11-21-2022

Spacecraft Systems & Navigation

Christopher Vanacore

This textbook is steered towards higher educational course entailed in Commercial Space Operations. This textbook will be covering in detail Orbital Satellites, and Spacecraft. These topics are discussed according to their application, design, and environment. The power system, shielding and communication systems are reviewed along with their missions, space environment and limitations. Any vehicle, whether manned or unmanned, intended for space travel is a spacecraft. A spacecraft's required systems and equipment depend on the information it will acquire and the tasks it will perform. Although their levels of sophistication vary widely, they are all subject to the harsh conditions of space. Depending on the missions that each spacecraft is designed to carry out, they can be broadly classed.

Christopher Vanacore is pursuing a continuing education in a Master of Science in Space Studies from Embry-Riddle and a Master of Science in Space Systems Engineering from The Johns Hopkins University. His Scientific Research involves a concentration in Orbital Mechanics, Guidance Navigation & Control, Attitude and Satellite/Spacecraft Systems.

9 786205 514573

Spacecraft Systems & Navigation

Christopher Vanacore

Christopher Vanacore

Spacecraft Systems & Navigation

LAP LAMBERT Academic Publishing

Christopher Vanacore

Spacecraft Systems & Navigation

Christopher Vanacore

Spacecraft Systems & Navigation

LAP LAMBERT Academic Publishing

Imprint

Cover image: www.ingimage.com

Publisher:
LAP LAMBERT Academic Publishing
is a trademark of
Dodo Books Indian Ocean Ltd. and OmniScriptum S.R.L Publishing group
Str. Armeneasca 28/1, office 1, Chisinau MD-2012, Republic of Moldova, Europe
Printed at: see last page
ISBN: 978-620-5-51457-3

The Author

Christopher Vanacore holds a Bachelor's Of Science in Space Flight Operations from Embry-Riddle Aeronautical University. He is pursuing a continuing education in a Master of Science in Space Studies from Embry-Riddle and a Master of Science in Space Systems Engineering from The Johns Hopkins University.

His Scientific Research involves a concentration in Orbital Mechanics, Astrodynamics, Space Debris Studies, and Human Factors research for Long Mission Duration in a confined environment, Plasma Physics, Heliophysics.

Additionally, Christopher Vanacore is an Analogue Astronaut Candidate with project PoSSUM. Christopher Vanacore also successfully completed the Suborbital Space Flight Human Centrifuge Training which involved the Virgin Galactic Space Profile undergoing up to 6.3 G-Forces.

Author Aknowledgments

I must start by thanking my parents, who provided me with an awesome life and extended assistance while in college pursuing my education and other extracurricular activities.

Next, I would like to thank my partner, who has always supported all my decisions and future endeavors.

Being an Astronaut is not easy, and having that mindset requires a strong cold personal character and knowing there is always the risk of a mission failure. However, I am incredibly grateful to have supportive people

Next, I would like to thank one of my favorite professors, Program Coordinator, and Research Advisor – Lana Laskey, who has been there next to me to provide many helpful ideas and assistance in my Academics while at Embry-Riddle Aeronautical University.

Finally, I thank my closest friends, who have always supported my decisions and ideas.

Having all these people surround me made me a stronger person and potential candidate to become the next future Astronaut for one of the Space Companies.

Off we got to Mars! Space to Inspire…

Contents

Chapter 1: Basic Spacecraft Systems

This chapter will sketch the basic systems found on typical satellites/spacecraft. Later chapters will cover the same material in more depth. Those systems and subsystems include the following:

I. Propulsion

II. Communications

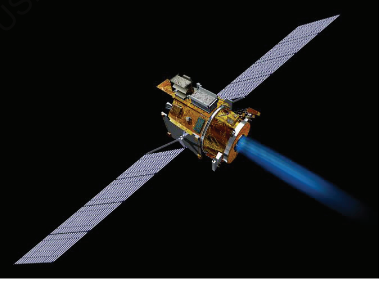
III. Command and Data Handling (C&DH)

IV. Electrical Power

V. Guidance, Navigation & Control (GN&C) VI. Thermal Control

VII. Structures

I.PROPULSION



Propulsion systems can be classified in many ways because of their diverse applications, but are generally assessed by thrust, thrust efficiency and total thrust duration, or energy. These characteristics are often associated with the propulsion system applications. A short list of uses follows.

Different categories of propulsion systems also refer to the fuel type or the energy source for thrust. Those characteristics will be outlined briefly below, and in more detail in the propulsion section. Modern propellants are being researched all around the world, generally with the goal of increasing their operability, manufacturing cost, or management hazard rather than always improving the performance of the engine.

According to several research, LOX-kerosene propulsion can be improved by adding solid particles to traditional fuel or using novel hydrocarbon chains, which are occasionally highly unsaturated. This activity may not be effective for huge rockets and is mostly advocated for missile applications.

Today's most effective element is molecular hydrogen (H_2), which produces the highest gas velocity exhaust for a given temperature. However, because the recombination of atomic hydrogen into molecular hydrogen is so exothermic, this efficiency might even be increased by adding a little amount of atomic hydrogen to the LH2.

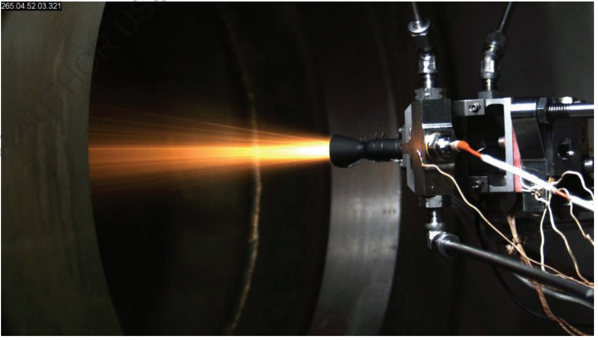
The following below are some of the main usages for propulsion:

- **Launch:** It is limited to high thrust vehicles (from Earth, or another planet/moon)
- **Booster** including multi-stage operation, transfer orbit, or orbit insertion/exit
- **Attitude control** related to small thrust operations for regaining the Orbital Altitude
- **Orbit Maintenance** in relation to station keeping, and for atmospheric drag reboost operations
- **Boost Separation** (Solid Rocket Booster from External Tank after burnout)

The following are the different types:

Chemical

The Chemical Propulsion can be base out of a Liquid propellant, Monopropellant or Bipropellant. The substances that make up chemical propellants belong to a unique class. Typically, some of the people in this class aren't seen to be significant enough to merit a separate discussion in chemistry classes. This makes it seem worthwhile to give a list of exemplary chemicals with a concentrate on the unique qualities that make them suitable as rocket propellants. A bipropellant engine allows for better performance but adds the complication of a propellant supply mechanism. The two greatest oxidizers, oxygen and fluorine, must be kept in a cryogenic condition for storage. The most typical oxidizer utilized in liquid bipropellant rocket propulsion systems is oxygen. It has a concentration of 1.32 g/cm3 and a typical boiling point of 90 K. Its critical pressure is 5.0 MPa and critical temperature is 155 K. A typical oxidizer that can be stored and is hypergolic with various fuels is nitrogen tetroxide. Chemical propulsion fuel known as a monopropellant doesn't need an additional oxidizer. Instead of a fuel and an oxidizer line, a monopropellant-based rocket engine simply needs one fuel line. The word "mono" in

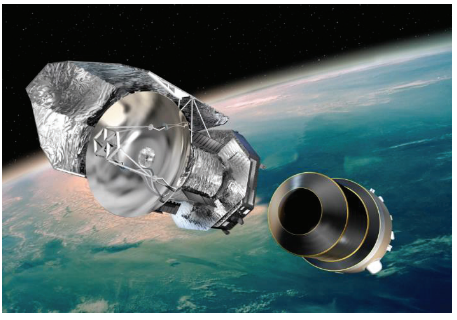


monopropellant refers to a fuel that can run on its own. A bipropellant is a chemical propulsion mechanism that combines like hydrogen and oxygen. Due of the oxidizer's bond with the atom itself, a monopropellant burns independently of any other fuel. This results in a lighter, more affordable, and more dependable rocket engine. Control thrusters frequently employ monopropellant designs, although not genuine propulsion components.

Solid

Solid propellants have a challenging range of operating and storage conditions to retain structural integrity under. For example, operational pressures for some tactical motors are frequently over 1000 psi, and operating temperatures for some strategic motors range from -60 °C to 65 °C. Particularly for case linked grains, the conditions impose large mechanical loads on the propellant. Concerning mechanical characteristics include strain capability at low temperatures under both high strain rates and extremely low strain rates (for cooling the motor during low-temperature storage) (ignition pressurization of a cold motor). In order to prevent excessive deformation under high-temperature ignition circumstances and creep during prolonged exposure to high temperatures, the modulus must be high enough. To withstand breakage or fragmentation upon failure at high stress or strain rates, durability is also crucial.

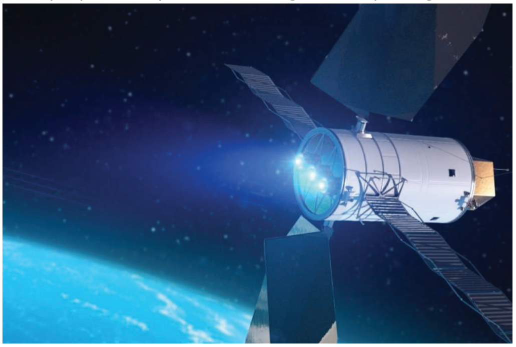
Hybrid



Hybrid propulsion systems combine two different engine types (often an internal combustion engine and an electrical engine). They can be series hybrids, in which case the vehicle is only propelled by the electrical engine, or parallel hybrids.

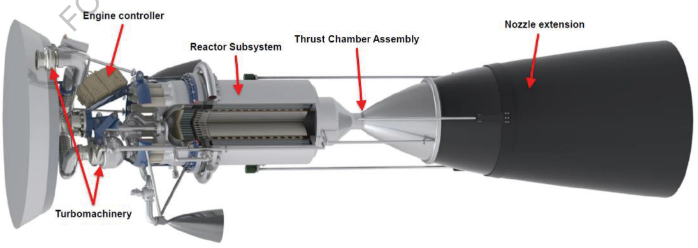
Electric

Electrothermal, electrostatic, and electromagnetic propulsion systems can be generically categorized as interplanetary propulsion systems. As the range of electric propulsion technology spans five orders of magnitude in power and around two orders of magnitude in specific impulse—the propellant exhaust velocity—each type of thruster serves a significant niche. The propellant is accelerated as quasi-neutral plasma in electromagnetic thrusters. As opposed to electrostatic thrusters, which speed up ions or electrically charged particles, this technique. Thus, unlike gridded ion thrusters, electromagnetic jets are not constrained by electric space charge. There are now several different kinds of electromagnetic thrusters under investigation. They consist of helicon, magnetoplasmadynamic (MPD), and pulsed thrusters. A significant point is that MPD thrusters have the potential to produce high thrust at high power—up to 10^5 W—and might thus be taken into consideration for cargo and human transportation missions to Mars.

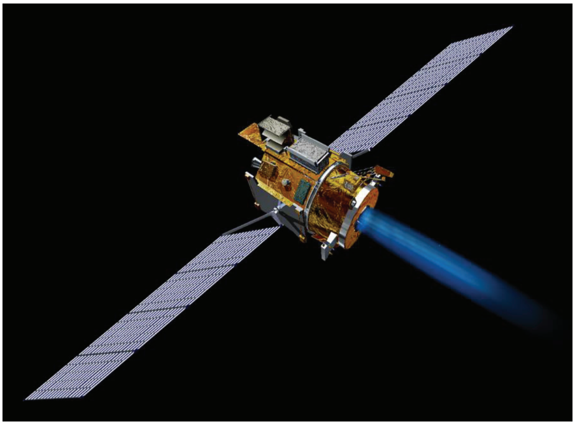


Nuclear

Humans can explore the solar system for extended periods of time thanks to nuclear propulsion. Compared to chemical or solar electric propulsion, its excellent fuel economy permits lighter transfer vehicles and speedier conveyance for Mars missions. In comparison to conventional, less efficient propulsion methods, the simpler transfer vehicles and speedier transport capabilities together offer safer overall transportation to Mars and other deep space sites. Transfer vehicles can be reused for several Mars trips thanks to the enhanced performance of nuclear propulsion, resulting in significant cost savings.



Solar



With the support of NASA's Solar Electric Propulsion (SEP) initiative, ambitious new science and exploration missions can be carried out for longer and with greater capabilities. Innovative propulsion methods like SEP might provide the ideal balance of economic benefits, safety, and improved propelling force to enhance a range of future travels to planets and locations outside of Earth orbit.

The electrically driven system, which is powered by the electric energy from on-board solar arrays, will consume 10 times less fuel than a comparable, conventional chemical propulsion system, like those used to launch space shuttles into orbit.

However, the sturdy power generated by the reductions in energy mass will enable robotic and crewed missions to travel far beyond low-Earth orbit, including sending exploration spacecraft to far-off places or transporting cargo to and from interesting locations, laying the groundwork for future missions, or refueling current ones. The creation of cutting-edge technologies the project is creating and showcasing, including as huge, light-weight solar arrays, magnetically shielded ion propulsion thrusters, and high-voltage power processing units, are being driven by mission needs for high-power SEP.

II.COMMUNICATIONS & TELEMETRY

The communications systems in spacecraft have corresponding communications equipment on the ground for two-way data transfer. The communications link must be maintained between the ground station and the spacecraft for successful spacecraft mission operations. Currently, the only reliable communications bands for ground-to-space communications through the Earth's atmosphere are in the microwave bands (1-10 GHz). Laser communications technology is being developed for high-bandwidth space communications through the atmosphere.

The basic communications system consists of the receiver, the transmitter, the antenna, and the command system.

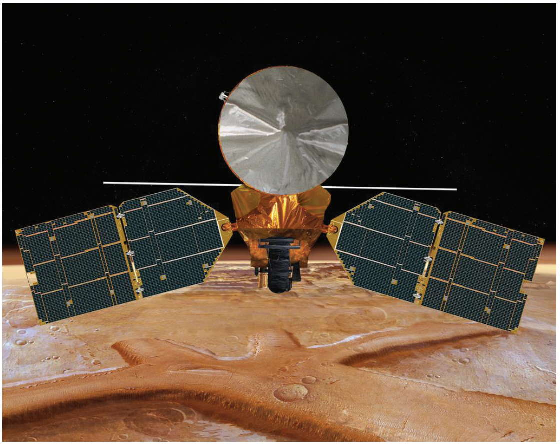
Receiver

The detector and decoding subsystem that reproduces the original data sent from the transmitter. The noise contribution of each component of a spacecraft must be reduced in order to listen to its signals, starting with the antenna and moving on to the RF receptors. Currently, noise-free systems are achieved at temperatures below 20 degrees Kelvin. The receiver set, which consists of an amplifier located in the antenna's focal point and a receiver positioned in the control room with the other electrical devices, is a crucial link in this chain. The amplifier's goal is to boost the antenna's weak signal as much as possible without introducing noise. The achieved magnification ratio ranges from 100,000 to 700,000.

Transmitter

The signal coding and transmission subsystem which converts spacecraft data into a microwave signal.

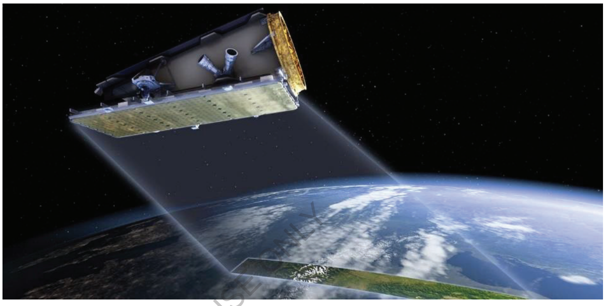
Antenna



An antenna is the receiving and/or transmitting element that couples the signal to space. Two microwave antennae are used by the spacecraft for telecommunication, while one VHF antenna is used for telemetry, beaconing, and commanding purposes. Broadband signals from a ground transmitter are received by one microwave antenna centered at 6 gc, and signals are sent to a ground receiver by the other microwave antenna centered at 4 gc. The electronic receiver and transmitter are coupled to each microwave antenna by a sophisticated precision feed system. Each microwave antenna is made up of several circularly polarized radiating elements that are evenly placed around the equator of the spacecraft. The VHF antenna is a tiny, multi-element helix that is fixed to the spacecraft's pole and emits a signal that is linearly polarized. Each antenna produces approximately isotropic antenna patterns, with the spin axis of the spacecraft serving as the axis of symmetry. The antenna systems passed rigorous electrical, mechanical, and thermal tests despite being made of lightweight but durable materials.

Command system

The command system is the command interpreter as part of the data management system controls data flow and commands throughout the spacecraft. The only point of entry or exit for data into or out of the spacecraft is a system coupled to RF transmitter and receiver modules. A space link is a communication channel between two spacecraft or between a spacecraft and the ground system with which it is connected. A frame relay known as a space link protocol is one that is intended to be utilized over a space link or in a network that has one or more space links. Telemetry and telecommand signals make up the bulk of the data flowing via a space link. As a result, the TC uplink and TM downlink serve as a channel of communication between the spacecraft and the ground controllers.



In addition, there are several concepts that are common to spacecraft communications.

Up/down link

Uplink and Downlink are both considered to be communication paths between ground station and the spacecraft. The uplink is separate from the downlink to allow simultaneous communications with the spacecraft. Uplink is transmission to spacecraft (transmitter on ground & receiver on spacecraft. On the other hand, Downlink is transmission from spacecraft (transmitter on spacecraft, receiver on ground station). Systems may also include a relay or link to other satellites, to other spacecraft, or to other communications systems such as the TDRSS network.

On the Uplink portion, commands are split into Application-Specific Commands and Direct commands to the Spacecraft for reconfiguration. The Downlink portion telemetry can be split into Spacecraft HK Data, Orbital Positioning Data, Payload Data which is mainly considered as scientific related Data, Telecommand Reception Status (also known as CLCW).

Bandwidth

Bandwidths are the width of the frequencies in the total system or in each communications channel and they determine how fast data can be transferred.

Channels

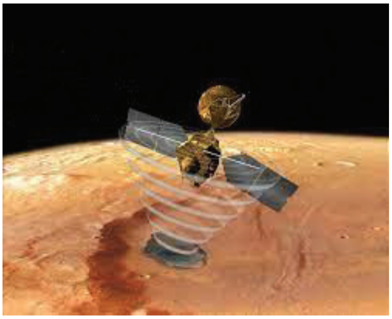
Channels are defined as the number of separate frequencies that the system can communicate on. On a communications relay satellite (for example, television video satellites), each channel is a separate transmitter and receiver called a transponder. The necessity and starting point for the design of a communication system is channel capacity. Shannon's communications theory is utilized to analyze the channel capacity using the parameter estimates of the signal-to-noise ratio acquired by the communication link budget approach in order to determine the effects of a severe plasma environment on the communication system of reentry vehicles. First, a stratified medium finite-difference time-domain approach is utilized to determine the attenuation brought on by the plasma sheathing for typical S, C, and Ka telemetry frequencies in a typical reentry plasma environment. The channel capacity is then estimated using common telemetry transponder characteristics. Data suggest that plasma degradation at the C-band is marginally improved and that the S-band channel capacity is nearly zero at an elevation of 30 to 40 km. The blackout occurrence is still clearly visible, though. The Ka-band signal can pass through the plasma sheath layer with the least amount of attenuation, thus enhancing the channel's capacity and potentially meeting telemetry requirements. Even channels have a bandwidth which is the frequency bandwidth of the individual communications channel.

Center frequency

The center frequency of a filtration or channel is a measurement of a centre frequency between the upper and lower cutoff frequencies in electromagnetics and telecommunications. Typically, it is described as either the geometric or arithmetic mean of the lower and higher cutoff frequencies of a band-pass or band-stop mechanism.

III.COMMAND & DATA HANDLING

The command and data handling (C&DH) system provides the processing, data distribution and executive control functions within the spacecraft data and communications elements. These functions are normally allocated to several computer systems with redundant components for necessary reliability.



The Command and Data Handling System is comprised of different components.

Central Processor Unit (CPU)

The central processor unit handles internal data manipulation in the individual computer units. Although CPUs' shape, design, and functionality have evolved throughout time, their basic function has remained mostly same. The arithmetic-logic unit (ALU), which completes mathematical operations, processor registers, which provide operands to the ALU and store the results of ALU operations, and a control unit, which coordinates the coordinated operations of the ALU, registers, and other components, are the main parts of a CPU. The control unit concocts the fetching (from memory), decoding, and execution (of instructions).

Command Processor

The Command Processor is used for interpretation, validation, and relay of data from the receiver to the proper interface circuit or subsystem.

Memory

The various types of memory required for data operations include random access (RAM), bulk (disk, tape, etc.), and circuit (hard) memory used for CPU functions.

The Command and Data Handling System has a variety of functions and are the following: Processing commands from receiver and onboard executive controller, retrieve data for transmission to ground, store data from sensors and command processor for later transmission, process telemetry and instrument (sensor) data for relay or storage, control spacecraft system/subsystem functions, format data for onboard operations, relay and/or transmission, check redundant operations for errors and possible error recovery procedures and maintaining basic communications and operations functions in case of errors/failure.

Organization

The overall system organization includes communications, commands and computing. The design guidelines provide for maximum flexibility with a minimum or hardware & software resources:

- Data distribution topology used to maximize speed and accessibility
- Network topology must be readily accessible by all components
- Communications link must support computing, data transmission and storage needs
- Operating systems must be capable of handling all functions as rewired
- Redundancy necessary to provide for eventual errors and partial failures
- Reconfiguration must be built into system for changes in mission

Design requirements:

The Command & Data Handling System must provide sufficient processing speed for sensor data (this is an absolute necessity), it must have sufficient data storage to allow expected communications interruptions while sensors (or receivers for communications or relay satellites) are generating data, it must be adaptable to changing mission requirements – reconfigurable, it must be autonomous in case of errors or partial failures, it must be redundant to test and recover from errors or partial failures, it must be hardened and/or shielded against expected radiation levels and it must have uninterrupted electrical power.

IV.ELECTRICAL POWER SYSTEMS (EPS)

A reliable power source is needed for any spacecraft, even for a mission lasting only a few minutes. The EPS is used to power all of the spacecraft systems, and is a major factor in the overall spacecraft design.

The success of a space mission depends on an uninterrupted, dependable power source. In Earth orbit, the Sun produces approximately 1.4 kilowatts of power per square meter, a plentiful resource that spacecraft designers strive to utilize. This is why the bulk of spacecraft either have solar arrays stacked across their exterior or in the shape of wings.

These are made of photovoltaic cells that are networked together and generate an electrical current when exposed to light, operating similarly to regular light emitting diodes (LEDs) but in the opposite direction. It's fascinating to see how closely solar cells and LEDs use the same technologies.

V.GUIDANCE, NAVIGATION, AND CONTROL

Spacecraft flight is programmed, measured, and controlled by the combined functions of the command & data handling system and the guidance, navigation and control system (GN&C). Various subsystems on the spacecraft determine the trajectory or orbit of the spacecraft, measure the actual flight path, or rely on ground-based measurements and computations to correct or maintain the flight path.

Some of the basic functions of the GN&C are:

Guidance - Spacecraft trajectory control during the thrust phase

Navigation - Determination of spacecraft position and velocity relative to a specified reference frame

Control - Spacecraft attitude (orientation) and trajectory control. These functions can also be separated into two categories as follows:

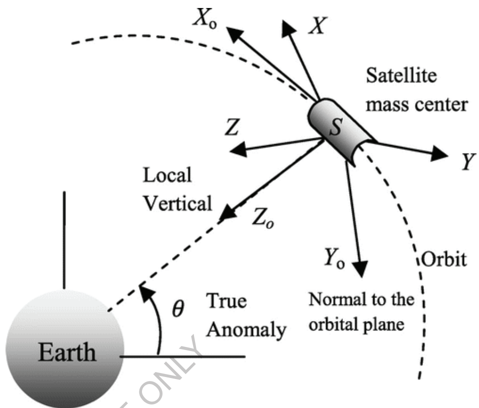
- **Attitude and trajectory determination** - Determination of spacecraft trajectory and/or orientation with respect to specified reference frame or object(s)
- **Attitude and trajectory control** - Control of spacecraft trajectory and/or orientation with respect to specified reference frame or object(s)

Attitude and motion sensors

For small satellites, determining attitude and motion with a small number of sensors is a desirable choice. A persistent trend has seen every satellite subsystem mass, dimensions, and power consumption reduced to even dozens of grams for picosatellites. As satellites get smaller, full attitude estimate with just one sensor becomes a useful capability. However, some tiny satellite missions may not be viable for them because to their cost and required power. Sun sensor measurements are unavailable in the shadowed region of the satellite's orbit thus cannot fully observe the attitude motion. To determine the attitude and floating measurement bias, the angular velocity sensor needs additional vector measurements. In non-equatorial low Earth orbits, three-axis attitude motion determination is provided by magnetometer data alone.

There are two main methods for spaceship attitude control stabilization:

Setting the spacecraft spinning allows for spin stabilization, which uses the rotating spaceship mass's gyroscopic action as a stabilizing mechanism. Only seldom are the thrusters of the propulsion system fired to modify the spin rate or the spin-stabilized attitude as needed. Thrusters or yo-yo de-spin can be used to stop the spinning, if necessary.



An alternate technique for spacecraft attitude control called three-axis stabilization keeps the object stationary in the proper orientation while preventing spinning. One approach is to continuously move the spaceship back and forth inside a deadband of permitted attitude error using small thrusters. Other names for thrusters are mass-expulsion control systems and reaction control systems (RCS). Utilizing electrically powered reaction wheels, also known as momentum wheels, installed on the spacecraft's three orthogonal axes is another way to achieve three-axis stability. They enable the exchange of angular momentum between spacecraft and wheels. The response wheel on a certain axis is accelerated in the opposite direction to rotate the vehicle around that axis.

Tracking systems

The attitude and articulation control subsystem (AACS), often known as the guidance subsystem, is one of any satellite's or spacecraft's most important subsystems. The AACS is in charge of the appropriate navigation of the spacecraft; its sensors are set up to scan the skies, which act as a reference map for the route of the spacecraft. Additionally, the AACS is in charge of supplying inertial measurements that will be utilized to keep the satellite in a stable position so that it can carry out its primary mission of acting as a communication platform. Each AACS sensor is a highly developed control assembly whose only responsibility is to ensure the correct satellite attitude.

Ground tracking (tracking from the ground) system for communications and/or position determination for spacecraft.

Ground control - control networks located at specific sites for spacecraft launch, flight operations, and data distribution.

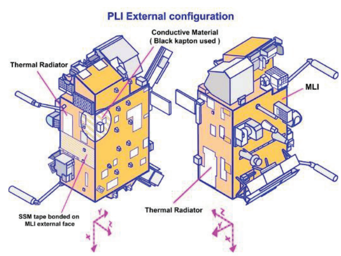
Communications networks - the spacecraft communications networks that are used for spacecraft communications and control

Deep Space Network (DSN) - developed by NASA to track and communicate with deep-space missions

Spacecraft Tracking and Data Network (STDN) - TDRSS satellite network used to communicate with satellite in Earth-orbit

VI.THERMAL CONTROL

Because of the temperature extremes in space, all spacecraft must be provided with thermal control components to moderate the extreme temperatures for heat sensitive components. The thermal control elements may be passive, active, or for most spacecraft, a combination of the two.



Passive systems - no active (powered) devices required

Below is a listing of different passive systems for Thermal Control: Coatings, Radiator panels/fins, Heat pipes, Insulation, Conductive structures & components, Louvers, Sun shields System, Cryogenics, Liquid heating/cooling loops, Electric cooling (thermoelectric), Electric heaters, Shutters, Heat Inputs, Spacecraft internal heat.

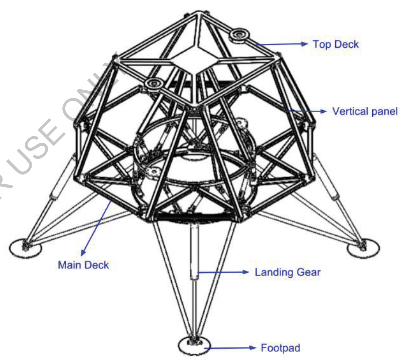
Surface coatings often very important and may be answer for much of the thermal control in small or simple spacecraft.

- Dark surface - absorb & radiate visible and IR radiation well
- Light surface - poor emitter and absorber
- Polished metal/film - reflects heat with very little absorption or emission

Near Mercury (0.3 AU)	800 K (980oF)	50 K (-280oF)
Near Venus (0.7 AU)	500 K (440oF)	50 K (-280oF)
Near Earth (1.0 AU)	400 K (260oF)	50 K (-280oF)
Near Mars (1.5 AU)	300 K (80oF)	20 K (-280oF)
Near Jupiter (5.3 AU)	100 K (-280oF)	20 K (-420oF)
Near Pluto (40 AU)	20 K (-420oF)	20 K (-420oF)

VII.STRUCTURES

Spacecraft structures are subject to many limitations because of the extreme conditions in space and the cost of getting a spacecraft into space. The primary structure, sometimes referred to as the bus, must be lightweight as well as strong. Each of the supporting structural elements must be capable of withstanding the launch vibrations and accelerations, and operate after reaching the hostile space environment - high radiation levels, extreme temperatures, and vacuum. Of course, dust and debris in space presents collision hazards that are rare but can easily be fatal.



Common structural components are the following:

1. **Bus (main) structure** - provides support for the main spacecraft assemblies

2. **Booms** - antennae and sensors are often located some distance from the main structure and need a rigid, lightweight attachment. Booms are often extendable, long and subject to vibration problems

3. **Instrument platforms** - these are sometimes rotating or movable, requiring sophisticated lubrication/bearings

4. **Solar panel extensions** - the solar panels found on many spacecraft have large surface areas and require a rigid structure to support the panels against spacecraft motion and vibration, and any drag that may occur in the upper atmosphere

5. **Attachment points**

Common structural materials: **Aluminum alloys** - most common since it is strong, lightweight, inexpensive and an easily machined metal*,* **Titanium** - strong, lightweight, high temperature metal*,* **Composites** - strong, durable, easily cast*,* **Beryllium** - lightweight, strong, expensive*,* **Steel alloys** - strong, dense, easily machined.

Common requirements for many structural materials:

Strength - must be high, **Stiffness** - usually required to be high, **Low density for low mass** (lower launch costs, fuel requirements), **High** (heat dissipation) **or low** (heat retention) **thermal conductivity**, **Low thermal expansion** - critical for close fitting components such as connections, bearings, optical devices etc.,**Corrosion resistance** – high, **High conductivity** to reduce charge buildup, **Ductility** to reduce cracking, **Fabrication ease.**

References & Resources

1. Brown, C. D., Spacecraft Mission Design, 1991, AIAA, Washington, D.C.

2. Griffin & French, Space Vehicle Design, 1991, AIAA, Washington, D.C.

Chapter 2: Space Environment

Background

Electromagnetic radiation is present from the microwave through gamma radiation frequency bands throughout the solar system and between stars and galaxies, as is particle radiation. This radiation is generally low intensity and therefore does not act as a heat source because it is so diffuse. However, in the vicinity of stars and larger planets, heat energy is significant and may affect spacecraft operations.

