered the 'how' of planetary motion, rather than the 'why'.

B6. **Isaac Newton**. Twelve years after Kepler's death, and in the same year as Galileo's, Issac Newton (1642-1727) was born near Grantham.[2] It would take too long to detail all his accomplishments – suffice it to say that his work in almost any one of his areas of interest would have guaranteed a historical reputation of some kind. He was separately a mathematician (who co-discovered calculus), physicist (with significant discoveries in both optics and mechanics), theologian (in his lifetime, his publications on religion exceeded those on science), politician (he was twice MP for Cambridge University), civil servant (he was Master of the Royal Mint in the latter stages of his life and took an active role in prosecutions for counterfeiting) and an enthusiastic alchemist (his death was possibly due to mercury poisoning, which may have been due to his alchemical research). He was also active in the Royal Society and in University affairs in Cambridge.[3] His contribution to planetary dynamics was to formulate the Law of Universal Gravitation, proposing an attractive force between any two bodies governed by their distance apart and their mass. This had two immediate implications for Kepler's laws:

---

[2] Some authorities cite Newton's birth as being in 1643, due to England not at that time having adopted the reforms of the Julian Calendar used elsewhere in Europe. Newton's birthday was on Christmas Day 1642 in the old calendar.

[3] He was the second Lucasian Professor of Mathematics at the University, a position later held (among others) by Dr Steven Hawking (the 17th incumbent).

a. Newton was able to demonstrate that Kepler's laws were a natural consequence of gravitation. If you assumed that gravitation worked as Newton described it, Kepler's laws followed; they could in fact be derived theoretically without any observations.

b. Since gravitation was universal, rather than a special property of planets or the Sun, it followed that any bodies moving under gravity alone would follow Keplerian principles. This allowed Newton to propose orbital motion, essentially using the argument outlined in Chapter 1, and explains why Kepler's Laws apply to artificial satellites as much as natural ones.[4]

B7. **Modern Refinements**. Since Newton, there has been little need to refine the laws of planetary motion. More distant planets were discovered based on subtle variations in known planets' motions – essentially the limits of the 'two-body' assumption were demonstrated by unexplained deviations, which were then used to predict the location of a body that might be causing them. Minute variations in the orbit of the planet Mercury were also used in the early 20th century as a confirmation of Einstein's Theory of General Relativity.[5] Anyone interested in following Newton's logic regarding Kepler's laws might also be interested in a lecture by Richard Feynman, who achieved fame (and a Nobel Prize) for research in quantum physics, but who as a 'novelty' lecture for undergraduate students at Caltech in 1963/4 reconstructed Newton's work using only high school geometry to prove Kepler's laws. Although the original transcript was lost at the time, it has been recovered and reconstructed and is available in print.[6]

Illustration B.1 – Three Founding Fathers of Orbital Mechanics. From left to right Brahe (date uncertain), Kepler (pictured in 1610) and Newton (in 1689).

---

[4] Subject to the constraints of the 'restricted 2-body' assumption outlined in Chapter 1.

[5] Just like a satellite orbiting the Earth, the line of apsides of planets precess (see Paragraph 140), for reasons including the oblateness of the Sun and attractions from other planets. Because of its small mass, these effects are particularly noticeable for the planet Mercury. Most of the causes of precession for Mercury's orbit were known to classical astronomers, but a small deviation remained obstinately inexplicable into the 20th Century. Applying corrections attributable to General Relativity, which allow for the distortion of 4-dimensional space-time caused by the Sun, accounted for the remaining error, providing an early validation of Einstein's work.

[6] Richard Feynman (auth), David and Judith Goodstein (eds), *Feynman's Lost Lecture: the Motions of the Planets Around the Sun*, (New York: Norton, 1996). The explanations are at once elementary (strictly geometry, no calculus) and at the same time challenging, but serve as a tribute both to Newton and to Feynman.

(INTENTIONALLY BLANK)

ANNEX C – FULL ORBITAL ELEMENTS

C1. **Defining an Orbit Precisely**. This Annex provides a more detailed explanation of the 6 classical Keplerian elements than is provided in the main text. The purpose of the orbital elements is to define exactly the orbit of a satellite around a body such as the Earth in a standardised way.

C2. **The Size and Shape of the Orbit**. Two elements define the size and shape of an ellipse exactly. The convention is to use the **semi-major axis** and the **eccentricity** of the ellipse. These define the ellipse itself exactly, and Kepler's First Law says that the centre of the Earth lies at one focus, but apart from that, the ellipse could at this stage be located in any orientation in space. It might lie over the Earth's equator, over its poles, or anywhere else, anchored only by one focus. We know, however, that the orbit will lie in a 2-dimensional **orbital plane**.

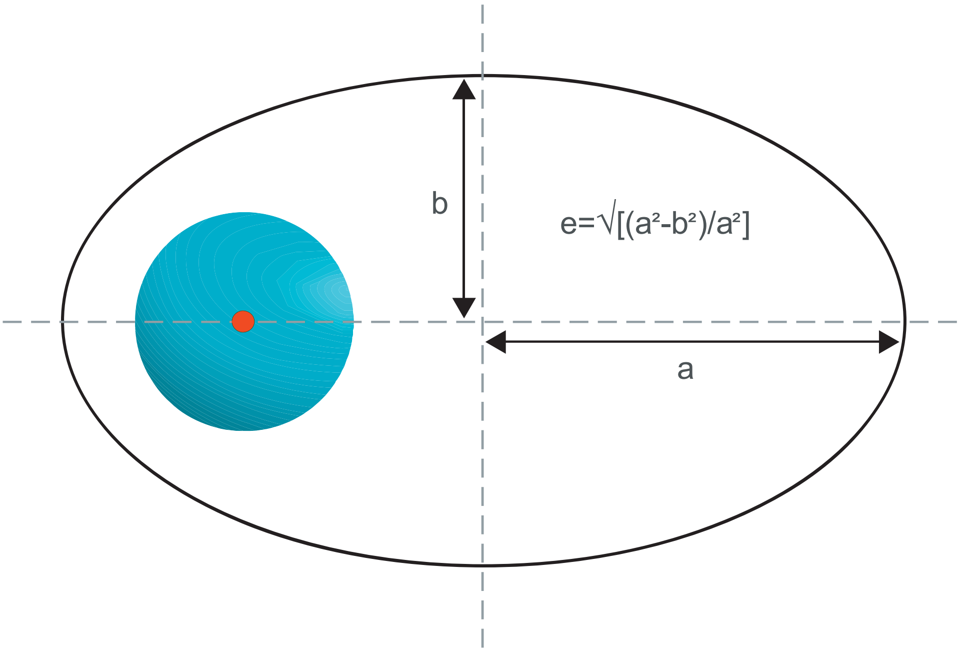


Illustration C.1 – Eccentricity and Semi-major Axis. The eccentricity and lengths of the semi-major and semi-minor axes are related. Any two define the third and are sufficient to describe the ellipse uniquely. The convention is to use 'a' and 'e' as orbital elements.

C3. **Location of the Orbital Plane**. Two elements define the orbital plane exactly. These are the **inclination**, and the **longitude of the ascending node**. To define these, we need a **reference plane**. The convention for earth-orbiting satellites is to use the plane of the Earth's equator as the reference.

a. **Inclination**. Inclination is the angle between the equatorial plane and the orbital plane. For an equatorial orbit, the inclination is thus 0°. Inclinations are measured between 0° and 180°, with inclinations between 0° and up to 90° indicating that the satellite is orbiting in the same direction as the Earth's rotation. An inclination of exactly 90° indicates a polar orbit, passing directly over the North and South Poles, and inclinations between 90° and 180° indicate that the satellite is orbiting in the opposite direction to the Earth's rotation. An inclination of exactly 180° would indicate an equatorial orbit, but with the satellite orbiting in the opposite direction to Earth rotation.

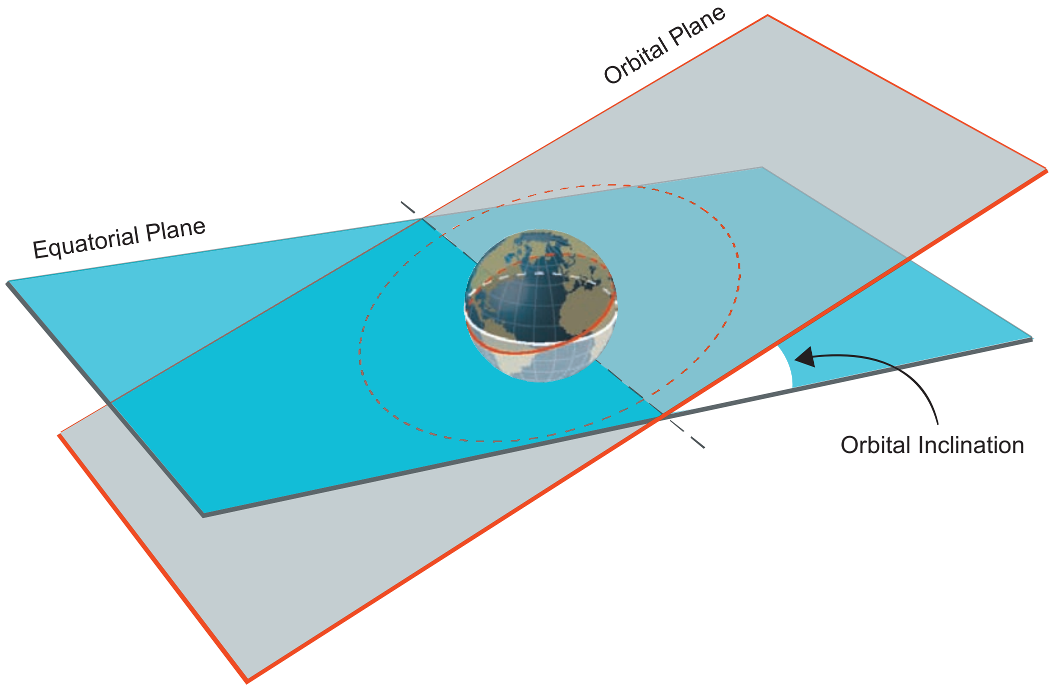


Illustration C.2 – Orbital inclination. This repeats Illustration 1.17. Inclination is the angle between the orbital plane (grey) and a reference plane. The convention for Earth-orbiting satellites is to use the equatorial plane (blue).

b. **Longitude of the Ascending Node**. The ‘longitude of the ascending node’ describes how the inclined orbital plane is oriented in space. Assuming the orbit is not equatorial, the intersection of the orbital plane and the equatorial reference plane is a straight line, extending into space and passing through the centre of the Earth (this line is known as the ‘line of nodes’). To orient the orbital plane, we need to define the direction of this line. For a given orbit, the line will cross the orbit at two places, on opposite sides of the Earth. The satellite will therefore cross this line twice per orbit too. On one occasion, it will be ascending, i.e. passing from the Southern Hemisphere into the Northern Hemisphere. The direction from the centre of the Earth to this ascending point is what we define. We do so with reference to a direction in space, agreed on by astronomers and space operators, known as the **vernal point**. You can draw a line from the centre of the Earth towards the vernal point, lying in the reference (equatorial) plane. The angle between that line and the line to the ascending node is the longitude of the ascending node.

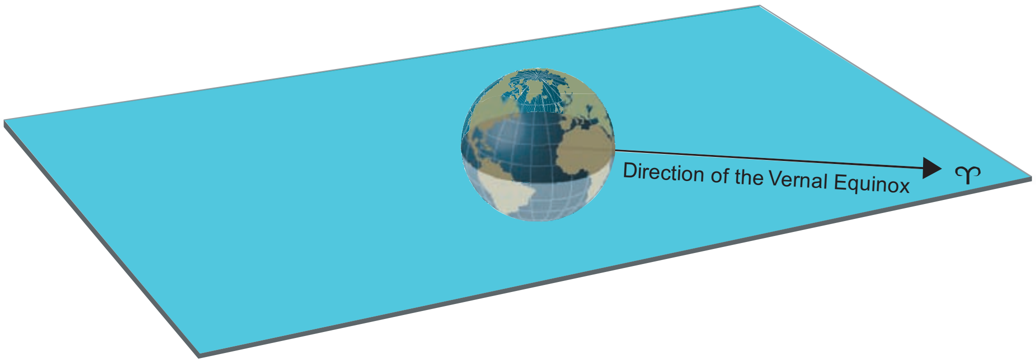


Illustration C.3a – Longitude of the Ascending Node 1. To orient the orbital plane in space, we firstly need to define a reference direction. By convention, the direction to an astronomical feature known as the Vernal Equinox, sometimes also referred to as the ‘First Point of Aries’, is used. It is often denoted in diagrams by the astrological symbol for the constellation Aries,

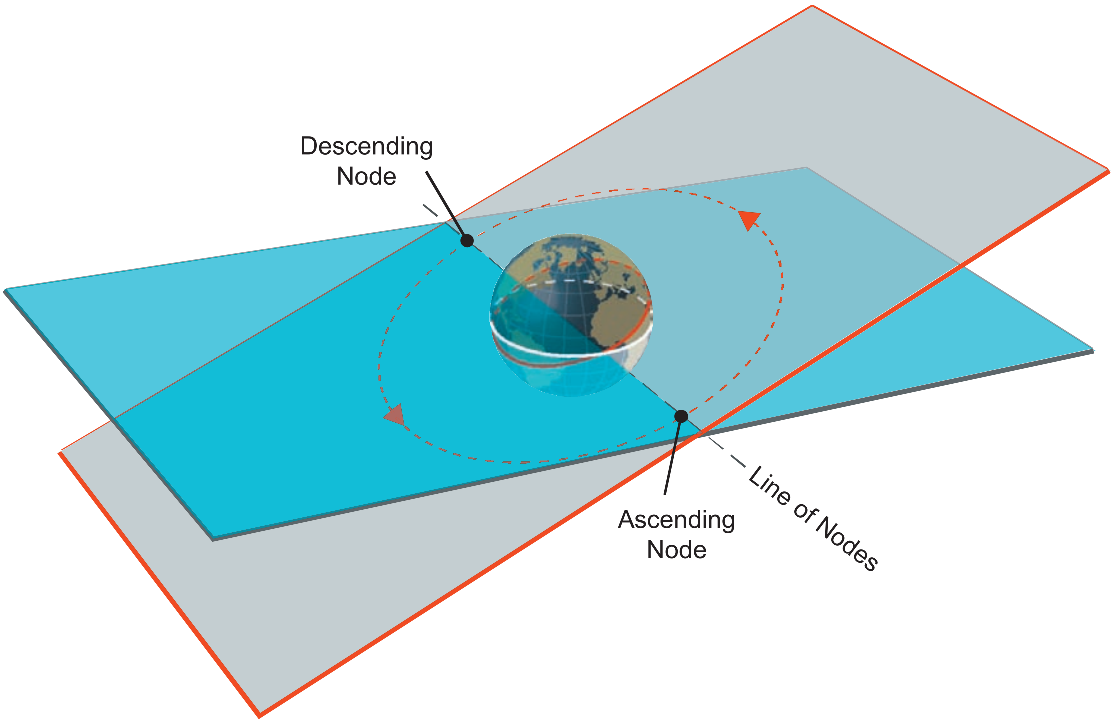


Illustration C.3 b – Longitude of the Ascending Node 2. As previously stated, the orbital plane of an inclined orbit will intersect the equatorial plane, defining a line, the 'Line of Nodes' (which will pass through the Earth). We need to distinguish between the two possible directions this defines. In one instance, the satellite will cross the line 'ascending', in the other 'descending'. The convention is to define 'up' and 'down' using the direction of rotation of the planet (so 'up' for the Earth has the North Pole at the top), and to use the direction of the ascending node as the line to be measured.