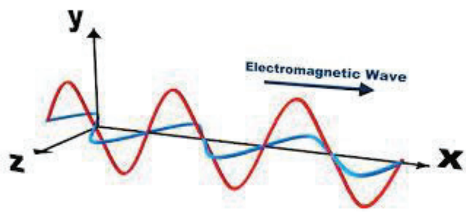
Earth Atmospheric Operations, Ascent and Reentry

Thermal inertia precludes significant thermal changes because the ascension phase after launch lasts just 30 to 60 minutes. The spacecraft is thermally insulated from the aerodynamic scorching during the initial phase of ascent by the rocket fairing (or the shuttle-bay doors), achieving some 100 km altitude and about 3 km/s speed in less than 5 minutes from lift-off. Since air-drag has reduced significantly from its peak at about 5..10 km altitude (about 1 minute after lift-off), and aerodynamic heating has reduced beneath solar heating parameters, the fairing is removed at that height to save acceleration mass. Reentry is fundamentally different from the climb phase in terms of the range of speeds associated, as the crossing of the Karman line at 100 km altitude throughout ascent changes from 3 km/s to 8 km/s to up to 11 km/s when coming from higher orbits (and kinetic energy is proportional to the square of the speed). The temperature reached by dynamic retardation (for example, the temperature in the shock layer before a blunt reentry body) increases proportionally to speed thanks to gas disintegration at high temperatures, which dramatically increases thermodynamic efficiency.

Thermal

In addition to the much more intense solar radiation, electromagnetic energy is produced by the large and small bodies, and by dust interacting with EM radiation throughout the solar system. As the energy radiates outward from the Sun or any other star, it decreases in strength as the inverse square of the distance. Warm bodies such as Jupiter (internal heat) and the inner planets (heated by the Sun) radiate IR energy primarily, but in small quantities compared to the Sun. Even at significant distances from the Sun or any other star (several AU for a small star), the heat energy becomes diffuse and too weak to provide significant heat for spacecraft to use for thermal conditioning. Beyond the orbit of Mars, for example, spacecraft require internal heating for most of the systems because of the extremely cold environment.

Although electrical heaters are simple, nuclear thermal plugs can provide a continual, relatively inexpensive heating source that lasts for years.



Kelvin background (microwave)

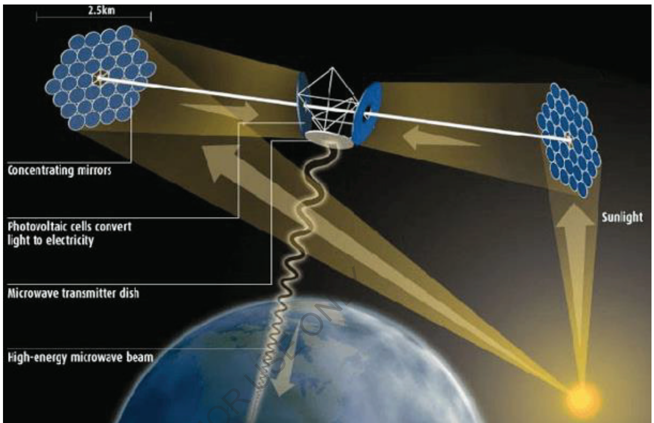
The heat energy produced by the initial Big Bang formation of the universe has not gone away, nor been absorbed. It continues to expand and cool from its 10,000 K temperature when hydrogen formed from protons and electrons. This is an important epoch and an important temperature because the ions (p+ & e-) were before that time interacting strongly with the photons which made up a substantial part of the universe. After the condensation of hydrogen at roughly 400,000 years following the Big Bang, the photons were free to expand at the speed of light. As the EM radiation did expand, it also cooled, but at a slightly different rate than the material.

The expansion and cooling of the universe, both energy and matter, continued for approximately 14 billion years and continues today where the temperature of the background photon radiation is now 3 K. That radiation is equivalent to approximately 90 GHz in frequency which is observable in the microwave region.

This is an important background radiation, although extremely diffuse. It not only provides important data that serves as much of the foundation of the Big Bang theory, but also establishes the interaction background for other physical process observed throughout the universe. One of those phenomena is the maximum energy of cosmic rays arriving from distant sources measured in millions and billions of light years. A limit to the maximum energy for the proton nuclei cosmic rays from energetic events is about 1020eV due to the interaction between the protons and diffuse 3 K radiation. This maximum has been observed, although there are much higher energy nuclei that are thought to be higher-mass nuclei.

High energy sources

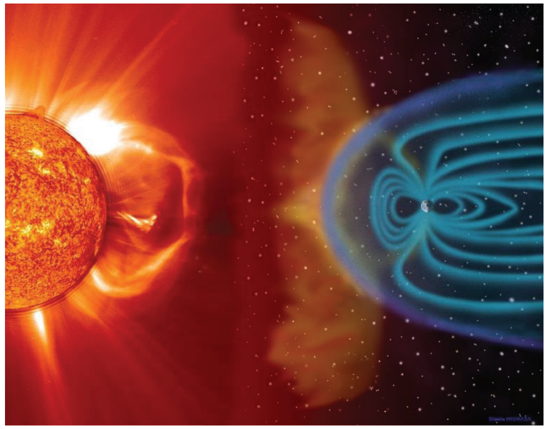
Other sources of energy throughout the galaxy, and the universe, produce high energy particles and electromagnetic radiation. This very energetic radiation can be measured most easily above the Earth's atmosphere, and on the Earth, indirectly. Both particle and electromagnetic radiation are found entering the solar system from other parts of the universe. Gamma radiation and particle cosmic rays above 1020eV have been measured from sources beyond our galaxy, but the intensity levels are much too low to be a concern for spacecraft heating, or even crew radiation exposure, with some exceptions. The energetic particles, however, can be damaging for sensitive microelectronics found on spacecraft. Single-event upsets and permanent damage to semiconductor circuits become important considerations when planning spacecraft missions, even in low-Earth orbit.



Diffuse background radiation is also present in the solar system as infrared radiation emanating from the low-density dust and particles throughout several regions in the solar system. There is also weak x-ray emission from the solar system interior, with a less-established source, that is extremely diffuse, but present. Radiation mechanisms such as these are important in understanding the physical cause of these processes throughout our solar system, and in other star systems, and because the spacecraft being built in the future will be capable of detecting even weaker sources of energy, providing us with greater insight into the physical universe.

Particle radiation heating is important in spacecraft heating and cooling only from relatively intense levels. This would include regions close to the Sun, and in radiation belts such as the van Allen radiation bands, and Jupiter's strong radiation belts. Obviously, this level of radiation would also include consideration of exposure to the electronic components and crew.

Solar Radiation



The ultimate source of radiation from the Sun is the nuclear fusion furnace at its core. This solar fusion reactor provides the energy for the particles, electromagnetic radiation, and magnetic fields that also generates most of the energy throughout the solar system. The exceptions are the extreme gravitational fields of the large gas planets.

Although the energy source at the Sun's center is thought to be relatively constant, the radiation, particle and electromagnetic, and magnetic field variations, are not. A cycle of activity from the Sun representing particle emission, electromagnetic radiation, and magnetic fields varies dramatically over a roughly 11-year cycle. Variations in the energy output from the Sun's upper interior and surface are closely related to the magnetic field characteristics, although representative models are still being developed. Solar magnetic field measurements have confirmed the eruptive activity-magnetic field change relationships, and more recently, the dramatic increase in particle temperatures from the Sun's surface.

Conditions within the central 20% of the Sun are dense and hot enough to fuse protons (hydrogen nuclei) into helium nuclei through a process of successive nuclear buildup. The first stage is the fusion of two protons into a nuclear-bound proton and neutron called deuterium. To change one of the protons into a neutron, the nuclear (protons are repulsive without neutron moderation) and electric field (both positive) repulsive forces must be overcome. This is accomplished with the speed of collision from extremely high temperature protons, necessitating high density and high temperature conditions for fusion.

p+ + p+ --> 2H (p+ + no) + e+ (e+ is a positron, or antielectron) + ve (electron neutrino) e+ + e- --> 2 photons (gamma rays)

Step two is the fusion of deuterium (p+ + no) plus another proton into two protons and one neutron, or tritium, another isotope of hydrogen (isotope meaning a constant proton number but a change in neutron number).

2H + p+ --> 3He + photon (gamma ray)

The third step is to fuse two tritium nuclei into one helium and two proton byproducts. The end result is the fusion of four hydrogen nuclei (protons) into one helium nucleus, and two gamma ray photons, and two electron neutrinos, and kinetic heat energy.

3He + 3He --> 4He + 2p+

The net result is a helium nucleus, 6 gamma-ray photons with a net energy of 26 MeV, two electron neutrinos, and kinetic heat energy. Its representation would be

p+ + 2 e --> 4He + 2 neutrinos + 6 photons

The process actually employs other light elements within the fusion reaction of smaller stars, described as the p-p, or proton-proton cycle. For larger stars with somewhat higher core temperatures, lithium burning takes place which also involves the following simplified version of the more complex p-p II reaction:

3He + 4He --> 7Be + photon 7Be + electron --> 7Li + neutrino 7Li + p+ --> 2 4He

For even higher core temperatures, the alternate process also includes the p-p III reaction

7Be + p+ --> 8B + photon

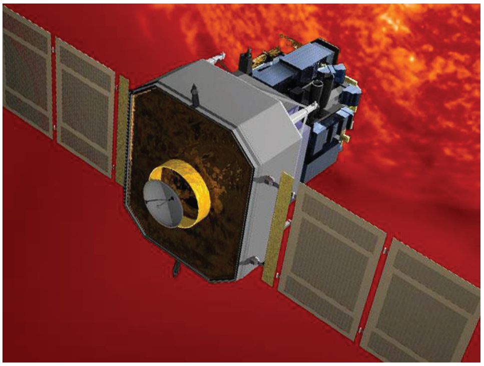
8Be --> 2 4He + antielectron + neutrino

Larger stars employ a carbon-nitrogen-oxygen (CNO) cycle that does not consume the CNO components, but requires the conversion products as catalysts.

Electromagnetic Radiation

Radiation from the Sun begins as gamma-ray photons deep in the interior, slowly shifting in energy to lower-frequency visible and infrared bands, and includes lower levels of radio, x-ray and gamma-ray emissions. A look at the different frequency bands shows distinctly different features near the solar "surface." Since there is no surface on the Sun, the term generally represents the most visible layer called the photosphere. Images of the solar radiation from various satellite and ground observatories are shown below to distinguish features seen in the separate electromagnetic bands. Click on each of the smaller images to see the higher resolution details. One angstrom, a measure of wavelength, is equal to 10-10 meters. Shorter wavelengths are equivalent to higher energy.

SOHO spacecraft UV image of the Sun at 171 angstrom (from WWW-1, 9/3/99)



Energetic UV emissions show link between high energy UV and low energy (soft) x-ray coronal holes shown in the image below. Coronal holes are the primary solar wind ejection regions from the Sun, appearing larger just after sunspot maximum.

The particle wind, or solar wind streaming from the Sun is composed primarily of high energy protons and electrons ejected from the surface - especially near the high energy coronal holes. The speed of the particles decreases slightly with distance from the Sun, with the approximate speed near 1 AU at 400 km/s. These ionized particles follow the magnetic field lines of the Sun, concentrated along the equatorial sheet as seen in Figure 3-2. The density, energy, and speed of the particle wind varies dramatically as the activity of the Sun varies. The most active period of the Sun, the peak in the sunspot number, has a corresponding peak in particle energy and flux (particles per area per second), translating into much higher radiation levels. A time plot of the proton flux is shown below in Figure 3.9 indicating the radiation levels seen during solar maxima and minima.

Proton radiation flux correlating with the solar cycle. The proton flux is shown on the left in protons per cm2. The Zürich sunspot number (smoothed) is shown on the right (Wagner & Tascione, 1987).

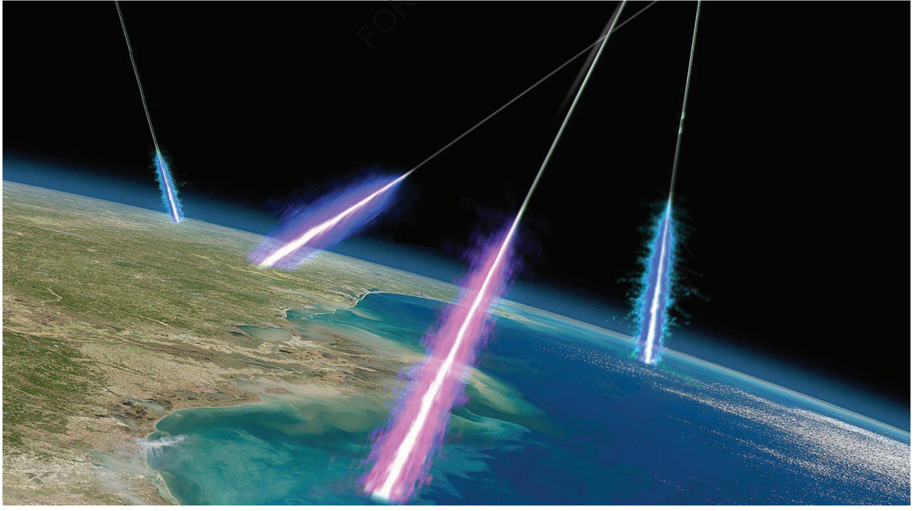
The Sun's varied eruptions produce some increase in high-energy photons, but generally more copious amounts of high energy protons and electrons, as well as other ions in much lower abundance traveling outward along the solar magnetic field lines. The travel time before intercepting the Earth ranges from several hours to several days. Higher energy protons and electrons (several hundred MeV) may take only a few hours to reach the Earth, while the normal travel time for the solar wind is of the order of three days. The fastest particles that are by definition the highest energy particles represent the most hazardous radiation levels for manned missions, and for spacecraft electronics and terrestrial communications. The alert time for these energetic protons may be a little as an hour after being observed on the Sun's surface from Earth, or by satellite. Below is a diagram of the typical travel time and evolution of solar particle and electromagnetic energies observed on Earth. Although electrons and protons travel at the same approximate speed in the solar wind, higher energy bursts of protons travel without the corresponding cloud of higher energy electrons.

Solar Features

Visible layers of the Sun are shown on the diagram at the top of the page including active and normal features. Several of those features are shown below.

Cosmic Rays

In addition to the protons and electrons flowing from the Sun, higher energy ionized particles, or ions, including energetic protons, come from the Sun, other stars, supernovae, the galaxy center, and even from energetic events outside the galaxy. These high energy particles include helium, iron, carbon, oxygen and many other nuclei; the most abundant being protons from the Sun. Solar cosmic rays are generally lower in energy than the galactic and supernova varieties, and follow the solar wind outward from the Sun, with some enhancement along the Sun's equatorial plane. Cosmic rays outside the solar system arrive from nearly random directions and approximately follow the weak interstellar and solar magnetic fields. The highest energy cosmic ray particles come more directly from their original source and in some cases can be traced back to the approximate location in the sky by geometrical calculations using the location of sensors. These particles were a puzzle when first observed earlier in the twentieth century, and were therefore called "cosmic rays."

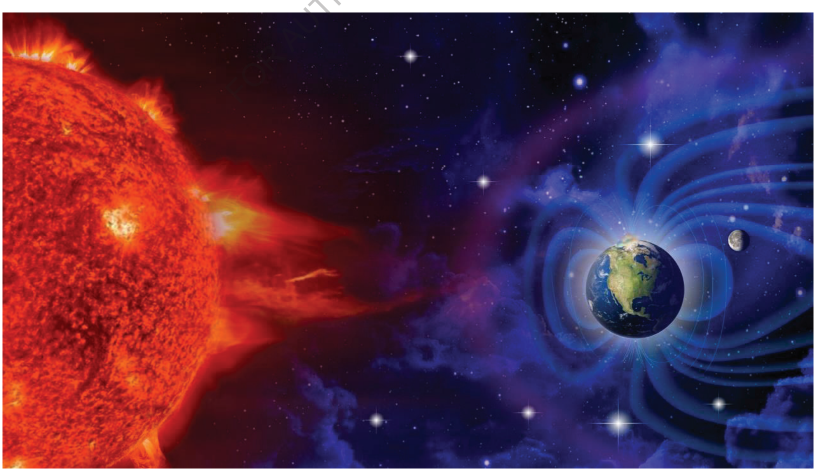


Cosmic ray particles are the highest energy particles found in the solar system. They arrive at a much lower frequency than the solar wind and solar cosmic rays, and therefore do not present the same dose rate of radiation. Nevertheless, these particles are important in calculating hazards to spacecraft and crew's sine the dose and dose rate of particle and electromagnetic radiation is a combination of rate (intensity) and energy. These high-energy particles do, however, present a higher potential for damage of microelectronic circuits. Because of this, the sensitive microprocessors and the solid state memory on spacecraft must be shielded and hardened from particle radiation.

Space Weather

The sum of radiation and magnetic fields from the Sun can be visualized by going to the Space Weather pages at several Internet sites, including the one listed below. Of great importance are the solar flares, eruptions, and mass ejections that come from the Sun at almost any time in its 11-year phase of activity. Mild to extreme solar events are shown, along with a variety of charts and graphs, as well as criteria for alerts for the more active outbursts.

Included in the space weather data are sunspot numbers and implications for the sunspots, if there are any. Sunspots generally follow the Sun's activity in its 11-year cycle, however, individual spots may range anywhere from insignificant (smaller spots) to indicative of coming flare activity (large).



Lastly, Earth-crossing asteroids are described along with a listing of the potentially dangerous objects that could possibly impact Earth in the future. These data can all be found.

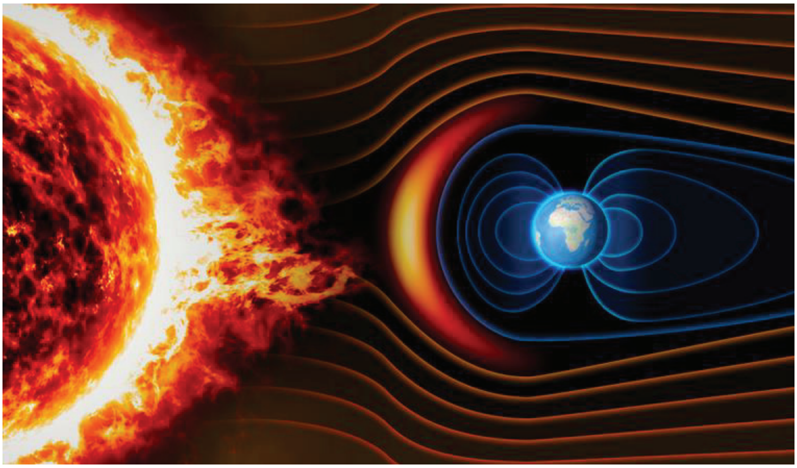
A newer source of solar activity and solar dynamics is available online with data and interpretation from the Solar Dynamics Observatory (SDO) spacecraft. High-resolution images of the Sun in a wide variety of frequencies are shown, along with animated videos indicating the importance of changes (dynamics) in the Sun's character. SDO home page

Solar radiation levels vary widely, from mild (although hazardous) emissions from the Sun's continual solar wind, to extremely damaging particle outbursts during eruptions and flares that include Solar Proton Events (SPEs) and Coronal Mass Ejections (CMEs). While the radiation hazards from these particle blasts may concern spacecraft operators and space flight crews, there is much greater protection from electromagnetic and particle radiation underneath the Earth's atmosphere and magnetic field. Even so, we can see evidence of solar eruptions from the Earth's surface in bright aurora, in satellite interruptions and failures, or even electrical transmission failures. Outside Earth's protective magnetic field and atmosphere, energetic eruptions from the Sun can have serious consequences to space crews.

One of the large recent eruptions that began on August 2, 1972, occurred between the flights of Apollo 16 and 17. A total dose of 400 rem (4 Sieverts) took place following a huge solar eruption. In a prompt, or acute, exposure, the Apollo astronauts could have received a fatal dose of particle radiation. However, the radiation outburst from the Sun took place over roughly a week, producing an extended, or chronic, dose. This time-extension of the absorbed dose is important since it can have a less serious consequence because of the greater time the tissues have to recover from the radiation damage. The Apollo astronauts would have had a lower probability of potential biological mutation over a one-week exposure versus a several minute outburst, and therefore a lower probability for radiation-induced cancers (see Radiation and Biology).

Solar Magnetic Fields

The magnetic field of the Sun has a complex structure that is generated by the motion of the plasma (ionized gas) in its interior. The stellar core and deep interior are dense, high temperature gasses that produce currents of charged particles from the outflow of heat from the core's energy generation mechanism - the fusion of hydrogen into helium. Those ion currents produce magnetic fields throughout the interior of the Sun, which extend outward, far beyond the "surface" features. The relatively rapid rotation of the Sun also generates circulation currents in the ionized gas from the Coriolis effect. These forces produce a complex rotating, and radial (inward and outward), circulation.



Since electrical currents generate magnetic fields, these strong ion currents within the Sun, and near its surface, produce strong internal and external solar magnetic fields. However, the Sun's rotation is not uniform over all latitudes. The Polar Regions rotate more slowly than the equatorial region, causing the magnetic field lines which are tightly bound within the strongly ionized gas in the interior, to become "stretched". The stretching of field lines has a loose analogy to elastic bands that have some give, but will eventually break with increasing tension. The magnetic field lines "break" and rejoin, creating a release of energy, and intense outflow of particles, mostly proton and electrons - the most abundant and lightest materials in the Sun.

Measurements of the solar magnetic field, from near the surface to deep space, have been made from the Earth, and with orbiting spacecraft. Surface measurements of the Sun derived from the interaction of ionized atoms with the magnetic fields (Zeeman Effect, for example) provide insight into the intensity and orientation of the complex solar magnetic field. Spacecraft measurements of the extended solar magnetic fields have been made since the earliest space flight missions. All but one of these missions have been in the ecliptic plane of the solar system, however. The ESA/NASA Ulysses Solar Polar Explorer spacecraft has completed a dual orbit flight around the poles of the Sun at slightly more than 1 AU. Information on the solar wind, magnetic fields, and cosmic ray influence has been made available to researchers since the October 6, 1990 launch on the Space Shuttle. The Ulysses Solar Polar Explorer found increasing solar wind speeds at higher latitudes (almost double the outflow speed), but lower outflow density. The figure below shows the X-ray image of the Sun taken by the Japanese Yohko spacecraft with the solar wind speed and density measured by the Ulysses spacecraft below.

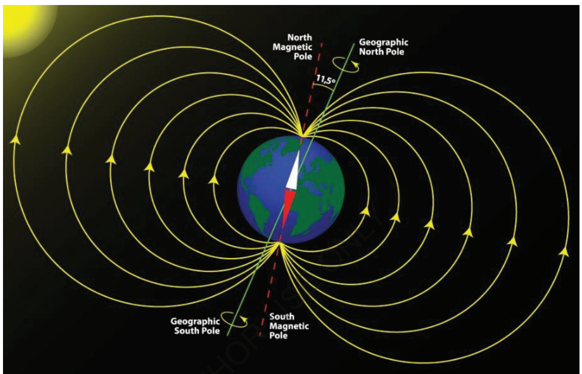
The internal magnetic fields of the Sun are highly variable, showing periods of relative quiet and periods of violent eruptions. The period of overall activity is related to the magnetic field lines stretching, then reattaching, with an average of eleven years in a cycle - the solar sunspot cycle. The sunspot number used to define the cycle comes from the number of dark, cool spots found on the photospheric "surface." The greatest number of spots occurs in the most active, violent period of particle and magnetic field eruptions. These active periods produce high levels of particle radiation that can be hazardous to manned space flight, as well as to satellites and spacecraft electronics, and to terrestrial communications. The magnetic field orientation reverses in these spots, as well as in the overall magnetic field, each eleven years.

As the solar activity level increases, the number of spots and the associated magnetic fields increase in number. Both a North and South Pole magnetic field appear as a pair associated with most of the spots, while larger field structures span the Sun's diameter, or may be even larger in scale. In addition, a weak but large-scale magnetic field structure extends to the edge of the solar system, actually defining the edge of the Sun's influence - the heliopause. This structure, which is the largest in the solar system, directs the outflow of solar wind particles, and creates a current of protons and electrons near the equatorial plane. The wave-like sheet feature of the largest solar magnetic field is sketched below, showing the irregular shaped disk which the planets travel through. As the planet crosses the current sheet, the particle radiation levels increase, which also pushes on the magnetosphere, injecting particles stored in the radiation belts surrounding the planet into the upper atmosphere near the magnetic poles. On the Earth, these aurora are generally confined to the poles, but more active solar bursts can produce aurora at lower latitudes.

Earth's Magnetic Field

The Earth's magnetic field is produced within the interior liquid metal core resulting from Coriolis-induced currents that generate a dipolar field. The rotation of the Earth induces the Coriolis circulation, which in turn, induces the highly conductive metal core to produce an electrical current flow, and the resulting magnetic field, The geomagnetic field model is based on the size of the core (strength of magnetic field related to planet's mass), and the rotational speed of the planet. Smaller planets (Mars, Mercury), or slower rotating planets (Mercury, Venus) have much smaller fields, while larger planets (all of the gas planets), especially Jupiter with a rotational period faster than any other planet, have a much larger magnetic field strength. There are some irregularities in this model, however.

The Earth's magnetic field is a dipolar field with the geomagnetic poles offset from the rotational poles by nearly twenty degrees. The magnetic field strength and orientation are not fixed, but change on both a short-term and a long-term basis. The reversal of the Earth's magnetic field orientation can be measured on the order of tens of millions of years from volcanic rock dating. Orientation of the axis of the magnetic field is continually changing, an indication of the dynamics in the Earth's molten core region that produces the changes. A diagram of the long- term (secular) orientation shift in England is shown below, Inclination is from horizontal, and declination is measured from North.

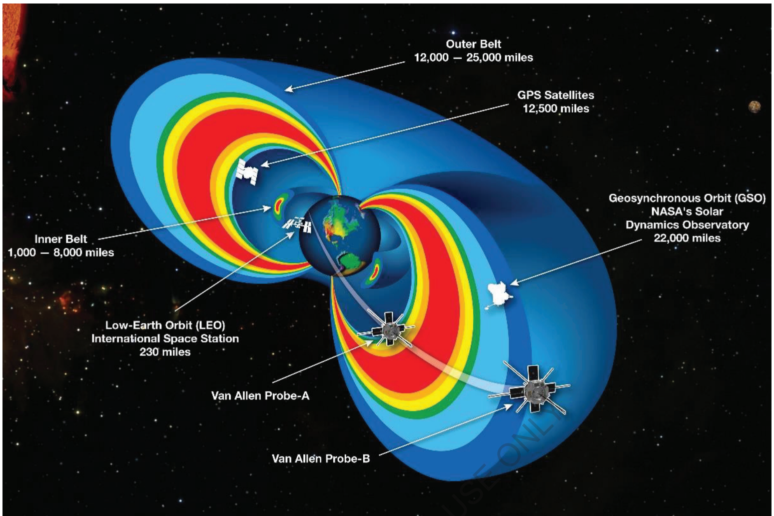


In addition, the solar magnetic field and solar wind pressure modulates the strength and orientation of the geomagnetic field on a short-term and long-term scale. The strength of the Earth's magnetic field varies from 10-4 Tesla (1 nT or nano Tesla is 1 gamma, or 100,000 nT is 1 Gauss) to much lower values far from the Earth's center.

The Earth's radiation belts are magnetic field lines which store charged particles from the Sun's particle wind stream. The highest intensity magnetic field regions correlate with the highest radiation levels near the Earth because of the ability of the magnetic field to hold higher energy particles. The diagram below shows the relative position of the magnetic field lines and the van Allen radiation belt, and the influence of the solar wind pressure on the shape of the magnetosphere.

Trapped Radiation

The density of ionized particles trapped in the Earth's magnetic field varies with altitude and particle energy. The peak density of electrons is different in number and radius than the proton belt because of the lower mass of the electron. Figure 3.3 shows a comparison of the radial distribution of electrons in Earth radii (1 RE = 6.4x106m) during solar activity minimum and maximum.



The general motion of the trapped particles in the planetary magnetic fields is a complex path. The spiraling, oscillating path also has an overall drift in opposite directions for the electron and proton belts because of their opposite charges (ring current). Figure 3.6 below shows the general spiraling motion of a particle moving under the influence of a magnetic field. The oscillation occurs from the increasing magnetic field strength near the poles. The particles slow, then reverse direction as they approach the poles. This increases the density of particles near the poles, as well as some of the kinetic energy. Hence, the highest particle radiation levels are observed near the polar regions. The magnetic poles are also the regions where most particles Interact with the atmosphere since the charged particles must travel along the field lines in helical (spiral) motion.

Radiation levels found in the Earth's magnetic field, the van Allen radiation belts, vary with altitude and latitude. In addition, there is some variation with longitude because of the variation in the magnetic materials contained in the upper mantle and crust of the Earth. The South Atlantic Anomaly represents a large concentration of ferromagnetic ore which brings an increased magnetic field strength near the Earth's surface and a resulting charged particle increase (higher radiation levels).

Radiation from the particles stored in the van Allen radiation belts varies with altitude and longitude, and diurnal (daily) phase as described above. There is, however, significantly greater radiation in orbital flight than near the Earth's surface because of the protective atmospheric layer, and the Earth's magnetic field protecting us from much higher solar wind radiation levels found outside the magnetopause.

To reduce particle radiation levels in spacecraft for sensitive electronics and crews, shielding can be used to slow particle speeds, thus reducing the energy of the particle. Upon slowing, however, the charged particles emit electromagnetic radiation equivalent to the kinetic energy lost, and proportional in energy to the slowing time. This secondary radiation can be hazardous, and must be considered when shielding is considered. A plot of accumulated radiation in rads (1 erg of energy per gram of material) for various shielding thicknesses in a one-year accumulation.

Earth's Atmosphere

The Earth's atmosphere is a thin layer of gasses that provides a boundary between the extremes of space and the Earth's surface. The atmosphere's composition and characteristics provide the inhabitants on Earth with protection from the vacuum conditions and radiation hazards of space, as well providing as critical life support elements. Earth's atmospheric features are important in understanding space operations for several reasons. Launch and reentry vehicles must pass through the atmosphere, requiring a detailed understanding of the physical composition and dynamical motion. Attenuation of radiation from space by the atmosphere, both particle and electromagnetic, is also important in planning shielding requirements in low-earth orbit where the atmosphere interacts strongly with radiation in the Earth's van Allen belts, and varying solar radiation.

The composition of the space environment near the Earth's upper atmosphere is determined by the gasses in the Earth's upper atmosphere, and its interaction with solar radiation. The diurnal (daily) motion and seasonal variations in the atmosphere also produce a continual variation in the space environment near the Earth. Similarly, solar activity produces a variation in the near- Earth space environment, and the solar system in general. Higher activity reflects higher energy coming from the Sun, and greater energy absorbed within the atmosphere. This higher energy means a higher temperature and therefore a higher density of gasses in LEO, and a higher drag on satellites operating below 1000 km. The higher density gas at higher temperature during the active solar periods also changes the chemistry in the upper atmosphere, and increases the atomic oxygen abundance, as well as other gas species. Spacecraft operations and design, therefore, includes determining the atmospheric effects on spacecraft launch, operations, and related hazards.

The Earth's atmosphere may extend several thousand kilometers above the surface in an active solar period, but generally ends at approximately 1,000 km. Because of the greater speed of lighter particles, lighter gasses are found at higher altitudes, with hydrogen and helium being a major component in the upper exosphere. In addition, atomic oxygen is found in the higher atmosphere and is a very active ion of oxidation. This ionized atomic oxygen presents a problem for exposed surfaces on spacecraft over periods longer than several days because of the reactivity of ionized oxygen.

Other ions in the upper atmosphere also present problems for spacecraft operations because of the charged particle environment and the conductivity of the spacecraft surfaces. Charge buildup on spacecraft represents a significant problem if charges in different regions of the spacecraft accumulate, allowing a discharge or unusual current flow in sensitive electrical systems. An overall charge buildup does not represent a hazard to spacecraft since the relative potential (voltage or charge) is not important except in the presence of an opposite charge. Because of this, a highly conductive surface on spacecraft is desirable. Composite materials, which are becoming more common on aircraft and spacecraft structures, are generally nonconductive, and when used in spacecraft outer structures, must be made conductive to avoid charge difference buildup.

Radiation and Biology

Radiation interacts with matter by the exchange of energy between the particle or electromagnetic photon radiation and the material receiving the radiation. Higher-energy radiation with sufficient energy to ionize atoms or molecules is known as ionizing radiation, and also represents potential tissue damage. The measure of damage from exposure to radiation includes both the radiation energy and the intensity, or number per unit time, of particles or photons. Thus, both the radiation energy and the intensity are used in specifying radiation damage. This combination converted into the amount of radiation energy absorbed is the accumulated energy absorbed, or dose.

Radiation dose is total energy (particle/photon energy) times intensity or number (rate), of the energy absorbed from the radiation. A more meaningful measure of radiation dose, or damage potential, is the absorbed dose times the radiation weighing factor, which produces the equivalent dose. The radiation weighting factor is dependent on the type of radiation for biological tissue. Potential damage of radiation to human tissue includes the weighting factor for electromagnetic (EM) and particle radiation. The lowest weighting factor is 1.0 for EM radiation, while the highest is for neutrons and alpha (helium nucleus) particles. The product of the absorbed dose times the radiation weighing factor is called the equivalent dose.

Because damage to biological tissue from radiation varies with the type of tissue, the equivalent dose is multiplied by a tissue damage factor called the tissue weighting factor. This produces the effective dose and forms the measure of radiation absorption and its potential damage to specified human tissue.

The duration of radiation is an important gauge for estimating potential biological damage, since radiation damage is cumulative. Maximum allowable safe levels of radiation in absorbed doses are listed in several of the following tables for acute and chronic radiation of body parts and whole-body exposure.

Roentgen

The amount of electromagnetic radiation (X-ray or gamma-ray) that ionizes 0.001293 grams of air

Absorbed Dose The amount of radiation absorbed in the medium specified. Based on the amount of radiation (Roentgens) times the effectiveness of absorbing medium. RAD (Radiation Absorbed Dose). Radiation corresponding to 100 ergs (cmgs units) of energy absorbed in one gram of any medium. Medium normally specified (e.g. tissue rad, Al rad).

RBE Relative Biological Effectiveness which is a measure of how damaging a given type of particle is compared to an equivalent dose of x-rays REM (Roentgen Equivalent, Man).The absorbed dose of any ionizing radiation which produces the same biological effects in man as those resulting from the absorption of 1 Roentgenof X-rays.

Equivalent to absorbed dose in RADS times the RBE.

Grey 100 rad (J/kg) The Joule-Kg absorbed dose

Sievert (Sv) 100 rem (J/kg) The Joule-Kg radiation equivalent (1 Sv = 1 Grey x RBE)

Radiation encountered in space produces biological and health effects classified in two different categories: acute, or early effects; and late or delayed effects (Nicagossian et al., 1989). The symptoms of the early effects are measured within hours, days or weeks following the radiation exposure. Latent effects are manifested months or years after exposure and include tissue damage, reduced fertility, induced cancers, genetic alterations and development abnormalities in the newborn

Atomic and subatomic particles are capable of interacting with the electron shell (chemistry) and the nucleus (species and stability) of molecules, materials, and cells, making them a hazard to humans, and to biological life in general, as well as to spacecraft.

Protons

Are consisted of a Hydrogen nucleus and are designated as p+, they are found in lower-earth orbits (to 3-5 Re) depending on energy of protons. Damaging to tissue because of high energy (higher mass = higher kinetic energy)

Electrons

The mass of an electron is approximately 1/2000 of the proton. Generally is lower in energy than protons because of it low mass. It's designation is e- or beta radiation. It is easily stopped but produce x-rays when energetic electrons are accelerated/decelerated. Electrons are found at much higher in radiation belt around Earth. They orbit in opposite direction of revolution of Earth. The electrons have a maximum flux (concentration) near geostationary orbit altitude. Low density material is sufficient for absorbing lower energy beta (e-) radiation

Alpha particles

Alpha Particles are consisted of Helium nuclei (2p + 2n), they are designated as He++ and are heavy (2 atomic mass units) and usually energetic. They seem to have a low penetration power except at very high energies. Biological damage is due to the alpha mass roughly similar mass to tissue mass. The most efficient energy transfer (deposition) is when masses of projectile and target are equal. They are extremely hazardous if alpha-emitting isotope is inhaled or ingested since the alpha particles can reach DNA and other susceptible parts of the cells. Alpha particles represent the greatest potential damage to tissue.

Cosmic rays

High energy ions from space called cosmic rays enter our solar system. They are necessary for the creation of cosmogenic nuclides in rocks at the Earth's surface, which we use for cosmogenic nuclide relationships, as well as the generation of 14C in our atmosphere, which is utilized in radiocarbon dating.

High Atomic Energy (HZE)

HZE ions, which must be the nuclei of heavier elements than hydrogen or helium, are the high-energy nuclei component of galactic cosmic rays (GCRs), which have an electric charge of +3 e or larger.

The letters "HZE" stand for high (H), atomic (Z), and energy (E). The nuclei of all elements heavier than hydrogen, which has a +1 e charge, and helium, which has a +2 e charge, are included in HZE ions. Since there are no orbiting electrons in the nucleus of each HZE ion, the charge of the ion is equal to the atomic number of the nucleus. Although its origin is unknown, supernova explosions are regarded to be their most likely cause. HZE ions, for instance, make up only 1% of GCRs compared to protons' 85%, which makes them uncommon. Like other GCRs, HZE ions move almost as quickly as light.

HZE ions are created by the Sun in addition to HZE ions from cosmological sources. HZE ions are occasionally created in small quantities during solar flares and other solar storms, along with the more common protons, but their energy level is significantly lower than HZE ions from cosmic rays.

Neutrons

Neutral particles that have potential tissue damage because of high mass of neutrons and lack of electrostatic repulsion in atoms. Energy (damage potential) transfer for this particle requires low mass absorber. Water is the best inexpensive moderator because of the hydrogen atoms.

Small Bodies, Dust & Debris

Solid material found throughout the solar system is increasingly abundant with decreasing size. The material from which planets were formed, planetesimals, range in size from kilometers to centimeters in diameter. The composition of the planetesimal material ranges from rock and metals to the ices found in comets and Pluto. The number density of these mid-sized objects is relatively low, but more common in the asteroid belts and near the numerous Lagrange points in the solar system.

Smaller, sand and dust sized material is much more abundant throughout the solar system and becomes a hazard near the Earth because of the very high relative velocity of the material in space. The orbital speed of the Earth around the Sun, for example, is approximately 30 km/s. This is much greater than the 8 km/s Low Earth orbit (LEO) speed and is representative of the meteor speeds of material coming from outside the near-Earth environment. Other small particles entering the Earth's atmosphere are from LEO sources.

Other material in the near-Earth space includes spacecraft debris that is classified as man- made. This type of material is rapidly increasing in number density and has become a serious concern for the USAF, NASA, Russia, and other organizations interested in space operations. One small sand-sized particle can penetrate some of the shielding on the Space Shuttle, or almost any other spacecraft, making space debris a true hazard. Even smaller particles are more abundant, but have less potential risk to spacecraft, and are also found near other planets. The lightest materials, and the most abundant, are the gas molecules and atoms found throughout the solar system.

Larger material, such as boulder sized ices and rock to asteroids and comets, present a much greater hazard to spacecraft than smaller objects, but are often large enough to track with radar or by telescope. These objects are part of a data base maintained by several countries, and by the NORAD Air Defense Organization. The separation of spacecraft and these natural objects is

somewhat easier with improving technology, but the hazards are still present because of the difficulty in measuring small, distant objects at high speeds.

The density of material in space has always been a concern, and was one of the first three experiments that was flown on the first U.S. spacecraft, the Explorer 1. More recent measurements of the LEO space environment has focused on the debris hazards to the International Space Station.

Comets

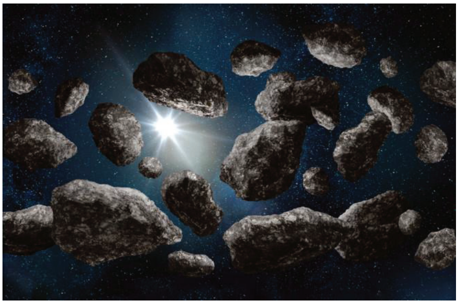
Comets that enter into the inner solar system have more eccentric and larger period orbits than the asteroids, on the average. Their ice and dust composition is actually a combination of ices, including CO2 ice, methane ice, ammonia ice and water ice, and dust particles made up of silicates, graphite, metals and oxides. The list of materials in the dust is a list of the materials in the terrestrial planets, from the lowest

atomic mass to the highest, except for the gases. The ice character of the comets makes them structurally weak, volatile, and easily broken up in near- encounters with the planets. The outer shells are easily vaporized if their orbit takes them inside the orbit of Earth or Venus.

A source of comets resides beyond the planetary region of the solar system in a spherical cloud of nebular material called the Oort cloud. The random-orbit objects can be perturbed by the Giant planets and brought into the planetary region. Once inside the orbit of Earth, these extremely cold objects begin solar heating and sublimating cloud of gas. The Sun's photon pressure stretched the cloud away from the Sun, giving it the traditional long tail.

Asteroids

Asteroids, also known as minor planets, are formed within the inner regions of the solar system which include the terrestrial or rocky planets. Their composition is somewhat similar to the terrestrial planets and represents the material that survived the Sun's early stages that heated the interior solar system, pushing the lighter materials outward and vaporizing the volatile materials, such as the ices. A simplified composition of the asteroids is rock and metal, and the various combinations include mostly rock and mostly metal. These metals are primarily iron and nickel. The C-type asteroid includes more than 75% of known asteroids. It is extremely dark (albedo 0.03) and is similar to carbonaceous chondrite meteorites. The S-type asteroid represents 17% of the population and is relatively bright (albedo 0.10-0.22). The primary composition is metallic nickel-iron mixed with iron- and magnesium-silicates. The M-type asteroid makes up most of the rest: of the population and is the brightest with an albedo of 0.10- 0.18). Its composition is pure nickel-iron.



Albedo = percentage of reflected light (reflected light to incident light ratio). The albedo range is 0-1, and bright surfaces would have high numbers while dark surfaces would be low. The Moon which is one of the darkest objects in the solar system has an albedo of 0.12 (Mercury is 0.11). The Earth's albedo is 0.37, while Venus' is 0.65 because of its thick cloud cover.

The smallest objects

Smallest of all of the solar system population would be the atomic and molecular gas, dust and solar wind particles. And although these are the greatest in number, the potential for spacecraft damage is usually superficial. However, small particles the size of sand grains to pea-gravel have velocities of 40 km/s (90,000 mi/hr) and a resulting kinetic energy (1/2 mV2) large enough to penetrate even shielded spacecraft. Fortunately, these are small in number except in the gravity wells near planets and moons. One example being the natural and man-made space debris in Earth orbit. With such high relative velocities and greater abundance than larger far more destructive material, the small sand grain-sized particles represent the most common material capable of destructive or erosive impact.

Review Questions & Problems

1) What object or factor mostly affects the Space Environment?

2) Make a comparison and contrast between solar flares and solar wind.

3) List at least six hazards found in Space Environments, with description of each and the possible drastic effects on spacecraft.

4) Describe the South Atlantic Anomaly including it's definition and what are the different effects on the spacecraft.

5) Please describe the mechanism that protects the Earth from the effects of the Solar and Cosmic charged Particles.

MATLAB Coding Exercise

- Assume that the coordinates p0, p1, p2, and p3 are in three dimensions.
- Starting at location p0 at time t0, your spaceship was travelling at a steady speed.
- At time t1, your spaceship is anticipated to arrive at a star at point p1.
- You just learned over the radio that an asteroid at the location p2 at time t0 has been identified.
- The asteroid travels at a steady speed.
- At time t1, the asteroid is anticipated to arrive at another star at point p3.
- The minimum distance between your spaceship and the asteroid must be greater than the distance d throughout the time interval between t0 and t1, else return false, according to the MATLAB function "safetrip."

Chapter 3: Orbits

Orbits describe the motion between two or more bodies interacting through an attractive force. The relative motion of the two objects plays an important part in the orbit features, as does the total energy. Most of the orbital characteristics can be determined by the two types of energy contained in the relative motion of two or more masses. The two relevant measures of energy are

Kinetic energy (Ek) =1/2mV2 Potential energy (Ep) = -GMm/r

If one of the two energy values is greater than the other, we have entirely different orbit conditions than for the opposite case. Gravitational potential energy is negative by convention, and kinetic energy positive, so adding the two energies to get the total orbital (Ek+Ep) will show a positive, negative or zero net value. Greater binding (potential) energy than relative speed (kinetic) energy means that the negative potential energy is greater that the kinetic. The total energy is therefore negative and the orbit is bound. Faster relative motion, where kinetic energy is greater than potential energy, means that the total energy is now positive, and the orbit is unbound; that is, the two bodies will continue to separate. If the energy total is zero, the potential and kinetic are exactly equal, and escape velocity has been reached by the smaller body.

Potential energy = -GMm/r12 = Ep

where:

G = 6.67x10-11Nm2/Kg2

M = mass 1 (same convention as m1, m2) m = mass 2

r12 = distance between position of mass 1 and mass 2

Kinetic energy = ½mV2 = Ek where V = relative velocity between mass 1 and mass 2

Total Energy = Etotal = Ep + Ek = Etotal

There are three types, or conditions, of orbits based on the total energy. Those conditions are bound, unbound, and neither bound nor unbound (escape condition).

- **Bound orbits** - Elliptical orbits - potential energy greater than kinetic energy
- **Unbound orbits** - Hyperbolic orbits - potential energy less than kinetic energy
- **Parabolic orbits** - neither bound nor unbound - potential energy is equal to kinetic energy (escape condition)

Table Energy relation and eccentricity for orbit type

Orbit	Circular	Elliptical	Escape/Parabolic	Hyperbolic
Condition	Bound	Bound	Neither	Unbound
Total Energy **C3 =** $E_k + E_p$	Negative ($E_k < E_p$)	Negative ($E_k < E_p$)	Zero ($E_k = E_p$)	Positive ($E_k > E_p$)
Eccentricity	e = 0	0<= e < 1	e = 1	e >1

Elliptical (bound) orbit

Negative total energy (negative potential energy is larger than kinetic)

This means that the relative speed is lower than escape speed (smaller object moving too slow to escape gravitational pull of larger object)

For circular case (circular orbits are a specific case of an elliptical orbit)

Vcircular = (GM/r)1/2

For elliptical orbits:

Velliptical = [G(M+m)(2/r-1/a)]1/2

Parabolic orbit

Total energy is zero (Ep + Ek = 0 which is equivalent to Ep = -Ek) This means that the relative motion (speed) is exactly escape speed

$$-GMm/r + 1/2mV2 = 0 \quad V2 = 2GM/r$$

$$V_{escape} = (2GM/r)1/2$$

Hyperbolic orbit

Since a hyperbola is an open curve with a negative semimajor axis, its eccentricity e is more than one and its total energy is positive. Spacecraft fleeing Earth's gravity at the start of an interplanetary journey and a flyby encounter with a target planet are two examples of hyperbolic orbits. Only one of the different branches of the hyperbola's conic section corresponds to the physical trajectory. The asymptotic velocities of the spacecraft at either end are calculated from the energy because the arrival and departure routes of the hyperbola are along two straight-line asymptotes.

Kepler's Laws

As seen from above the north pole of the Sun, the planets orbit the Sun counterclockwise, and their orbits are all parallel to what astronomers refer to as the ecliptic plane. The planets' orbits, which geometers refer to as ellipses, are stretched or flattened circles, not circles. As a result, Kepler was eventually forced to acknowledge this. Mars' movement presented particular challenges for Brahe because its orbit was the most elliptical among the planets for which he had comprehensive data.

Using the metric system, the period p is in seconds, mass (both M and m) is in kg, and G, the gravitational constant, is 6.67x10-11Nm2/kg2

Orbit ellipse

The formula e = c/a, where c is the separation between either focus or the elliptical orbit's center, determines an orbit's eccentricity. For an ellipse, the range for eccentricity is 0 e 1; the circle is a particular instance, with e = 0. For an elliptical orbit, semimajor axis and is positive, hence the total energy is negative.

The apse points are the most extreme places along the orbit's primary axis. Periapsis refers to the point that is closest to the attracting body, while apoapsis refers to the point that is farthest. These extreme positions are referred to as "perigee" and "apogee," accordingly, for orbits around the earth. Since apoapsis corresponds to a true anomaly of 180 degrees, the true anomaly is calculated from the periapsis orientation.

e = eccentricity = [1 - (b/a)2]1/2

which can range from zero for a circular orbit where b = a, to less than one where a>>b.

- **Bound orbits** 0 <. e < 1 (e is greater than or equal to zero and less than 1)
- **Unbound orbits**e > 1
- **Escape (parabolic) orbits** e = 1

The length of the minimum separation of two bodies in orbit is called the periapsis, or periastron (perigee for Earth-orbit, perihelion for solar orbit).

This is designated rp rp = a (1 - e)

The length of the maximum separation of two bodies in orbit is called the apoapsis, or apoastron (apogee for Earth-orbit, aphelion for solar).

This is normally designated ra. ra. = a (1 + e)

Velocity in orbit

The velocity in orbit is dependent on the separation and masses of the two orbiting bodies which varies for elliptical, parabolic, and hyperbolic orbits, but not for circular orbits. In the elliptical orbit, the relative velocity increases as the two bodies approach the periapsis, and slow as they approach apoapsis. Because energy is conserved, increasing velocity as the two bodies approach periapsis creates an increase in kinetic energy, hence the gravitational potential energy decreases (becomes more negative).

Expressing these velocities in terms of mass and separation shows:

Vcircular orbit = [G(M+m)/r]1/2 Vescape = [2G(M+m)/r]1/2 Velliptical = [G(M+m)(2/r-1/a)]1/2

Specification of an orbit and the position in that orbit requires a reference system. The most common reference system is the Keplerian orbital elements, or just, orbital elements. This reference has a fundamental origin as the vernal equinox. Also known as the first point of Aries, the vernal equinox represents the position in the sky at which the Sun appears to ascend through the plane of the ecliptic

(Sun-Earth plane, or Earth-orbit plane). Orbit variables is listed below, followed by a diagram of the orbital variables (also called elements or parameters) and the orientation.

- Plane of orbit - the plane in which the orbiting object travels
- Plane of the ecliptic - plane of orbit for the Earth around the Sun
- Eccentricity (e) - the flatness of the orbital ellipse
- Inclination - the angular difference between the orbital plane and a reference plane
- Nodes - the two intersecting points between an orbit and a reference plane
- Line of nodes - the line intersecting the two nodes, (also the intersection of the orbit plane and the reference plane)
- Periapsis (or periastron) is the closest approach distance for a 2-body orbit

2 – Line Elements

Orbital data for satellites is available in several formats, including Cartesian coordinates and Keplerian coordinates, and can be easily converted from one to another. Each orbital data set has a reference epoch to correct for the changing reference frame (background star positions) due to the Earth's axis precession. The three most common formats shown in several examples of NASA data below are the Cartesian Mean Position 1950 (M50), the Julian Date 2000 (J2K), and Keplerian (perifocal or orbital) elements.

ISS TRAJECTORY DATA

Given:

- Area (sq ft) : 9000.0
- Drag Coefficient (Cd) : 2.36
- 90 day mean solar flux (Jansky) : 179.0
- 12 month mean earth geomagnetic index : 2.35 Coasting Arc #1 (beginning on orbit 695)
- Position vector Time (GMT): 2001/295/09:39:30.000 Position vector Time (MET): N/A
- Weight (lb): 301906.0

M50 Cartesian   M50 Keplerian

Y	=	93016.17 meter	E	=	.0005995
Z	=	1623097.82	I	=	51.71101
XDOT	=	1349.159356	Wp	=	75.92851

YDOT =	-4915.420616 m/s	RA =	167.95779 deg
ZDOT =	5732.889437	TA =	301.84176
		MA =	301.90011
		Ha =	214.988 n.mi
		Hp =	211.192

M 50 Cartesian		J2K Cartesian	
X =	-21580200.01	X =	-6586082.76
Y =	305171.15 feet	Y =	19435.18 meter
Z =	5325124.08	Z =	1591115.33

XDOT = 4426.375841 XDOT = 1376.152176

YDOT = -16126.708058 feet/sec YDOT = -4900.187019 meter/sec

ZDOT = 18808.692378 ZDOT = 5739.510821

The mean element set is posted at the UTC for which position is just north of the next ascending node relative to the listed state vector time.

Two Line mean element set example ISS

125544U 98067A 01295.46347873 .00057998 00000-0 70152-3 0 9007

2 25544 51.6368 168.5225 0008202 233.9470 126.0925 15.57755304 6968

Satellite: ISS Catalog Number: 5544

Epoch time: 01295.46347873 = yr day.frac day

Element set: 900

Inclination: 51.6368 deg

RA of node: 168.5225 deg

Eccentricity: .0008202

Arg of perigee: 233.9470 deg

Mean anomaly: 126.0925 deg

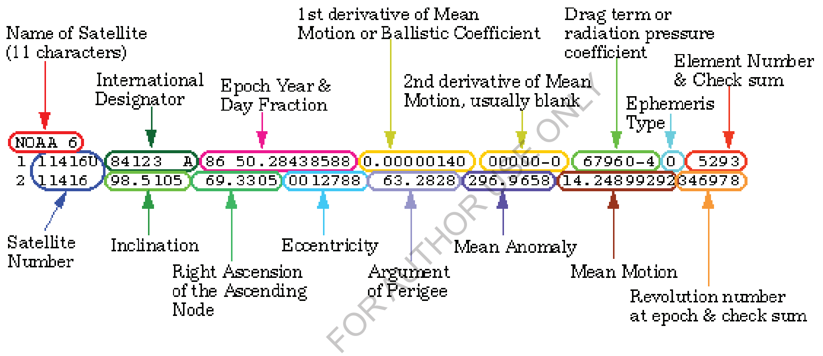
Mean motion: 15.57755304 rev/day

Decay rate:5.79980E-04 rev/day^2

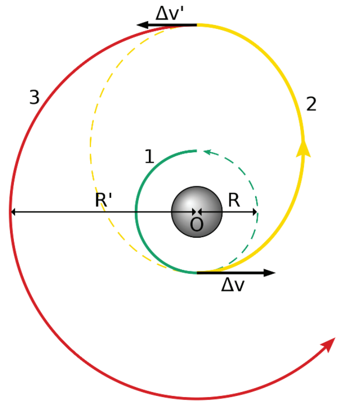
Epoch rev:696

Checksum:334

A useful graphic representation of the 2-line element set is shown below:



Transfer orbits



Changing from one orbit to another requires a change in orbital velocity which represents a change in kinetic energy and a change in potential energy. Additional change in energy/velocity is needed if a change in inclination plane is required

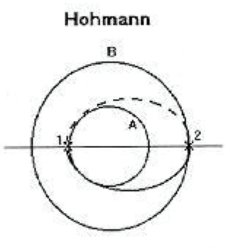
Hohmann transfer

Consider first that the spacecraft is already in solar orbit as it sits on the launch pad if you want to send a spaceship from Earth to an outer planet like Mars with the least amount of propellant. In order to send the spacecraft to Mars, the solar orbit must be changed: The perihelion (closest approach to the sun) and aphelion (farthest distance from the sun) of the intended orbit will be at the distances of Earth's orbit and Mars' orbit, respectively. Known as a Hohmann Transfer orbit, this orbit. The spacecraft's trajectory is the part of the solar orbit that takes it from Earth to Mars.

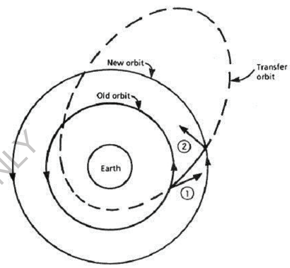
Spiral transfer

Direct transfer - this is a short period, high thrust transfer that reduces transfer flight time.

Note that energy, hence thrust, is required to change the semimajor axis of an object in orbit.



Type 1 transfer



High energy (direct) transfer

Transfer orbits with shorter periods than the Hohmann transfer are possible, but at the expense of higher thrust/energy, and often cost. The Type I transfer still has a two-boost transfer, but has a larger semimajor axis to the transfer ellipse. Rather than using one- half of the elliptical transfer trajectory, the Type I covers a fraction of the half-ellipse

Hohmann transfer - 180o 2-burn transfer

Type II - greater than 180o, and less than 360o

Type III - greater than 360o, and less than 540o

Calculations for a Hohmann transfer period from Earth

atransfer = ½ (aA + aB)

A period calculation uses the same third law of Kepler, but only one-half since the second half represents a return path to the departure planet.

ptransfer = ½ (atransfer3/2)

Start with the semimajor axis values for the planets aEarth = 1.00 aVenus = 0.72 AU aMars = 1.52 AU

aJupiter = 5.20AU aSaturn = 9.56 AU aUranus = 19.22 AU aNeptune = 30.11 AU aPluto = 39.55 AU

Earth-Venus pHohmann = ½ [(aHohmann)]3/2 = ½ [(aA + aB)/2]3/2 = ½ [(1 AU + 0.72 AU)/2]3/2 = 0.40 yr

Earth-Mars PHohmann = 0.72 yr

Earth-Jupiter PHohmann = 2.73 yr

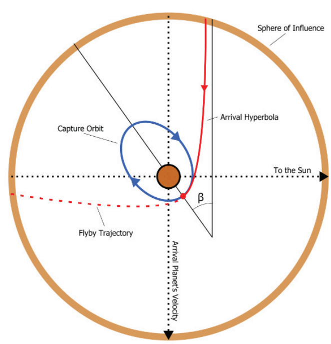
Earth-Saturn PHohmann = 6.07 yr

Earth-Uranus PHohmann = 16.1 yr

Earth-Neptune PHohmann = 30.7 yr

Earth-Pluto PHohmann = 45.7 yr

Planetary Transfer Sequence



Typical orbit transfer procedures for boosting a spacecraft from one orbit to another orbit using either a Hohmann (elliptical) transfer, direct transfer, or spiral transfer orbit. However, two transitions are needed when traveling from one planet to another since the gravitational influence of the individual planets and the Sun must be taken into account. These transitions take place at the departure from the original planet when transitioning to the orbit transfer orbit point (first burn in Hohmann transfer), Sun's gravity field between the planets and the Orbital injection point of the target planet (second burn in Hohmann transfer).

Take, for example, a spacecraft sent from the Earth to Venus. Inside the Earth's gravity field that dominates the space close to Earth, the spacecraft will need sufficient velocity to leave the Earth's gravitational pull in order to reach the Sun's gravitational influence. This is the concept of the sphere of influence. The planetary sphere of influence is usually tens or hundreds of planetary radii from the planet. The Sun's sphere of influence extends throughout the solar system because of its dominant gravitational mass (99.9% solar system mass), except near each of the planets and their moons.

An estimate of the radius of the sphere of influence can be made with the Laplace expression:

$$\mathbf{R_{sphere} = R(M_{planet}/M_{Sun})^{2/5}}$$

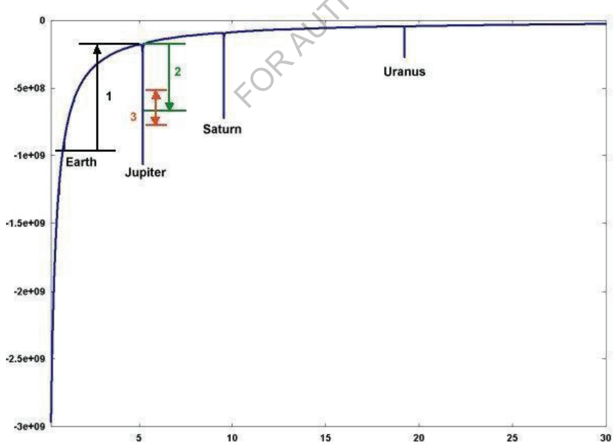
with R_{sphere} the sphere of influence radius, R = the mean orbital distance of the planet from the Sun

M_{planet} = the mass of the planet

M_{Sun} = the mass of the Sun

The Earth's sphere of influence, for example, is 9.27×10^8m (almost 1 million km), or 140 R_{Earth}.

A plot of the gravitational potential energy in the solar system can be used to demonstrate the transitions for interplanetary space flight graphically.



Plot of the solar system gravity potential in arbitrary units on the left an distance in Astronomical Units on the horizontal scale. Jupiter, the largest planet with the deepest gravity well, is followed on the right by Saturn and Uranus. If the negative sign of the potential is removed, the graph would be inverted. Getting to Jupiter from Earth is shown as a three step process. The first step is not shown, but represents getting out of the Earth's potential well which is too small to show up on this chart. Second (arrow 1) is the transfer orbit to Jupiter's sphere of influence. The third is a retrograde boost to go into orbit around Jupiter shown as an arbitrary altitude indicated by the length of arrow number 2 (green). Notice that getting closer to Jupiter requires more energy, meaning a larger retro-booster. A compromise would be to reach a lower altitude by making the Jupiter orbit elliptical as shown with arrow 3 (red). The more eccentric the orbit, the closer the spacecraft can get to Jupiter in its periapsis, but the less time it spends in that region. This plot is also a representation of the escape energy required to get away from the Sun, or from the outer planets. For the Sun, which is the solar system, the escape energy value is the energy difference between the spacecraft position and the top line (0 line). For escape from the planets, it would be the energy difference between the position in orbit (or on the surface) to the transition at the Sun's curve which is the sphere of influence. Notice how deep the gravitational well is for Jupiter. Note that it would take the same amount of energy to reach Jupiter from its sphere of influence to its surface or lower orbit that it would to each the sphere of influence from the surface or lower orbit position.

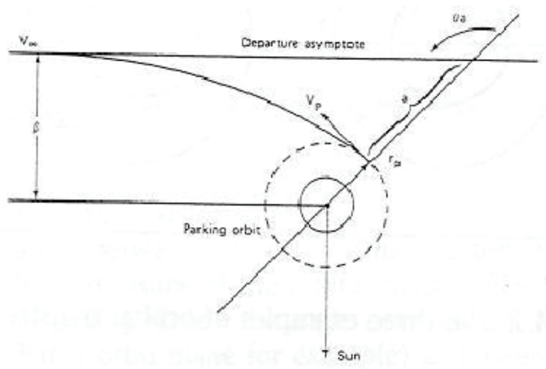
Simple interplanetary mission

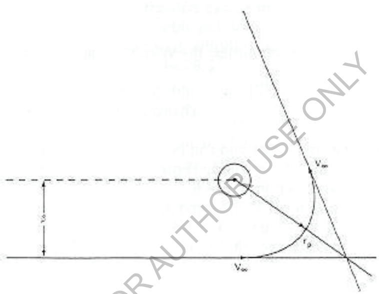
To reach Venus from Earth, the first step would be to place the spacecraft in a parking orbit around the Earth in preparation for orbit. This would be a simple circular orbit launch to low-Earth orbit.

Next would be a departure orbit from Earth to reach the sphere of influence at zero velocity plus the velocity to reach Venus in the transfer orbit. This is called hyperbolic excess velocity and takes the spacecraft to Venus and no farther.

Next is to slow the vehicle as it reaches Venus in the reverse method of leaving Earth. A transition boost is needed to slow the spacecraft and place it in orbit around Venus called orbit insertion. This is the transfer orbit velocity minus the departure (arrival in this case) velocity for Venus.