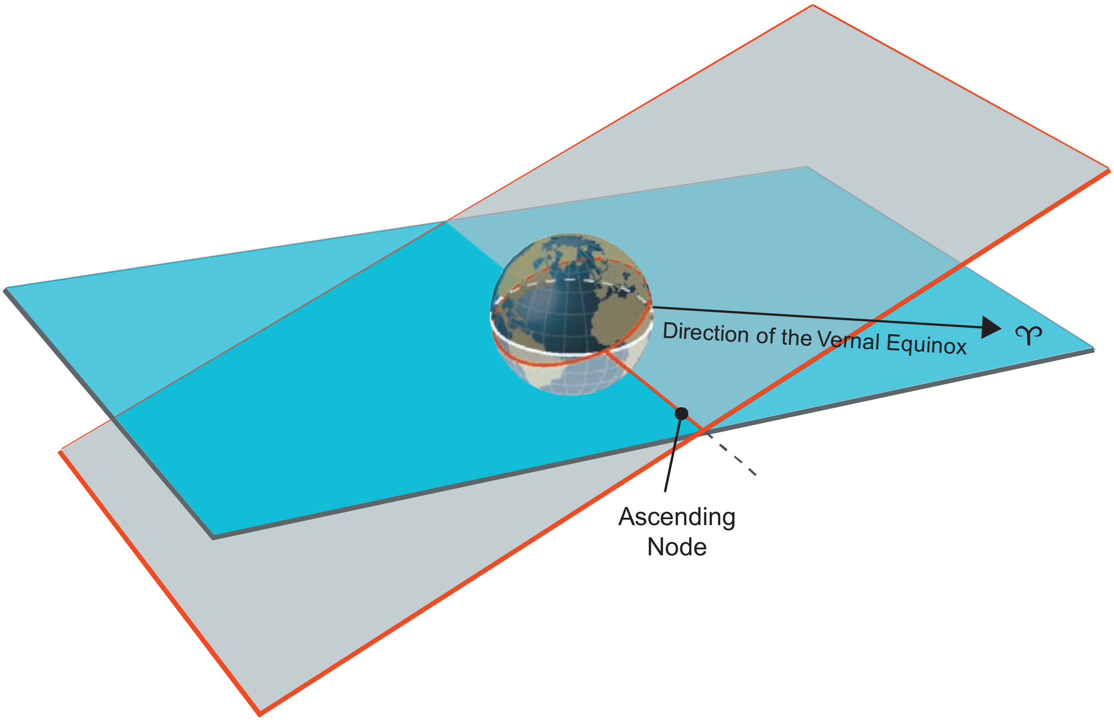


Illustration C.3c – Longitude of the Ascending Node 3. The conventions outlined above simplify the diagram; the direction of the ascending node and the direction of the vernal equinox both lie in the equatorial plane.

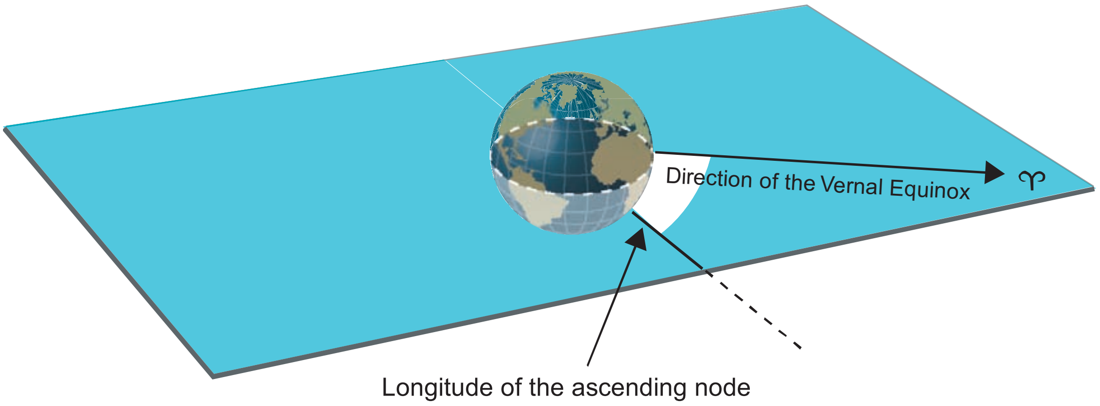


Illustration C.3d – Longitude of the Ascending Node 4. The longitude of the ascending node is the angle between the two reference lines. There are standard conventions to make sure it is measured in the correct direction.

c. **Practical Use of the 'Longitude of the Ascending Node'**. Where a constellation of multiple satellites exists, for example the GPS constellation, you will often find that subsets of the constellation follow identical orbits (same size and inclination), but that the orbit of each subset is separated in space from the others. It is in fact this longitude of the ascending node that varies between the subsets, to space the orbital planes out around the world.

C4. **Placing the Orbit on the Plane**. We need a parameter to define how the orbit is located on the orbital plane. To visualise this, imagine a plainly elliptical orbit at a moderate inclination. At some point in the orbit, the satellite will lie on the semi-major axis of its orbit. The angle between the semi-major axis and the line of the ascending node (which we also used above) locates the ellipse on the plane. This angle, which is another orbital element, is known as the **argument of perigee**.[1]

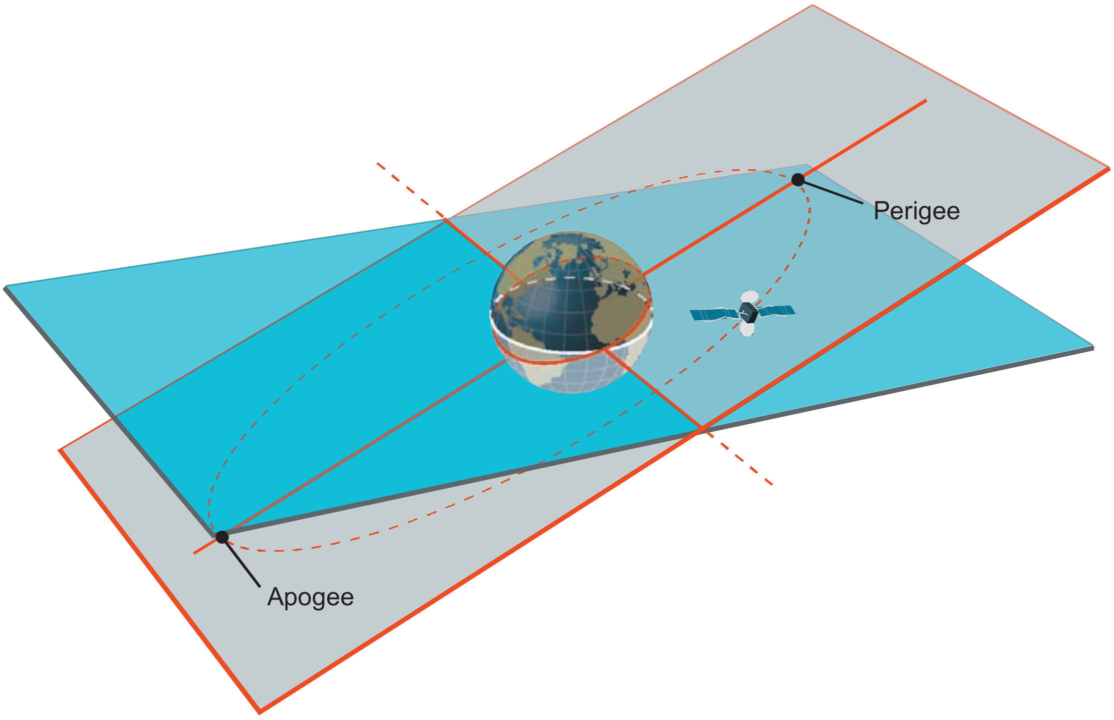


Illustration C.4a – Argument of Perigee 1. The orbit lies in the orbital plane, as previously defined. This means that the major axis of the orbit also lies on it, and hence also the perigee and apogee.

---

[1] See Paragraphs 129-130 in the main text for the implications of the '-gee' suffix and the definition of the term 'apsides'.

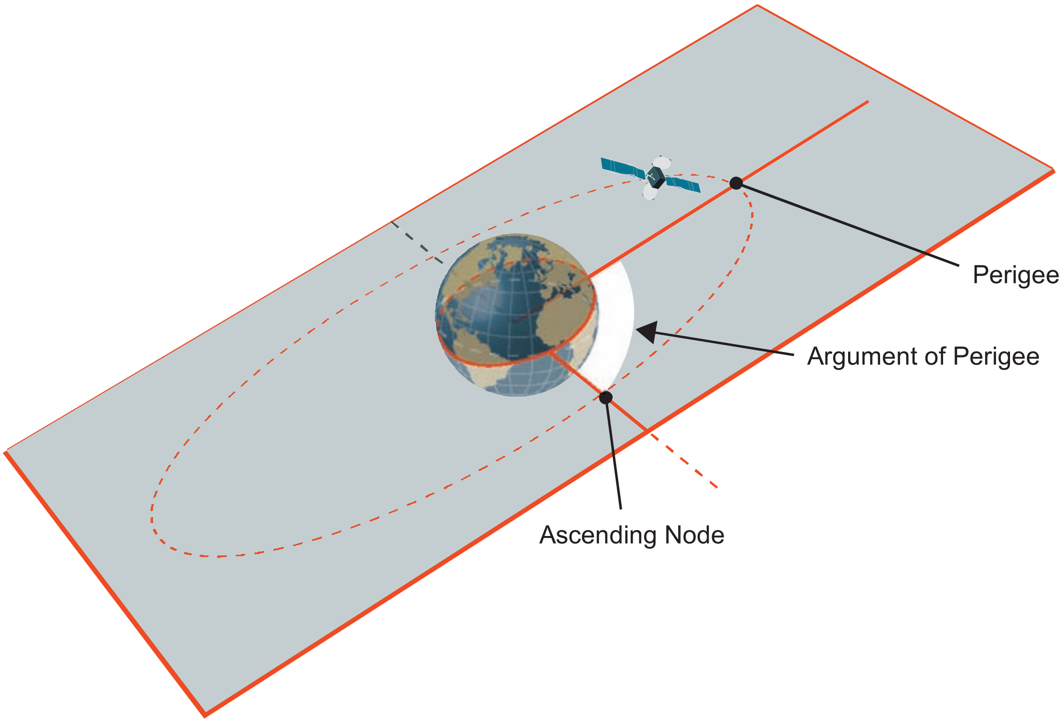


Illustration C.4b – Argument of Perigee 2. We have also previously defined the line identifying the ascending node. This lies in both the equatorial and the orbital planes (since it was defined by where they intersected). The angle between that line and the line pointing to the perigee defines the **argument of perigee**.

C5. **Placing the Satellite on the Orbit**. We have now defined the orbit in space, but at any given instant, the satellite must be somewhere on the orbit. We need to define a start point at a given instant to set the satellite moving. The generic parameter for this is the '**anomaly**'. Because the satellite may not move round the orbit at a constant speed (it will not if the orbit is elliptical, though Kepler's Second Law tells us what the variation in speed will be) there are mathematical conventions about how to define this anomaly. There is a quantity called the **mean anomaly** that varies regularly with time (akin to an average speed for the whole orbit), and another called **true anomaly** which gives the instantaneous position of the satellite. True anomaly is quoted as the orbital element, given at an instant known as the **epoch** for the orbit. The epoch also defines the time system used and some of the astronomical constants (which are not truly constant, but only change very slowly) that measurement of the elements depends on. True anomaly is measured as an angle (the position angle of the satellite) at a given epoch, measured between the satellite, the centre of the orbit, and the line of periapsis. Where multiple satellites in a constellation lie in the same orbit, their true anomaly will be adjusted to space them out around the orbit.

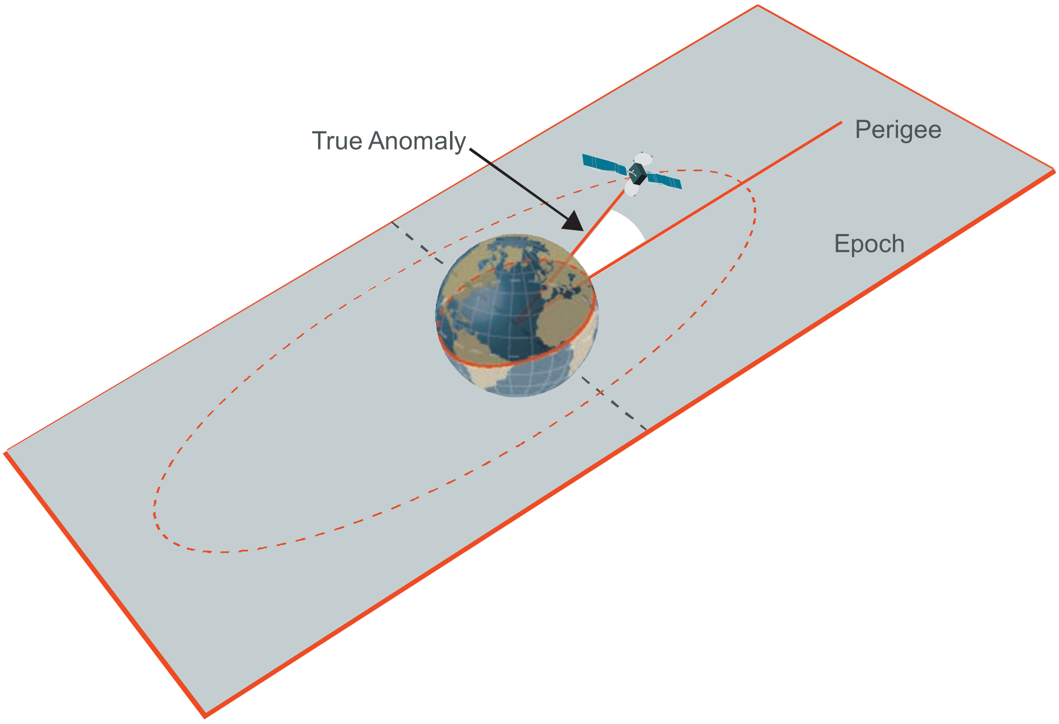


Illustration C.5 – True Anomaly. The final orbital element, the *true anomaly*, simply fixes the position of the satellite at an instant in time. It is measured as an angle around the ellipse, measured at the centre, between the direction of perigee (as used above) and the position of the satellite. The term **epoch** defines the instant that the angle is measured.

C6. **Undefined Elements**. We have already noted in the main text that for some orbits, some of the elements are undefined, for example the semi-major axis of a circular orbit. There are various conventions for how these problems are overcome mathematically. The curious reader will find a good explanation of one possible solution in Sellers '*Understanding Space…*', which is listed in the Bibliography.

C7. **A Seventh Element**. Orbital dynamics depend on the mass of the body being orbited. For artificial satellites around the Earth, that can be assumed to be a constant, and it is not therefore stated. Orbital elements for satellites orbiting other heavenly bodies will include details of the value used.

ANNEX D – ‘MANY-BODY’ PROBLEMS AND LAGRANGIAN POINTS

Dynamics of the Many-Body problem

D1. **The Restricted Two-Body Problem**. When we reviewed Newtonian dynamics in Chapter 1, and then looked at how that led to Kepler’s Laws, we made several assumptions. The two most important of these were that we only accounted for two bodies – the Earth and the satellite – and that one was much more massive than the other. This spared us from having to consider the gravitational attractions due to external bodies or other satellites, and allowed us to treat the Earth as stationary, even under the gravitational attraction of the satellite. To turn the resulting equations of motion into detailed predictions of satellite position is still a complex task. Some of the equations can only be solved by approximation. Nonetheless, the equations exist, and at least in principle, solutions can be found for them. The solutions are stable,[1] and give us the detail we need for prediction. This combination of assumptions and solutions are described as the **restricted two-body problem**.

D2. **Many-body Problems**. If we move away from the immediate vicinity of the Earth, we can no longer ignore the gravitational attraction of other bodies in the solar system. ‘Many-body’ problems do not, in general, have stable solutions. Where we need to consider them, for example in plotting the positions of the planets themselves, or in navigating between the planets, the equations may have to be solved by dividing the problem into many small steps in time and re-constructing the system after each step. There are, however, isolated special cases where stable solutions exist.

D3. **The Restricted Three-Body Problem**. Where 2 relatively massive bodies interact with a third of negligible size, we have what is known as the **restricted three-body problem**. In 1764, a Franco-Italian mathematician called Joseph-Louis Lagrange discovered that in such a system, 5 stable solutions existed. These are now known as **Lagrangian Points**, and are often referred to as L_1 to L_5. They are illustrated at Illustration D.1. L_1 to L_3 are the various cases where the three bodies lie in a straight line. L_4 and L_5 refer to configurations where one massive body orbits the other, and the small body follows the same orbit, but leads or lags the large body by 60° (so that all three bodies lie at the corners of an equilateral triangle). Essentially, for each of the Lagrangian points, the gravitational attractions of the two large bodies on the small one either cancel out or reinforce each other by precise amounts. L_4 and L_5 are truly stable, so an object there will tend to hold a steady position without expenditure of fuel.[2] L_1 to L_3 are in fact slightly less stable, though they still compare favourably with random configurations. Overall, a satellite at a Lagrangian point can hold a steady position in the system without continuous manoeuvre and excessive expenditure of fuel.

Applications

D4. **Actual Applications of Lagrangian Points**. Lagrangian points in the Earth-Moon and Earth-Sun systems have found uses in scientific research. They are useful as locations for sensors such as space telescopes, for example to give a constant view of the Sun’s disc, without it being eclipsed by the Earth or Moon or where it is necessary to keep a telescope in a roughly fixed position, but also distant from the Earth.

---

[1] In this context we mean that we can put a satellite into an orbit where the orbital elements will only change gradually, in a broadly predictable fashion. Small disturbances thus lead to small changes in the orbit, which can then be corrected if required.

[2] One of the better known instances of the L_4 and L_5 points are the families of **Trojan** asteroids. These follow roughly the same orbit around the Sun as Jupiter, but leading or lagging Jupiter by 60° and forming Lagrangian Sun-Jupiter-Asteroid combinations Analogous families of asteroids exist for other planets, but the Jovian ones were the first discovered and are the best known.

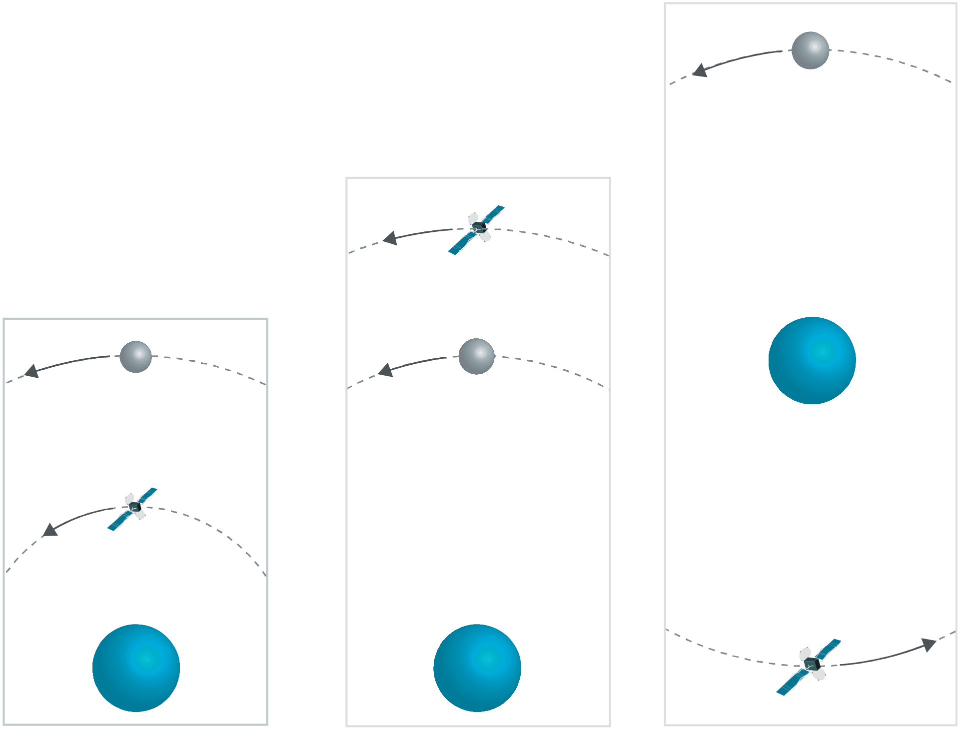


Illustration D.1a – Lagrangian Configurations L1 to L3. This diagram shows configurations L1 to L3. In the first, the satellite orbits between the primary and secondary bodies (e.g. the Earth and the Moon). In the second, the satellite orbits beyond the secondary body, and in the third, the satellite and secondary body orbit the primary, separated by 180º.

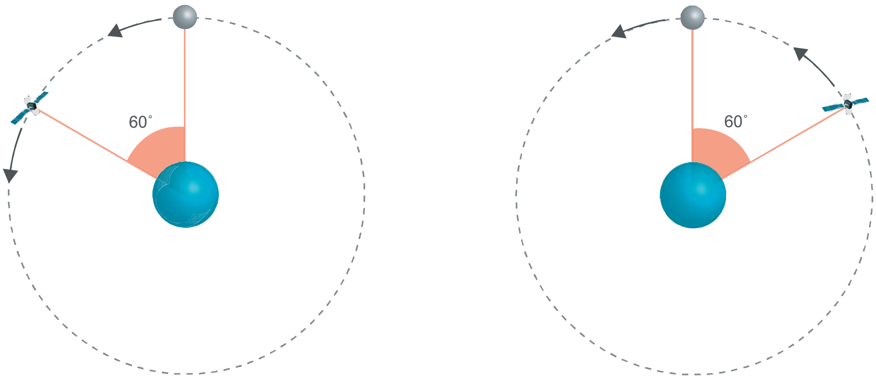


Illustration D.1b – Lagrangian Configurations L4 and L5. In this diagram, the two truly stable Lagrangian configurations are shown. The satellite orbits the primary body in the same orbit as the secondary, either leading or lagging it by 60º. Real applications of this configuration have included solar observations at an Earth-Sun Lagrangian point. The Trojan asteroids orbit the Sun at Sun-Jupiter Lagrangian points, and smaller collections orbit in conjunction with other planets.

D5. **Potential Applications for Lagrangian Points**. The Lagrangian point in the Earth-Moon system between the Earth and the Moon (the L_1 point) could provide a useful location for re-supply of extended manned spaceflight. The L_2 point beyond the Moon could provide telecommunications coverage for the far side of the Moon. At present, such points are probably too far from the Earth to be useful for surveillance, though their future use should not be completely discounted.

ANNEX E – ROCKET ENGINE FUEL CYCLES

E1. **Outline**. In Chapter 1, we compared the various basic types of rocket motor that can be used to launch, propel and manoeuvre space vehicles – specifically, we described solid rockets, liquid-fuelled rocket engines and hybrid designs. Because of their prevalence, we then examined the various design options for liquid-fuelled engines, contrasting pressure-fed and pumped designs. High-thrust designs are invariably pumped; this is currently the only practical way of generating the required thrust levels. There are, however, a variety of ways of powering the fuel and oxidiser pumps, and the choices made are often then used to describe the overall design. In this Annex, we describe some of the more common varieties that may be encountered.

E2. **The Generic Pumped Cycle**. (Illustration E.1) Turbo-pumps are large pumps to move high volumes of liquid to the combustion chamber. They gain their name from the large turbine impellers that move the liquid. Since in a rocket-motor fuel and oxidiser are invariably required simultaneously or not at all, the separate fuel and oxidiser turbo-pumps are usually mounted on a common shaft alongside a turbine wheel, which shares the shaft and drives the impellers. Providing a gas flow through the turbine thus drives both impellers together. The rate of flow of the fuel and oxidiser is controlled by valves downstream of the turbo-pumps. The various design options involve how the turbine is driven.

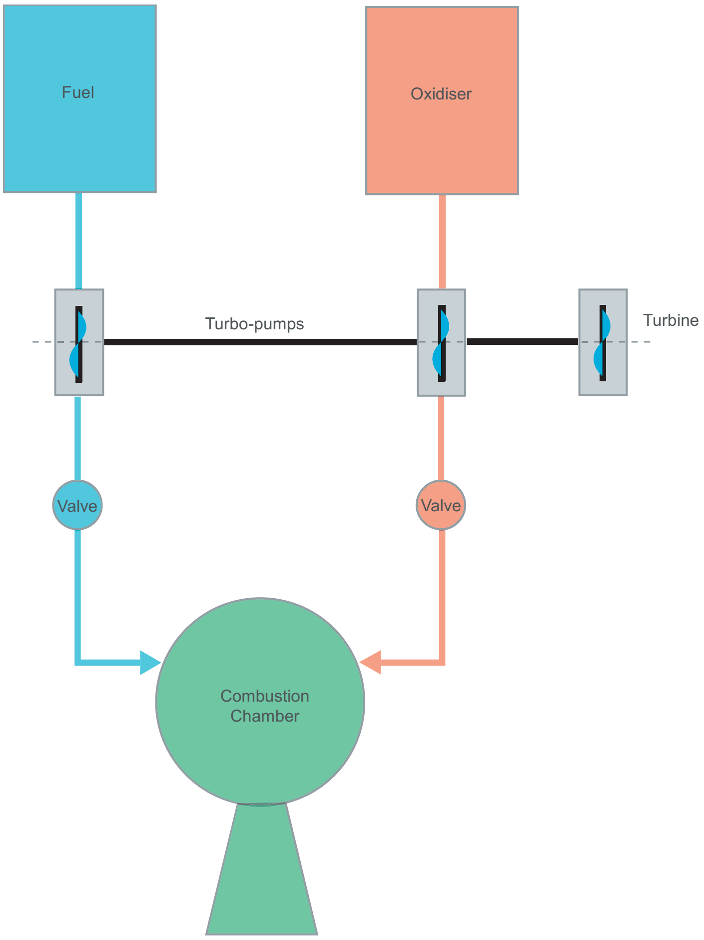


Illustration E.1 – A Generic Pumped Engine. This illustration shows the common elements of almost all pumped engines: tanks hold fuel and oxidiser, turbo-pumps on a common shaft provide fuel and oxidiser at high pressure to the combustion chamber, with flow controlled by valves. The turbine on the common shaft powers the two turbo-pumps.

E3. **Auxiliary Fuel Systems – The V2**. (Illustration E.2) The German V2 rocket used an auxiliary fuel system, where hydrogen peroxide was stored under pressure then decomposed by a catalyst to drive the turbo-pumps, which then fed liquid oxygen and a water-alcohol mixture to the main chamber.

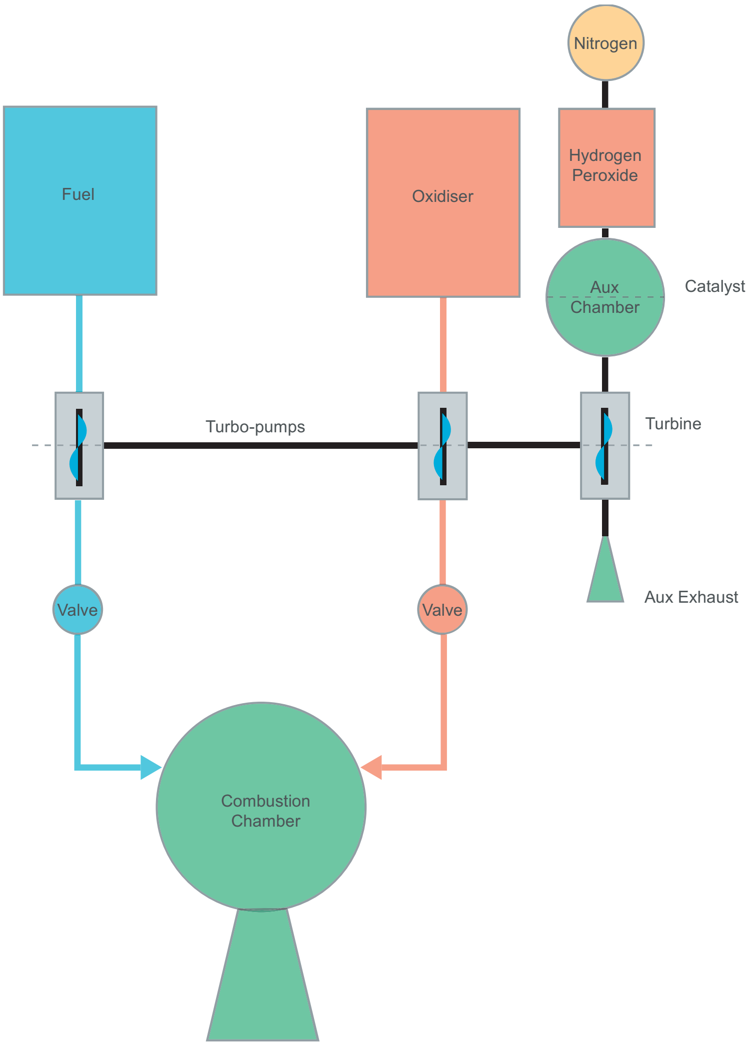


Illustration E.2 – The V2 Rocket Engine. The German V2 rocket of World War 2 was the first production design to use turbo-pumps. The fuel and oxidiser (alcohol and liquid oxygen) were pumped by an auxiliary cycle using hydrogen peroxide decomposed by a catalyst. The advantage was that hydrogen peroxide decomposition was a process already understood by German engineers, and was capable of developing the power required. The disadvantages included the need to deploy an additional unstable chemical in the field for launch.

E4. **Expander-cycle Engines**. (Illustration E.3) An expander-cycle engine captures waste heat from around the motor and exhaust nozzle, uses that to expand and vaporise one of the working fluids (typically the fuel), which then drives the turbine in its un-burnt state en-route to the combustion chamber. This design requires some auxiliary method of starting, as until combustion is established, there is no heat to achieve the vaporisation.

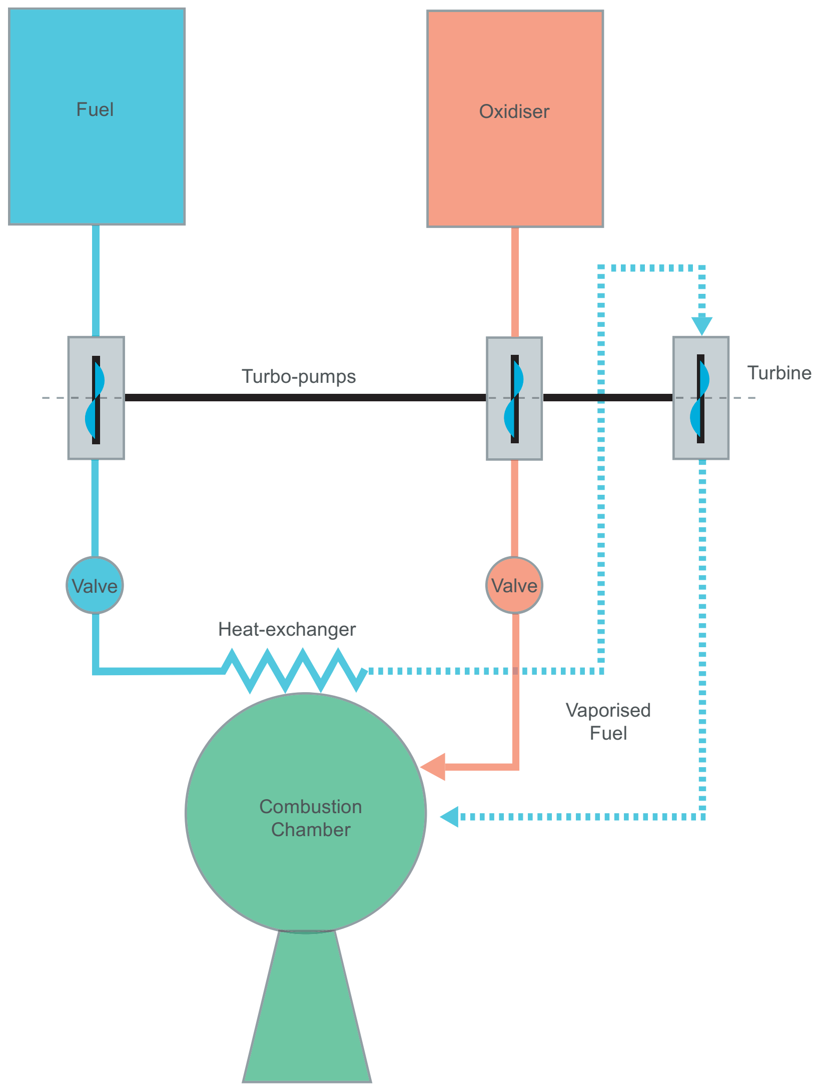


Illustration E.3 – Expander Cycle Engines. If (as is often the case) the fuel is inherently cold and vaporises easily, waste heat from the combustion chamber can be used to drive the power turbine. The fuel is fed around the chamber where it becomes gaseous. Because its volume increases, so does its pressure and it can then drive the turbine en-route to the combustion chamber. There may also be advantages in cooling parts of the combustion chamber. A variation uses a small quantity of vaporised fuel to run the turbine, then dumps the resulting exhaust overboard. Although this wastes fuel, it is actually more thermodynamically efficient to run the turbine with an open exhaust, so for some applications this option is preferred.