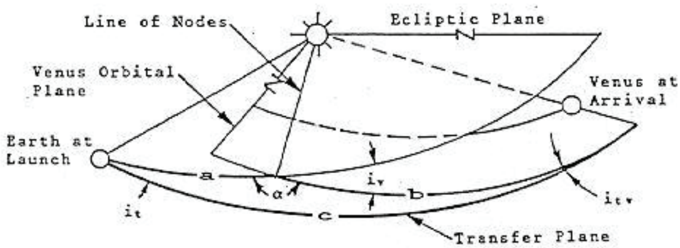
A final circular orbit around Venus would likely be a circularization burn, but that's icing on the cake.





Notice that these transitional hyperbolas assume a circular orbit at the departure and target planets. This is not necessary, but is simpler, certainly for this example.

Since Venus and the Earth are not on the same orbital plane, an inclination plane change must be made. This is normally done at one of the nodes of Venus, except that it is not necessary since the transfer orbit plane will provide a simple path that is independent of the Earth's orbit and Venus' orbit, and hence, does not need a plane change. With the requirements that the position of Venus be correct at the arrival of the spacecraft, a calculation of the inclination difference of the spacecraft at departure from Earth and arrival at Venus produces an approximately 4o plane change. The most effective way to accomplish this is to time the launch and departure hyperbola boost with the correct alignment of the two planets. This alignment will describe the launch dates and some of the launch window criteria.



Inclination plane change

To change inclination planes, the velocity change is expressed as

$$\mathbf{\Delta V = 2Vo\sin(\Theta/2)}$$

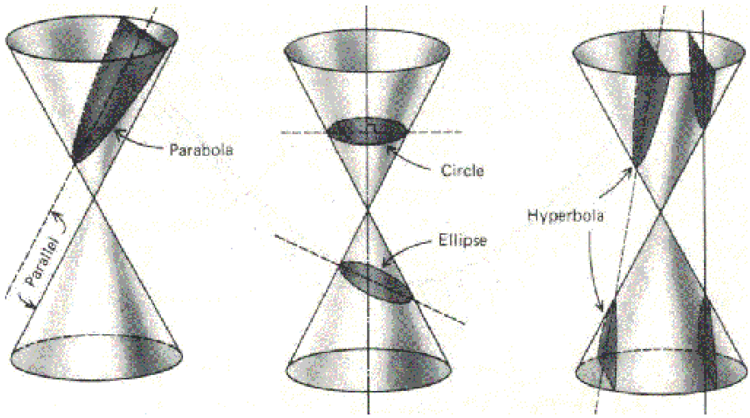
with Vo the orbital velocity at the time of the change, and Θ is the inclination angle change.

This usually requires a large fuel/energy budget and is performed at or near one of the two nodes (ascending or descending). Those nodes are shown in Figure 1 with another example shown in Figure 9.

The procedure for changing orbital planes is often executed during the transfer burn to reduce the propulsion fuel needed for both burns. This is optimal at the apogee in Earth orbit since that is the lowest speed in the orbit (where Vo is smallest).

Conic Sections

Conic sections which define the shape of the 2-body orbits are called that because the intersection of a cone and a plane at various angles produce the geometrical equivalent of an orbital path. The cone-plane intersection can reproduce circular, elliptical, parabolic, and hyperbolic geometries that can then be used to piece together the trajectory of an Earth-orbit, lunar, interplanetary flight. The exception is the launch to orbit trajectory that is more complex than any of the four conic shapes for several reasons.



Circular satellite orbit

This is an easy case because it describes the simple circular 2-body orbit. An actual circular satellite orbit around a planet or moon is only approximated by a circle because of the asymmetries in the planets and their moons, and the gravitational perturbations from other objects.

Departure orbit

Traveling to another planet requires escape from Earth's gravity to reach an elliptical transfer orbit. This departure orbit segment is a hyperbola, the path from Earth or an Earth parking orbit that matches with the transfer ellipse segment If entering orbit around the target planet or moon, the process is just the opposite, where the arrival is on a hyperbolic trajectory until a circular orbit entry boost is executed.

Transfer orbit

Transferring from one orbit to another, or from one planet to another on an interplanetary mission is accomplished with an elliptical orbit that has its perihelion at the departure orbit (if traveling to planets outside the Earth's orbit for example) and an aphelion at the target planet's orbit. The reverse applies if traveling to planets inside the departure planet's orbit.

Arrival orbit

The arrival orbit for an orbit entry or a swing-by gravity assist is also a hyperbola.

Actual planetary orbit transfer - Mars

The actual ΔV needed for the transfer hyperbolic excess velocity is not the same as the ideal value described above since the orbits of Mars and Earth are not circular nor coplanar orbits. To minimize the total required ΔV, a compromise must be made between a departure at one of the nodes, a departure at aphelion and arrival at perihelion, a plane change at one of the nodes, and an ideal departure 180o from arrival prescribed for the Hohmann transfer.

Non-circular orbits add broad variations in the actual minimum ΔV which makes the results difficult to describe in any meaningful way. Nonetheless, a graphic representation of the ΔV values based on departure and arrival dates can be plotted to find ΔV (or C3) minima. From these minima velocity/energy values comes the required launcher size, and/or the maximum payload mass. These plots are called pork chop plots because it describes the shape of the plotted data (sometimes called butterfly plots).

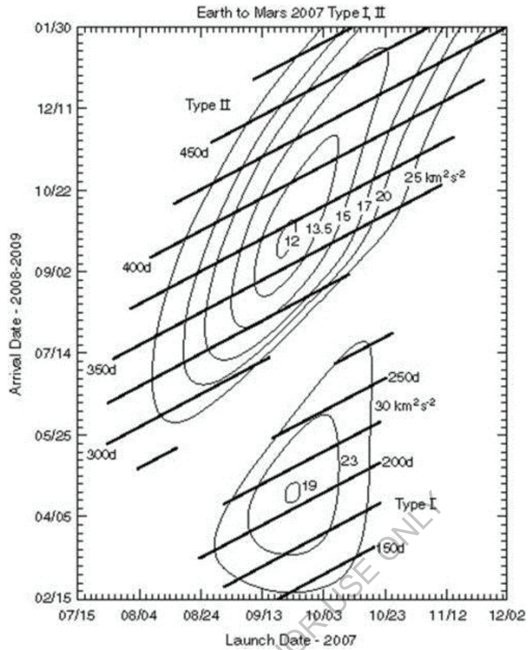
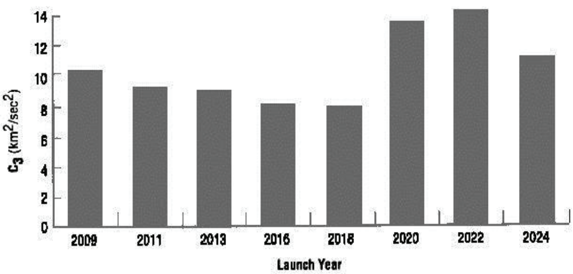
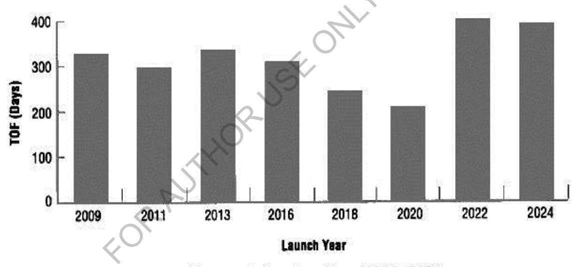


Figure: Plot of the C3 values for varying departure dates (horizontal axis) and arrival dates (vertical axis) is used for initial planning of an Earth-Mars mission for the 2005-2006 opportunity. Red lines are time-of-flight values in days. Type I orbits appear at the bottom, while Type II are shown in the collection above the Type I. The minimum C3 energy can be found in the center of the closed blue curves (isophotes) which are listed in km2/s2. The gap between the two regions is often the demarcation that surrounds one of the nodes. Note in the figure above that the minimum energy (C3) values are more than the C3 calculated earlier for Earth and Mars in circular, coplanar orbits. For realistic computations, the lowest possible transfer energy would be for a launch from the aphelion of the Earth's orbit and an arrival at the perihelion of Mars' orbit. In the ideal case, the Earth also be at the spacecraft departure position while also at the Earth's aphelion (perihelion of transfer orbit). In addition, Mars would ideally be at the correct arrival position of the spacecraft as it reached its orbital perihelion. Alignment of these four variables, as well as the optimum alignment of the line of nodes (ideally at the departure point at Earth's orbital aphelion), are extraordinarily rare.

Because of their scarcity, the coincidence of these five variables can be better visualized not in single pork chop plots plotted near the synodic conjunction, but in a representative plot of the lowest C3 transfer energy values for each of the conjunctions (2005 conjunction data is shown in the pork chop plot above). A histogram plot of the approximate C3 minima for future Earth- Mars conjunctions is shown below.



Cargo mission departure energies, 2009–2024



Cargo mission durations, 2009–2024

Figure: Histogram plot of the minimum energies and corresponding time-of-flight values for the Earth-Mars conjunctions listed. C3 energies differ because of the changing positions of the two planets in their orbits with respect to their apsides (aphelion and perihelion) that changes over time. Even though these minima repeat on a 30-year cycle, the repetition is not perfect and long-term effects change the repetition pattern of these values. Cargo missions presented here are calculated for minimum boost energy (maximum payload) which are normally Type II orbits. Crew mission analysis would emphasize minimum time-of-flight (TOF) and look distinctly different. In general, C3 energies would be greater, and TOF would be shorter for crew missions at the same conjunctions. Note the 2.1-year conjunction/synodic period (NASA TM-208533).

Orbit types

The various types of orbits can describe the orbital energy and the orbital shape (eccentricity mostly), or reference orbital orientation, orbital period or planetary surface coverage for the orbiting satellite. Communication, remote sensing, and surveillance all require specific orientation throughout the satellite operation. A number of the types and uses for these orbits are given below.

1.Bound (elliptical) orbit

A bound orbit, which is also an elliptical orbit, has relative kinetic energy less than the combined gravitational potential energy.

2.Unbound (hyperbolic) orbit

A hyperbolic orbit is unbound, meaning that the relative kinetic energy is greater than the combined gravitational energy

3.Escape (parabolic orbit)

A parabolic orbit is neither bound nor unbound since the relative kinetic energy is exactly equal to the combined gravitational potential energy. The parabolic orbit conditions are the same as escape velocity.

4.Prograde orbit

A prograde orbit has an inclination angle less than 90o which follows the same direction as the Earth's rotation

5.Retrograde orbit

A retrograde orbit has an inclination greater than 900 which travels in reverse direction to Earth's rotation

6.Polar orbit

A polar orbit has an inclination of 90o which allows world-wide coverage over a period of hours to days depending on altitude. This is an orbit commonly used for meteorological and surveillance satellites.

7.Geosynchronous

A geosynchronous orbit has an orbital period equal to the Earth's rotation period of 24 hours (23h56m4.09s). The semimajor axis of this orbit is 42,164 km. The inclination for this orbit is assumed to be 0o.

8.Geostationary

A geostationary orbit is also geosynchronous, but has an equatorial orbit (i = 0o), with an eccentricity of zero. This provides a fixed communications platform with respect to the Earth, and is used for communications, remote sensing (whether satellites, for example.), and surveillance.

9.Sun-synchronous

A sun-synchronous satellite orbit maintains a constant orientation between the Sun and Earth which is useful for several applications. The most common are the remote sensing applications satellites (Landsat, for example) and astronomical observation satellites (IRAS, for example). These applications require either full back-illumination, or complete shadowing from the Sun. To accomplish this, a nearly polar orbit is used. If the orbit were polar, the orientation of the orbit would be fixed with respect to the Sun, but not the Earth. This would show a 0.98o per day change in orientation as the Earth does an make the Sun-satellite orientation seasonal, as the Earth is. Hence, the spacecraft needs a -0.98o per day retrograde motion in the orbital plane to counter the Earth's orbital motion around the Sun. To do this, an orbital retrograde rotation can be made by using the oblateness of the Earth to place a torque on the orbit and produce a precession of the nodes (rotation of the orbital plane). A range of altitudes and corresponding inclinations are available. Landsat satellites use a = 709 km and i = 98.2o.

10.Molniya orbit

Russia has traditionally used communication satellite orbits that are highly-inclined with high eccentricity for their high-latitude ground stations with the extended orbit (apogee) over the higher latitudes. The Molniya's 12 hour orbit period also allows for communications between the Asian and North American continents, since it is one-half of the sidereal day (43,082 sec). The inclination for this orbit is 63.4o, with a semimajor axis of 26,562 km, and with varying apogee and perigee values that satisfy the desired period and semimajor axis.

11.Tundra orbit

A tundra orbit is an eccentric, high-inclination (63o) orbit similar to the Molniya orbit but with a period twice as long (one sidereal day). Like the Molniya orbit, this tundra orbit is used for communications at latitudes far from the equator.

12.Parking orbit

A parking orbit is a temporary orbit commonly used for spacecraft checkout operations before departure from Earth.

13.Graveyard orbit

A graveyard orbit is a permanent, higher-than-normal orbit used to remove defective or aging spacecraft from the busy geostationary region

14.Walking orbit

A walking orbit gets its name from the orbit's rotation or precessional motion due to the planet's asymmetrical shape. An example is the Earth's oblate shape (larger equatorial diameter than polar diameter) due to its relatively rapid rotation.

15.Halo orbit

A halo orbit is found not around a celestial object but around either the Lagrange L1 or L2 stability regions (objects within or close to the L1 and L2 regions are not in a stable position, while objects in halo orbits remain in a stable orbit).

Satellite orbital coverage

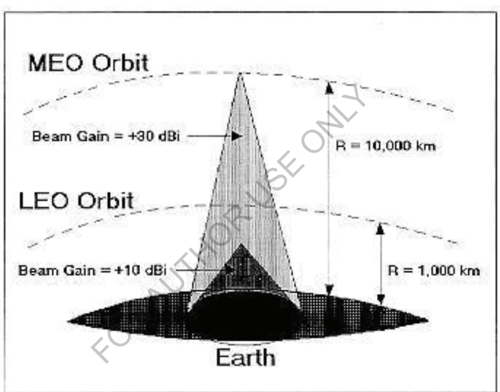
The surface coverage (exposure) of satellites has three critical aspects. One is the communication coverage area, another is the observational coverage (approximately equal), and the third is the communication power required at the satellite's altitude. As shown in the diagram below (Figure 4.7), the surface coverage depends on the orbit's altitude, with a maximum of one-half of the planet's disk covered if the satellite is located at a very distant point. The number of satellites required to cover the globe (the low and moderate latitudes) is inversely related to the orbital altitude. For example, three geostationary satellites can cover the Earth, while several dozens of communications satellites are required at a 300 km orbit.

In addition, the satellite power required to maintain communications is a function of the altitude (squared). Lower altitudes require less power, more satellites, and multiple inclined orbit planes. Figure 12 below shows the relationship between the orbital altitude and the number of satellites for global communications coverage.

Intermediate orbits are useful for communications and navigation because of the lower number of satellites required compared to lower altitude satellites and the more significant number of satellites required for communications or navigation needs than for the geostationary set of three. For example, the Global Positioning System (GPS) navigational satellites are composed of 24 satellites with six satellites in each of 4 orbital planes with an orbital period of just under 12 hours. This arrangement provides coverage roughly lower than 80o latitudes, with a minimum of 5 satellites in view for 95% of the time.

Lower-altitude communication satellites include the IMARSAT, IRIDIUM, and others. The less expensive, lower-power satellites are helpful because of the greater tolerance of motion than navigational satellites and the more significant number of channels available from the individual satellites. Another primary concern with orbital altitude is the radiation dosage above approximately 600 nm.

Communication power loss is also a concern with satellites and can be a determining factor in the orbital altitude decision for the constellation. A beam gain factor (loss) diagram is shown below for a constant antenna size and gain/r2.



Orbital launch velocities

While orbital velocities decrease as the inverse square root of distance, the instantaneous velocity required to reach orbit increases with increasing distance; this launch velocity, or ΔV, is a measure of the gravitational potential energy that must be overcome to reach a specified orbital altitude. Because the ΔV value depends on the gravitational potential energy, giant planets require a higher ΔV than more minor planets. The enormous ΔV energy requirement would be departing from a near-solar orbit. Conversely, the velocity increase of a spacecraft at a specified orbital distance to the orbited object (or a lower orbit) is the same as the ΔV required to reach orbit (or higher orbit). For example, an enormous booster would be required to launch a spacecraft from Mercury to the Earth because of the Sun's powerful gravitational pull (potential). This implies that the same-sized booster would be required to sufficiently slow a spacecraft from Earth orbit to reach an orbit around Mercury.

Orbit Precession

The orbit of an object around a spherically symmetric object and without other external forces will not change with time or process. However, an orbiting object around an oblate-shaped object such as the Earth will precess if it is not in the plane of the equator in two ways.

1. The periapsis will precess
2. The line of nodes will also precess

The precession rate depends on altitude and eccentricity.

This implies that the orbit must have the same orbital period and precession rate to rendezvous or synchronizes motion with another orbiting object.

Precession occurs in the Earth's orbit around the Sun (the night sky changes) in 26,000 years. The Moon's precession in its orbit around the Earth has been slightly over 18 years (saros). This means that the eclipse which is due to the alignment of the Earth, Moon and Sun, repeats each saros period divided by two (9 years) because an eclipse can occur on either side of the Earth-Moon alignment with the Sun. These events occur in the plane described by the Earth-Sun orbit named the ecliptic plane.

Precession also occurs in most satellite orbits, but is much smaller for distant satellites and satellites on the equatorial axis of the orbited body.

Review Questions and Problems

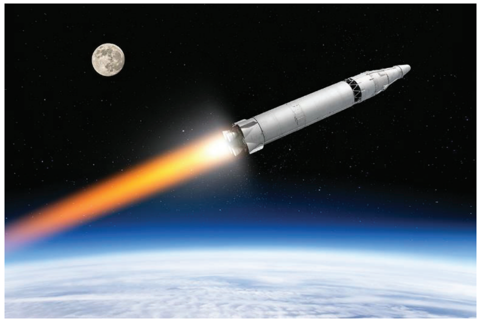
1) What are the two major types of satellites that can be found in the Geosynchronous/ geostationary Earth Orbit (GEO)?

2) A small satellite is in orbit around a planet at an average distance of 0.005 AU. If the orbital period of the satelliute is 40 days, what is the mass of the planet?

3) What is the average distance between the Earth and the Sun?

4) A rocket with mass M=12.0M=12.0 metric tonsmetric tons is moving around the Moon in a circular orbit at the height of h=100 kmh=100 km. The braking engine is activated for a short time to lower the orbital height so that the rocket can make a lunar landing. The velocity of the ejected gasses is u=1.00×104m/su=1.00×104m/s. The Moon's radius is RM=1.74×103kmRM=1.74×103km; the acceleration of gravity near the Moon's surface is gM=1.62m/s2gM=1.62m/s2. Suppose that, at point A, the rocket is given an impulse directed toward the center of the moon, to put it on a trajectory that meets the Moon's surface at point C (see right part of the figure). What amount of fuel is needed in this case?

Chapter 4: Propulsion



There are four general uses for propulsion on spacecraft. First, some spacecraft may not use internal propulsion systems beyond the launch vehicle, such as onboard thrusters or booster stages. Instead, most spacecraft use some propulsion as attitude control and/or orbital maintenance or trajectory control. This chapter will discuss some of those systems, emphasizing chemical rocket propulsion.

Newton's third law is the most crucial concept in rocket propulsion: an applied force reacts with an equal and opposite force. This principle applies to all of the propulsion systems except photon propulsion.

Three primary measures of propulsion system effectiveness and efficiency are the following.

1. **Thrust -** the force available for launch or flight
2. **Thrust duration -** the time available for acceleration
3. **Thrust efficiency -** measured by the specific impulse = Isp

Thrust

Propulsion thrust is used for launch, orbital change, maintenance, station-keeping operations, and attitude control.

- **Launch** - very high thrust required (10^6 -10^7 Newtons (note 1 Newton=0.2248 lb). Some of the examples are the Atlas, Delta, Proton, Ariane 5, Long March
- **Apogee and orbit boost** - moderate thrust needed (10^4 -10^4N). Examples are Upper Inertial Stage (IUS), Payload Assist Module (PAM), andCentaur upper stage.
- **Attitude control** - low thrust needed (1-10 N or less typical) - short duration bursts and frequent operational cycles.

Thrust Duration

A measure of the total energy available in the propulsion system. According to Newton's third rule of motion, thrust is a mechanical force that is created as a result of accelerating a mass of gas. The engine and aircraft are accelerated in the opposite direction while a gas or working fluid is propelled to the back side. We require a propulsion system of some sort in order to accelerate the gas. Let's consider the propulsion system as a machine that accelerates a gas for the time being.

There are three different durations: Short duration which consists of a time range between .01 and 10 seconds and is primarily used for attitude control; Intermediate duration which ranges between 10 to 1000 seconds and is used for launch and boost and lastly we have Long Duration wich ranges between 10^3 to 10^7 seconds and is used for deep space propulsion (including ion, and nuclear propulsion).

Thrust Efficiency

The efficiency of the propulsive system, also known as the effectiveness of the jet engine, is concerned with how effectively the thermal energy of the fuel is converted into kinetic energy, represented by the jet velocity, and how this energy is then used to best drive the aircraft forward.

Thermal or internal efficiency is the word for the effectiveness of converting fuel energy to kinetic energy. Like other heat engines, it is governed by the cycle pressure ratio and combustion temperature. Regrettably, the turbine's tolerance for thermal and mechanical stresses places a limit on this temperature. New methods and materials are always being developed to lessen these restrictions.

The quantity of kinetic energy lost by the propelling mechanism has an impact on the efficiency of converting angular momentum to propulsive effort, also known as the propulsive or external efficiency. [W(vJ - V)2]/2g is an expression for the waste energy wasted in the jet wake, which is a loss. (vJ - V) is the waste velocity. Therefore, the pure jet stream is less productive across this speed range than a propeller system since it spends a lot more energy there. Although the jet stream remains to emerge from the engine at a high velocity, its velocity relative to the surrounding atmosphere is reduced and the waste energy loss is minimized, hence this component changes as aircraft speed increases.

There are three different types of thrust efficiencies: Low (1-100 s) and it is used in attitude control or in small spacecrafts (for example cold gas, resistojets which are used for orbital maintenance on larger spacecraft); Moderate (100-500s) which is used for Launch and Boost operations and for attitude control on larger spacecraft (for example liquid monopropellant, liquid bipropellant and solid fuels); and High (500-5000s) used for Interplanetary propulsion, these are mostly in design stage of development. Some nuclear and Ion systems have been tested.

Basic Rocket Propulsion

Momentum = p = mass x velocity

Measures mass flow and/or propulsive mass reaction. Newton's third law states that there is an equal and opposite reaction for a reaction. Since thrust is more significant for a larger mass flow and higher velocity exhaust, high propellant flow and a high combustion rate, along with high-velocity exhaust, are desirable.

Force = F = dp/dt = d(mv)/dt = vdm/dt + mdv/dt

Force is momentum change or acceleration. When the mass is constant, this is often seen as F = ma (mass times acceleration).

Thrust = force (exact dimensions)

Thrust comes partly from the pressure difference between the rocket nozzle and ambient (surrounding) pressure and partly from exhaust mass reaction force (momentum forward equals exhaust momentum rearward).

Thrust from exhaust gas = mass rate flow times exhaust velocity

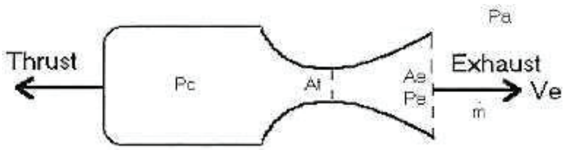
Thrust = T = Ve dm/dt

Thrust from internal (nozzle) and external (ambient) pressure difference

T = PeAe - PaAe = Ae(Pe - Pa)

The greater the force, the more significant the difference between internal pressure, Pe, and external pressure, Pa. Greatest when ambient pressure, Pa, is zero (in the vacuum of space).

Total thrust = Ve dm/dt + Ae(Pe - Pa)



- Pc = chamber pressure At = throat area
- Pe = exhaust pressure Ae = exhaust chamber area
- Pa= ambient pressure Ve = exhaust velocity dm/dt = mass flow rate

Exhaust Velocity

The chemical rocket engine/motor performance is determined, in part, by the exhaust velocity. Higher exhaust velocity improves thrust performance. Ve is often expressed by:

$$V_e = \sqrt{\frac{2\gamma R_o T_c}{(\gamma - 1)m}\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{(\gamma-1)}{\gamma}}\right]}$$

This shows that increased exhaust velocity (and performance) is available by: It increases the chamber-to-exhaust pressure ratio. Lower exhaust pressure in space makes the chemical rocket more efficient than in the atmosphere. A compromise is made between the performance of launch vehicle engine(s) at sea level and in space. It increases the combustion chamber temperature. The limitation on this temperature is based on the chamber and nozzle design and materials. Decreasing molecular weight - Hydrogen is the best fuel but needs an oxidizer. H_2 + O_2 is one of the best propellant combinations. Lower specific heat ratio - This is a property of the fuel chemistry and has a limited range available.

Specific Impulse

For rocket engines, specific impulse serves as a gauge of efficiency. You might say it is the change in momentum per unit of mass for rocket fuels, to be more precise. Simply put, particular impulse measures how much thrust (push) builds up during fuel combustion.

Impulse = I = Thrust (force) x time increment t, or I = T t

Specific Impulse = Isp = I/mg   m = propellant unit mass, g = gravitational acceleration at the Earth's surface

Isp = T t/mg for constant thrust and constant mass flow

A measure of efficiency either for fuel or engine/motor. Defined as the thrust per unit weight flow rate of propellant.

High Isp = high thrust efficiency per unit mass - essential, especially in mass-critical designs

For chemical fuels

Isp = K[Tc/m]1/2

K = constant

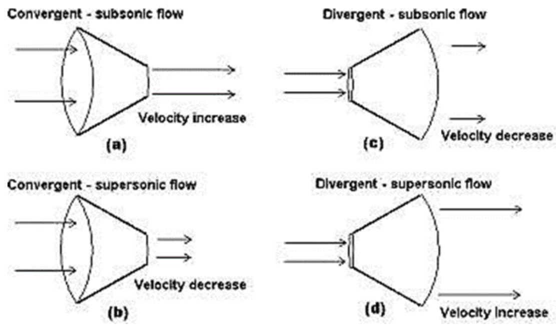
Tc = chamber temperature. More significant is better, but there are operating temperature limits.

m = molecular weight (mass) of exhaust gases. More minor is better, so hydrogen is best since its molecular weight = 1.

Nozzle Design

The exhaust nozzle is one of the most important design criteria for the chemical rocket motor. Exhaust velocity maximization and pressure difference optimization are developed within the nozzle parameters, from curvature to internal diameter ratios. First, the nozzle must generate the maximum exhaust velocity exiting the combustion chamber and entering the surrounding atmosphere or the vacuum of space. To accomplish this, the nozzle must constrict the exhaust gas flow to increase the exhaust velocity, then provide a divergent flow to accelerate the gas to an even greater velocity.

In subsonic gas flow (less than the speed of sound), a gas will accelerate if constricted in a convergent cone, as shown below in Figure 2 a. If subsonic gas flows through a divergent cone, the gas speed will decrease. For supersonic flow, the opposite applies. Gas flowing through a convergent nozzle will decrease in speed and increase when flowing through a divergent cone, as shown in the figure below.



If two sections from those shown above are arranged into a rocket motor nozzle, the highest exhaust velocity obtained from the arrangement would be for subsonic flow from the combustion chamber to section a (subsonic, convergent), then section d (supersonic, divergent). Such an arrangement would be shaped like the standard rocket nozzle but would have specific flow speed requirements.

1. From the combustion chamber to the throat, the flow is subsonic

2. From the nozzle throat outward through the nozzle, the flow is sonic or supersonic

A number of other parameters must be optimized for the nozzle design, including the shape of the throat area and the angles of convergence and divergence. Another important consideration in nozzle design for launch vehicle propulsion systems is the compromise between an optimum nozzle expansion for operation at sea level and in a vacuum (space). Since these rocket motor operations can span that pressure range, a nozzle optimized for operation in a vacuum would be inefficient at sea level pressures, and the converse would apply.

An important element in exhaust nozzle design is the area ratio and resulting gas expansion, which also represents the maximum exhaust velocity. This ratio is also important since the nozzle is matched to the combustion chamber output and the ambient pressure(s) at the nozzle exit. The area ratio is defined as the exit area divided by the throat area, or

e = Aexit/Athroat

Ideally, the nozzle expansion curve would produce the same pressure at the exit point along the wall as the ambient pressure at the nozzle exit. Although this suggests an extended nozzle, weight limitations and aerodynamic pressures and loads restrict the length of the nozzle, as does the cooling mechanism.

Propulsion System Performance

The performance of a rocket motor or engine can be measured in several ways, but the most common are Isp , ΔV, and Ve. ΔV is the velocity change, or acceleration times the thrust duration = a t.

$$\Delta V = \text{acceleration} \times t = F\,t/m = Isp\ g$$

this a dimensional relationship only. The more accurate representation is

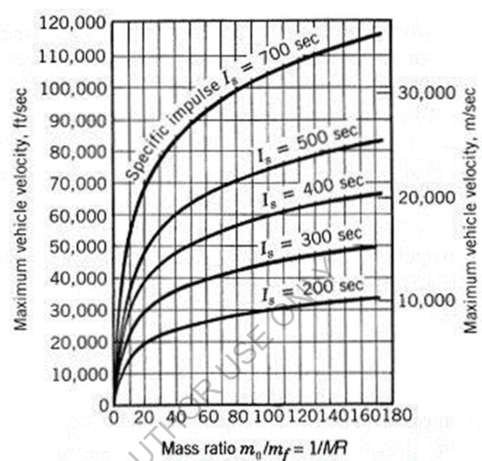
$$\Delta V_{\text{ideal performance}} = Isp\ g\ \ln[m_i/(m_i - m_p)]$$

Here, ln = natural log, mi = initial vehicle mass, mp = mass of consumed propellant, mi - mp = mf

= empty (final) vehicle weight - less is better

Orbit/Trajectory	**ΔV (m/s)**	**Percent Payload**
Launch to LEO (400 km) (excluding gravity, atmospheric drag)	7,750	100% (reference)
Gravity & drag during ascent through atmosphere (typical)	1,400	
Flight path correction from vertical to horizontal	350	
Launch to LEO (total)	9,500	100% (reference)
Launch to GEO	10,200	10-25%
Lunar impact	12,500	35-45%
Lunar landing (soft)	14-15,000	20-30%
Launch to lunar orbit	13,500	20-30%
Circumlunar mission with LEO return	16,000	25-35%
Lunar landing and return	16-18,000	1-4%
Mars	11,390	20-30%
Mars - landing and return	23-27,000	0.1-1%
Venus	11,450	
Venus - soft landing	23-25,000	

Mercury	12,500	
Jupiter	13,930	
Sun	30,450	
Earth Escape	12,700	
Escape Solar System	17,500	



From the above chart, the limitation of staged chemical rocket motors with a maximum Isp of roughly 450 s is approximately 20,000 m/s with a initial-to-final mass ratio of 150. Included in the final mass fraction, mf, are the vehicle inert mass (structure, tanks, engine, residual fuel, etc.), as well as the payload mass. The propellant mass fraction, expressed as 1 - mf/mi, would ideally be 1, although completely impractical since the entire vehicle would be propellant. The ideal propellant mass fraction of 1 would also mean that the final-to-initial mass ratio be zero, or the final mass is zero. Also, this is an obviously impractical design. Propellant mass ratios beyond 0.85 require careful design, with a practical limit near 0.95.