E5. **Gas-generator or Open-Cycle Engines**. (Illustration E.4) A gas-generator or open-cycle (both terms are used synonymously) engine draws small amounts of fuel and oxidiser from the main supply, burns them in an auxiliary chamber to drive the turbo-pump then dumps the exhaust from the turbine overboard.

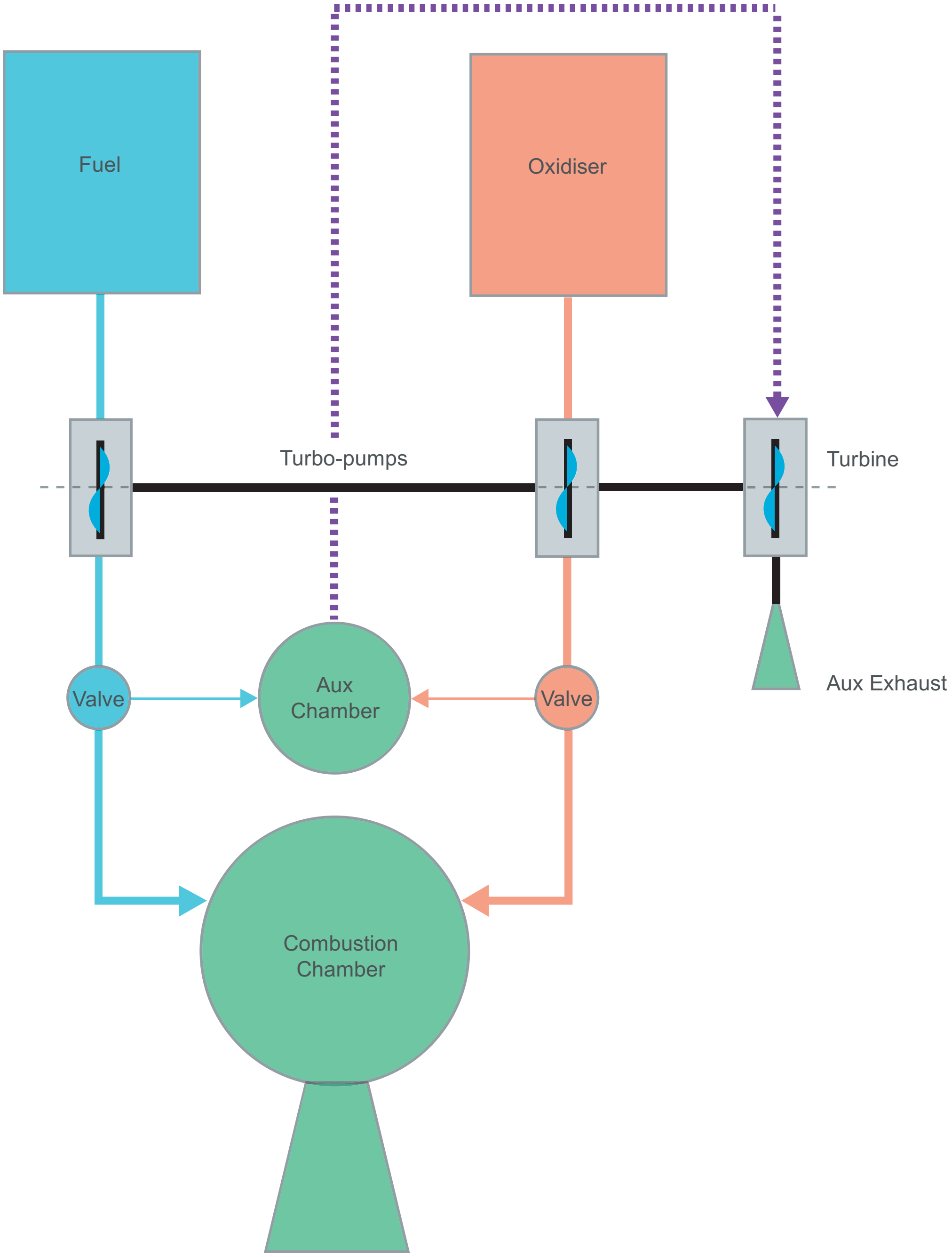


Illustration E.4 – Gas Generator Cycles. In this arrangement, small amounts of the main engine fuel and oxidiser are burnt in an auxiliary chamber, and the resulting combustion products are used to drive the power turbine. Strictly, an engine such as the V2 engine discussed above was a variation on the gas-generator cycle.

E6. **Staged-combustion or Closed-cycle Engines**. (Illustration E.5) A staged-combustion or closed-cycle engine takes all of one pumped liquid and some of the other (usually all the fuel and some of the oxidiser – the fuel-rich option), ignites it in a pre-burner chamber, uses that combustion to drive the turbine then sends all the part-burnt product to the main chamber where the remaining liquid is injected to produce the thrust. There have also been oxidiser-rich staged combustion engines, though these face severe material challenges.

E7. **Specific Examples**. At Illustration E.6, you will find example applications for each of the major design types discussed in the Primer. It makes no claims to be comprehensive, but shows how the various designs have been employed over time.

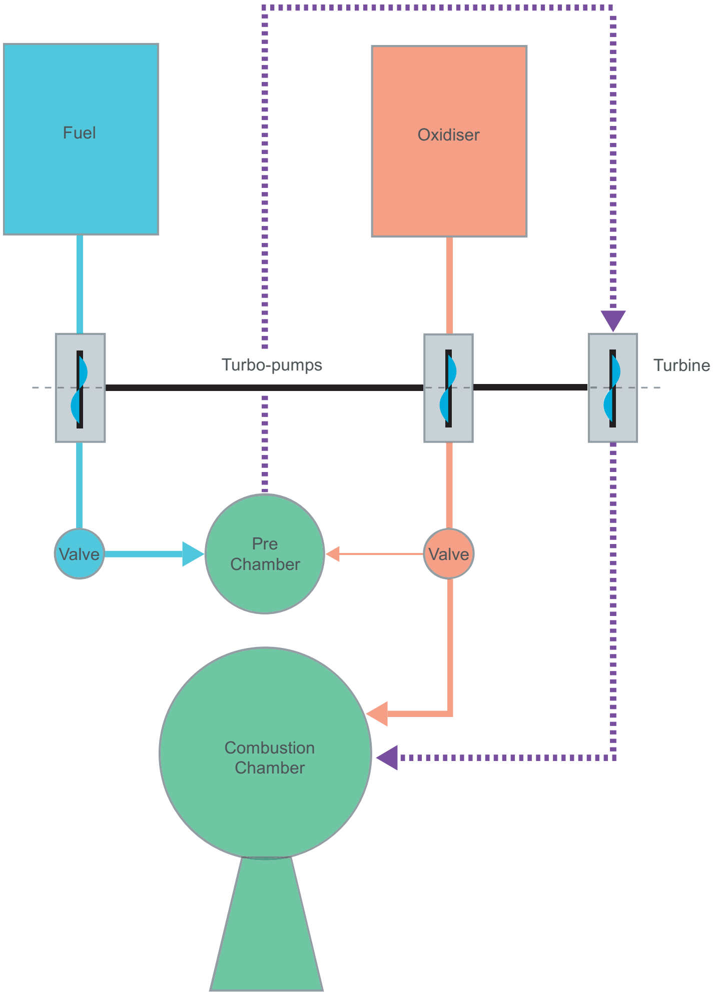


Illustration E.5 – Staged Combustion Cycles. In this arrangement, all of one component (usually the fuel) is partially burnt with a limited oxidiser supply, and the resulting product used to drive the turbine. It is then fed into the main combustion chamber where the remaining oxidiser is added. This completes the combustion and produces the thrust.

Rocket Motor Type	**Sub-type/design variation**	**Example Application**	**Remarks**
Solid-fuel	-	Space Shuttle Solid Rocket Booster.	Also most ICBMs, Ariane 5 strap-on boosters and numerous others, including air-launched missiles. See main text, Paragraph 156.
Liquid-fuel	Pressure-fed	Apollo Lunar Module ascent and descent stages, 'Kestrel' engine for SpaceX/Falcon 1 launcher.	Helium is the most common pressurising medium. See main text Paragraph 158 and Illustration 1.26a.
	Auxiliary Gas Generator	V2 Missile.	Nitrogen-pressurised hydrogen peroxide (H_2O_2) decomposed to power turbopumps.
	Expander cycle	Rocketdyne/Pratt&Whitney RL10 – widely used.	RL10 used in Delta IV and Atlas V launchers. First flew in 1963, still being used and upgraded in 2009.
	Gas-generator cycle	Snecma Vulcain motor.	Used in Ariane 5.
	Staged-combustion	Space Shuttle Main Engine, various Soviet/Russian designs.	Invented in Russia in 1949, used widely in USSR, and in legacy engines from that programme such as the RD-180 now used in US launchers.
Hybrid	-	SpaceShip One.	Butyl-rubber solid fuel, Nitrous Oxide (NO_2) oxidiser. See main text Paragraph 160 and Illustration 1.27.

Illustration E.6 – Examples of Rocket Motor Design Applications

ANNEX F – THE ELECTROMAGNETIC SPECTRUM

Electromagnetic Radiation – An Introduction

F1. **Radiation Everywhere**! Electromagnetic (EM) radiation is all around us, all the time. Our bodies rely on it to see things with (light), but even in the dark we are bathed continuously in it. Because it is so all pervasive, we are inclined to forget about its existence, but we need some understanding of it and the phenomena it exhibits to explain several aspects of space capability. This annex provides an overview of EM radiation to achieve that purpose.

F2. **Wave Theory**. The first principle to establish is that EM radiation displays **wave properties**. This does not mean that everything displaying wave properties is EM radiation – sound is transmitted as a wave, but it is **not** an example of EM radiation; waves in the sea are true waves too.

a. The common properties of all these waves are that they travel at a measurable **speed**, which will vary according to the material that the wave is travelling through (the **medium**), and that they have two linked characteristics that can be compared, **frequency** and **wavelength**. Speed, frequency and wavelength are linked mathematically. In summary:

frequency x wavelength = speed

b. Frequency tells us how many waves arrive somewhere in a given time (so many waves per second, minute, hour…). Wavelength tells us how much space a wave takes up (measured in feet, miles, kilometres, millimetres…). Speed just gives us the rate at which the wave moves. If you use everyday units, the relationship becomes obvious, so for example:

25 waves arrive in an hour… Each wave is 2 kilometres long…so the waves are travelling at 25x2 = 50 kilometres per hour.

EM radiation only complicates this slightly by using extreme (by everyday standards) units.

c. Finally, note that wavelength is not the same as, and is not linked to **amplitude**. Wavelength (linked to frequency) tells you how long the wave is, amplitude tells you how high it is (or frequency tells you how often a wave breaks on the beach, amplitude tells you if it is a breaker or just a ripple). If you prefer the sound analogy, wavelength/frequency tell you the *pitch* of a note, amplitude tells you its *loudness*.

F3. **EM Waves**. EM Waves are true waves. Thus, they have a characteristic speed, frequency and wavelength. As an indication that all EM waves are different manifestations of a common phenomenon, they all travel at broadly the same speed in any given medium. In a vacuum, they all travel at *exactly* the same speed, commonly referred to as the ‘speed of light’.

F4. **The Speed of EM Radiation**. The value of the speed of light is typically quoted as being about 186 000 miles per second, sometimes given as 300 000 kilometres per second, or (when quoted by an engineer) 3×10^8 metres per second.[1] This sounds unbelievably fast, but at the distances involved in space it is not. The radiation from the Sun travels at this speed, but still takes about eight minutes to reach the Earth. Scientists frequently use the symbol ‘c’ to denote this speed. To be

---

[1] The form ‘ms^{-1}’ is sometimes used for ‘metres per second’.

precise, the symbol 'c_0' is conventionally used to denote the speed of EM radiation in a vacuum, and 'c' without a subscript to refer to a measured speed in a given material and for a particular wavelength.

F5. **Changes in 'c'**. In anything except a vacuum, EM radiation is slowed down, by an amount that varies according to what the medium is, and to a limited extent by the wavelength of the radiation.

a. **Materials**. In the case of air, the change from c_0 to c is very small, of the order of 0.03% reduction. Even this variation can be important, for example in correcting the measurements that a GPS receiver makes of the signals from satellites.[2] In other transparent materials, such as glass, the change in speed can be more substantial, and in the laboratory, scientists have managed to produce materials that can almost bring EM radiation to a halt. These last examples are currently fairly exotic, rather than of practical use. Corrections to c are governed by a property of the material known as its **refractive index**. This is a very widely used quantity in the design of optical systems – it is in fact the change in speed that accounts for the path of light being bent in prisms and lenses. We will discuss this further below.

b. **Radiation**. The refractive index is not, however, simply a property of the substance concerned. It varies with the wavelength of the radiation too. This explains why a prism separates colours of light – one colour is being slowed more than another. The proper name for this phenomenon is **dispersion**.[3] It has practical consequences in designing optical systems, including those used in satellites.

c. **"Isn't the speed of light a constant"?** Having established that the speed of light (and other radiation) is subject to considerable variation, we should point out that the observation that 'c is a constant' does not contradict this. That comment refers to the apparent speed of arrival of radiation in moving systems; the implications of it are explored in Einstein's Special Theory of Relativity. Without delving into relativity, we look further at this below when we discuss the Doppler Effect.

F6. **Units**. Having established what the values mean, we now introduce the units we normally use to describe EM radiation. Speed – the constant we will hang the other quantities on – is usually given in **metres per second**. Frequency is measured 'per second', 'once per second' being described as **1 Hertz** (abbreviated **Hz**). Wavelength is traditionally measured in **metres**, or a multiple or fraction of them. Since we need to deal with both very large and very small multiples and fractions when discussing EM radiation, we use the customary metric prefixes such as 'kilo-', 'micro-' and so-on to keep the actual numbers manageable. For reference, a summary of the prefixes is at Illustration F.1. Remember that some of them hide very large or small multiples, so that frequency and wavelength multiply together to give 300 000 000 (3×10^8) metres per second.

---

[2] See Annex G for further details of how and why this correction is made.

[3] Overall, the correction from c_0 to c due to the refractive index of the material is usually bigger than that for dispersion. For practical purposes, dispersion can sometimes be ignored.

Greater than x1		Less than x1	
Prefix (abbreviation)	**Multiplier**	**Prefix (abbreviation)**	**Multiplier**
kilo- (k-)	x 1 000 (10^3)	milli- (m-)	x $\frac{1}{1000}$ (10^{-3})
mega- (M-)	x 1 000 000(10^6)	micro- (μ-)	x $\frac{1}{1000000}$ (10^{-6})
giga- (G-)	x 1 000 000 000 (10^9)	nano- (n-)	x $\frac{1}{1000000000}$ (10^{-9})
tera- (T-)	x 1 000 000 000 000 (10^{12})	pico- (p-)	x $\frac{1}{1000000000000}$ (10^{-12})
(Only the most common prefixes are shown)			

Illustration F.1 – Metric Prefixes

F7. **The EM Medium**. We discussed above the value of the speed of EM radiation in a vacuum. This begs an immediate question about how the radiation travels without a proper medium to travel through. We will not provide an adequate answer to it in this primer. Better simply to accept that for EM radiation, a vacuum can be the medium, and that propagation happens (and a good thing too, or the Earth would not be illuminated by the Sun and life on Earth wouldn't happen either!).

F8. **EM Wave Theory**. We need to say a (very) little about what an EM wave is, however. The theory of EM waves was worked out by James Clark-Maxwell at the end of the 19th century. His explanation depends on two 'fields' that can exist in a vacuum: an electric field and a magnetic field. EM waves are ripples in those fields. An individual EM wave is a pair of ripples – an electric one and a magnetic one – moving through space. The two components travel in the same direction and at the same speed, 'propping each other up'. The 'E' and the 'M' ripples propagate in separate planes, which are always at right angles to each other. The line where the two planes intersect is the direction the wave is travelling in. We will use this explanation when we consider polarization, a property with useful implications for surveillance.

F9. **Waves and Particles**. Anyone who has read any popular science about quantum physics will have encountered the theory that ascribes wave properties to particles and vice-versa. This theory yields useful results in many areas of science and engineering, but again we will not develop it further in this primer. We will treat EM radiation as waves because that best explains how it behaves, and treat particles as particles, even though the real explanation is more complex.