Limitations

For deep-space or interplanetary missions, a high payload capacity is desirable, hence a final- to-initial mass fraction would be relatively large. This would also require minimal remaining vehicle mass. Hence, the propellant mass fraction would be modest, on the order of 0.70 to 0.80, and the final-to-initial mass ratio would be perhaps 0.1 or 0.2 (1/MR = 10-20). This configuration implies several things, one of the most important being the Isp. Using the chart above, a 1/MR of 10 with an Isp of 400 would provide a ΔV maximum of approximately 12,000 m/s.

Another way to look at the ΔV limitation is by using the final-to-initial mass ratio.

m_f/m_i, = $e^{-\Delta V/gIsp)}$ = $e^{-(\Delta V/V_e)}$ with the exhaust velocity Ve equivalent to the Isp times g, which is accurate if we use equivalent exhaust velocity.

From this expression, the need to deliver a meaningful payload mass on an interplanetary flight means that the vehicle's exhaust velocity needs to be comparable to the mission velocity requirement (ΔV). Thus, a 400 s Isp vehicle would have difficulty delivering a sizable fraction of the total remaining vehicle mass as payload beyond Mars since g Isp is of the order of 4,000 m/s. This is far short of the 12,000-14,000 m/s ΔV requirement beyond Mars (see Table 1).

Yet another way to look at the limitation of specific impulse on mission payload is with the following graph.

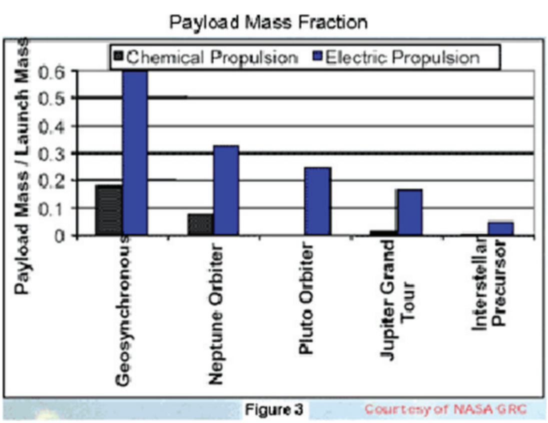


Figure 3 Courtesy of NASA GRC

As shown in this graph, the available payload mass becomes nearly insignificant beyond Jupiter, although higher Isp electric propulsion boosters are meaningful even on a solar system escape mission to our Galaxy's interstellar environment, already being explored with the Voyager craft. Missions beyond Jupiter would be difficult if not prohibited for chemical rocket boosters if not for gravity assist from larger planets, especially Jupiter. Because of the lower propellant requirement for the same thrust duration with electric propulsion, ΔV limitations do not represent a obstacle to exploration even to interstellar exploration.

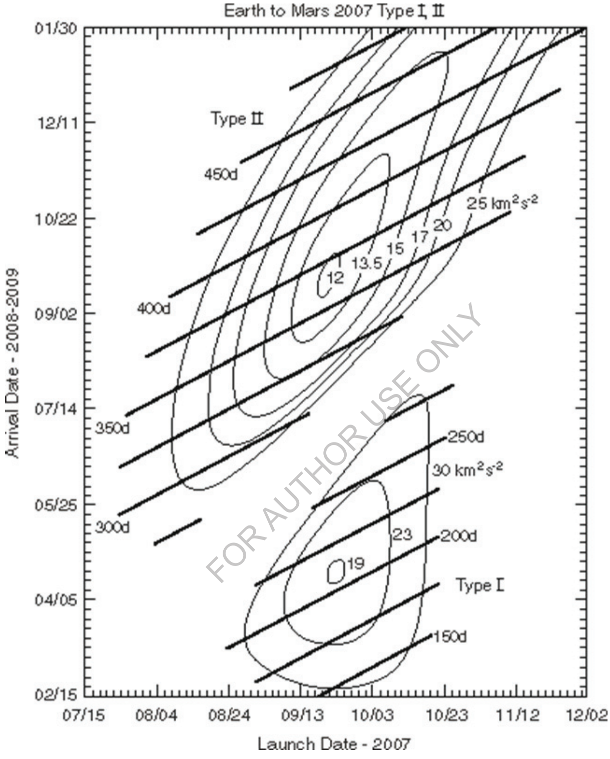
Design Application

Launch vehicle design is directed at maximizing both Isp and propellant mass fraction to overcome the Earth's large gravity well. In general, this increases both vehicle size and cost. In contrast, small orbital transfer vehicle design is directed towards optimizing ΔV capability for weight savings, which minimizes vehicle size and cost.

Final Solution

Mission design often pivots on the time and launch vehicle cost, which in turn depends on the location and distance of the target. Time limitations can be influenced by the payload, rendezvous requirements, or by restrictions due to human crews. And although a Hohmann transfer may present the lowest energy transfer from one orbit to another, non-coplanar and elliptical orbits prohibit that simple orbital transfer solution. A compromise for the time and energy variables in the mission flight plan can be represented in several ways. One of the most effective methods of presenting the data is with a time-of-flight vs.energy chart. Sometimes called the "porkchop" plot, the varying time-of-flight lines are superimposed on the energy- distance values between the launch position to the target, plotted as C3 (measured from C3=0, with units of km2/s2). The C3 energy units are in velocity squared, corresponding to the ΔV squared values, beginning at C3=0 (Earth-escape velocity). This corresponds to a ΔV of approximately 12,900 m/s from launch on Earth. A plot of the past 2007 Earth-Mars opportunity and the solutions for <180o (Type I, shorter transfer time), and >180o (Type II, longer transfer time) transfers are shown with C3 data. Note on the chart below indicating the minimum C3 energy appearing in the longer period Type II transfer region at approximately 370 days (C3=12 km2/s2). A shorter transfer period (Type I) orbital trajectory, however, comes at the cost of greater energy (C33=19 km2/s2). The total ΔV requirement for either would be the C3=0 value of 12.9 km/s plus the hyperbolic excess velocity to reach Mars. A circularization boost would have to be included if a Mars orbit is desired.

From a ΔV (or C3 energy) and time-of-flight requirement, a launch vehicle can be selected that is based also on payload mass and cost.



Propulsion System Types

Chemical

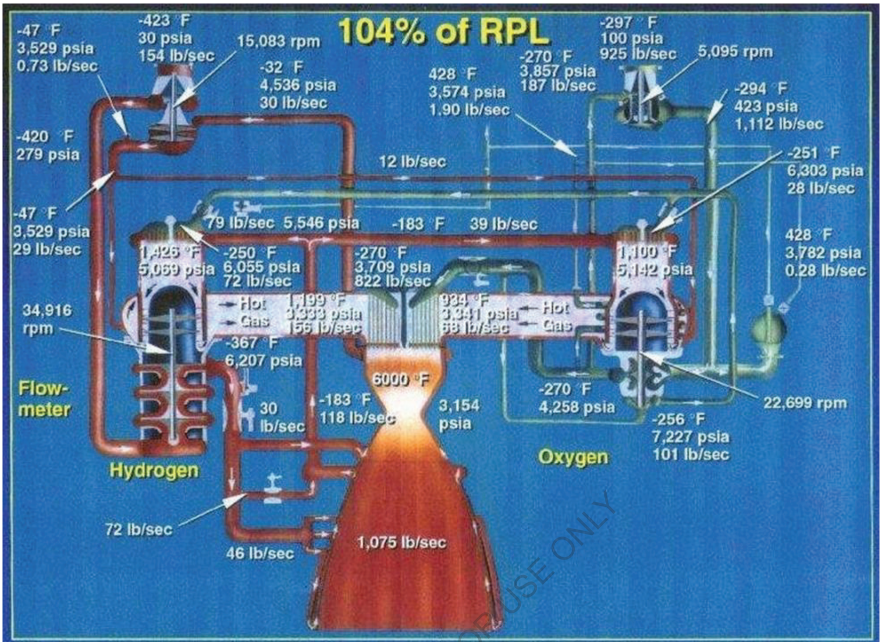
The liquid propellant is kept in tanks and supplied into a combustion engine as needed mainly in the chamber. This can be accomplished using either a pump or just gas pressurization (which is required to empty the entire propellant tank). Cooling of the combustion chamber and the nozzle is crucial since high temperatures are needed for high specific impulses. The propellant may occasionally be kept in a gaseous condition as well. This is only feasible for small engines though, as the gaseous state demands a lot more volume (and mass!) than the liquid stage does. Even research is being done on how to use a blend of solid and liquid fuel, or "slush," to further lower the volume.

Liquid

Bipropellant - this system has a separate fuel and oxidizer. Liquid fuels and liquid oxidizers are needed (gas not dense enough, solid doesn't flow). Oxygen is one of the best oxidizers, but must be kept at cryogenic temperatures, so these systems are efficient, high thrust and expensive because of the cryogenic storage, transfer, and engine components.

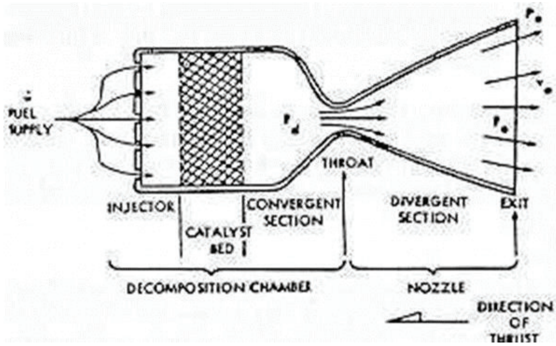
$H_2 + O_2$ is best because of the low exhaust weight, and generates clean combustion products (water and heat). Liquid oxygen (LOX, LO_2) and RP-1 (highly refined kerosene) is also used in a number of first stage booster engines, as are a number of other liquid fuels in combination with oxygen and fluorine.

An example of the highest performance liquid bipropellant engine, the Space Shuttle Main Engine, or SSME, is diagrammed below, along with the performance and operation values.



Other bipropellant combinations in use include fluorine and hydrogen, fluorine and hydrazine (N_2H_4), oxygen and hydrazine, and nitrogen tetroxide (N_2O_4) and hydrazine.

Monopropellant - This is a single fuel with a self-contained oxidizer. The chemical reaction still takes place and produces exhaust gas and heat (endothermic). The most common fuel is hydrazine N_2H_4 (goes to NH_3, N_2, H_2) and variations (monomethyl hydrazine, unsymmetrical- dimethyl hydrazine). This fuel is hypergolic when passed over a catalyst surface. The simplicity and moderate performance of this system make this useful for many spacecraft propulsion applications. The thrust and specific impulse limitations do not allow for monopropellant use on launch vehicle engines.



Propellant choice is based on more than performance and density, however. A variety of qualities are considered in the choice of a propellant, including safety and cost. This is a list of the more important considerations in the selection of rocket motor propellants.

The propellant combination that provides the highest I_{sp} and is both non-corrosive and non-toxic is hydrogen and oxygen. The drawback to liquid oxygen and liquid hydrogen is that they must be stored cryogenically which makes the propellants expensive to transport, store, pump, and handle. For large boosters, this combination provides the best performance and the one of the lowest weight propellant choices. These were design considerations recognized for nearly a century but first implemented in the Saturn V's 2nd and 3rd stages. Liquid oxygen and liquid hydrogen is and was the propellant of choice to get astronauts and cosmonauts to the Moon and back.

Oxidizer	Specific gravity	Boiling point (1 atm)	Characteristic s
Liquid oxygen (LOX)	1.14	90 K (-183°C, -298°F)	• Not hypergolic but can combust spontaneously with many materials at elevated pressures • Most commonly used rocket fuel oxidizer • Non-toxic and non-corrosive
Hydrogen peroxide (H_2O_2)	1.19	423 K (150°C, 302°F)	• Oxygen and heat are released by the decomposition of hydrogen peroxide into $H_2$0 + $_2$ • Decomposition is spontaneous with exposure to a catalyst such as platinum or iron oxide • H_2O_2 was used to generate gas to drive turbopumps in the V-2, X-1 and X-15
Nitric acid (HNO_3)	1.26-1.41	356 K (83°C, 181°F)	• Nitric acid and its variants are highly corrosive • Red fuming nitric acid is nitric acid + 5-20% nitrogen dioxide; more stable, less corrosive than pure nitric acid • Addition of <1% fluorine ion (HF) reduces corrosion (inhibited red fuming nitric oxide) • Used as propellant oxidizer with gasoline, amines, and hydrazine • Nitric acid is hypergolic when combined with hydrazine and amines
Nitrogen tetroxide (N_2O_4, NTO)	1.44	291 K (18°C, 64°F)	• Mildly corrosive unless mixed with water • Spontaneous combustion occurs when exposed to many materials • NTO is hypergolic when combined with most fuels • High vapor pressure requires relatively heavy tank • Used in numerous Russian rockets, the Titan booster series, and the Space Shuttle attitude control • Highly toxic - exposure limit < 5 ppm
Fluorine	1.11	83 K (-190°C, -310°F)	• Fluorine and fluorides have been proposed in various fuel combinations which are highly corrosive, difficult to handle, and toxic • No fluorine oxidizers have been used for production rocket engines

Fuel	Specific gravity	Boiling point (1 atm)	Characteristics

RP-1	0.80-0.815	420 K (147°C, 297°F)	• Highly-refined kerosene
			  ○ Developed as a fuel that could also be used for cooling high-temperature nozzles and combustion chambers
			  ○ Sulfur, aromatics, and unwanted isomers removed to permit use at high temps
			  ○ Greater stability, lower toxicity, less residue, higher performance than other hydrocarbons
			• High flash point 336 K
			  ○ Safer, less explosive than many hydrocarbon fuels including gasoline
			• Used in Russian R-7 booster and its derivatives, Soviet N-1, Atlas, Thor, Delta I-III, Titan I, Saturn I, IB, V (1st stage)

Liquid Hydrogen (H_2, LH2)	0.07 (requires large fuel tanks)	20 K (requires extensive insulation for tank and feed lines)	• High specific impulse • Highly flammable when hydrogen gas is mixed with air • Increased density possible with super cooled solid or slush hydrogen (not yet used) • Non-toxic (breathable gas; can replace nitrogen in an artificial atmosphere) • Nontoxic exhaust gas when reacted with oxygen • Hydrogen britalizes most metals, making turbo pump design more challenging than with other fuels

Oxidizer	Fuel	I_{sp} (theoretical,	I_{sp} (theoretical, 1 atm)

		vacuum)	
Liquid oxygen (LOX)	Liquid hydrogen (ILH2)	477 s	390 S
LOX	Kerosene (RP-1)	370 s	300 s
LOX	Monomethyl hydrazine	365 s	301 s
LOX	Methane (CH4)	368 s	296 s
Liquid ozone (O_3)	Hydrogen	580 s	
Nitrogen tetroxide (N_2O_4)	Hydrazine (N_2H_4)	334 s	292 s
Red fuming nitric acid	RP-1		269 s
Hydrogen peroxide (H_2O_2)	Monopropellant	154 s (90% H_2O_2)	
H_2O_2	RP-1		279 s
Fluorine (FI)	Lithium	542 s	
FI	Hydrogen	580 s	410 s

Liquid Propellant Systems Complications

Zero gravity fuel feed

The liquid fuels are not confined to any specific region within the tank while in low or micro gravity. To counter this, there are several ways to allow for positive fuel feed during low gravity conditions.

Capillary devices

Use surface tension to keep gas and liquid separated in the tank. Requires pressurization. Used on Shuttle and Viking.

Diaphragms and bladders

Physically separate gas an liquid with flexible lining made of elastomer or Teflon. Requires pressurization. Used on Magellan and Voyager.

Bellows

An expandable metal device to separate gas from liquid. Requires pressurization. Used on Minuteman.

Temperature extremes

The functional temperature range of liquid propellants is limited and is a function of pressure. To keep liquid fuels and oxidizers within these ranges, other requirements may have to be observed.

Cryogenic liquids

Must be kept at low temperatures, moderate pressures, and kept from significant heat sources until ready for the combustion cycle or preburner stages of combustion. This also means that the storage system, transfer and pumping system, and the combustion components must be able to reliably withstand very high temperature extremes (approximately 20K [storage] to 3,000K [combustion]). Hydrazine, its variations, and nitrogen tetroxide, do not require cryogenic systems for operation. However, they have to be kept from temperatures that are too low. This may require electric strip heaters.

Oscillations

Liquid propellants have the ability to oscillate or slosh while in the tanks, especially during high thrust during launch. This can seriously affect the tank structure as the fuel surges up and down. Baffles, reinforcements and overall tank design are used to reduce this effect.

Solid

Solid propellants used for rocket propulsion have a greater diversity in the combined chemicals used since many components are needed beyond the basic fuel and oxidizer. None of the solid fuels need cryogenic or cold temperatures for storage. In fact, the required properties for solid rocket fuels include long-term stability at relatively low, moderate, and elevated temperatures. While homogeneous solid propellants are more closely associated with modern gunpowder and used for separation motors or pyrotechnic charges, solid fuels used for rocket propulsion are typically composite mixtures containing separate granulated or powdered fuel and oxidizer, with a chemical binder, stabilizer, and often accelerant or catalyst added for improved performance and stability. The final mixture of solid fuel is a dense, rubber-like material that is cast into the combustion chamber. The cast fuel has a central cavity to allow burning throughout the length of the rocket motor that varies in shape from a simple cylinder to an elongated star.

Solid rocket fuel is typically identified by the type of chemical binder used; either HTPB or PBAN. HTPB (hydroxy-terminator polybutadiene) is a generally a stronger binder, more flexible, and faster curing, but suffers from a slightly lower Isp than PBAN (polybutadiene acrylic acid acrylonitrile) and uses fast-curing, toxic isocynates. PBAN, on the other hand, has a slightly higher Isp, is less costly, and less toxic which makes it popular for amateur rocket makers. PBAN Is also used in the large boosters including the Titan III, the Space Shuttle SRBs, and the new Constellation Ares I and Ares V launchers. HTPB is or has been used in the Delta II, Delta III, Delta IV, Titan IVB and Ariane launchers.

The Space Shuttle's Solid Rocket Boosters are powered by a PBAN-based ammonium perchlorate composite propellant (APCP) that develops a specific impulse of 242 seconds at sea level and 268 seconds in a vacuum. The SRB composition consists of:

- Ammonium perchlorate oxidizer (69.6% by weight)
- Powdered aluminum fuel (16%)
- Iron oxide catalyst (0.4%)
- PBAN polymer binder that is also a secondary fuel (12.0%)
- Epoxy curing agent (2.0%)

Aluminum is used for the SRB fuel and is also employed in a variety of other solid rocket motors because it has a reasonable specific energy density, and a high volumetric energy density.

Aluminum powder is also difficult to accidentally ignite. Ammonium perchlorate is an attractive oxidizer because of its oxygen content, stable crystal form, and relatively fast burning character.

Large solid rocket motors are also used for auxiliary boost on numerous expendable launch vehicles. These include the Delta launcher's graphite epoxy motors (GEMs), the earlier TItan III and IV solid rocket motors (SRMs), the Arianne V's solid rocket boosters (SRBs), the Chinese Long March strap-on boosters, and the Russian Proton strap-on solid boosters.

Components in a typical large solid rocket motor include:

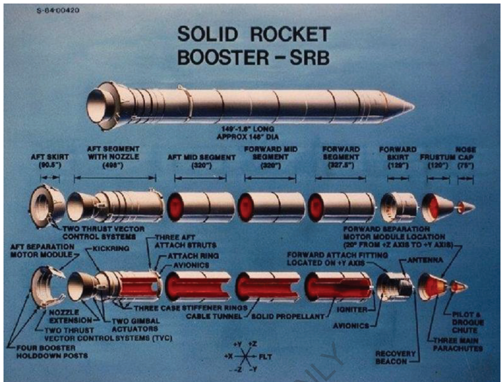
Case - the pressure shell containing the fuels and combustion pressures throughout the burn. Materials are commonly titanium, steel and Kevlar.

Liner - Used in the case-propellant boundary to insulate the case as the propellant burns. Usually a propellant binder.

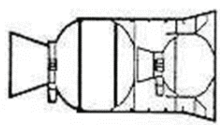
Nozzle - provides the converging-diverging sections which accelerates the exhaust gas after combustion. Since this is the highest temperature region, the materials are graphite epoxy, or carbon-carbon with a carbon throat.

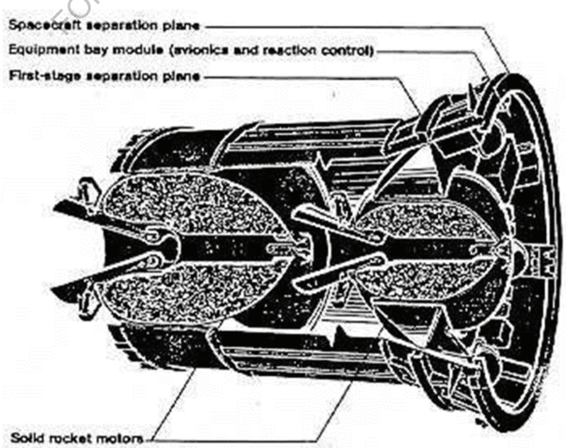
Nozzle closure - a membrane or surface in the nozzle area that protects the propellant from exposure to atmosphere and/or the hard vacuum of space.

Igniter - raises the propellant grain temperature to the ignition point and increases chamber pressure during start.



The basic two-stage IUS has two solid-propellant motors (one with 21,400 pounds of propellant, the other with 6,000 pounds), an equipment bay, and an interstage structure. This vehicle will transport spacecraft from the shuttle's park orbit (150 nautical miles) to higher Earth orbits. The large motor will send a spacecraft to apogee, and the burn of the small motor will place it in a precise orbit.





Cold Gas

Pressurized gas or heated liquid is used to expel gas through thrusters, usually arranged for 3- axis attitude control in smaller spacecraft. The energy is stored as pressurized gas, not chemical reactions, making it a low efficiency system. The thrust is low (.01 to 10 N) and the Isp is low (5-100). The simplicity, controllability and inexpensive features make this type of system usable in many attitude control applications. It is commonly used in small spacecraft attitude control, and to supplement other 3-axis attitude control systems (reaction control wheels, momentum wheels e.g.).

Electric

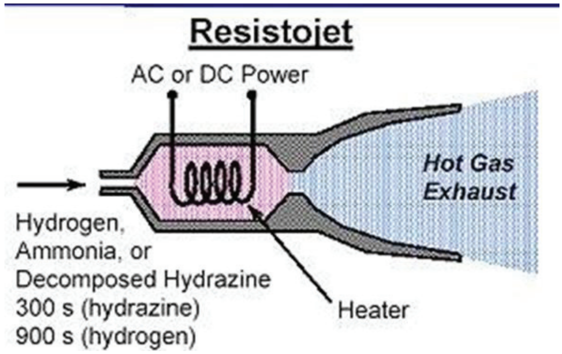
Electric propulsion comes in broad variants, and an even greater variety of fuels. What is common the electric propulsion (EP) systems is the electric or electromagnetic power consumed. The simplest of these is the heated gas thruster that increases the thrust of a cold- gas thruster by heating and expanding the gas, giving it greater exhaust velocity and mass flow. The resulting thrust efficiency multiplication can be considerable since Isp for neutral gases is roughly based on the gas temperature. In addition to electric power, gas heating can be generated by solar heat and microwave/radio frequency heating similar to a microwave oven, and by radioactive core heating. Several types of electrothermal engines are discussed below.

Electrothermal

Resistojet Thruster

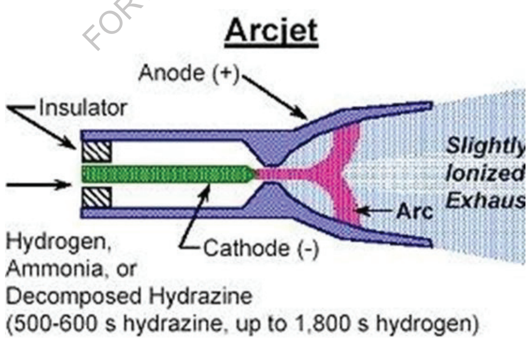
Compressed gas can be heated directly or indirectly by electrical energy, such as resistive heating from an electrical current. The heated gas expands in rough proportion to the temperature increase, increasing exhaust gas velocity and thrust efficiency. Isp increase may be a factor of 100 or more with external heating. The same concept is employed in the nuclear thermal engine, but with a radioactive core heater instead of the electrical heater.

Resistojet thrusters can not only be used with non-reactive gasses such as hydrogen or nitrogen, the thruster can improve reactive gases expansion such a hydrazine, producing Isp values as high as 300 s or greater. Resistojet thrusters were planned for use on Space Station Freedom for reboosting by using methane gas generated in the carbon dioxide reduction system.



Arcjet Thruster

A more efficient method of heating and expanding gas for improved thrust uses a high-current arc in a circular chamber. The arcjet thruster produces higher chamber temperatures than the resistojet but also requires greater power for operation. Arcjet improvements in efficiency can be as much as 600 s for hydrazine propellant, and 2,000 s for hydrogen. In addition, Arcjet thrusters require moderate power (1-100 kW) and are suitable for moderate-sized spacecraft propulsion.



Solar Thermal propulsion collects solar energy by concentrating sunlight inside an expansion chamber to heat the propellant. Similar in concept to the resistojet thruster, the solar thermal system expands pressurized gas in that exits through an expansion nozzle. The increase in Isp over a compressed (cold) gas system is a function of the internal chamber temperature, which is dependent on the solar concentrator efficiency and collection area. Although not flown in space, the solar thermal engine is expected to operate at an Isp of 500-800 s at modest power/thrust levels.

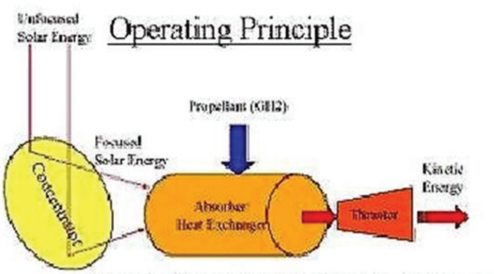


Figure X: Solar Thermal Propulsion Thruster

Microwave energy can be directed into a heating chamber similar to the solar thermal engine, producing the same gas expansion effect with either an inert (H_2, N_2) or reactive (hydrazine) propellant. Heating from the microwave excitation is dependent on the microwave energy and the resonance absorption efficiency of the energy by the propellant.

Electrostatic

Ion

Another type of EP thruster is the ion engine that forces charged particles out of an ionizing chamber by electrostatic fields. A negative grid at the exhaust end of the thruster pulls high-mass positive ions towards the mesh, passing through as exhaust. The high-mass ions are created in several ways, one being the bombardment of neutral gas by high-energy electrons. For example, xenon, commonly used as the propellant gas, is ionized, accelerated through a positively-charged grid, then neutralized with a beam of electrons as the ions exit to keep them from returning to the negatively charged screen. A second grid can also be attached with a relatively low voltage grid outside of the high voltage screen to repel electrons that may stray from the neutralizing electron beam. The energy of the ions is equal to the net electrostatic field (E) times the charge (q). The exit velocity is approximate:

$$\mathbf{V_{exit} = 2qE_{net}/m}$$

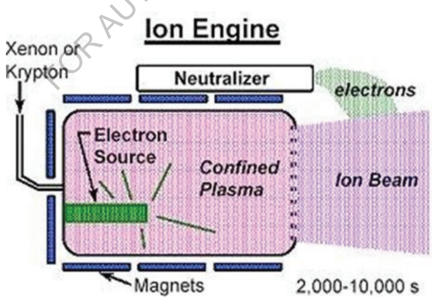
where m is the ion's mass and q is its charge.

This electrostatic thruster has exceptionally high efficiency but shallow thrust. Because of the low thrust, the applications are for long-term thrust, normally geostationary station keeping, or for orbit boost of smaller satellites from LEO to GEO. The very high efficiency also means that thruster is ideal for deep-space missions because of the low propellant mass required compared to chemical thrusters. However, this application would require an extended flight time due to the low thrust.

Exit velocity for the ions can be as high as 30,000 m/s, some 100 times that of conventional chemical rocket exhaust gas. Since the exhaust velocity is equal to the specific impulse times g, the efficiency is also 100 times greater than chemical rocket motors. More importantly, the fuel required for the same ΔV boost is 100 times less than a typical chemical rocket, although the ratios for both thrusters vary with propellants.

The momentum reaction is based on the charge-to-mass ratio of the atom. Higher mass atoms, multiply ionized are the best fuel. Fuel can be stored as liquid or solid (if easily sublimated)

Ejected ions must be neutralized or the cloud of charged particles will build up an electric field which will accelerate the ions back to the spacecraft without a net thrust. Electrons are injected into the ion exhaust to produce neutral atoms to counter this. It also requires a high vacuum for operation or testing.

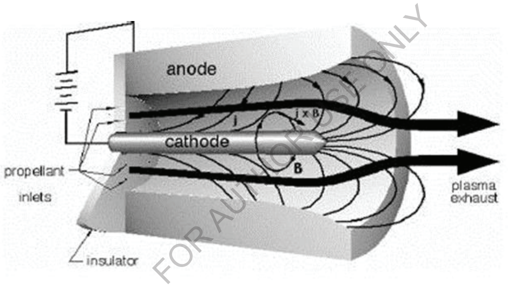


Electrodynamic

Electrodynamic engines use electromagnetic fields and currents to accelerate ions in similar fashion to the static field ion engine, but with different techniques and fuels. Magnetoplasmadynamic engines, for example, generate thrust by interacting a current of charged particles (a plasma) with its own induced magnetic field.

Magnetoplasmadynamic Thruster

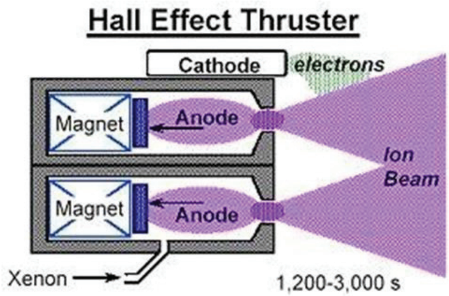
In addition to the induced magnetic field generated by the plasma current in the MHD engine, an external magnetic field is also used to increase the acceleration of the plasma particles which increases exhaust velocity (efficiency or Isp) and thrust (more plasma exhaust). These thrusters require high currents but also generate the highest thrust of the EP engines, making them ideal for sizeable interplanetary spacecraft. The MHD engine accelerates the ionized plasma as it passes through concentric electrodes at high voltage. In addition, the plasma current that is created self-induces a radial magnetic field, further accelerating the plasma Lorentz force (j × B). The electric and magnetic fields accelerate the plasma current because they are perpendicular to both. In addition, an external magnetic field can be superimposed to increase acceleration. These relatively high currents produce sufficient thrust to be used on larger spacecraft.



Hall Effect

The Hall Effect thruster has been in operation for several decades, mostly on Russian satellites. Thrust efficiency of the Hall thruster is generally less than the ion engine, but the thrust range is greater than ion thrusters. This makes the Hall thruster more useful for near-Earth applications where station keeping and orbit operations require greater thrust.

Hall thrusters use an electric field to accelerate ions, similar to Ion thrusters. This is combined with a radial magnetic field that generates Hall current and an accelerating force on the plasma. To date, the highest power level for a Hall thruster was produced at NASA's Glenn research facility at 95KW which generated 3.3 Newtons (Glenn).

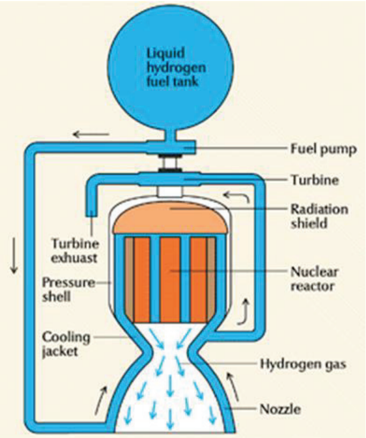


Nuclear

Nuclear fission material produces radiation products (particle and electromagnetic radiation) and heat. The heat can be used to heat gas to ionizing temperatures for acceleration by thermal expansion, and furher with electromagnetic fields. The optimum fuel or this system is the lowest atomic mass material - hydrogen which can be stored as liquid or lower temperature slush.

Nuclear Thermal

Liquid or gas fuel (hydrogen) heated by passing over nuclear fission "hot" surface or fission core, then expelled at high velocity. This is a more straightforward system than the nuclear ion, but still produces radiation byproducts, moderate thrust, and high Isp.



Nuclear Thermal Ion

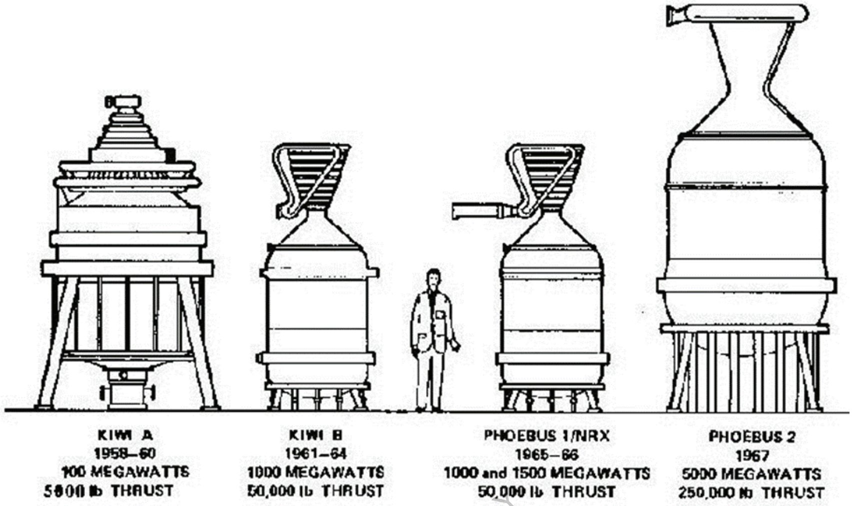
Liquid or gas fuel is ionized by heating as it passes over a surface heated by nuclear fission material (particles embedded in broad surface region or a large surface area core), then accelerated with an electric field.

- Hydrogen fuel optimum
- High temperatures require cooling or moderation
- Moderate thrust could be used for booster or long duration propulsion
- High Isp useful for interplanetary propulsion
- Radiation products can be severe hazard - not usable for launch nor manned missions

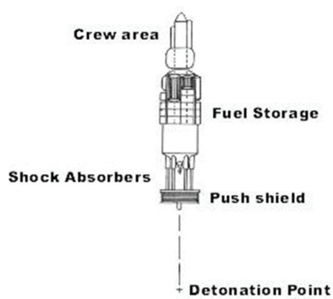
Nuclear Propulsion Systems

Between 1955 and 1972, the Department of Defense (DoD), the USAF, the Department of Energy (DOE), and later NASA sponsored research into nuclear fission propulsion for moderate to high thrust engines. These designs were to be used for defense and exploration applications, to be used for surface launch and boosters. Reactor fuel was highly concentrated U235, with liquid hydrogen as the stored propellant. Reactor cooling was accomplished by basic internal moderator and reflected neutron techniques. Thrust levels were 25,000 to a whopping 250,000 lb. Reactor power in the largest of the nearly two dozen models reached 4,000 MW.

Problems arose in a number of areas including hydrogen corrosion and brutalization of the metal components in contact with the propellant. Later project reactor designs that included the NERVA, required high-pressure, high-speed turbopumps for the liquid and gas hydrogen propellant. These research and development programs led to the F-1 booster design of NASA's Saturn V engine, not surprisingly, since Rocketdyne was contracted to develop both engine turbopumps (Gunn).



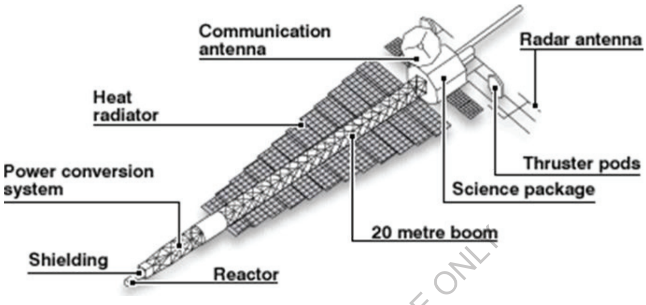
Pulsed Thermonuclear



In the late 1950's and early 1960's the U.S. Government began a thermonuclear propulsion system under the name Project Orion. Small thermonuclear explosions were to propel the leading vehicle by shock impulse. Project Orion was conceived for exploration and manned missions to the Moon and Mars. Tests of the shock impulse propulsion took place with chemical explosives, but little hardware was actually developed for the bizarre project. Although it was reported to have the support of Wernher von Braun, the future of the project was doomed by a wide variety of forces that included the nuclear test ban treaty, NASA's reluctance to pursue a radical and risky program, the growing Apollo program, and basic reasoning.

Project Prometheus

NASA managers have long wanted to improve propulsion for interplanetary missions because of the inefficiency of the chemical rocket motor. The Agency's efforts in earlier nuclear propulsion systems were pulled off the back burner when the Bush administration considered the nuclear propulsion concept an advancement in exploration. Project Prometheus, a combined nuclear reactor and nuclear propulsion vehicle, was promoted quietly while concept and preliminary development started under NASA chief administrator, Shawn O'Keefe. Headwinds for this project were strong, and Congressional budgets were slashed before substantial resources were committed to the program which was affiliated with an attractive exploration mission, studying the outer solar system moons. Prometheus, also known as the Jupiter Icy Moon Mission, could also have provided technology for Mars cargo missions.



Project Pluto

An air-breathing version of a nuclear propulsion system was under development in the late 1950s and early 1960s under the name Pluto. A ramjet fueled reactor would power the cruise missile to altitudes of 35,000' at Mach 4 or greater, although the efficiency and reactor cooling compromise suggested low-altitude operations in cruise. A combination of extreme heat generated in the reactor, radiation dissipation along the cruise missile's path, extreme radiation hazards on impact from the reactor, destructive sonic booms along its path, and a few other complications sank the Pluto (also known as SLAM) project before its construction. A scaled nuclear reactor power plant was built and tested but the missile was never designed to completion before cancellation in 1964. Even those swept up in the Cold War weapons mania could not accept possibly the most sinister weapon conceived, which would have spread lethal radiation along its flight path over allied countries, with the potential of destroying population centers from radiation at its impact area, or anywhere it would have crashed by accident.

Photon Propulsion

The solar radiation coming from the Sun can be harnessed for interplanetary propulsion by using large surface area sails to transfer momentum from solar photons to the spacecraft.

Photon (solar) propulsion

Solar photon pressure provides a small force capable of propelling a small payload mass with a large surface area. Useful to approximately the orbit of Mars, no prototype yet flown in space.

However, photon pressure has been used for attitude control and orbit changes on the Mariner 10 mission to Mercury. Photon pressure can also produce measurable drag in near-solar missions.

Laser momentum delivery

Concentrated laser light can be used to propel a spacecraft away from the Earth without the need for a large sail structure, although there are a number of limitations on the process.

Although the laser light does not decrease as 1/r2 as with solar electromagnetic radiation, the collimated light alignment with the spacecraft reflection surface is critical. Intense laser light also has propagation anomalies through the atmosphere which disrupt the collimated beam.

Technology for laser momentum propulsion may not be able to surpass the ion/electric propulsion systems that are proving useful in many applications, from Earth-orbit attitude and orbit control, to deep-space missions reaching interstellar flight.

Staging

Spacecraft launch systems require more than one stage out of necessity. A single stage would have an enormous structural skeleton at the end of the boost, making the final acceleration much lower than if separable boost stages were used. If too many stages are used, however, much of the mass in the launch vehicle is taken up with engine and tank structures. Most launch vehicle designs use three or four stages for reaching Earth orbit, based primarily on payload capacity.

Basic rules:

1. Stages with higher Isp should be above stages with lower Isp.
2. More ΔV should be provided by stages with higher Isp.
3. Similar mass stages should provide similar ΔV
4. Each succeeding stage mass should be smaller than its predecessor.

Review Questions and Problems

1) List the different types of Spacecraft Propulsion

2) A spacecraft's engine ejects mass at a rate of 30 Kg/s with an exhaust velocity of 4,100 m/s. The environmental pressure at the Nozzle Exit is 5 kPa and the Exit Area is equivalent to 0.7 m^2. What is the thrust of the engine in a vacuum?

3) A 5000 Kg spacecraft is in Earth orbit travelling at a velocity of 8000 m/s. Its engine is burned to accelerate it to a velocity of 13,000 m/s placing it on an escape trajectory. The engine expels mass at a rate of 10 kg/s and an effective velocity of 4000 m/s. Calculate the duration of the burn

Chapter 5: Space Communications &Telemetry

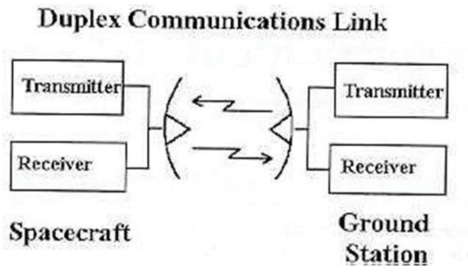
Communication systems that cover distances beyond wire or fiber optic links are universally based on signals transferred in the electromagnetic spectrum, from low-frequency radio bands and even lower, very low frequencies used for underground or submarine communications to visible light frequencies. Spacecraft communications have additional restrictions due to signals traveling through the atmosphere. Those limitations are primarily in frequency bands with less interference with the ions in the upper atmosphere or the absorption due to several molecules.

Communications systems for spacecraft must also compensate for the continual relative motion between the spacecraft and the ground station. This includes both in antenna pointing and for Doppler (relative motion) shift of the signal. Another critical design consideration is the deficient transmission and reception error rate needed for reliable communications over the conditions expected for the mission life.

The communications link for a spacecraft system must have a two-way (duplex) link to a ground station for simultaneous information flow to and from the spacecraft and the ground station. The two signal paths are the uplink (to the spacecraft) and the downlink (from the spacecraft).

Uplink is the transmission to the spacecraft (transmitter on the ground & receiver on spacecraft). These are entirely separate systems using two different frequencies.

Downlink is the signal path from the spacecraft to the ground station



Spacecraft communications systems may include a relay link at higher or lower frequencies to other satellites or spacecraft, or to other relay communications systems such as NASA's Spaceflight Tracking and Data Network (STDN) and the Tracking and Data Relay Satellite System (TDRSS) satellite network. Descriptive classifications for these two space communications networks are the ground network, or GN, and the space network, the SN.

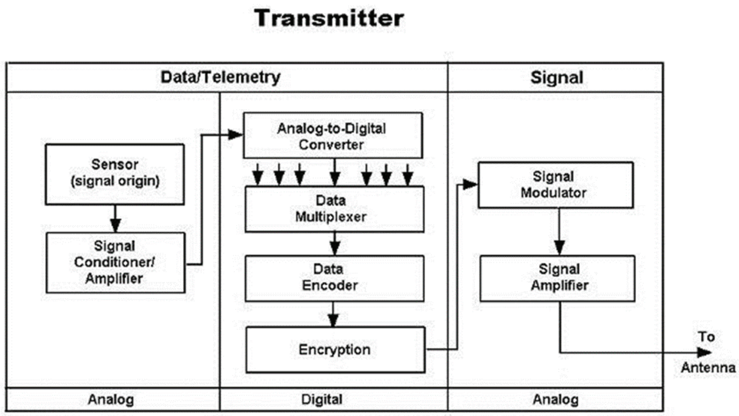
The spacecraft and ground communications systems are composed of four basic units, the transmitter, the receiver, the data management system, and the antenna. Individual components of the transmitter, receiver, and data management unit, vary dramatically between spacecraft because of the different missions and environments.

Basic Radio Frequency Communications

The transmitter and receiver are similar in their signal processing in that signal propagating through the transmitter components must be reversed by the receiver components, or balanced, to reconstruct the original information.

Transmitter

The transmitter is the signal production, signal encoding and power amplifier circuitry of the communications link. Functional components of the transmitter will depend on the application, modulation, and coding scheme used for the spacecraft. Established standards for modulation and encoding reduce these variations between spacecraft, but the vast differences in missions result in unique communications system design for each spacecraft. A schematic of the major components is shown below for a simple transmitter configuration.



Sensor

The sensor instrument converts the data input (light, pressure, magnetic field, acceleration, biomedical device, etc.) into an analog electrical signal. The sensor output is generally a very low voltage (microvolts or millivolts) and/or current (microamps or milliamps) that correspond to minor variations in the sensor/instrument output.

Signal conditioning

The signal conditioning is used to convert the low voltage or low current levels or varying frequency from the sensitive sensors to a higher voltage level (fractions of a volt to several volts) for digitizing into binary code.

Signal conversion

This process translates the analog voltage output data from the signal conditioner into digitized binary data for more effortless transfer and manipulation by the onboard computer. This is very often an analog-to-digital (A/D) converter, producing 8, 12, 16 or 32 binary bits at the output.

Multiplexer

The Multiplexer takes sequential input from several sources of data (in parallel) and combines them in a sequence (serial) to send to later signal stages of the transmitter (encoding & modulation). Multiplexing scheme is usually time-division multiplexing or frequency-division multiplexing.

Encoder

The encoder combines the digital data with spacecraft and sensor identification, time and noise reduction and error correction codes to improve the signal when decoded by the receiver. A part of the encoder function includes a formatter to transform data into data frames (commutation).

Encryption

The encryption module codes the signal for secure transmission & reception to prevent unwanted commands or instructions sent to/from the spacecraft.

Modulation

The two-part modulation process combines the data signal and the carrier (radio frequency or RF) signal at the center frequency that best suited for transmitting (and available for spacecraft communications). The second step of modulation is the alteration of the frequency, phase or amplitude of the RF signal to correspond to the data (frequency, phase, pulse or amplitude modulation). The most common type of spacecraft signal modulation is Pulse Code Modulation (PCM).

Power amplifier

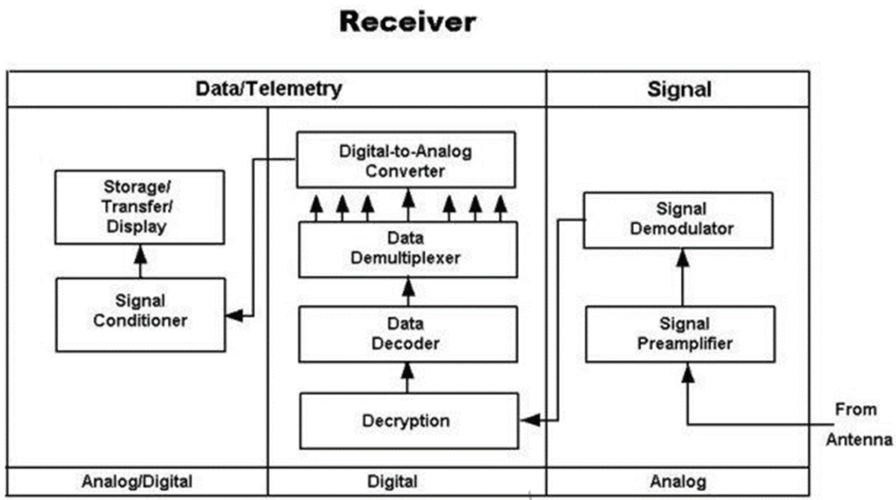
The power amplifier increases the power of the final signal (milliwatts) to several watts (kilowatts for ground stations) for transmission to a distant receiver.

Antenna

The antenna directs the electronic signal to the receiver and matches the outside environment (atmosphere or space).

Receiver

The receiver detects and decodes the transmitted signal in the reverse order of the signal coding and conditioning performed by the transmitter.



Preamplifier

The received signal level from the spacecraft antenna is typically feeble and carries background noise. To improve the signal information, a low-noise preamplifier is placed after the antenna and before the signal detection stage of the receiver. This is the most sensitive stage for noise input, hence the most sensitive, low-noise amplifier practicable. Very low- noise pre-amplifiers are cooled to reduce thermal noise.

Demodulator

The demodulator removes the carrier (center) frequency from the received signal and extracts the valuable signal from the transmitted coding scheme. The signal output from the demodulator is encrypted binary data. In short, this module reverses transmitter modulation.

Decryption

According to the encryption recipe, the binary data stream is rearranged to provide the data in a specific order. As suggested by its name, this module reverses the encryption process.

Decoding

The resulting binary data is separated, or parsed, into individual data data packets corresponding to the transmitted data sources, as well as spacecraft identification, time, error coding, and other frame information at this stage. This module reverses the transmitter encoding process.

Demultiplexing

The serial data stream is segregated according to the source or sensor data on the ground station receiver, the spacecraft's command instructions, and the appropriate time or sequence identification. Segregated data is then divided, or demultiplexed, into the appropriate number of parallel data channels at the ground station, or routed to components/systems as operating instructions. This module is the reverse of data multiplexing or combining.

Conversion & calibration

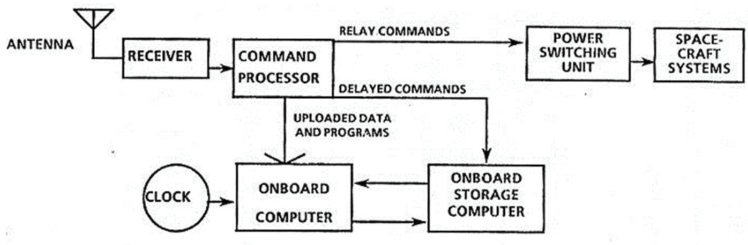
For telemetry coming from the spacecraft, the individual source data are compared with reference data and calibrated. The data can then be formatted or converted to provide accurate values representing the corresponding spacecraft telemetry sensor data. The command and control data received from the ground station is converted into command or control instructions for routing to the various onboard components and systems.

Storage/transfer/operations

Calibrated values can be displayed, transferred, and/or stored for future use at the ground station. In addition, data received by the spacecraft can be transferred, routed, stored, or converted into instructions by the command and data handling system.

Command Data Processing

An essential function of the spacecraft communications system is controlling, interpreting, and distributing command flow to and from the spacecraft. As an integrated function of the spacecraft's data handling and RF communications systems, the command and data processing element is a semi-autonomous operation that can execute instructions and commands without ground station contact. This requires separate data storage and processor for these operations, as diagrammed below.



Antennas

Antennas are used to send and receive communications signals, which implies operations at high and low power that correspond to transmitter and receiver signal levels. At the microwave frequencies, antennas vary in size and shape. Antenna size relates directly to the capacity to collect electromagnetic signals as waves, or using slightly different definitions, collecting photon particles. Because the antenna shape has innumerable effects on the signal, the design incorporates optimal signal transmission.

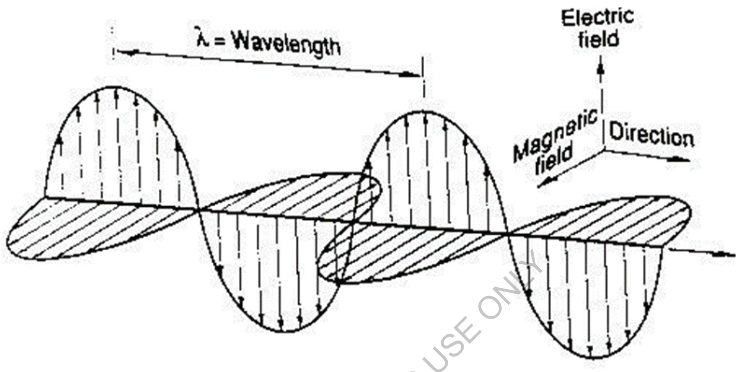
An essential aspect of the antenna size is signals' collection capacity, or the gain. Gain or multiplication of signal is measured using a standard chosen as a simple dipole. More on that later. One of the most common methods of increasing the signal collecting area is with a reflector surface such as a parabolic dish. A less common is a refractor, a lens in the microwave band which increases the wave/photon density, increasing the signal strength.

As with an optical lens, the reflected/refracted signals must be kept in phase to improve signal strength. Phase irregularities from irregular surface of an optical lens produce blurred or unfocused images. Similarly, irregularity in the reflector or refractor surfaces decreases an RF antenna's signal strength. For example, a parabolic reflector, or dish, must have surface irregularities kept below 1/10th the center frequency/wavelength to avoid significant losses. For a wire mesh antenna reflector, this spacing must be less than 1/10th of the signal wavelength to be an effective reflector. That also means that a wire mesh can be used to either exclude or contain electromagnetic waves with wavelengths more significant than 10 x the mesh spacing. Look at your microwave oven and note the size of the metal mesh spacing.

That screen is what contains the microwave energy from exiting the front of the oven through the screen.

Electromagnetic Waves & Antennas

Antenna width or diameter is also important because it limits the range of frequencies/wavelengths that can be transmitted or received. An antenna with a length perpendicular to the signal alignment (usually the electric field plane) smaller than 1/4 of the signal wavelength will have reduced signal capacity. So, in addition to antenna size, antenna orientation concerning the signal is essential. More considerable antenna lengths/widths increase signal strength, but many exceptions exist.



The basic antenna element, the dipole dual split wire, has a span of the opposing wires 1/4 of the signal wavelength or large, and optimally in increments of the quarter wavelength. However, this split wire element has a limited area for intercepting electromagnetic signals. Therefore, to improve the signal collection capability of the dipole, a reflector, such as a parabolic dish or refractive bars (VHF or Yagi antennas seen on outdoor television antennas) increase the signal concentration at the dipole element.

To improve an antenna signal capacity, a reflective or refractive surface must be used to focus the signal on the fundamental antenna element, the dipole. Moreover, a larger antenna area will generally result in more significant signal gain over a dipole because of the signal-gathering surface area. The surface area is measured perpendicular to the signal. A wrong orientation to the antenna can result in little or no signal.

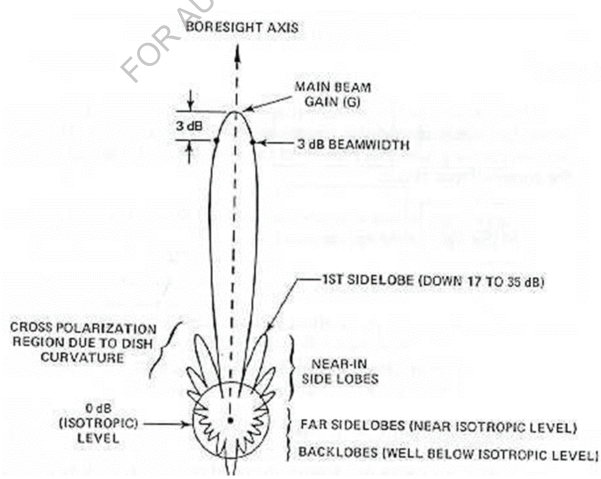
Gain

Gain is generally expressed as signal multiplication compared to a uniform or isotropic radiator. This would be an idealized antenna that radiates uniformly in all directions. Neither the dipole nor any other antenna element is capable of this, but the simple expression is convenient.

Antenna Beam Shape

Antenna beam shape, also known as a directional response since an antenna does not have a uniform response in all directions, can be visualized as a signal beam coming from the center axis of the antenna. The response would be measured from all directions to define the response curve. This response shape would be the same for transmitting and receiving, even though the signal levels are dramatically different.

In addition to signal E-field alignment, the signal power received by an antenna is dependent on the direction of travel of the signal relative to the direction of the antenna center, or pointing axis.. We will use a parabolic antenna (reflector plus dipole) for these examples to show how wide or narrow the beam of the antenna response is and how the beam width depends on the antenna diameter and signal frequency.



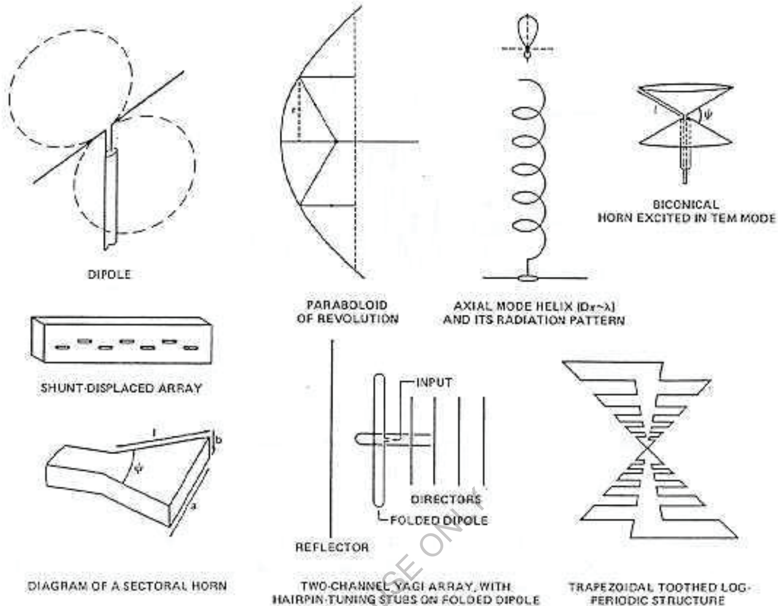
Antenna Types

Dipole - This is arguably the most common type of antenna and consists of two straight-wire elements with a gain of 1.5 above isotropic. The gain is most significant perpendicular to wire axis, and least along the wire axis. The traditional auto FM/AM antenna is an example of a dipole element. Appearing as just a single element, the vertical antenna and the metal body form a dipole arrangement with the second vertical element generated by the reflective-plane character of the car's metal skin. In addition, some cars use a very thin, painted dipole conductor on the windshield.

Parabolic - A parabolic antenna is a parabolic-shaped reflector that focuses arriving signals within the beam onto the dipole element located at the parabolic focus. The parabolic shape provides a gain above isotropic of $\eta\ (\pi Df/c)2$

Helical - This is a spiral conductor with a reflecting plane at end opposite to the signal travel. Gain for the helical antenna is proportional to surface area, including diameter and length.

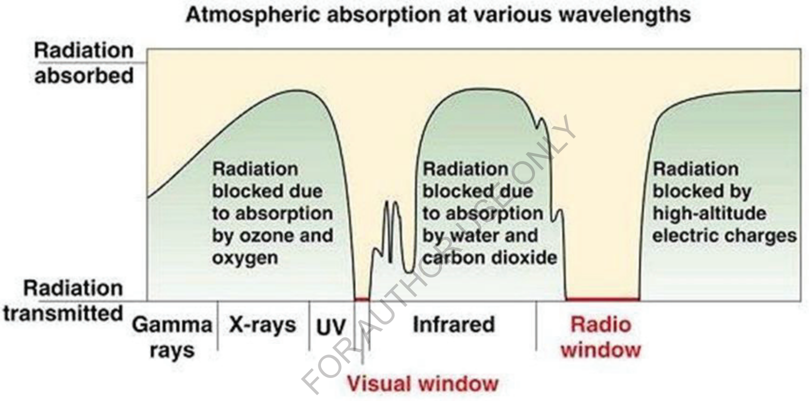
Horn - A horn antenna consists of a horn-shaped reflector connected to a rectangular cavity with the dipole pickup near the closed end. Gain for this antenna is 10 A (f/c)2 where A is the entry throat area.



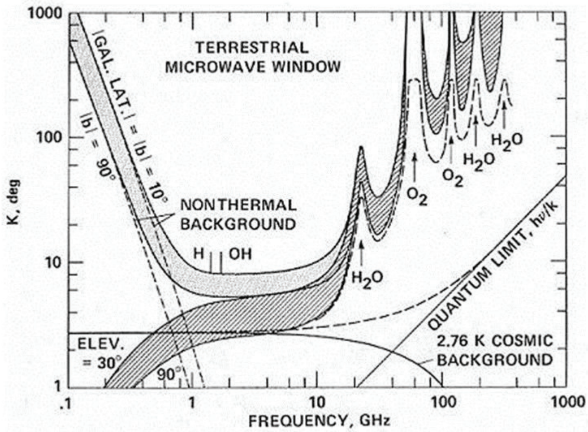
RF Communications – Signal Propagation

Electromagnetic waves can be described by a sine wave function with three variables. First is amplitude (power); second is frequency (cycles per second); and third, is wavelength. Electromagnetic waves travel at the speed of light, through the atmosphere, space and some materials that do not strongly interact with the EM wave (glass, water, and some gases are examples in the visible band). Propagation of EM signals through the Earth and other planetary atmospheres is complex, and depends on a variety of variables, including time of day and angle of incidence between signal and atmosphere. For signal propagation through relatively thick atmospheres, many EM frequencies are attenuated by absorption, scattering, or refraction. This limits EM communications channels to the low-attenuation bands, especially for spacecraft. Windows through a relatively thick atmosphere will depend on the atmosphere composition, temperature and ionization level, gas circulation, and more. Fortunately, the solid planets in the solar system with significant atmospheres have transmission windows for communications.

The Earth's atmosphere has several EM windows, although clouds and precipitation do attenuate/scatter/absorb visible frequencies. A small window also exists in the Earth's atmosphere in the infrared, although thermal heat in many areas of the background sky and on the Earth almost completely eliminate the possibility of space communications in the IR band. The Earth's upper atmosphere also refracts and interacts strongly with lower frequencies, which limits the lower bound of radio communications. Consequently, the microwave band is the only reliable communication window for spacecraft. The figure below shows the attenuation of electromagnetic radiation throughout the electromagnetic spectrum.



The Earth's atmospheric composition allows relatively little interference in several areas of the microwave band, a frequency range that is well suited for spacecraft communications. The lowest interference and noise for microwave communications is located roughly in the 1 to 10 GHz region, as shown in the graph below. The graph depicts temperature or thermal noise versus frequency. Optimal space communications frequencies are found at the lowest noise levels. On this chart, that floor is indicated by the lowest background (noise) temperatures.



Two essential characteristics of electromagnetic waves used for communications are propagation speed and energy. The energy of the electromagnetic wave will determine its interactivity with atmospheric gases, liquids, and reflective solids. It will also limit the maximum bandwidth available for communications since higher communications frequencies have higher potential bandwidth. In addition, the speed of light is essential since an atmosphere's refractive index can reduce the propagated speed of light. A material's refractive index reduces signal speed by the square root of the index of refraction, n.

$c_{material}$ = c / n where c is the speed of light (2.998×10^8 m/s) and n is the index of refraction (1 or greater).

Electromagnetic radiation has both wave and particle (discreet energy) properties

Wave properties:

c = f λ

c = speed of light = 2.998x108m/s

f = c / λ

f = frequency in Hertz (Hz)

103Hz = 1 kHz 106Hz = 1 MHz

109Hz = 1 GHz

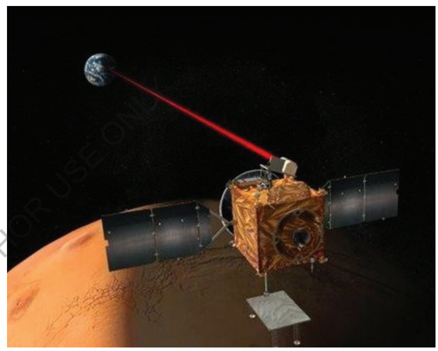
λ = c / fλ = wavelength (in meters)

Particle properties:

The energy of electromagnetic radiation appears in quantized (discreet) units and is proportional to the frequency. This minimum quantity of energy is described as the photon energy. The number of photons represents the intensity of the EM radiation.

$E = hf$, E is energy in Joules, h is Planck's constant = 6.626×10^{-34}Js (Joule-seconds), and f = frequency in Hz.