Varieties of EM Radiation

F10. **Sorting the Waves**. Having said a lot about things that are not EM waves, we need to describe things that are. We have implied that light is EM radiation (which is true), but there are many other varieties. Radar transmissions, X-rays in a hospital and microwaves in an oven are all examples. We need some way of relating them to each other, however, to make sense of the variety.

F11. **Sorting by Wavelength or Frequency**. We noted above that in any given medium, the speed of light is a constant. We also related wavelength and frequency to speed. This gives us a system for sorting radiation. We can order it using either frequency or wavelength; as one quantity gets bigger, the other will get smaller, to keep 'c' at a fixed value and satisfy the constraints of Paragraph F2b.

F12. **Long Wavelength to Short Wavelength**. We will survey the varieties of EM radiation starting with the largest (longest) wavelength (i.e. smallest/lowest frequency) moving to the shortest wavelength (highest frequency). When organised in this way, the categorisation is referred to as the **EM Spectrum**.

a. **Radio Waves**. Radio waves are the longest wavelength EM radiation. Wavelengths can reach up to kilometres (which give frequencies as low as kilo-hertz – the 'hertz' on its own is not a very useful unit for EM radiation), and down to centimetres (frequencies in the giga-hertz range). The radio-frequency (RF) portion of the EM Spectrum is so large that it has its own sub-divisions. These are listed in Annex G.

b. **Micro-waves**. Micro-waves fill the gap between radio and Infrared (IR) radiation. Their uses overlap with both. Principally, they are used as the radiation in radar, as a communications medium,[4] and domestically, for micro-wave heating/cooking.[5] Note however, that some radar systems use radiation that would otherwise fit in the radio subdivision, and that at the boundary of micro-waves and IR (in the area sometimes referred to as **tera-hertz radiation**), there may be potential to use the radiation for directed-energy weapons. Roughly speaking, wavelengths are in centimetres and millimetres.

c. **Infrared Radiation**. After the micro-wave band comes IR radiation. This is very similar to visible light, it just has a slightly longer wavelength – slightly too long for our eyes to detect. In all other respects, however, it displays similar properties. IR wavelengths are measured in micrometres also sometimes referred to as **microns**.

d. **Visible Light**. The EM radiation we can detect most easily is visible light, as this is what our eyes are sensitive to. The visual spectrum spans from red light (longest wavelength, lowest frequency) through green (where our eyes are most sensitive) to blue-violet (shortest wavelength, highest frequency). Wavelengths lie in the nanometre range. For historical reasons, wavelengths in the optical range are also sometimes measured in **angstroms**. One angstrom is 1/10th of a nanometre, or 10^{-10} metres.

e. **Ultraviolet Light**. Just beyond visible light lies Ultraviolet (UV) light, again just outside our eyes' detection range. Rather like micro-waves, UV light vibrates at frequencies that can cause molecules and atoms to vibrate in sympathy with it. We see the result (but not the UV radiation itself) as **fluorescence**.

f. **X-Rays and Gamma Rays**. Beyond UV light lie x-rays and gamma-rays. X-rays are obviously familiar for their medical uses, and are also a component of solar radiation. Gamma-rays are more often thought of as a dangerous variety of radioactivity.[6] Wavelengths are the shortest of all, measured in nanometres and picometres.

---

[4] 'Microwave communications' would normally be taken to refer to relatively high-power systems used for trunk-communications, but some domestic Wi-fi devices also use very low-powered microwaves.

[5] Micro-wave cooking depends on the fact that the frequency of the radiation happens to coincide with the frequency that makes a molecule of water vibrate. The water molecule – two atoms of hydrogen linked to one of oxygen as 'H_2O' – is like a very small tuning fork with a characteristic frequency. Because the frequencies coincide, if you subject the molecule to the radiation, it vibrates, and we measure that vibration as heat.

[6] Strictly, the definitions and characteristics of x-rays and gamma-rays overlap. The rigorous definition distinguishes how they were generated; x-rays are generated by the motion of electrons, gamma rays within the atomic nucleus. Practically, the distinction will be of little importance in this Primer.

Properties of EM Radiation

F13. **Similarities and Differences**. To make practical use of the EM Spectrum, we not only need to know what EM radiation has in common in all its forms, but also where it differs. We may not realise where we are seeing differing instances of the same effect.

F14. **Transparency**. Transparency is crucial to where EM radiation can travel, but is a remarkably hard property to pin down. The only truly transparent medium is a vacuum, where all EM radiation can travel without loss *en-route*. We know from everyday experience that air and glass are effectively transparent to visible light, but that brick walls are opaque to it. If you think deeper, however, the situation gets very complex very quickly. Some examples will illustrate this:

> a. A mobile phone uses radio-waves, and your phone works indoors. This implies that ordinary walls are reasonably transparent at radio wavelengths. However a simple wire mesh of the correct dimensions is totally opaque to the same waves, so the phone will work in one building but not another, depending on how the walls are constructed.
>
> b. Water vapour in the atmosphere generally blocks IR radiation (another interaction between radiation and molecules, just like microwaves in the oven), but at two specific bands in the IR range, it is relatively transparent. This allows IR-guided missiles to see their targets. In the same way as (say) green glass allows green wavelengths of visible light through, while blocking other colours, the atmosphere does the same thing to IR, allowing two 'shades' through while blocking the rest.

F15. **Refraction**. In the discussion about the speed of EM radiation above, we noted that it varies in different media. We do not sense this slowing down as a delay in seeing things; the speed of light remains incredibly fast in all everyday media for all practical purposes. What we do see is bending of its path if it strikes a denser medium at an oblique angle. This is how lenses and prisms work. One standard model explanation of refraction is a body of troops marching obliquely onto soft ground. Those who encounter it first are slowed down, while those at the other end of the rank proceed unhindered. The effect is a 'left' or 'right-wheel' towards the soft ground, until everybody is on it, when uniform, slower, motion resumes. The analogy is a beam of light being deflected towards the centre of a lens. All optical systems rely on this effect, but, for example by a similar argument, radio transmissions from space are distorted as they pass into the atmosphere. The blue sky and the red sunset and sunrise are evidence that the effect also varies with wavelength (i.e. dispersion, which we described above).

F16. **Propogation Loss**. If the vacuum of space is transparent, why does range from a transmitter matter? The answer depends on how the signal (radio, visible light, radar) spreads out. If a signal is travelling in all directions from its source, such as when it is emitted by a simple antenna or an unshielded light source, the further away you or the receiver are, the fainter the signal you detect. The missing signal has not been absorbed en-route. Rather it has been spread out across a larger and larger area, in a similar manner to a balloon or bubble being inflated – the 'skin' of the signal gets thinner and thinner as it gets bigger, and harder to detect as a result. Do not take this analogy too far, however. EM signals do not 'burst' at a critical range; they just get fainter and fainter.

F17. **RF Propagation**. A combination of transparency, refraction and reflection (which we will not explore further) allow us to explain and predict how radio-waves travel through the atmosphere – a subject of immense practical importance. The very large range spanned by RF frequencies and wavelengths complicate matters, however. Suffice for now to note that refraction effects vary with frequency/wavelength, so some RF signals travel in straighter lines than others. The reflective properties of different layers of the atmosphere also vary with frequency, so some signals 'hug the ground' while others pass out into space. We noted in Chapter 1 the interaction between solar radiation and the atmosphere, and one effect of this is that the reflective and refractive properties of its upper layers vary. You have probably seen the practical effect of this if you have ever listened to an old-fashioned (medium or short-wave) radio at night. The changing properties of the atmosphere in the absence of the sun's radiation allow signals to travel much further than in daytime. Space weather also affects this phenomenon significantly, which is one reason why a space weather forecast can be of immense practical importance to communications staff.

F18. **The Doppler Effect**. The Doppler effect describes the influence of speed on our perception of frequency or wavelength. If the source of an EM signal (of any kind – radio, visual or anything else) is moving towards or away from the detector (receiver, our eye, etc), common sense might suggest that the speed of the signal seen at the detector would be altered. In fact, this is not generally the case, as the speed of propogation is constant in a particular medium; instead the frequency and wavelength vary. One common illustration of this is the change in perception of the pitch of a siren on a vehicle passing a bystander (the use of the word 'perception' is deliberate – for someone in the vehicle, there is no apparent variation in pitch however fast or slowly the vehicle moves). The problem with relating this to EM radiation is that everyday velocities are so small compared to the speed of light that convincing demonstrations are hard to come by. There are different approaches to analysing the Doppler effect mathematically too, depending on whether one needs to take account of relativistic corrections due to speeds approaching 'c'. The relative speed affects the EM radiation in either case, however. Aircraft radar can exploit the Doppler effect to separate moving and stationary targets, even though aircraft speeds are miniscule fractions of 'c'; Doppler-based radar navigation systems made similar use of it. The effect still depends on the speed as a fraction of the speed of light, so it is not detectable in everyday circumstances (the police-car siren's tone changes as it drives past you, but you do not see the colour of the headlights or beacon change. The car is travelling at a measurable fraction of the speed of sound, but not of light). In space applications though, where satellites orbit at speeds of kilometres per second, these are real effects that must be allowed for by designers.

Generating EM Radiation

F19. **Black-Body Radiation**. We can not only *detect* radiation with our eyes and skin, we also *generate* it. This has important consequences for some satellites, particularly those conducting surveillance. All matter generates some EM radiation as a function of its temperature, and the term **black-body radiation** is applied to this phenomenon. For practical purposes, most black-body radiation is emitted in the IR, visual and UV bands. The fundamental causes are complex, but the effect is simple. If you heat a piece of iron or steel in a forge, it at first glows dull red. As it gets hotter, the glow becomes more orange-yellow, and eventually becomes a blue-white. This colour – which is EM radiation emitted by the body – is the black-body radiation. We might think that it only started to radiate it when we put it in the forge, but in fact it was emitting it all the time, even at room temperature. The waveband where the peak emission occurs is a function of temperature. At room temperature, the peak emission is in the IR band, where our eye cannot see it. A warmer body (iron in a forge) emits its peak in the visual band, initially in the red area of the visible spectrum (shorter wavelength than IR). By the time it is glowing blue-white, it is emitting a portion of the radiation in the UV band and the peak is somewhere at the blue end of the visible spectrum (even shorter wavelength). The only way to control the radiation is to control the temperature of the body in question. Although we used iron as our example black-body, the effect is independent of the material. Our own bodies radiate in the same way, predominantly in the IR band because of our body temperature. This is how a thermal imager or NVGs detects people – by focussing the IR energy emitted using special lenses, detecting it, and displaying it at a wavelength our eyes can see.

F20. **Generating Other Radiation**. We plainly have more practical methods of generating EM radiation than heating iron in a forge (though effectively, that is what happens to the filament in a light-bulb). At the low-frequency end of the scale (radio, radar and micro-waves), electrical power can be converted directly into radio-waves in a transmitter, and those waves can be transformed to generate radar and micro-waves. These are relatively straightforward processes. There is little need in space applications to generate UV, x-ray or gamma radiation, though all are generated to an extent in the Sun and form part of its radiation. The one exception to this might be the need to generate UV radiation for directed-energy weapons. This can be done by a variety of means, including controlling chemical reactions that emit UV radiation spontaneously. This is one of the means by which chemical lasers work.

Other EM Radiation Phenomena

F21. **Radiation and Energy**. You can infer from the paragraphs above that some forms of EM radiation are inherently more energetic than others, and that frequency/wavelength has some relationship to this. Although the explanation 'why' depends on quantum theory, which we will not explore further, the basic observation is correct. EM radiation travels in 'wave packets', and the amount of energy that can be 'fitted into' a wave-packet depends on frequency/wavelength. Higher frequencies/shorter wavelengths equate to more energetic packets. Thus radio waves carry little energy, while micro-waves can boil water. Visible light can be bright, but UV radiation is energetic enough to cause sunburn (and to disrupt skin-cells malignantly too). X-rays and gamma-rays are inherently very energetic, capable of penetrating tissue and consequently are even more dangerous.

F22. **Lasers and Coherent Light**. When a hot filament in a light-bulb emits visible light, the individual wave packets are generated at various frequencies/wavelengths clustered around the red/orange/yellow area of the spectrum, but the packets themselves are unrelated. One frequency will be the most common – the peak frequency – but it will be accompanied by near neighbours. If wave-packets were soldiers, they would be in different uniforms, though with one regiment or corps predominating, and milling in a general direction, but in an uncoordinated fashion. Laser light is a more disciplined version of this phenomenon! Because of the way it is generated, it is emitted at a single 'spot' frequency. This means that the wave packets can all stay in phase (the technical term is that they are '**coherent**'). There are great practical advantages to this. Because of coherence, the light beam tends not to spread out as it travels, making it more efficient for carrying a signal, or more concentrated in delivering energy.[7] The 'soldier' analogy is obvious. Laser light is all the soldiers in one uniform, marching in step.

F23. **Polarisation**. Another property of wave motion that we may wish to exploit is **polarisation**. We described in Paragraph F8 how EM radiation consists of electric and magnetic field vibrations, occurring in planes at right-angles to each other. Normally, those planes of vibration are uncoordinated between wave packets, so within a beam of light (say), there is no relationship between the planes in one packet and in the next. They are at right angles to each other **within** a packet, but uncoordinated **between** packets. Polarisation is the condition where all the planes are similar or the same between packets.

---

[7] We do not circumvent the frequency/energy relationship here – a high-power IR laser concentrates more power than an eye-safe red laser pointer, despite its lower frequency. The individual red packets are potentially more energetic, however, and blue or UV ones would be even more so, but the IR laser generates many more of them in a given area because of its design.

a. In everyday life, polarised sunglasses filter out light where the wave planes are in a particular orientation, and allow light through where they are in another at right angles to the first. This has a double effect. Firstly, this reduces the total amount of light passing through, but more interestingly, it exploits the fact that much glare in the everyday world is reflected off smooth surfaces – the surface of water for example. That reflected light/glare is already polarised, because reflection favours light where one of the fields is parallel to the reflecting surface. Polarised sunglasses selectively filter that light in preference to any other, reducing glare.[8]

b. The example above used visible light, but any other EM radiation can also exhibit polarised characteristics. In particular, radar, which depends on reflection, can make use of polarisation information to improve resolution and discrimination.

F24. **Resolution and Diffraction – Apertures and Antenna Size**. Whether working with radio waves or with optical wavelengths, a satellite operator often generates a picture of activity on the surface. Accuracy, usually described as **resolution**, is often at a premium, and two fundamental constraints affect it from the EM point of view. One is the wavelength used for the observation; the other is the physical size of the antenna or viewing device. This latter quantity is usually quoted as **aperture**.

a. **Wavelength**. Resolution improves with reducing wavelength, so for example all else being equal, you can resolve finer detail at optical wavelengths than you can with radar.

b. **Aperture**. Resolution also improves with increasing aperture. The reason for this is due to a phenomenon known as **diffraction**. This is an effect caused at the edge of an imaging system, where the path of a wave is distorted as it encounters an abrupt edge. The distorted wave from the edge of the aperture blurs the overall image resolution. As an aperture gets bigger, the distorted 'edge-wave' becomes a smaller and smaller part of the overall signal and resolution improves.

c. **Measuring Resolution**. Note that although resolution is often quoted as a distance or dimension: '*…the system has one metre resolution…*' it is actually measured as an angle – the apparent separation at which it is just possible to distinguish between two objects. The actual distance apart of the objects will depend on the range that you are observing from, so for example if a device had a resolution of one degree, and two targets were at a range of sixty miles, they could be distinguished if they were one mile apart. At 120 miles range, the same system could only distinguish them at two miles separation, but the resolution would still be one degree. Of course for an orbiting satellite, the observation range may well be fixed, so the resolution distance is meaningful.

d. **Diffraction Criteria**. The strict measurement of resolution, which describes the effect both of wavelength and of aperture, is governed by the **Rayleigh Criterion**. This states that:

---

[8] That is also why two polarising filters held at right-angles appear almost opaque. They are filtering in both possible orientations.

$$\alpha = 1.22\lambda/d$$

Where α is the resolution angle in radians, λ is the wavelength and d is the aperture of the system (λ and d measured in consistent units).[9]

F25. **Bandwidth**. Bandwidth is conceptually the signals equivalent of resolution in reconnaissance, although the word means subtly different things in different areas of signal engineering. For us, it describes the limitation on the amount of data (including voice signals and data streams such as digital information or video signals) that can be fitted onto a given communication channel. An example may make this clearer.

a. A communications signal, such as a voice message, cannot be conveyed directly by radio. Voice messages are not the same thing as the signal carrying them. In order to transmit them by radio, we need to use the voice signals to modify (the usual term is **modulate**) the radio wave in a logical fashion.

b. At the receiving end, we look at the radio signal and using our logical scheme, somehow extract the message content. An unavoidable consequence of modifying the radio signal is that we 'smear' its frequency so that it occupies more of the RF spectrum than a *pure* radio wave (one not carrying information).

c. Where two signals are being transmitted in the same physical area simultaneously, they must be separated sufficiently in frequency for the receiver to distinguish between them. If not, the receiver will receive both signals simultaneously, and will either be unable to decode them, or will decode them in some unintelligible fashion. The degree of separation required, which dictates how many different signals you can fit into a given portion of a frequency band, is driven by the amount of 'smearing' caused in modifying the radio signal.

d. Different kinds of signals smear the carrier radio wave by different amounts, essentially related to the amount of information contained in them. Morse code, for example, which is a very simple signal, hardly smears the carrier at all, so multiple Morse signals can be squeezed close together in a frequency band. Voice and other signals need more 'frequency space', the exact amount depending on the quality of voice required. A video signal, which conveys a vast amount of information compared to audio, requires vastly more bandwidth too.

e. Bandwidth is just the aggregate capacity of a communications channel. Even relatively old communication channels could 'stack' multiple signals together, and modern digital encoding is vastly more efficient at compressing multiple signals together than older analogue techniques. Bandwidth still limits capacity, however, so the user ultimately has to decide between transmitting a few 'bandwidth-greedy' signals or multiple lower-grade (slower or containing less information) alternatives.