RF Communication – Terminology



The concepts and terminology associated with communications systems are simple but not commonplace. A brief discussion of the concepts and terminology often encountered in communications systems follows. Center frequency the middle or center frequency of the electromagnetic signal. Terrestrial spacecraft communications are found almost exclusively in the 1-10 GHz microwave band because of the atmospheric interference outside that range. However, optical and infrared frequency systems are possible in certain atmospheric conditions. NASA is pursuing a laser communications network development, first with a Mars mission, expanding to future deep-space missions. The Mars Laser Communication Demonstration (MLCD) is expected to be completed by the end of the decade, although signal attenuation in clouds and rain remains challenging since the laser's operating frequency is in the visible band. Today, spacecraft communications are limited to the microwave band, a collection of bands between 300 MHz and 300 GHz. The more common microwave band designations are as follows:

L-band 1-2 GHz

S-band 2-4 GHz

C-band 4-8 GHz

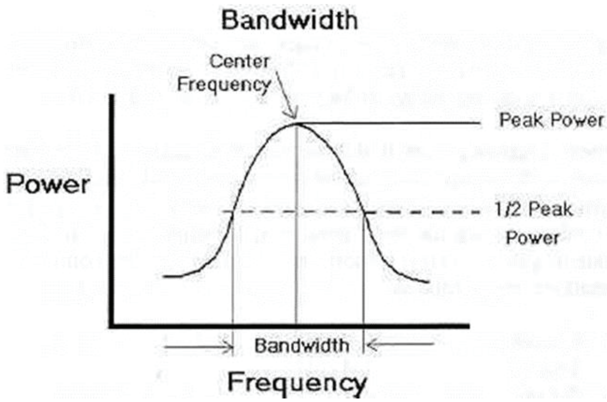
There is no logical link between the band designation and the frequency range or nearby bands since the choice of frequency band designations was selected to confuse the Axis enemies during development of radar technology in WW-II.

More general electromagnetic frequency designations are given below.

ELF	Extremely Low Frequency	30-300 Hz
VF	Voice Frequency	300-3000 Hz
VLF	Very Low Frequency	3-30 kHz
LF	Low Frequency	30-300 kHz
MF	Medium Frequency	300-3000 kHz
HF	High Frequency	3-30 MHz
VHF	Very High Frequency	30-300 MHz
UHF	Ultra High Frequency	300-3000 MHz
SHF	Super High Frequency	3-30 GHz
EHF	Extremely High Frequency	30-300 GHz

Bandwidth

The number or width of frequencies that make up the primary signal in a communications system, or in each communications channel. The bandwidth is customarily measured at the 1/2 peak power point (-3dB), as shown below. Higher bandwidth communications links inherently contain more fantastic signal information. The power contained in the signal bandwidth is proportional to the area under the power-frequency curve, as shown below. Corresponding transmitter power is more significant for higher bandwidths. Noise content is also more incredible for higher bandwidth signals.



Channels

The communications system uses a number of separate but sequential frequency bands. Each channel has a separate transmitter and receiver called a transponder on a communications relay satellite.

Channel bandwidth is the frequency bandwidth of the individual communications channel.

Bandwidth (total) = Nchannels x channel bandwidth

Power

The electromagnetic power received or transmitted as a signal or as noise (or both). This is usually measured in Watts (Joules/sec).

The power of a signal is often expressed as a ratio with respect to a reference power. This ratio can be extremely small or large. Exponential expression of power is easier to use with very big or very small numbers. Base 10 is used for the convention of the decibel which is used to express a power ratio.

Free Space Loss

A 1/r2 decrease in the amount of energy per area as the communications distance increases.

This frequency-dependent loss represents the relative power (or energy) loss from the signal expanding as it propagates through space. The same calculations can be used for approximating the loss through the atmosphere although there is an even greater loss from scattering and absorption.

Loss factor = the magnitude of the loss (actual loss will be the inverse of this) = (4π f d/c)2 with f = frequency in Hz, d = separation in meters, and c = speed of light = 2.998x108m/s

Problem 1: Calculate the loss factor and the loss for a signal between the Earth and Moon with a separation distance between the receiver and transmitter (d) of 385,000 km (Earth-Moon distance) at a frequency of $3.99\text{x}10^9$ Hz (3.99 GHz)?

Solution: Loss factor = $(4\pi\ f\ d/c)^2 = (4\pi \times 3.99\text{x}10^9\text{Hz} \times 3.85\text{x}10^8\text{m}/\ 2.998\text{x}10^8\text{m/s})^2$

Loss factor = $4.140\text{x}10^{21}$

In dB this is:

$$10\log_{10}(4.140\text{x}10^{21}) = 216 \text{ dB}$$

However, this is a loss factor and needs to be converted into a loss. This is done by inverting the loss factor. For the dB value, take $10\log_{10}$ of the inverted value, or simply make the dB value of the loss factor negative (same result).

$$\text{Loss} = -216 \text{ dB}$$

Problem 2: Calculate the loss factor and the loss for a signal to/from Pluto with a semi major axis of 39.4 AU (39.4 AU x $1.49\text{x}10^{11}$ m/AU)

Solution: Loss factor = $10\log_{10}(4\pi\ 3.99\text{x}10^9\text{Hz} \times 5.87\text{x}10^{12}\text{m}/\ 2.998\text{x}10^8\text{m/s})^2 = 300$ dB

And hence the loss = -300 dB

Signal To Noise Ratio

The signal power can easily be overwhelmed by ambient noise if the source and receiver are widely separated since the signal can drop off as 1/r2. The noise can also be a problem if the strength is equivalent to the signal, or if the noise source comes from multiple sources. To find the needed level of signal power to noise power needed for reliable communications, a simple relationship can be used that represents a ratio of power that allows reasonable signal detection. That ratio is often quoted as 3dB, or twice the power available in the signal as in the noise (within the same frequency band).

Other signal detection techniques can be used to separate the signal from internal and external noise with a much lower power ratio than 3dB. However, those techniques usually involve signal averaging (requires repeated signals) and signal placement (spectrum spreading/coding).

Signal improvement can also be implemented with larger antennas (higher gain) (directional antennas usually pick up more signal than noise) and higher sensitivity (lower temperature) receiver preamplifiers.

Polarization

The signal waveform shown in Figure 6.5 represents the propagation of electromagnetic energy through space. Actually, the electromagnetic character of the wave requires an electric component and a magnetic component to the wave, making it indeed an electromagnetic wave. The magnetic wave (field) portion is always perpendicular to the electric wave (field) and oscillates exactly 90o out of phase. The E-M relationship is always perpendicular, although the wave itself can and does rotate as it propagates through the ionized atmosphere and deep- space environments.

Polarized signals can have linear polarization, circular polarization or a combination of the two. Most spacecraft communications systems use a purely linear (horizontal = H, or vertical = V) or purely circular (right = R or left = L) polarization to reduce noise since noise is a primarily random source of interference.

The table below shows the received power (response) from a transmitted polarized signal (vertical axis) and the receiver polarization (horizontal axis). Polarization is produced by the physical properties and geometry of the antenna.

Transmitter	Receiver			
	Circular R	Circular L	Linear H	Linear V
Circular R	1	0	1/2	1/2
Circular L	0	1	1/2	1/2
Linear H	1/2	1/2	1	0
Linear V	1/2	1/2	0	1

RF Communications – Links & Power Budgets

The communications link must have enough signal power to overcome the losses and noise from the transmission, propagation and reception processes. The power level above a standard is called a gain, and below a reference, a loss. The sum of losses throughout the system and the sum of gains throughout the system must add up to approximately +3 dB above the noise level for a simple system (clever encoding and noise reduction or using a repeated signal can reduce the +3dB value significantly).

RF Signal Modulation

A signal coming out of the encoding equipment in the transmitter must be converted to microwave frequencies for propagation through the atmosphere. To accomplish the data signal upconversion to microwave frequencies, a mixer combines the frequency of a local oscillator at the desired center (carrier) frequency with the signal. This is called continuous wave (CW) signal modulation. A modulator is used to code the signal according to a chosen modulation format to make error checking and correction possible in a pulse modulation scheme. This is pulse modulation.

CW Modulation - is the process of conversion of the data into a higher frequency RF signal. Common schemes such as FM or AM are familiar to most of us because of those techniques used in commercial radio.

- FM - frequency modulation
- AM - amplitude modulation

Pulse and phase modulation are less common, but used in many applications in communications. In this technique, the data is represented by various forms of pulsed data at the transmitter output.

PM - pulse modulation

- PAM - pulse-amplitude modulation
- PWM - pulse-width modulation
- PPM - pulse-phase modulation
- PCM - pulse-code modulation

Information cannot be transferred if no change is made to the continuous signal. This requirement establishes two essential components in the communications signal: the carrier (center) frequency and the modulation (data embedding within the signal).

1. Amplitude modulation

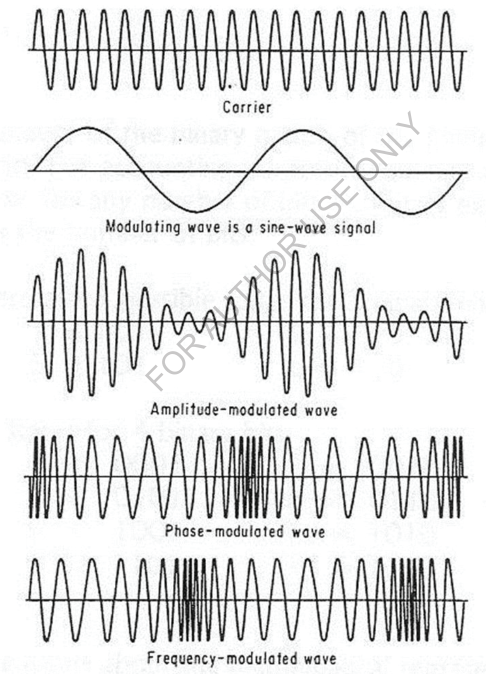
The data can be introduced into the signal, producing a combined carrier and modulated data signal, by changing the amplitude of the signal's waveform as shown in the figure below.

2. Frequency modulation

Changing the carrier signal frequency is a modulation method with a much larger bandwidth (data capacity) than AM modulation. FM also has a greater versatility in data handling than the phase modulation method. The modulation effect is shown below.

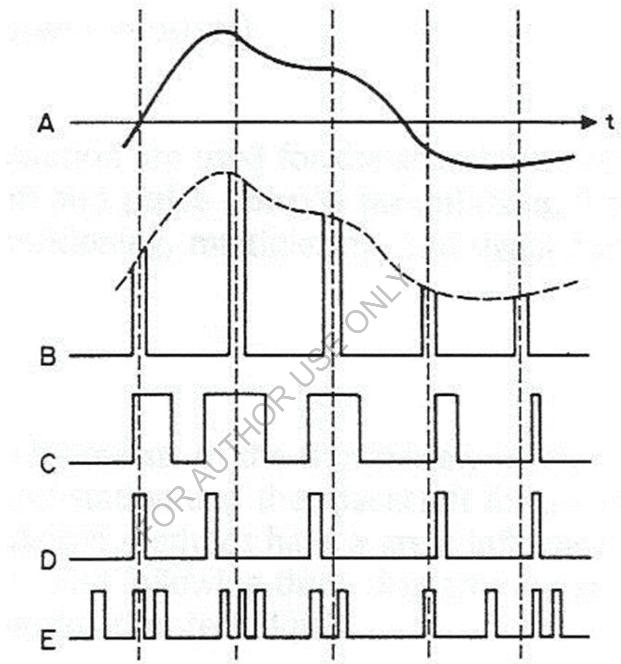
3. Phase modulation

Phase modulation of the signal consists of the changing phase relationship in the signal's sine wave shape. The frequency and the amplitude parameters of the signal are essentially unchanged. A sketch of the various phase modulation techniques is shown below.



Pulse Modulation

Pulse modulation techniques are well suited for digital data transmission and space communications because of the digital nature of the pulsed data properties. A diagram of the three most common types of pulse modulation are shown below.



Binary representation - Because of the binary nature of the computer, data representation is most natural in binary form. For accounting purposes, you can use the simple 3-bit representation shown below for any number of bits by simply expanding the total number of values by 2^n where n is the number of bits.

$2^3 = 8$ meaning there are 8 possible states for 3 binary bits as shown below

0 = 000	1 = 001	2 = 010	3 = 011
4 = 100	5 = 101	6 = 110	7 = 111

$2^4 = 16$ 16 possible states for 4 binary bits as shown below

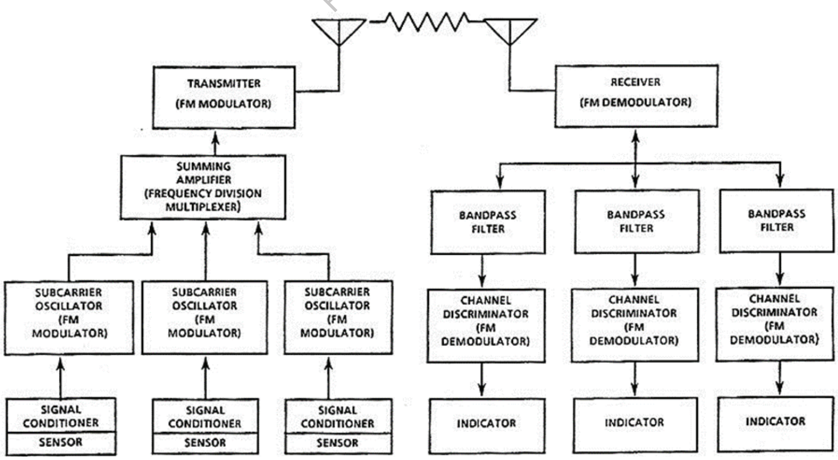
0 = 0000 1 = 0001 2 = 0010 3 =

4 = 0100	5 = 0101	6 = 0110	7 = 0111
8 = 1000	9 = 1001	10 = 1010	11 = 1011
12 = 1100	13 = 1101	14 = 1110	15 = 1111

Modulation Circuits

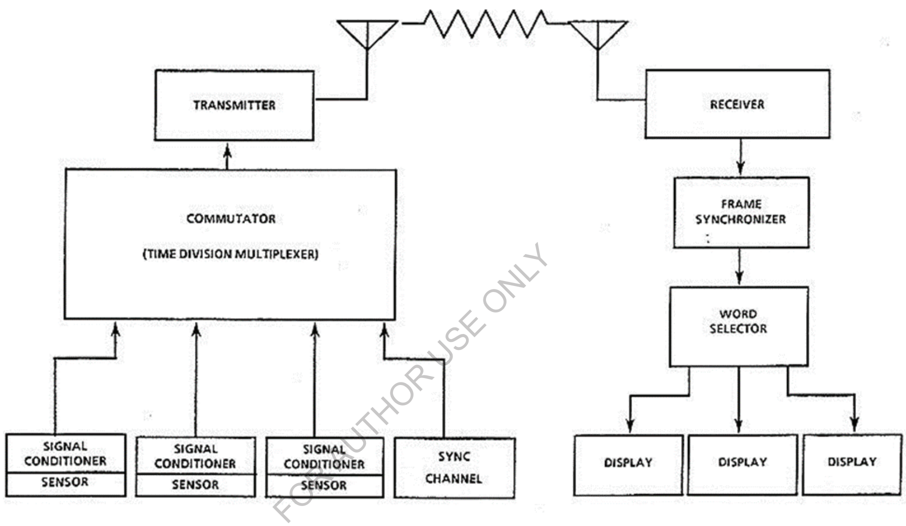
The two basic types of modulation are used for the space-borne communications systems are the frequency domain modulation and pulse domain modulation. These two types of signal encoding can be used for signal conditioning, multiplexing, and signal encoding. The various modulation methods greatly influence the communications and data system hardware, however. The following three diagrams depict communications modules for the three more common modulation techniques.

Frequency Domain Modulation



Pulse Domain Modulation

There are two types of pulse modulation: pulse code modulation (PCM) and pulse amplitude modulation (PAM). Pulse code modulation is used in vast majority of spacecraft applications. The two diagrams below show a diagram of the two different functions.



Spacecraft Telemetry

Project Mercury introduced several innovations in manned space flight, not the least of which were spacecraft ground control and telemetry communications. With much uncertainty about the physiological impact of space flight on humans, a great deal of weight was placed on in-flight measurement of both vehicle and biomedical parameters. Although the quantity of data transmitted from the Mercury vehicle was far less than today's standards, only 97 parameters, the methods used to multiplex and transmit data were innovative, setting refined and improved standards in later manned programs.

All of Mercury's 97 data parameters were analog. This simple character allowed them to be applied directly to the modulator of an FM transmitter. Seven of the parameters were dedicated to their own individual carriers, and the other ninety were commutated or sampled so they could be time-sequenced onto a single carrier. All eight carriers were then mixed into a single RF downlink. Many of the critical parameters were confirmed by voice during each pass over a ground station to verify telemetry system integrity. The FM signal containing all the sampled parameters was continuously recorded onboard the spacecraft during the entire mission for post-flight analysis, although there was no way to playback the recorded data inflight. The ground system consisted of seventeen stations, each communicating with Goddard Space Flight Center (GSFC) via teletype lines. GSFC in turn relayed the telemetry data to the newly completed Mercury Control Center at Cape Kennedy for near-realtime monitoring (Muratore).

During Project Gemini, telemetry was advanced dramatically from the conversion to a completely digital system, converting each analog parameter measurement to an 8-bit digital Pulse Code Modulated (PCM) value. The new system monitored a maximum of 338 parameters (193 analog, 120 bilevel, and 25 digital), of which less than 300 were used, with their primary purpose being malfunction detection. The parameters were sampled at various rates from 0.416 samples per second to 640 samples per second to produce a composite data stream of 51.2 Kbps. During Loss of Signal (LOS) periods, portions of this data stream were recorded for delayed downlink. In addition, the ground system received several major upgrades, including a completely new Mission Control Center (MCC) at the Manned Spacecraft Center (MSC) in Houston and additional sites and ships for 23 telemetry receivers. The new MCC had enough online computing power to perform automatic, high-speed telemetry data processing for the first time. In addition, the network was upgraded to handle up to 40.8 Kbps of wideband data from six ground stations to the MCC, with 2 Kbps data and 100 wpm teletype being relayed from the remaining stations to GSFC and then to the MCC (Muratore).

During the development of Project Apollo, the much greater distances and extended flights introduced more reliable and versatile communications equipment. However, there were no corresponding changes to the telemetry system used for the Gemini missions. Retaining the same 51.2 Kbps downlink, Apollo added a reduced rate 1.6 Kbps downlink when the high rate link was not usable. The number of instrumented parameters increased to 403 (365 analog and 38 digital) in the nominal data stream and 134 (100 analog and 34 digital) in the reduced rate data stream. The network support from Gemini was complemented by instrumentation aircraft, reentry ships, and stations capable of communicating with the spacecraft while in the vicinity of the Moon (Muratore).

Communications equipment for the Skylab program used the same telemetry capabilities as the later Apollo flights, but was upgraded to include data from the Orbiting Workshop, Apollo Command Module, and Apollo Telescope Mount on separate radio systems. The MCC retained its Apollo capabilities, but was extended to support missions of up to 56 days in duration.

Current telemetry designs

With new technology, advanced data processing, and instrumentation, spacecraft telemetry processing for the Space Shuttle program was expanded ten-fold. The increase in instrumented parameters from the 403 of Apollo to nearly 4000 for the Space Shuttle required higher data rates, although not all of them could be downlinked continuously. Consequently, distinct telemetry formats are developed to select and downlink only pertinent data during each Space Shuttle mission phase. The telemetry formats are defined in software prior to each mission, providing tremendous flexibility to tailor the telemetry to requirements.

The new data format requirements came with a substantial increase in the complexity of reconfigurations. Telemetry data from the Orbiter incorporates time-division-multiplexing into either a high rate 128 Kbps data stream or a low rate 64 Kbps data stream. As in previous programs, data is recorded onboard during LOS periods and then played back at a higher rate during later Acquisition of Signal (AOS) periods. During the first seven STS flights, data was transmitted directly from the Orbiter to the ground stations where it was simultaneously recorded and relayed to GSFC via the Spaceflight Tracking and Data Network (STDN) and then to the MCC. On STS-8, the Tracking and Data Relay Satellite (TDRS) was checked out and a new era of spacecraft telemetry began. When using the TDRS, the spacecraft is in continuous communication with the MCC for slightly more than half of each orbit, and as long as the orbiter's Ku-band antenna has a line-of-sight view of the TDRS satellite, the system can simultaneously relay orbiter telemetry (128 Kbps), two channels of voice (32 Kbps each), and up to 50 Mbps of experiment data from the Orbiter to WSGT and then to the MCC (Muratore).

In contrast, the International Space Station is instrumented to monitor over 25,000 individual parameters. The telemetry processing is distributed among the various systems' Standard Data Processors (SDP's) which communicate with each other via a high speed packet data network. The ground systems distribute packets of telemetry data to literally hundreds of users located around the world, and the users must be able to handle packetized data, extracting pertinent data and processing it as their needs dictate.

Telemetry Modulation Operations

All PCM (pulse-code modulation) ground stations have to perform the same basic front end functions

- Find where the bits periods start and stop (extract the clock) and make a decision on the value of each bit (bit sync)
- Find the word boundaries by finding the sync pattern (frame detect)
- Find the minor frame boundaries by finding multiple sync patterns between the correct number of words (frame sync)
- Find the subframe counter and verify that it is cycling properly and that it can be used to identify minor frames (subframe sync)
- Extract the parameters (decommutate), calibrate them, convert them, limit check and combine them for processing and display

Signal Format Sequence – Receiver

Demodulator → Bit Synchronization → Frame Sync → Subframe Sync → Decommutate → Process → Display

Bit Synchronization

The ground receiver removes a data stream from the RF carrier but where do the bit periods start and stop? The typical bit synchronizer utilizes a matched filter - an integrate and dump filter - to determine the alignment of bit periods. May use analog integrator (capacitor) or digital integrator. Based on the concept that when bit periods are appropriately aligned, the area under the curve of the wave form should be 0 for a 0 and pulse width times power for a 1. Frame synchronizer compares output of bit sync with expected frame pattern.

Subframe sync and Decommutator

Subframe (minor frame) sync checks the counter or other subframe (minor frame) sync method

Decommutator uses subframe identification (minor frame) to extract telemetry parameters using frame number and word offset

Space Communications Networks

The communications networks that control space operations and data are combined ground tracking and communications stations, communications links, and command and control centers. The Department of Defense manages the primary space communications networks in the U.S. (DoD) and NASA, and their affiliates. Other space data and communications networks are available for communications and meteorological satellites, Earth-observation satellites and research satellites.

Primary Networks

NASA has two primary spacecraft communications networks. They are the ground networks and the space network. These networks are operated by NASA's Office of Space Communications (OSC), which offers government and participating users spacecraft communications, data transfer and translation, spacecraft command, spacecraft control, spaceflight analysis, project planning, and project and flight development functions.

Ground Networks

This set of networks are globally distributed ground tracking stations that communicate with Earth-orbiting and deep space missions, as well as fixed and mobile facilities for NASA's aeronautics, balloons, and sounding rockets (AB&SR) programs. The GN encompass all the ground station resources of the Deep Space Network (DSN), Satellite Tracking and Data Network (STDN), Wallops Flight Facility (WFF), and Western Aeronautical Test Range (WATR). The WFF is the lead center for NASA's balloon and sounding rocket programs and provides management and technical oversight for NASA's ground stations at the White Sands Missile Range (WSMR), the National Scientific Balloon Facility (NSBF), Poker Flat Research Range (PFPR) and McMurdo, Antarctica. The WATR is the lead organization for NASA's aeronautical flight testing activities and provides services to the Space Shuttle/Space Transportation System (STS). The WATR's three California ground stations are managed by Ames Research Center (ARC) Dryden Flight Research Facility (DFRF). Together the WFF and WATR provide the principal services to NASA's AB&SR programs (Muratore).

Space Network

This is a constellation of geostationary satellites and associated ground-based systems that provide the primary communications with low-Earth-orbit missions. The SN element includes all facilities and systems associated with operation of the Tracking and Data Relay Satellite System (TDRSS) Ground Network

As mentioned before, these networks, and the communications functions of NASA are handled by the Office of Space Communications. The operations and functions of the communications networks described in this material are focused on the OSC because of its responsibility.

Deep Space Network

The primary mission of the DSN is to support deep-space, interplanetary, and radio science missions. In addition, DSN facilities provide launch and emergency backup support for the Space Shuttle missions and other Earth-orbiting missions. The DSN is managed by the Jet Propulsion Lab (JPL). The DSN consists of three main elements, the Deep-Space Communications Complexes (DSCC), the Network Operations Control Center (NOCC), and the Network Spacecraft Test and Launch Support Facility).

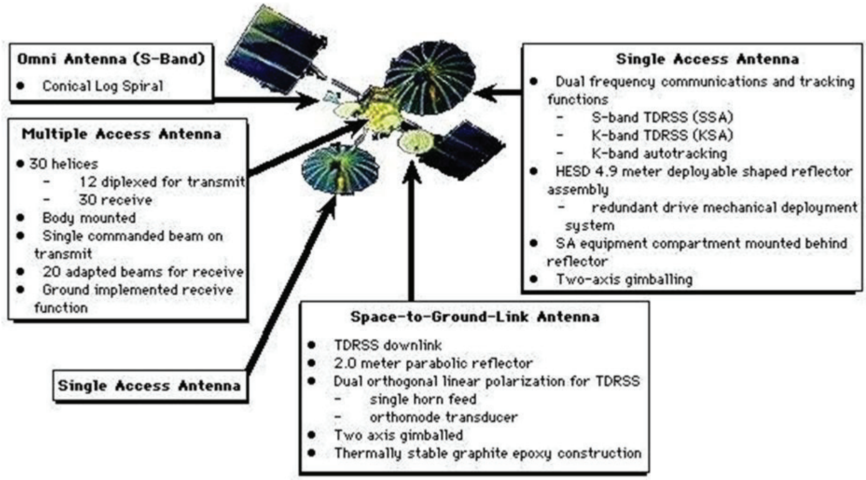
The DSN ground sites (DSCC) consist of three ground station complexes with multiple communications antennas and communications links that allows NASA and the spacecraft customer to control spacecraft and transfer/access data to/from specific spacecraft. These three stations are located in Canberra, Australia; Madrid, Spain; and Goldstone, California. Each Earth station supports four separate RF antenna subnets covering three different RF bands.

Spaceflight Tracking & Data Network

The STDN is one of the Space Net's principle components, providing forward and return telecommunications services between low-Earth orbiting satellites and the TDRS. The three elements of the STDN are the Tracking and Data Relay Satellites (TDRS), the tracking stations, and the Network Control Center (NCC).

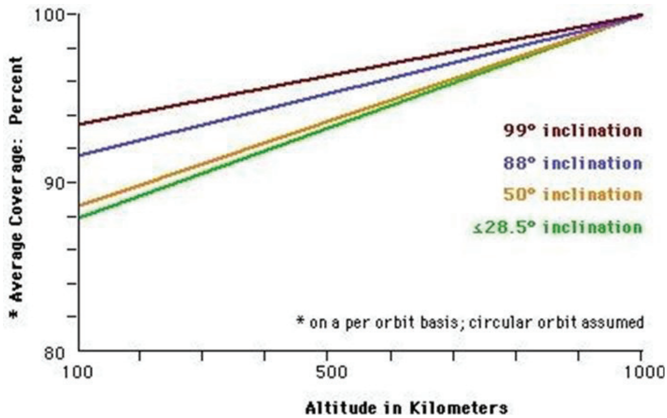
TDRSS

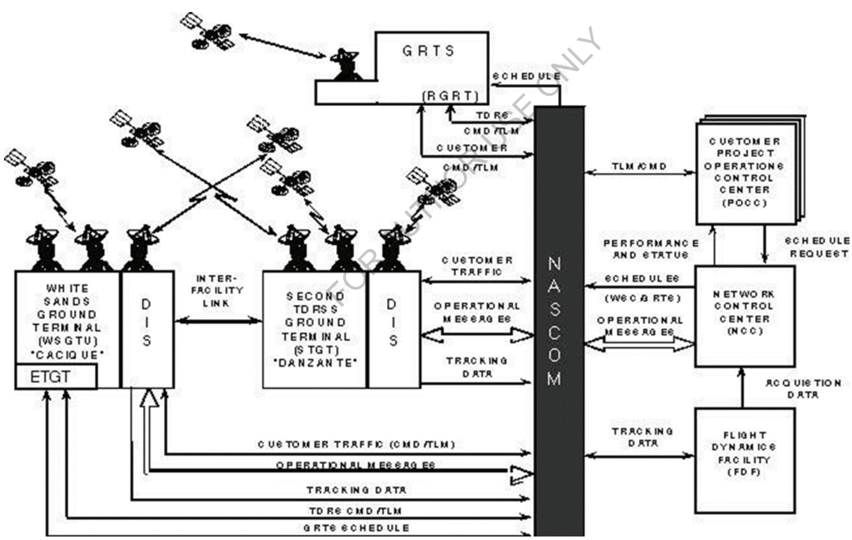
The Tracking and Data Relay Satellite System (TDRSS) consists of six on-orbit TDRS's located in geosynchronous orbit. Three of these are available for operational support at any given time. The other three provide operational backup in the event of a failure of any of the three primary satellites.



Ground Segment

The TDRSS ground segment is located at White Sands, New Mexico, and consists of two ground terminals, a control center, TDRS satellites, and the network. The low-Earth shown orbit satellite network provides nearly continuous coverage for these satellites, and greater coverage for higher-orbit satellites.


100
90
80
* Average Coverage: Percent
99° inclination
88° inclination
50° inclination
≤28.5° inclination
* on a per orbit basis; circular orbit assumed
100
500
1000
Altitude in Kilometers


GRTS
(RGRT)
SCHEDULE
TDRS CMD/TLM
CUSTOMER CMD/TLM
TLM/CMD
CUSTOMER PROJECT OPERATIONS CONTROL CENTER (POCC)
PERFORMANCE AND STATUS
SCHEDULE REQUEST
SCHEDULES (WSC & RTS)
OPERATIONAL MESSAGES
NETWORK CONTROL CENTER (NCC)
ACQUISITION DATA
TRACKING DATA
FLIGHT DYNAMICS FACILITY (FDF)
N A S C O M
WHITE SANDS GROUND TERMINAL (WSGTU) "CACIQUE"
ETGT
D I S
INTER-FACILITY LINK
SECOND TDRSS GROUND TERMINAL (STGT) "DANZANTE"
CUSTOMER TRAFFIC
OPERATIONAL MESSAGES
TRACKING DATA
CUSTOMER TRAFFIC (CMD/TLM)
OPERATIONAL MESSAGES
TRACKING DATA
TDRS CMD/TLM
GRTS SCHEDULE

Communications & Data Systems

The two branches of data services are the NASA Communications (NASCOM) system and the Program Support Communications Network (PSCN).

NASCOM is designed to provide highly reliable communications with secure access to protect the U.S.'s space assets interconnects all domestic and foreign tracking and data acquisition stations, launch complexes, mission control centers, network centers, field installations, and international partners.

The PSCN is designed to provide links between NASA headquarters, NASA installations, remote facilities, contractor facilities, and universities. The network handles nonoperational data and provides administrative communications support, including services such as computer networking, high-speed data transfer, electronic mail, and voice and video conferencing.

OSC Services

At the lowest layer of the OSC model are the Communication Services, which provide the basic communication paths between customer resources and OSC resources. For the customer's flight platform, these services are provided through radio links that support the transfer of information between the flight platform and OSC resources. This includes the coding and modulation of signals onto the radio frequency (RF) carrier for transmission information to the flight platform and the demodulation and decoding of received signals to re-create the data stream transmitted by the flight platform. The communication services are provided through data network connections to OSC resources for the customer's ground facilities.