---

[9] The Rayleigh criterion is based on the theoretical diffraction patterns created by a pair of point sources (typically of light), with resolution being that condition where the central maximum of one pattern just coincides with the first minimum of the other. There are other empirical resolution definitions in use in some fields, but the Rayleigh definition is widely accepted and yields practical results consistent with the theory.

(INTENTIONALLY BLANK)

ANNEX G – RF FREQUENCY BANDS – FULL LISTINGS

ITU FREQUENCY BANDS

Band Name	Abbreviation	Frequency	Wavelength	Remarks/Typical Use
Extremely Low Frequency	ELF	3-30 Hz	100000-10000 km	Very little practical use, due to enormous wavelengths and resulting large antennae sizes.
Super Low Frequency	SLF	30-300 Hz	10000-1000 km	
Ultra Low Frequency	ULF	300-3000 Hz	1000-100 km	
Very Low Frequency	VLF	3-30 KHz	100-10 km	Communications with submarines. Long range radio-navigation.
Low Frequency	LF	30-300 KHz	10-1 km	Navigation, Commercial Broadcasting.
Medium Frequency	MF	300-3000 KHz	1 km-100 m	Commercial (MW) broadcasting.
High Frequency	HF	3-30 MHz	100-10 m	Long-range communications.
Very High Frequency	VHF	30-300 MHz	10-1 m	TV, Radio broadcasting, communications.
Ultra High Frequency	UHF	300-3000 MHz	1 m-100 mm	TV Broadcasting, GPS, mobile phones, wireless networks, communications. SATCOM. Some radar.
Super High Frequency	SHF	3-30 GHz	100-10 mm	Microwaves and Radar, SATCOM.
Extremely High Frequency	EHF	30-300 GHz	10-1 mm	Some microwave applications.

Illustration G.1 – ITU Frequency Bands

IEEE RADAR BAND NOMENCLATURE (IEEE 521-2002)

Band	Frequency	Remarks
L	1-2 GHz	Overlaps with UHF.[1]
S	2-4 GHz	Partly overlaps with UHF.
C	4-8 GHz	
X	8-12 GHz	
K_U	12-18 GHz	
K	18-26 GHz	
K_A	26-40 GHz	
V	40-75 GHz	Not to be confused with 'VHF' – this is actually in the EHF band.
W	75-111 GHz	

Illustration G.2 – IEEE Radar Band Nomenclature

---

[1] UHF=300-3000 MHz (i.e. 0.3-3 GHz) in ITU compliant definition.

ANNEX H – SATELLITE NAVIGATION – GPS

H1. **How Global Positioning System Works**. In this annex, we look at the inner workings of a Global Positioning System (GPS) receiver, to enable a deeper understanding of a satellite navigation system, and also to give a practical example of how a constellation of satellites is constructed to provide a specific capability. Although we look specifically at the NAVSTAR/GPS system operated by the USA, the other systems use broadly similar principles.

H2. **Navigation by Precise Timing – The One O'clock Gun**. GPS position is based on the receipt of precise timing signals. We will develop our explanation from an analogous land-based method.

a. If you stand anywhere on the Royal Mile in Edinburgh at lunchtime, you will probably be able to hear the firing of the One O'clock Gun from the battlements of Edinburgh Castle. The gun is fired every day (except Sunday) at 13.00 hours. The one thing you can be sure of when you hear the gun, however, is that the time is *not* exactly 13.00. You are hearing the sound of the gun firing, and sound travels through air at about 760mph at sea-level, which equates to about 1 mile in 6 seconds. Since the Royal Mile starts just outside the Castle and leads away from it in roughly a straight line, the precise time you hear the shot determines your range from the gun. If you hear it almost instantly, you are at the Castle entrance, if you hear it at 13.00.06, you are at the other end of the Mile, near the Palace of Holyrood House.

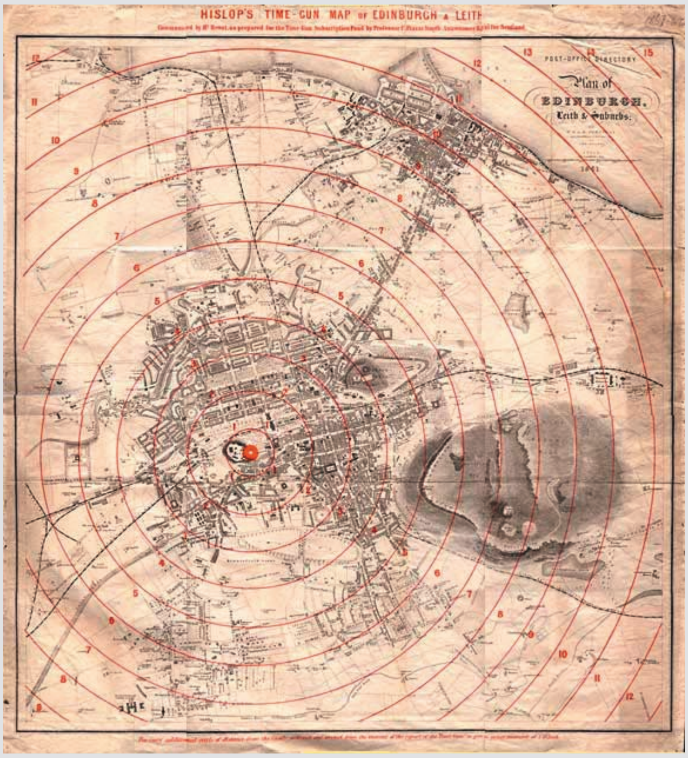


Illustration H.1 – 'Hislop's Time Gun Map of Edinburgh and Leith' (1861). The origin of the time gun signal in Edinburgh lay in allowing ships' navigators to set their chronometers accurately prior to a voyage. Edinburgh terrain and weather did not assist the traditional method of viewing a visual signal from the nearest observatory (in this case the Royal Scottish Observatory on Carlton Hill), and the time gun provided an audible substitute, but only if the error in timing due to the sound propagating could be corrected. This map indicated the appropriate correction. (Image supplied by *Edinphoto*, used by permission)

b. GPS does not rely on deriving position from an intersection of timing and street location; imagine instead that there were multiple One O'clock guns, but that you could somehow distinguish between them. For each shot that you heard, you could work out a range from the respective firing position. Where the range-circles crossed, you could derive your position. Two guns would give you two range circles, though these could cross at two positions. You would either need other information to allow you to discount one of the solutions, or you would need a third gun to resolve the ambiguity. This principle of using extra information is very important to GPS, as we shall shortly see.

c. GPS does not use sound – it uses radio time-signals, which travel faster and further. Radio signals propagate at the same speed as light, roughly 300 000 kilometres (186 000 miles) per second. The analogy would be timing sight of the muzzle-flash from the One O'clock gun, which also propagates at the speed of light, instead of listening for the sound of the shot. Note that it is still a finite time after 13.00 when you see the flash, albeit not one that could be resolved with a wrist-watch. The GPS constellation transmits multiple time signals – each satellite repeatedly transmits its own identity and time signal, referenced to an extremely accurate master atomic clock, along with additional information about the overall constellation. The GPS receiver decodes the timing signals, compares them with its own clock, and uses the delay in reception to deduce position by calculating its range from several satellites simultaneously.

d. There are flaws in the system as described so far. Our Edinburgh analogy assumed that our watch was set in exact agreement with the Castle Gunner's, and that we knew exactly where the gun was located, since that is where we measure our time-position from. Both these issues raise practical problems.

> (1) Atomic clocks are expensive, and are not readily transportable. Our GPS receiver needs to compare the time it receives a given signal with true time, ideally using a robust, affordable and portable clock, which will inevitably be less accurate than the master clock used by the satellites. It overcomes this by combining several satellite signals. A signal from a single satellite fixes our position in one dimension only, in the same way as one gun fixed our position along the length of a street only. Two guns would fix our position in two dimensions (there would be ambiguous solutions, since two circles can intersect in two positions, but we assume we can resolve this). Three signals fix our position in three dimensions, but GPS resolves timing error by taking a fourth signal from a fourth satellite. Although four signals do not tell you directly what the receiver's clock error is, the GPS receiver can safely assume that since the satellite clocks are all referenced to each other, at any instant the timing error is the same for all four signals. This allows the error to be calculated. This co-incidentally also means that the GPS receiver has access to very precise time data, which it can display, along with its position. We will see that for many applications, timing information is as important as position to the user.
>
> (2) If we are going to fix position by calculating our range from the GPS satellites, we are assuming that we know the positions of the satellites to a high level of accuracy. This information is included in the additional signal content that the satellites transmit. A GPS receiver contains almanac information for the GPS constellation,[1] but we have noted that satellites drift from their ideal location. GPS

---

[1] Essentially, the almanac is a full set of orbit elements for each of the active satellites in the GPS system, from which the receiver can determine which satellites should be in range for any given approximate time and position. Each satellite also transmits its

satellites may also suffer from small degrees of drift in their internal clock. The corrections for these errors are transmitted to the receiver along with the timing signal.

e. To determine the errors in satellite position and clock time, the GPS constellation is arranged so that individual satellites pass within range of designated radar dishes on the ground at discrete intervals. Each time they do so, their true position can be measured very accurately, and any clock error they are suffering from determined. Appropriate corrections can then be added to the data transmitted to the receivers.

f. There are still errors inherent in the system we have described. We have assumed that light (and the satellite radio signal) travels at a constant, known, speed, but this is only true in a vacuum.[2] The speed is reduced slightly when the signal travels through air, and the path lengths of the signal through the effective vacuum of space and the steadily thickening atmosphere nearer the ground, must therefore be allowed for separately for each signal. The greatest part of this error is caused by passage through the layer of the atmosphere known as the ionosphere; the error is thus called **ionospheric error**. The receiver carries a mathematical model of the atmosphere, which allows a correction to be calculated.[3] Receivers can also suffer from **multi-path errors**, due to receiving direct and reflected signals from the same satellite, typically caused by reflection from buildings in urban areas. Moving systems in vehicles suffer relatively less from multi-path errors, as the reflected signals vary rapidly as the vehicle moves, causing the false signal to be rejected. For slow-moving or static systems, multi-path errors are a greater problem, though they can be mitigated in the receiver by signal-processing techniques and careful antenna design.

H3. **The GPS Satellite Constellation**. The requirements outlined above lead to the design of the constellation.

a. From H2d(1), we need to ensure that a minimum of four satellites are visible (i.e. above the horizon) simultaneously on a global basis. This favours satellites in higher-altitude orbits, which have larger footprints and thus achieve the coverage with a practical minimum number of satellites. It also implies inclined orbits to provide visibility at higher latitudes.

b. Stable orbits minimise the effects of perturbations and consequent need to manoeuvre to hold station. This too favours higher-altitude orbits, immune from atmospheric drag.

c. Although we have not discussed it above, accuracy is also improved if the satellites are in different directions as seen by the user. This improves the geometry of the intersections between the range measurements. Range information from two or more

---

'ephemeris' – a high precision description of its own orbit in sequence with the overall almanac. See Paragraph H6b for further details.

[2] See Paragraphs F5c, and F18.

[3] In Paragraph H6f, we look at individual components of the GPS signal, which are transmitted on slightly separated frequencies. By comparing the delay in arrival between the differing components of the signal (atmospheric effects vary slightly with frequency), a more accurate correction for their total effect can be derived.

satellites that lie in the same direction provides a relatively poor quality fix. This implies placing the satellites in multiple orbital planes.[4]

d. The orbits chosen must finally have ground tracks that bring the satellites into the range of fixed ground radar stations on a regular basis. This implies some closed ground path. In practice a semi-synchronous orbit is used.

The combination of all these factors dictates the constellation. A minimum of 24 satellites are required for full operation, with 6 satellites evenly spaced in each of 4 orbits, separated by 90° of longitude. The orbits are circular MEOs, inclined at 55°.[5]

H4. **Differential GPS**. We have looked at the various corrections applied to the ‘raw’ GPS signal to create the accuracy of the final system, from updating the satellite positions and the clock errors, through to modelling the atmosphere, but there are still residual errors that cannot be removed. These are typically due to residual satellite position or clock errors, and un-modelled atmospheric fluctuations. One way of addressing this is **differential GPS**. This relies on transmitting a local correction signal separate from the GPS transmissions. A GPS receiver is placed in an accurately surveyed location. The position it receives from the GPS signal is compared with the known surveyed position and the resulting difference is transmitted on a low power line-of-sight radio. This assumes that the residual errors are roughly constant for a given area, allowing their local correction.

H5. **Wide Area Augmentation**. In some areas of the world, GEO satellites transmit the differential GPS correction signal. In the US, the **Wide Area Augmentation System (WAAS)** corrects for GPS signal errors caused by ionospheric disturbances, timing and satellite orbit errors, while providing vital integrity information regarding the health of each GPS satellite (a sudden change in the apparent GPS position of a fixed site might well be an indication of a satellite fault). WAAS consists of approximately 25 ground reference stations positioned across the United States that monitor GPS satellite data. Two master stations, located on either coast collect data from the reference stations and create a GPS correction message. This correction accounts for GPS satellite orbit and clock drift caused by the atmosphere. The corrected differential message is then broadcast through one of two GEO satellites. The information is compatible with the basic GPS signal structure, which means that any WAAS-enabled GPS receiver can read the signal. Other governments are developing similar satellite-based differential systems: in Asia the Japanese **Multi-Functional Satellite Augmentation System (MSAS)**, in Europe the **Euro Geostationary Navigation Overlay System (EGNOS)**.

H6. **GPS Signal Format and Anti-Jam Considerations**. Users think of a GPS Satellite as transmitting ‘the GPS signal’. In fact, the satellite transmits several over-lapping, mutually-dependent signals, which support the varying levels of service provided for different users. Together, they enable the highest level of system accuracy for selected users while at the same time providing a measure of jamming and spoofing protection. The GPS signal format continues to evolve; a user with a need to understand the very latest capability should seek advice either in the

---

[4] If this does not seem clear, imagine putting all the GPS satellites e.g. in the equatorial plane. The closer the user was to the equator, the nearer the range measurements would resemble straight lines running north-south. These would provide very shallow cross-cuts, where a very small fluctuation in the measured range would give a large variation in apparent position.

[5] There are typically more than 24 satellites in orbit at once, both as part of routine constellation maintenance and to provide redundancy in the event of temporary or permanent failure of a satellite. There have been up to 32 satellites in orbit simultaneously in the past. Recent (2009) public debate about the USAF’s commitment to maintaining the GPS Constellation has included the assertion by senior USAF leadership that 27 satellites provides well over 95% confidence in maintaining the required system accuracy and availability at all times.

professional literature or from an appropriate MOD authority, such as Dstl or the Air Warfare Centre. In order to give some idea of the system's complexity, while at the same time illustrating the principles employed, we look now at the various components of the signal.

a. **Signal Content**. The GPS signal includes two main pieces of information; the navigation message containing ephemeris information and other data about the satellite constellation, and ranging information to allow calculation of the distance from the satellite to the receiver. Both signal components are complex, and both are subject to improvement programmes.

b. **The Navigation Message**. The content of the navigation message is conceptually simple. It tells the receiver four key pieces of information: the time of transmission of the message, the ephemeris information, the state of health of the broadcasting satellite and an almanac message for the whole GPS constellation. Typically, ephemeris information is highly perishable, with a useful life measured in hours. Almanac data is longer-lasting, typical validity is measured in weeks. A potential drawback of the navigation message, from the system designer's point of view, is that it is lengthy, and is transmitted relatively slowly. If the receiver misses any portion of it, and has to wait for that element to reappear in the cyclical message, either there is a delay in providing position data, or a reduction in accuracy. For demanding GPS applications, this can be a challenge. Enhancements to the navigation message centre on providing higher data rates, improved error correction, and possible support of future constellation upgrades, including greater numbers of satellites. Originally known simply as the **'NAV'** component of the system, the latest upgrade, supported by recently-launched GPS satellites, is known as **'CNAV'**. There is also an enhanced message for military users known as **'MNAV'**.

c. **Ranging Signals**. The ranging signal generated at the satellite uses what is known as a **pseudo-random number (PRN)** code (sometimes also referred to as **pseudo-noise (PN)** code). This is a sequence of numbers that appear at first sight to be random, (for example individual digits will appear evenly over a long-enough period of time), but which in fact is generated in a mathematically predictable way. Each separate GPS satellite has a unique PRN code assigned to it.

(1) The key or 'seed' for each satellite's broadcast PRN code is its onboard clock time. At the receiver end, one can generate a similar PRN code based on the receiver clock time, which will not match the received satellite code because of the time of flight of the signal from the satellite (simplistically, for example, if you are generating the 13:00:00 hrs code on the ground, but comparing it to the 12:59:59 code sent by the satellite one second previously, these will not match). Now assume you can change your receiver's clock time to incorporate a delay. By trying different offset clock times to generate the receiver-end PRN code, you will suddenly arrive at a match between the generated and received codes. The nature of the PRN code ensures that you either get a minimal or an exact match between the signals, and the amount of clock-offset required to generate the match is the time of flight of the signal. In our fictitious example above, when you tried a 1 second delay, so that you generated the 12:59:59 PRN code at exactly 13:00:00, the 13:00:00 code at 13:00:01 and so on, you would get a good match, and deduce that the time of flight of the signal was 1 second. As the user and the satellite move relative to each other, and the time of flight changes, you will lose the correlation and need to try another delay length to regain it. In the real case, you plainly need to

use a much faster correlation check and clock-step, as in practice you will be trying to measure flight times of milli-seconds or less, and detect changes of nano-seconds or less.

(2) As a further refinement, each individual GPS satellite's PRN code will not match other signals in the constellation (the signals are said to be **orthogonal**). This means that the multiple signals from the GPS satellites can all be transmitted on the same frequency. To extract the signal of a specific satellite from the competing signals and noise, all the receiver needs to do is look for something that matches that particular satellite's PRN code. The orthogonality of the codes means you do not get a false match from another satellite's transmissions.

d. **C/A Code, P-code, W-code and P(Y)-code**. The original GPS signal transmitted from the satellites consisted of two components: the **Coarse/Acquisition** code (**C/A code**) and a **precision code** (**P-code**). C/A code deliberately used a relatively simple PRN generator, which meant that the total number of possible combinations of random digits repeated relatively quickly. The process (algorithm) used to generate C/A code was, and is, public knowledge, the coding aspect of it being employed to allow receivers to reject noise, and discriminate between multiple satellite signals. The short sequence repetition interval meant that GPS receivers could generate a match with the received signal relatively quickly. P-code is a more sophisticated version of C/A code, using a more complex algorithm with a much longer interval between repetitions.[6] P-code algorithms are then protected by an extra layer of encryption, known as **W-code**, which is superimposed on P-code to generate **P(Y) code**. P(Y) code is what is transmitted by a GPS satellite, and W-code, the details of which are classified, is what makes P(Y) code available only to selected (principally military) users. The dual signal also enabled a military receiver to acquire rough position and time information from C/A code (hence the 'acquisition' tag), and use that to generate a correlation with the P(Y) code.

e. **M-code**. In the same way as CNAV and MNAV messages are improvements on the original NAV message, M-code is an enhancement of P(Y) code. It is designed to improve anti-jam and anti-spoofing aspects of P(Y) code, and at the same time allow direct access to enabled users without the need to acquire C/A position first. This is known as **direct-acquisition**. There are significant challenges in integrating the M-code signal with existing signals; there are also aspirations to incorporate a civil dual-signal mode similar to M-code, because of the technical advantages inherent in dual-frequency operations.

f. **Carrier Frequencies.** The GPS signals from the satellites are all carried on two adjacent radio frequencies, shared by the entire constellation. These are known as the **L1** and **L2 frequencies**.[7] Initially, the C/A signal was carried on L1 only, and the P(Y) signal on L1 and L2. This allowed receivers that could read the P(Y) signal to generate improved corrections for delay in signal transmission through the Earth's atmosphere by comparing the P(Y) signal on each frequency, since the atmospheric effects were slightly frequency-

---

[6] Among other things, this makes the range signals unambiguous at very long ranges, typically out to the limits of the Solar System. Every time the C/A code cycle repeats, a potential correlation exists, but for normal applications, the receiver can discount them. In our numerical example, if we suppose that C/A code repeated every 10 seconds, we would not only get a match for a 1-second delay, we would also get it at 11, 21, 31… seconds. Given the speed of light, however, for terrestrial ranges we could safely assume the shortest match. At longer ranges, we could not necessarily discount those higher order delays. The real C/A code repeats every millisecond, and the real P-code repeats about once per week.

[7] A third frequency, perversely to be known as the **L5 frequency**, will be transmitted by new satellites to support certain safety-of-life signals.

dependent. The various NAV messages are superimposed on the C/A and the P(Y) signals. M-code will use novel techniques to accommodate the extra signal within the existing frequencies, and eventually the second civil signal will have to be accommodated too.

g. **Jamming and Spoofing**. A GPS receiver has to extract a relatively faint signal from a noisy environment, which includes the signals from all the other satellites in the constellation as well as random interference. We noted above how the signal format allowed receivers to distinguish between signals from different satellites carried on the same frequency. As well as serving to extract competing satellite signals, this also provides limited protection against low-power jamming. Plainly, however, it is not a complete panacea. Note also that because first-generation military receivers need to acquire C/A signals before being able to lock onto the P(Y) code, jamming the C/A code would prevent them operating. Direct acquisition of M-code signals will thus have significant anti-jam advantages. Spoofing a GPS receiver, particularly one that only used the C/A signal, would also be a challenging process, although again well within the capability of a technically adept adversary. There are several discussions of possibilities and practicalities in the technical literature; a web-search should reveal the latest public information.[8]

---

[8] Technical discussions include plausible intentional self-deception. For example, fishing vessels are regularly fitted with sealed data-logging devices that use GPS receivers to track their route and fishing patterns when at sea. Unfortunately for the regulating authority, some boats acquire GPS-spoofing devices that feed an erroneous signal to the logging devices, constructing a track that satisfies the regulations, while bearing little resemblance to where the boat is actually fishing.

(INTENTIONALLY BLANK)

BIBLIOGRAPHY

Bib 1. **Introduction**. This Primer has avoided strict referencing conventions, but readers may wish to pursue a specific topic in greater depth than has been possible here. This bibliography should provide a useful starting point. However; it makes no claim to being comprehensive; it also inevitably reflects personal preferences, and inclusion or omission of a work should not be construed as constituting Ministry of Defence (MOD) or Development, Concepts and Doctrine Centre (DCDC) endorsement or censure. Website access has been checked during January 2009, but no guarantee can be provided as to content, currency or status, and again inclusion does not constitute official endorsement of the content.

Bib 2. **General Texts**:

Lyn Dutton, David De Garis et al, *Military Space,* (London: Brassey's (UK) 1990) – A UK military book; the authors worked in the Department of Air Warfare at RAFC Cranwell. Published in 1990; Chapters 2 and 3 match Chapter 1 of the Primer, 4 matches Chapter 2 of the Primer, 5-9 and 11-12 match Chapter 3 of the Primer, and 10 covers part of Primer Chapter 4. Almost no mathematics, and Chapter 2 provides a particularly good summary of orbits and ground tracks.

Walter A McDougall, …*The Heavens and the Earth: A Political History of the Space Age*, (Baltimore: Johns Hopkins University Press 1985) – The definitive history of early and Cold War space policy, won the Pulitzer Prize for History in 1986. Most relevant to themes in Chapter 4, but also puts elements of Chapter 1 in a historical context.

***The Cambridge Encyclopedia of Space: Missions, Applications and Exploration,* (Cambridge: CUP 2003)** – Covers much of the same ground as this primer, including a chapter on military missions. It includes a comprehensive list of further printed and online references. CUP have produced several volumes with similar but not identical titles over the years, and may well replace this one with another related volume.

Bib 3. **Chapter 1 – Space Fundamentals**.

Jerry Jon Sellers et al, *Understanding Space: An Introduction to Astronautics* (3rd Ed), (New York: McGraw-Hill 2005) – A good general introduction to the material covered in Chapter 1, though with no bias towards military applications, and consequently covering interplanetary space in some depth. Also contains good historical illustrations, though most topics are covered more mathematically than here (up to about A-level). It can be read easily without following the maths. Part of a series endorsed by US DoD and NASA, and edited by a member of the faculty at the USAF Academy. There is also a dedicated website to support the book.

Thomas F Tascione, *Introduction to the Space Environment* (2nd Ed), (Malabar, Fl: Krieger Publishing 1994) – A university-level introduction to all aspects of the Space environment, including such matters as its effects on radio propagation. It can be used for reference, and it puts some of the scientific content in a historical context too.

Peter Fortescue, John Stark, Graham Swinerd (eds), *Spacecraft Systems Engineering* (3rd Ed), (Chichester: Wiley, 2003) – A UK undergraduate engineering text covering all major aspects of spacecraft design, such as power systems, attitude control, thermal management. Also covers the design implications of the space environment, and the principles of rocket motors and engines. Assumes undergraduate mathematics to understand the detail, but could be read for a specific topic without following the calculations.

Roger R Bate, Donald D Mueller et al, ***Fundamentals of Astrodynamics*** **(New York: Dover Publications, 1971)** – The USAF Academy text on astrodynamics. Not for the mathematically faint-hearted, but if you ever want or need to prove any of the results 'plucked from thin air' in Chapter 1, they will be in here somewhere.

James R Wertz and Wiley J Larson, ***Space Mission Analysis and Design*** **(3rd Ed) (New York, NY & Hawthorne, CA: joint publishers Springer & Microcosm Press, 1999)** – A heavyweight text covering all aspects of planning and designing a satellite, starting from the mission analysis, through payload, satellite and orbit selection up to production engineering. Endorsed by both the US DoD and NASA; very comprehensive. The level of mathematical and technical sophistication assumed varies slightly across the chapters, according to individual chapter-authors' preferences, but as a 'one-stop shop' for everything except launchers it is hard to beat. The 4th Edition is in production, tentatively scheduled for late-2010 publication.

There are good online history resources for the Early Space-Age, which illustrate some of the systems discussed in Chapter 1. The White Sands Missile Range, where amongst other activities, the USA tested captured V2 rockets after World War 2, is well documented at http://www.wsmr.army.mil (the official US Army website for the installation – history is under 'public affairs') and at http://www.wsmr-history.org/History.htm which is the official base museum website. NASA's official history resources are extensive, and can be found at http://history.nasa.gov. Although NASA's focus is on Civil Space, this site provides considerable detail on US Space policy and technology through the years.

Readers may wish to explore the implications of orbits and ground tracks. Software packages to model these are available, though they vary between 'enthusiast' systems of variable reliability to professional tools of great complexity (and cost). **'Satellite Tool Kit'** (STK), produced by Analytical Graphics Industries (AGI) is commonly encountered. There is a basic version of the program available free for simple visualization, though even this demands some grasp of terminology. See www.agi.com for details. Other 'planetarium' type applications aimed at amateur astronomers may also incorporate satellite modelling. Any reader who would like to try tracking satellites visually should visit www.heavens-above.com, where accurate predictions of the brighter satellites can be generated for any location and time. The site also includes a continuously updated visualisation of the location of the International Space Station (a very easy target for visual tracking when it is above the horizon). Trying this even once provides a very quick appreciation of the duration of typical pass times from LEO and of fields of view, particularly if it can be seen on two adjacent passes (about 90-100 minutes apart).

Sellers (above) provides an introduction to rocket propulsion from a technical viewpoint. Readers more interested in the history of rocket development in the 20th Century, including policy considerations and some of the personalities involved, will find J D Hunley's two-volume history invaluable: **J D Hunley,** ***Preludes to US Space-launch Vehicle Technology: Goddard's Rockets to Minuteman III*** **and** ***US Space-launch Vehicle Technology: Viking to Space Shuttle*** **(Talahasee: University Press of Florida 2008)**

Chapter 2 – Space and the Law.

Malcolm Shaw, *International Law* (5th Ed), (Cambridge: CUP 2002) – A well-known textbook on International Law, Chapter 9 covers Air and Space Law. Other chapters are good for the fundamental principles of international law, and how treaties work, and additionally cover such topics as the Use of Force and Maritime Law.

The Judge Advocate General's Department of the USAF, *Air Force Operations & the Law: A Guide for Air and Space Forces*, (Washington DC, 2009) – a comprehensive (600 page) US Military Manual, the latest edition has just been published.[1] The USAF interpretation of Space Law is in Chapter 5, (alongside the Chapter covering Cyberspace). It also includes material in its Chapter 13on the Unified Command Plan and US C2 arrangements, which may be useful for understanding parts of Primer Chapter 4. The full text of the manual can be downloaded from www.afjag.af.mil/library/.

Bib 4. **Chapter 3 – Military Uses of Space**. There are very few comprehensive references for military applications in space, mainly for reasons of classification. Some material is available on-line, and some specific programmes have received academic attention, either as details have been de-classified over time, or via Freedom of Information Act requests. Space Control, and BMD and ASAT technologies, have been of particular interest. Online encyclopædias such as 'wikipedia' contain entries for most named satellites and programmes, though the usual caveats about accuracy apply.

a. **General Reference**. **Clayton K S Chun, *Defending Space: US Anti-Satellite Warfare and Space Weaponry,* (Oxford: Osprey Publishing 2006)**. Dr Chun is on the faculty of the US Army War College at Carlisle, PA. This is a US-focused analysis of military space, which provides brief but comprehensive coverage of most military applications. Copiously illustrated if you want to see what the hardware actually looks like. Only 64 pages, and relatively affordable! Its bibliography may provide further leads for specific topics.

Details of **Eligar Sadeh**'s publication ***Space Politics and Policy*** can be found below in the paragraph for Primer Chapter 4. Chapters 16 and 17 cover the development of Military and Intelligence systems, and provide a good US perspective on these areas.

b. **Ballistic Missiles**. **Mark Berhow and Chris Taylor, *US Strategic and Defensive Missile Systems 1950-2004,* (Oxford: Osprey Publishing, 2005)**. Uniform with Dr Chun's text on anti-satellite systems referenced above, so again short and well illustrated. Written for the general reader, though the bibliography includes more detailed works. Other volumes in the series cover specific 'rest of the world' systems.

There are also numerous volumes and journal articles at all levels of complexity about nuclear deterrence, which implicitly analyse the wider implications of ballistic missiles.

c. **Data Networks in Space**.

(1) **Communications**. There are numerous technical publications about satellite communications principles. Dutton's 'Military Space' devotes a chapter to it (current

---

[1] There are substantial layout differences between the 2002 and 2009 editions, so the Chapter references do not apply to the 2002 edition.

in 1990), and a detailed engineering description of a communications payload is in Chapter 12 of Fortescue, Stark *et al's* 'Spacecraft Systems Engineering', both referenced above.

DE&S Information Systems and Services, *Information Systems and Services Publication No 5 (ISSP 5)* (DE&S: Version 0.3 dated August 2008). Military readers can consult this publication, which is available on the Defence Intranet. It is 'owned' by the GCS IPT and is updated annually. It contains both technical and operational details of UK Satellite Communication Services.

www.paradigmsecure.com (*not* www.paradigm.com) contains some details both of the UK arrangement for Commercial/PFI provision of military satcom, and of the Skynet constellation, although the information is scattered across a fairly complex website.

www.inmarsat.com includes a page on 'maritime safety' that describes the GMDSS maritime safety regime and the various methods of compliance. Greater detail specifically of the GMDSS regime can be found by searching at www.imo.org.

(2) **Position, Navigation and Timing**. The major GPS receiver manufacturers distribute information about the operating principles of GPS.

www.trimble.com includes an on-line tutorial on GPS operations at about the level of this Primer. The 'military systems' page includes illustrations of typical modern military applications.

www.garmin.com includes a downloadable 9 page booklet on GPS principles, less detailed than this primer but providing a useful introduction.

(3) **Surveillance from Space**. Classification limits open discussion of operational systems, but references exist for basic principles.

Jeffrey Richelson, *America's Space Sentinels: DSP Satellites and National Security* (Kansas, MO: University Press of Kansas, 2001). This is an example of the academic/historical analysis alluded to above that is now available on early security missions. It provides a grounding in the principles of IR surveillance of Earth and the targets that might be detected, and includes analysis of the operational employment of DSP.

For many years, Dr Nicholas Short at the NASA/Goddard Flight Center at Greenbelt, MD conducted courses on 'Remote Sensing'. Dr Short retired and the courses ceased, but his course material is still available on-line, comprising an extensive on-line teaching resource. It is no longer updated, but much is timeless and it provides a comprehensive background to the principles of reconnaissance from space. Dr Short's contact details are maintained via the website, and he will supply DVD copies of the package at cost. See also the caveats on the website about unauthorised commercial copies in circulation. The tutorial is at: www.rst.gsfc.nasa.gov

A hard-copy companion to the GSFC material is **Dr George Joseph, *Fundamentals of Remote Sensing* (2nd Edition), (Hyderabad: Universities Press (India), 2005).** This is an undergraduate-level text (assumes some physics and mathematics), based on the activities of the Indian Space Research Organisation. It is a very readable introduction to the principles of remote sensing, with a bias towards earth-resources applications, but providing excellent background on aspects such as sensor performance, multi-spectral and hyper-spectral imaging and some of the engineering implications of these roles.

Readers needing details of synthetic aperture radar (SAR) principles will find multiple sources available. Radar theory is taught at the Air Warfare Centre, as part of various courses, including the GD Aerosystems Course.

There was a good introduction to SAR in Chapter 14 of the *second* edition of **Merrill I Skolnik, *Introduction to Radar Systems* (New York: McGraw-Hill, 1980).** Unfortunately, it has been eliminated from the latest edition **(New York: McGraw-Hill HE, 2001)**. The second edition may still be found in some libraries.

John C Toomay and Paul J Hannen, *Radar Principles for the Non-specialist,* (New York: SciTech Publishing, 2004), also provides an introduction to the topic, though both Toomay and Skolnik assume some mathematical and engineering background.

The principles of Cospas/Sarsat are at www.cospas-sarsat.org. The site includes both details of the fixing principle and the satellite constellation that enables the mission, as well as regular updates on the services supported and mission results.

Bib 5. **Chapter 4 – Space in Modern Society**. Because of its content, Chapter 4, and particularly Sections V and VI, contain more direct references than the rest of the primer. These should provide points of departure for detailed exploration of topics, and the bibliographies of some of those publications will provide onwards references. Additionally, the following may be useful for general or specific needs:

a. **General Reading.** As a general introduction to Chapter 4 themes, **Eligar Sadeh (ed), *Space Politics and Policy: an Evolutionary Perspective,* (Dordrecht: Kluwer Academic Publishers 2002)** may be helpful. (This reference is the European printing – the Springer (US) edition may also be found) This heavyweight academic tome provides much detail on Chapter 4 themes, primarily from a US viewpoint. Part 1 covers the historical background to space policy formulation, Part 2 details the various organisational players in US Space Policy and Part 3 covers various areas of operation in space – civil, commercial, military and intelligence based. Specific chapters are highlighted where relevant, but overall, this is also a sound generic introduction to the area, with contributions from many authors.

b. **Section II – Thinking Strategically about Space.** The Section contains direct references to cited texts, which are not repeated here. Additionally, two academic journals cover the subject area exclusively, *Space Policy,* and *Astropolitics*; other academic journals covering strategic studies may also yield relevant articles. The USAF School of Advanced Air and Space Studies at Maxwell AFB, Alabama is also a regular source of Masters-level theses on space topics, many of which are accessible on-line, either via a public search engine, or via the built in search facility at: http://www.au.af.mil/au/schools.asp

c. **Section III – Dependencies on Space. UKspace, *Space secures prosperity: the Security and Defence Case4space* (London: Newsdesk Communications, 2008)**. This report was issued in 2008 (launched at a RUSI Conference on the National Security implications of Space). The section on 'Critical National Infrastructure' provides a UK view on overall dependencies. For access to its content, contact UKspace through their website www.ukspace.org.

d. **Section IV – Commercial Uses of Space**.

(1) **BNSC, *UK Civil Space Strategy: 2008-2012 and beyond*, (London: BNSC, 2008)**. This publication provides a good overview of the UK's perspective on commercial space activity. It is one of a series updated occasionally. Copies can be ordered and/or downloaded via www.bnsc.gov.uk. BNSC publish a magazine **space:uk** which covers all sectors of UK space activity. Finally, the website is well supplied with news, statistical detail and other reports.

(2) **Space Foundation Inc, *The Space Report 2009* (Colorado Springs, CO: Space Foundation, 2009).** In the USA, the Space Foundation tracks global space activity, and among many other activities publishes an annual review. Although covering all sectors of space at some level, it concentrates on civil and commercial space, including business and financial analysis. The executive summary is available free online, and the full report can be purchased at the Space Foundation website, www.spacefoundation.org. The US weekly journal *Space News* (published by Imaginova Inc – see www.space.com/spacenews) also provides much commercial and civil space coverage.

(3) Sadeh, Chapter 10-11 covers the economic and commercial aspects of space, though note the publication date. Chapter 14 is a very detailed account of 'Space Business', covering the development of the US rocket launcher industry as well as the business implications of defence and intelligence contracts. This coincidentally provides background on missile and military applications, relevant to Primer Chapter 3 above, and additionally explores the implications of the growing international capability in space.

e. **Section V – 'Civil' (Scientific) Space**. Most of the references and online resources for commercial space also cover civil space. Details of specific missions or research areas can also be explored via online encyclopaedias. Active NASA and international missions invariably have their own websites providing news and links to results. White Sands Missile Range and its associated museum maintain separate websites which provide details of the science programs associated with the exploitation of captured V2s. See www.wsmr.army.mil and www.wsmr-history.org for details.

f. **Section VI – Military Space Doctrine and Organisation**. The relevant doctrinal publications are highlighted throughout this section. Unclassified US publications can be located on the internet. UK Doctrine publications can either be found on the Defence Intranet, or, if publically accessible, via the DCDC internet website. See also the reference to the USAF Operational Law manual above for Chapter 2, which outlines US Space C2. Congressional and Presidential review panels have received media analysis and reporting in the Aerospace and Defence specialist press such as *Space News* and *Jane's Defence Weekly.*

g. **Section VII - Civil Space Capability and Governance**. Most of the organisations and companies cited in the text maintain websites, which should be consulted for details of current activities. The BNSC and Ukspace annual publications highlighted above provide commentary on significant contract awards, deliveries and launches in the UK and Europe, and the Space Foundation reports cover similar ground in the USA.

ACKNOWLEDGEMENTS

This book was written at the UK Development, Concepts and Doctrine Centre during 2008-10. I am grateful for the support and encouragement afforded to my by the senior staff who tolerated me spending long periods in consultation, fact-finding and writing, despite the fact that the output did not fit any known template for Military Concepts or Doctrine and consequently required a large leap of faith that it would fill a need and find a readership. I hope that the results justify that faith.

I could not have written it alone, and wish to thank those who assisted by commenting on drafts, sharing their expertise and in one case giving it to their 12 year old daughter to read! There is not enough room here for me to name them all, but I must extend special thanks to Mr David Beard from Dstl, Dr Stuart Eves at SSTL, Wing Commander Martin Littlehales from the Directorate of Air Staff in the Ministry of Defence, Mr Chris Otley from HMGCC, and Squadron Leader Mal Stainforth from the Air Warfare Centre at Cranwell, all of whom gave freely of their time and made specific contributions to the text.

Finally, Ms Lauren Jones from the SERCO Graphics Section at Shrivenham and Mr Robin Jennings from DCDC produced the elegant diagrams that (hopefully) clarify the impenetrable parts of the text.

Gerry Doyle

(INTENTIONALLY BLANK)

LEXICON OF ACRONYMS AND ABBREVIATIONS

NOTES:

1. **Generic and Specific Acronyms/Abbreviations**. Some abbreviations listed below are generic, with wide meaning; others are specific, usually to a nation or a particular service or organisation. The latter case is indicated by prefixing or including the appropriate qualification in brackets in the description.

2. **'Named' Space Missions**. Names given to space missions, or series of missions are often descriptive (e.g. CORONA, Hubble Space Telescope), rather than serving as abbreviations or acronyms, though there are exceptions (e.g. TDRSS). Strict abbreviations are listed below, mission names are listed in the index where appropriate.

A	
AAA	Anti-aircraft artillery
ABL	Airborne Laser (US technology demonstrator program)
ACAS	(Royal Air Force) Assistant Chief of the Air Staff
AFB	(United States) Air Force Base
AFDD	(US) Air Force Doctrine Document
AFM	(US) Air Forces Manual
AIAA	American Institute of Aeronautics and Astronautics
ASAT	Anti-satellite
B	
BCE	Before Common Era (relating to dates)
BMD	Ballistic Missile Defence/Defense
BNSC	British National Space Centre
C	
c	Symbol used to denote the speed of light.
C/A	Coarse Acquisition (code or signal within the GPS system)
CCD	Charge-coupled device (variety of imaging electronics)
CE	Common Era (relating to dates)
CIO J6	UK MOD staff division acting as Joint User for Military SATCOM
CMOS	Complimentary metal-oxide semiconductor (variety of imaging electronics)
COPUOS	(United Nations) Committee on the Peaceful Use of Outer Space
D	
DARPA	(US) Defense Advanced Research Programs Agency
DCDC	(UK MOD) Development, Concepts and Doctrine Centre
DLOD	Defence Line of Development (UK construct for military capability development).
DSP	Defense Support Program (US missile-launch warning system)

DSTL	(The) Defence Science and Technology Laboratory (UK MOD Organisation)
E	
EGNOS	European Geostationary Navigation Overlay System (relating to SATNAV)
EHF	Extremely High Frequency (ITU Frequency Band Designation)
ELF	Extremely Low Frequency (ITU Frequency Band Designation)
EM	Electromagnetic (radiation)
EMP	Electromagnetic pulse
ESA	European Space Agency
ESSS	European Space Surveillance System
F	
FAI	Federation Aeronautique International
ft	Feet
G	
GEO	Geo-stationary Orbit
GMDSS	Global Maritime Distress and Safety Service
GPS	Global Positioning System
H	
HEO	Highly-elliptical orbit
HF	High Frequency (ITU Frequency Band Designation)
HMGCC	Her Majesty's Government Communications Centre
hrs	Hours
HST	Hubble Space Telescope
Hz	Hertz (measurement of frequency) (for EM radiation, usually qualified with a metric system prefix, e.g. kHz, MHz…)
I	
ICAO	International Civil Aviation Organisation
ICBM	Intercontinental ballistic missile
IEEE	Institute of Electrical and Electronics Engineers
IN	Inertial Navigation
INF	Intermediate-range Nuclear Forces (treaty)
IR	Infrared (radiation)
IRBM	Intermediate-range ballistic missile
ISAR	Inverse synthetic aperture radar
ISS	International Space Station
ISTAR	Intelligence, surveillance, target acquisition and reconnaissance (UK terminology)

ITU	International Telecommunications Union
J	
JARIC	(*previously*) abbreviation for the (UK) Joint Air Reconnaissance Intelligence Centre. Now used as a title within 'JARIC – the National Imagery Exploitation Centre', part of the UK Intelligence Collection Group but describing the continuing organisation.
JFCC	(US) Joint Functional Component Command (qualified by specific role, e.g. JFCC-IMD for Integrated Missile Defense)
JP	(US) Joint Publication
JPL	Jet Propulsion Laboratory (division of NASA)
JSpOC	(US) Joint Space Operations Center
K	
km	Kilometre
KSC	Kennedy Space Center (NASA facility)
L	
LEO	Low-earth orbit
LF	Low-frequency (ITU Frequency Band Designation)
LW	Long-wave (description of radio frequency band)
M	
MDA	(US) Missile Defense Agency
MEO	Medium-Earth orbit
MF	Medium frequency (ITU Frequency Band Designation)
mins	Minutes
MLRS	Multiple launch rocket system
MOD	(UK) Ministry of Defence
MRBM	Medium-range ballistic missile
MSAS	Multi-function Satellite Augmentation System (relating to SATNAV)
MTCR	Missile Technology Control Regime
MW	Medium-wave (description of radio frequency band)
N	
NASA	National Aeronautics and Space Administration (US government agency)
nm	Nautical miles
NOAA	National Oceanographic and Atmospheric Administration (US government agency)
NORAD	North American Air (*later* Aerospace) Defense Command
NSSP	(UK) National Space Security Policy
NUDET	Nuclear detonation (detection)

O	
OMS	Orbital Manoeuvre System (component of US Space Shuttle)
ORS	(US) Operationally Responsive Space (project and associated organisation).
OSC	Orbital Science Corporation (US business)
OST	Outer Space Treaty
P	
PAROS	Prevention of an arms-race in Outer Space
PD	(US) Presidential Directive
PN	Pseudo-noise
PRN	Pseudo-random number
Q	
R	
RF	Radio Frequency (component of the EM Spectrum)
RORSAT	Description of (cold-war era) Soviet radar reconnaissance satellite
RTG	Radio-thermal generator
S	
SAA	South-Atlantic anomaly
SALT	Strategic Arms Limitation Talks/Treaty
SAM	Surface to Air Missile
SAR	Synthetic aperture radar
SATCOM	Satellite communications (generic term)
SATNAV	Satellite navigation (generic term)
SBIRS	Space-based Infra-red system (US missile-launch warning system)
SDI	(US) Strategic Defence Initiative
SHF	Super-high frequency (ITU Frequency Band Designation)
SLBM	Submarine-launched ballistic missile
SLF	Super Low Frequency (ITU Frequency Band Designation)
sm	Statute mile
SMG	(UK MOD) Space Management Group
SpSA	Space Situational Awareness
SRBM	Short-range ballistic missile
SSAP	(UK) Space Strategy Action Plan
SSTL	Surrey Satellite Technology Limited (UK company name)
START	Strategic Arms Reduction Talks/Treaty
STS	Space Transportation System (correct designation of the NASA ‘Space Shuttle’)
SW	Short-wave (description of radio frequency band)
SWG	(UK MOD) Space Working Group

T	
TDRSS	Tracking and Data Relay Satellite (US system)
TIO	Targeting and Information Operations (grouping within UK MOD)
TLE	Two-line elements
U	
UAV	Unmanned air vehicle
UCP	(US) Unified Command Plan
UHF	Ultra-high frequency (ITU Frequency Band Designation)
UK SpOCC	UK Space Operations Coordination Centre
ULF	Ultra Low Frequency (ITU Frequency Band Designation)
UN	United Nations
USAF	United States Air Force
UV	Ultraviolet (radiation)
V	
VHF	Very High Frequency (ITU Frequency Band Designation)
VLF	Very Low Frequency (ITU Frequency Band Designation)
W	
WAAS	Wide Area Augmentation System (relating to GPS)
WMD	Weapons of Mass Destruction
X Y Z	

(INTENTIONALLY BLANK)

INDEX

All Index entries are to paragraphs, and not to page numbers. Index entries beginning with an upper-case letter are in the relevant annex.

(INTENTIONALLY BLANK)