Data Services

The OSC Data Services provide higher-layer capabilities for processing data streams created aboard the flight platform, including removing communications artifacts, error detection, and auditing. These services also prepare data to be transmitted to the flight platform compatible with the onboard data processing systems. Four functional layers within this service provide the following capabilities:

Data Processing

The data Processing system detects and correct errors, restores time sequence, generate data sets, and correlate data with flight platform position, and attitude.

Packet-Level Functions

The packet-level function process and manages data using a packet structure.

Frame-Level Functions

The frame-level functions manage information on the frame level. A frame is a structure containing a block of data where the block can consist of a number of packets or a number of bits.

Digital-Bit Level Functions

Process data at the bit level. This form of data processing is currently being replaced with more modem and cost-effective.

Time Services

The OSC Time Services disseminate a reference time and frequency to a customer's flight platform and facilities to ensure the coordination of mission operations and data.

Navigation Services

The OSC Navigation Services provide four hierarchical services, each built upon the products of the lower service:

Navigation Aids

A broad set of capabilities that use products from the other navigation services for creating onboard navigation and attitude control data, fuel use monitoring, sensor occultation periods, ground traces, and contact point.

Attitude Functions

The capability for processing data transmitted from the flight platform to determine platform attitude system calibration, and attitude control.

Flight Path Functions

The capability for processing low-level tracking data to create a definitive flight path history and predict and control future flight path. These capabilities are available for all forms of flight, including powered, ballistic, and orbital.

Tracking Measurement

The capability for providing radiometric information gathered by OSC RF (radio frequency) resources for computing range, range rate, and/or angle information.

Mission Operations Services

The OSC Mission Operations Services provide capabilities for performing general mission operations functions and also for coordinating and scheduling the lower level OSC services. The five OSC Mission Operations Services are:

Planning and Scheduling

The capability for transforming a mission science plan into a series of flight platform events and then scheduling these events along with OSC services.

Real-time Command and Control

The capability that provides real-time transmission of commands and/or command files to the flight platform and verification that the commands are received and processed correctly

Health and Safety Monitoring

The capability for monitoring and formatting the flight platform telemetry for analysis by the customer's ground controllers.

Board Processor Management

The capability for creating the command and software loads for the flight platform and around systems as dictated by the mission events developed under Planning and Scheduling Services.

OSC Services Coordination

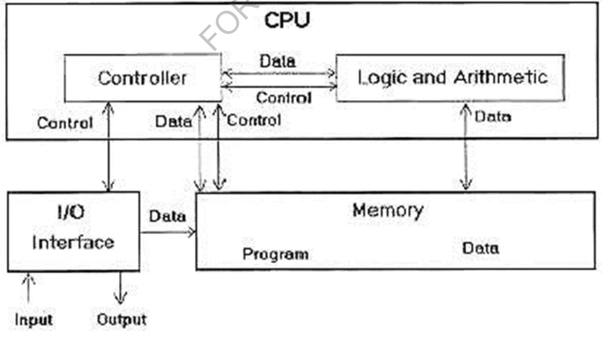
The capability for interacting with OSC organizations for scheduling OSC services and performing real-time configuration changes on behalf of the customer.

OSC resources are available to manned and unmanned spaceflight programs, expendable launch vehicles, suborbital flight projects, aeronautics, and non-flight projects such as radio astronomy, geodynamics, and solar system radar experiments. Customers include both NASA and non-NASA programs.

Chapter 6: Data Processing Systems

The digital computing systems required for onboard processing and data handling on the typical spacecraft are often multiple specialized processors with multiple redundancy. The amount of data acquired, transferred and stored with today's spacecraft command and data systems are often so great that dedicated subsystems are used to manage the flow of the very diverse data. Large spacecraft like the Space Shuttle, or the Hubble Space Telescope contain dozens of data processors, although only a handful may actually be defined as digital computers since they can be programmed.

Current computing systems on spacecraft could be divided into six main functional blocks or subsystems. Those subsystems are the primary components of the command and data handling system, including data storage and management, and instrumentation data and telemetry functions. Because command instructions pass through the communications system receiver from the ground link a part of the spacecraft communications system is dedicated to the command and data handling functions - traditionally the command processor.



Data Processing Systems Terminology

Bit - the binary information used in digital logic and computers consists of two (binary) states which can be used to make numerical data (telemetry for example) and non-numerical data (instructions for example) with multiple bits. Those multiples can be bytes, words or other bit sets - always in 2n.

CPU - Central Processing Unit is the heart of the computer, executing the programmed instructions and processing data through the data and memory buses.

Command System - the entire spacecraft command system includes the communications and telemetry system as well as the C&DH system.

Command Processor - the coded commands coming from the ground control to the spacecraft receiver are sent to the command processor for interpretation, validation, and relay to the proper interface circuit or subsystem. The command processor is commonly a microprocessor-based unit for current spacecraft.

Commutation - the data that comes into the multiplexer for serial distribution (time division multiplexing) has source identification attached to the data, and is arranged to produce a specific "frame" or multiple bit pattern. There are generally minor frames making up the major frames in the complete pattern.

Frame - a formatted data set containing specific command or data information along with the source identification and time in some applications.

Memory - the data and instructions that are required to operate the logical instructions in a computer are stored in several ways according to the specific use and needed speed for access.

The processing instructions for the CPU are written directly into the processor circuits permanently. Any loss or alteration of this memory is fatal to CPU operation.

Specific software instructions for the mission can be coded in programmable memory for repeated use and rapid access. This programmable memory must be protected from alteration or loss, but can be corrected under certain conditions.

Random Access Memory (RAM) - the data and instructions for the software programs operating on the CPU are stored in this directly accessible memory storage. This memory must be accessed rapidly for reasonable CPU speed. Loss or alteration of the memory is not a significant problem if error detection and recovery routines are provided.

Bulk (Mass) Memory - the large (mass) memory is needed for large amounts of data storage or complex software routines. The greatest mass memory requirement is usually for the spacecraft sensor data which is often stored for playback when a communications link is made for data transfer to ground stations.

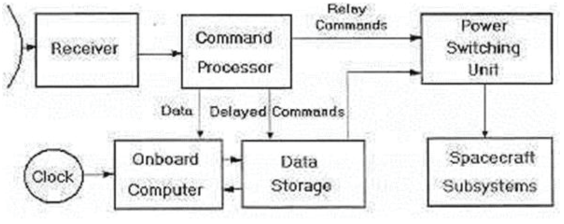
MIPS - Millions of Instructions per Second - this measures the computer processor's processing power and speed since the instructions are synchronized to the master clock. The length of instruction, the width of the processing registers, and the speed of the CPU clock combine to produce the execution power of the computer processor, or the number of instructions executed in one second.

Multiplexer is a multiple access device that places several input channels on a single output line in either a time sequence or a frequency band. A demultiplexer is a device for the reverse operation (segregating serial input into multiple outputs for the respective data).

Sampling - the measurement of analog data from a sensor which is converted to binary information for placement on the data network (usually through the multiplexer.

Resolution - the range of measurement that a sensor output is capable of, usually measured in the number of output data bits between minimum and maximum sensor range.

Nyquist sampling - the minimum number of samples taken during a signal change to assure reasonable data representation. Minimum is 2 times the data frequency, more practical sampling is on the order of 10x data frequency.



Computer & Command System Structure

- System Organization
- Design guidelines
- Distribution topology and networks
- Communications link
- Operating system
- Redundancy
- Reconfiguration

System Organization

The C&DH system (sometimes called the command and data system, the computer command system, or the flight data system) is organized into functional blocks or subsystems.

Executive control - software and hardware command and control, resource control, and scheduling. This is the ultimate management unit

Command processing - this data processor is responsible for interpretation/decoding, validation, verification, and distribution of uplink and onboard commands for the spacecraft.

Computations - the processor components and instructions that are responsible for data manipulation, scaling, calibration, formatting, compression, processing, and analysis.

Data acquisition - The data from sensors and instruments normally is processed through an A/D (analog-to-digital) converter and transferred for processing and/or storage. The data rates for some instruments can be extremely high and several processors may be dedicated to data acquisition.

Data storage - the bulk storage for payloads, instruments, communications (uplink and downlink data stored temporarily for backup & verification), telemetry, onboard operations, and systems status.

Communications interface - this function furnishes the uplink and downlink communications with the appropriate data manipulation and formatting, sequencing, encryption and decryption, coding and decoding, and protocol operations.

Mission specific operations - the payload and instruments have individual requirements that must be accommodated explicitly by the hardware and software design. The C&DH system must also have unique fault and error handling functions and reconfiguration capability. These are provided within the overall design of hardware, software, and (possibly) human input.

Design Guidelines

The cost and operations (acquisition, fabrication, testing) scheduling are primary concerns of the spacecraft and mission design efforts, including the C&DH elements.

C&DH considerations

- Size
- Weight
- Power required
- Processing speed
- Complexity

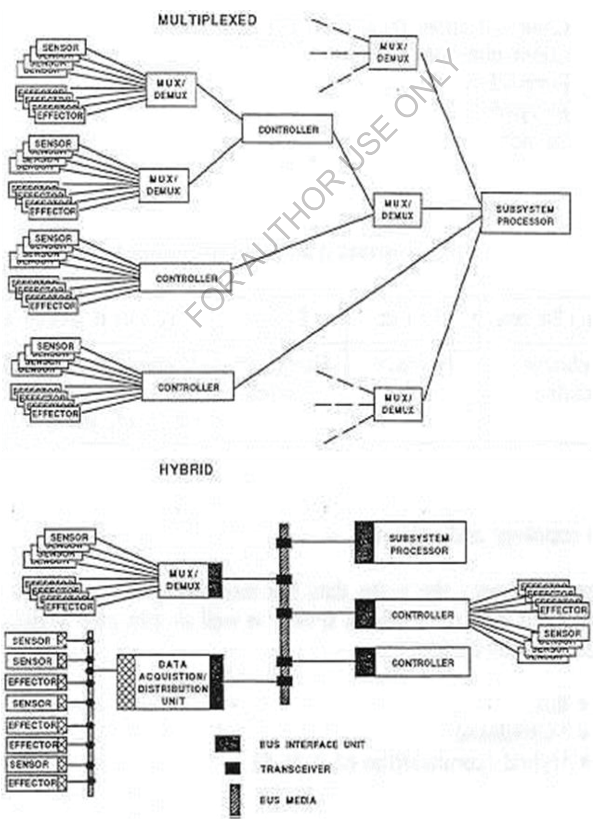
Data considerations

- Transmission and distribution speed
- Communications bandwidth & rates
- Communications network
- Formatting
- Encryption
- Storage

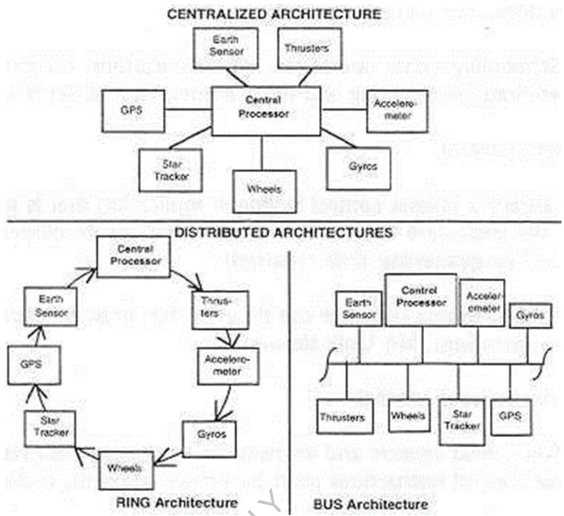
Distribution Topology and Networks

Hardware topology - this is the data bus interconnection structure for the system. Processing and data transmission speed, as well as data path width, are of primary importance in the design. The three primary types of hardware topology are:

- **Bus**
- **Multiplexed**
- **Hybrid (combination of A & B)**



Network topology - this represents how the processors and subsystems in the C&DH system are connected. The three primary types of network topologies are the ring, star, and bus.



Ring - circular connection allows simultaneous communications, software protocol more flexible.

Star - direct interconnect excludes simultaneous communications, centralized structure requires more cabling, less flexibility

Bus - bus connection allows simultaneous communications, multiple bus types, with more flexibility.

Communications Link

The data and commands to and from the spacecraft must be processed through the communications system. The speed of the data transfer is determined by the processing speed of the C&DH system, and the bandwidth of the communications link. Lower frequencies have a generally lower bandwidth and are often used for telemetry rather than payload data transfer.

Imaging payloads (video cameras) require high data communication rates and large bandwidths.

Typical transfer rates

- Small spacecraft - 10 Kbytes/sec or less
- Communications satellites - 200 Mbytes/sec or more

Operating System

The computer operating system controls high level-to-low level computer operations

Executive (overall) control

Usually a high level software operating system for comprehensive operational control of subsystems. Generally a large code with resulting slow functional operation. Unix, VMS are examples of larger Earth-based operating systems.

Resource allocation - CPU, data, command, system/component functions, and communications

Scheduling - data operations, communications, computations, payload operations, engineering and flight events, fault & error recovery tests, etc.

Subsystem control

Usually a unique control language application that is sometimes available for the subsystem control in modular form (or developed separately with lots of programming time required).

Fairly versatile language can be used, but must be capable of high data rate operations. No Unix is allowed here.

Sensor/instrument control

Since most sensors and instruments have large data rates when operating, these control instructions must be simple, efficient, dedicated, and fast.

Usually machine language. Time consuming programming often required but may be available from other space applications packages.

Redundancy

Computer system and memory redundancy are required in anticipation of the effects of radiation on microelectronic circuits and the resulting errors and component failures. The error recovery must be planned. The fault detection and fault tolerance must be designed into system and subsystem hardware and software. Lastly, the catastrophic event recovery - multiple CPUs and software capability needed in the design of the C&DH system.

Reconfiguration

Spacecraft C&DH system reconfiguration is necessary for error and/or failure recovery and future enhancements.

Radiation

Particle radiation can create microcircuit errors and failures. Permanent error/damage is a hard error, recoverable errors are called soft errors.

Spacecraft Digital Data System History

NASA's first major project was the manned Mercury program first launched with astronauts onboard in 1960. The Mercury spacecraft contained an attitude control system compressed of automated and manual inputs, but no orbital control. Mercury's only orbital thrusters were the deorbit engines under the command of automated control from the mission sequence system or ground-based instructions. Manual override was also possible, and was exercised during the program. Even so, there was no need for an onboard computer for navigation or complex mission tasks, hence no computer was integrated into the capsule. In stark contrast to the simple Mercury capsule, the Gemini capsule could change orbits, perform deorbit from various orbit configurations, and alter its orbit accurately for rendezvous maneuvers. Gemini's complex orbit operations demanded digital computing power for both position and trajectory determination and trajectory and orbit control. These versatile functions on a manned vehicle meant that the command and data handling system had to be a very reliable, sophisticated (for its era) digital computer with simple crew interfaces. It would also have to have the capability of taking over for the TItan launch vehicle's guidance and control functions in an emergency.

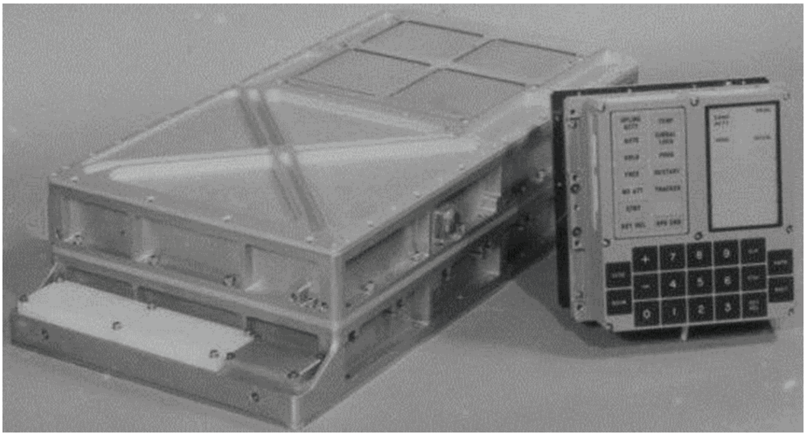
Apollo

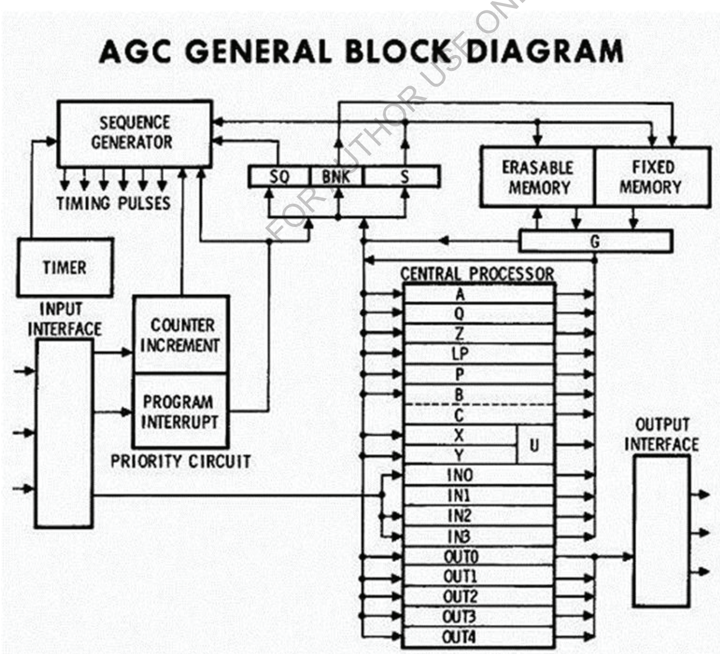
The Apollo capsule presented a much more demanding computer hardware and software design since the mission would involve complex maneuvers at, on, and near the Moon a quarter of a million miles away. Precise navigation accuracy, and unsurpassed reliability meant that the Apollo computer had to be far more advanced than anything available, and had to be compact since it had to fly on the Command Module (CM) as well as the Lunar Module (LM).

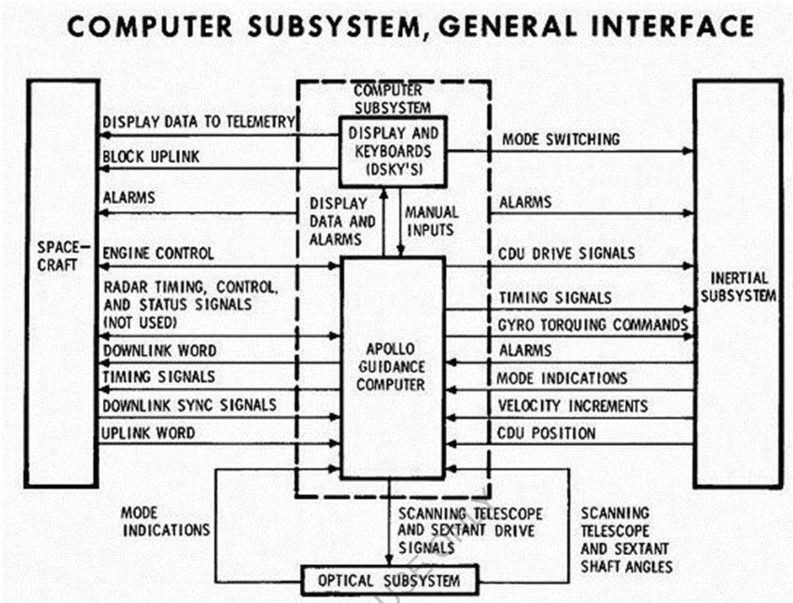
NASA handed the task of developing and constructing the Apollo computer to the MIT Instrumentation Lab in 1961. The assignment predated the first Gemini computer which presented enormous challenges to NASA, the computer hardware designers, and the software engineers. And even though the two computers on the separate command and lunar modules, a separate operating system had to be written for each since the operating environment and risk environment was different. Some of the more important initial requirements for the Apollo Primary Guidance, Navigation, and Control System (PGNCS) included: a second computer in the Lunar Module was included in the system design to take over in case the primary computer failed on descent, landing, or ascent. Ground systems had to be able to back up the CM computer and guidance system to manually guide the spacecraft if the CM system failed, based on data transmitted from the ground. To provide a safety margin for lost ground contact, the CM system had to have autonomous return capability.

Even with the new redundant capabilities, the final decision for the Apollo Command Module and Lunar Module guidance computations was to expand accuracy and reliability of the vehicles by using ground-based solutions that would be transmitted to the CM and LM. Only in a communications failure or similar critical event would the onboard computers furnish autonomous guidance and navigation calculations. The new integrated circuit components helped the Apollo computer system (PGNCS) control many system operations and monitoring functions previously managed by the crew and ground systems.

The legacy of the Apollo digital computer system was primarily the advancement of the integrated circuitry technology, the miniaturization of electronics. But within the 10 years from the first development efforts to the last Apollo lunar flight in 1972, the rapid advances in digital circuitry and computers left the Apollo digital hardware far behind.




AGC GENERAL BLOCK DIAGRAM
SEQUENCE GENERATOR
TIMING PULSES
SQ
BNK
S
ERASABLE MEMORY
FIXED MEMORY
G
TIMER
INPUT INTERFACE
COUNTER INCREMENT
PROGRAM INTERRUPT
PRIORITY CIRCUIT
CENTRAL PROCESSOR
A
Q
Z
LP
P
B
C
X
Y
U
IN0
IN1
IN2
IN3
OUT0
OUT1
OUT2
OUT3
OUT4
OUTPUT INTERFACE

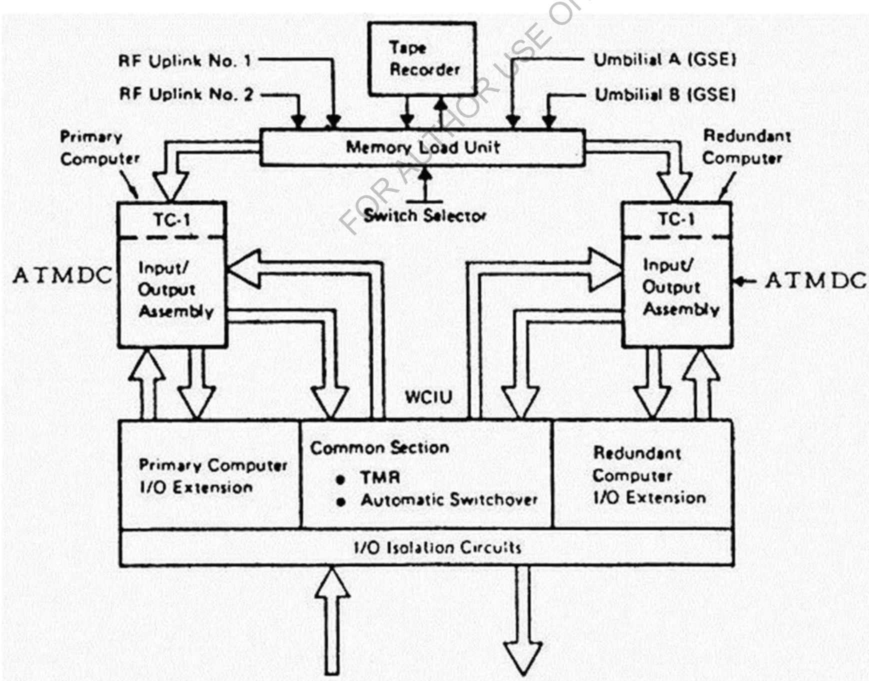


SkyLab

The design of the Skylab digital computer and data system was a step away from the custom design of the Gemini and Apollo computers towards a production processor and peripheral components. Skylab's computer design was the first fully-integrated digital system which allowed to manage the entire space station without crew intervention and without complete control from the ground computers.

IBM was awarded the contract to develop the Skylab digital computer and data system using its 4Pi descendant of the 360 minicomputer architecture. The processor was used in military aircraft for similar functions and was already proven and radiation hardened (NASW-3714). The 4Pi central processor was augmented with specialized input/output and peripheral devices to adapt the control and command systems of the station to the relatively advanced computer. A similar IBM 4pI architecture was used for the Space Shuttle redundant General Purpose Computers that were under development soon after the Skylab computer design was underway.

Skylab's digital computer consisted of two separate redundant units - one powered and one not. Each had triple-redundant circuitry running in parallel with voting circuitry to establish which computations were correct when inconsistent results were detected. The redundant polling technique proved effective in practice and would be adopted for the Space Shuttle digital processing system which used five redundant complete computers.



Also setting the stage for the later Space Shuttle digital processing system, the Skylab software was divided into two separate functions called the executive and the applications software.

Executive operations were generally the priority functions, interrupt processing, and time keeping. Redundant tape drives were used for program loading, also one of the carryovers to the Space Shuttle digital system. A keyboard and display interface also allowed the crew to interact and control the program operations, however, the autonomous system required little intervention during its use (NASW-3714).

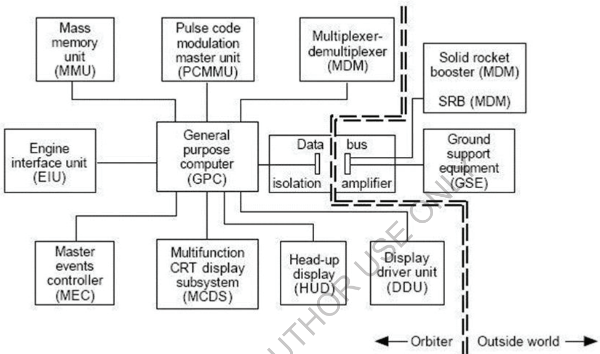
Space Shuttle Program

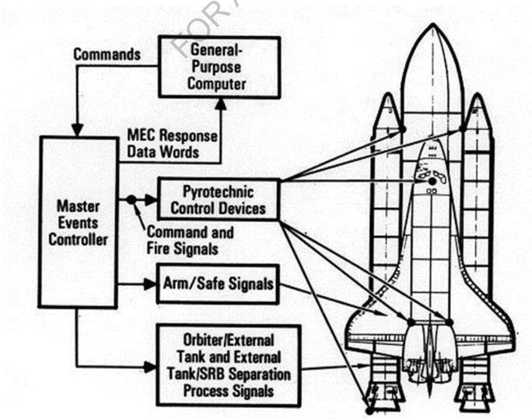
The Space Shuttle's Digital Processing System (DPS) was designed with a variety of attributes that evolved from preceding manned vehicle designs, including the program lessons learned. To save time and reduce expenses, the DPS was to be selected as an "off-the-shelf" item like the Skylab digital system. This led almost directly to an updated version of the Skylab's IBM 4Pi, designated the AP-101. Instead of the triple circuits and standby computer system, the Space Shuttle DPS design employed five independent computers with completely separate components, each running in parallel. Digital processing for the Space Shuttle was a much broader task because of the launch, reentry, and landing phases that not a part of the Skylab mission. More extensive software required more complex and longer programs for the vehicle operations from launch to landing, yet the computer memory was only marginally larger than the IBM model adopted for Skylab. The operational phases were divided into software loads that could be automatically or manually pulled from the dual tape drives to overcome this limitation.

While cumbersome, the procedure proved reliable enough to endure nearly 30 year of operation.

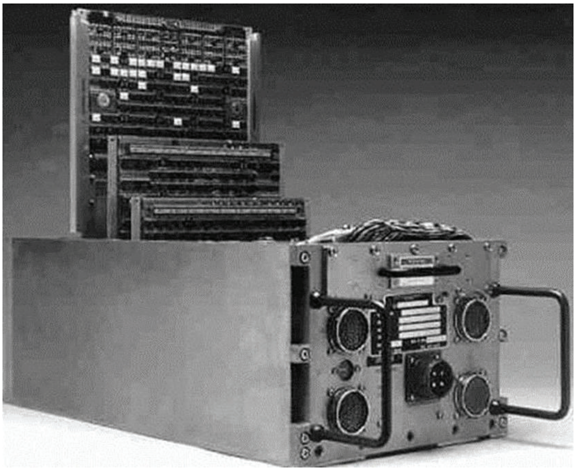
The Space Shuttle's Digital Processing System consists of key components, supported by a wide variety of interface and support elements including the communications link, command processor, and the power supply. Those key components include five General Purpose Computers (GPCs), 24 (20 Orbiter, 4 SRB) Multiplexers/Demultiplexers, two master events controllers, two Mass Memory Units, 24 Serial digital data buses, three SSME interface units, four multifunction CRT display systems, two data bus isolation amplifiers and one Master timing unit.

Similar to the Skylab software architecture, the Space Shuttle's DPS software is divided into two components - service and applications. Unlike Skylab, the two software functions that run simultaneously run on a backup GPC that executes a compact version of the primary software called the Backup Flight System (BFS). The BFS runs independently on a separate General Purpose Computer and can be activated to replace the primary software at the flip of a switch. This supplies a completely independent software control package, written in a different language, to avoid unforeseen software errors or catastrophic failures.





Master Events Controller



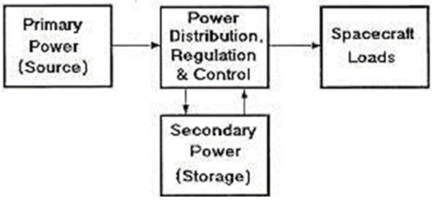
Chapter 7: Electrical Power Systems

The electrical power system is used to power all of the spacecraft systems and subsystems. Nearly all of the satellites and spacecraft have a primary power source and a secondary storage subsystem to compensate for variable sources and loads. Short- duration flights can use a simple primary source, such as batteries, but often have a redundant supply for increased reliability.

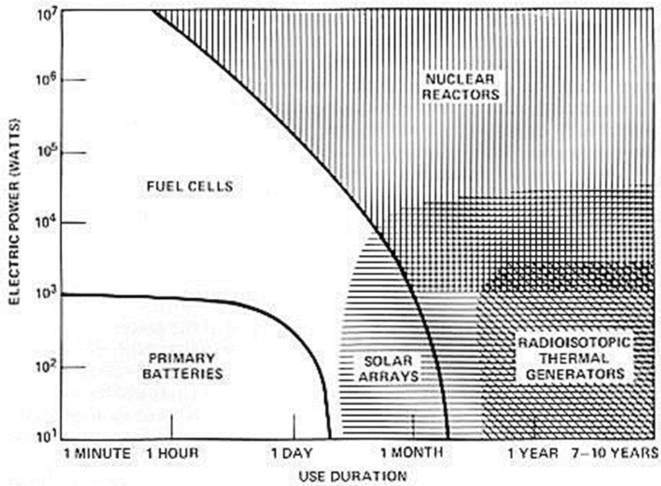
Major design requirements for the spacecraft power system:

Long term power supply for expected spacecraft lifetime, Reliable source of power while operating in space environment which includes high radiation levels and extreme temperature variations, type of orbit. Provide power for designed maximum and minimum load requirements. Payload requirements including bus voltage (unregulated, fully regulated, quasi- regulated), peak loads, redundancy. Satellite configuration, such as spinner, 3-axis stabilized, launch vehicle constraints. Area of the spacecraft affected by aerodynamic drag/torque, solar torque, Survivability, Customer (military, commercial)

Major components of a typical electrical power system for spacecraft are the primary source (which is considered to be the power generator), secondary source which is where the power is being stored and it is only needed for photovoltaic systems and it is not needed for battery supplied power, nuclear systems and fuel cells; then we have the power distribution, power regulation and control and lastly the electrical loads.



Because power systems for spacecraft must satisfy load requirements, flight duration needs, and space environment limitations, the choice of an electrical primary source is based on mission requirements and spacecraft criteria. A graph of the duration and load ranges for typical spacecraft is shown below. Note that solar/photovoltaic electrical power useful range is to the orbit of Mars but not beyond.



Solar Photovoltaic Power

Solar energy in the visible or higher energy band can mobilize electrons in conduction bands of semiconductors. This produces a current flow with a voltage equal to the energy required to overcome the band-gap energy for the electron. Typical energy values for semiconductors are 0.5 to 0.7 volts (listed as electron volts).

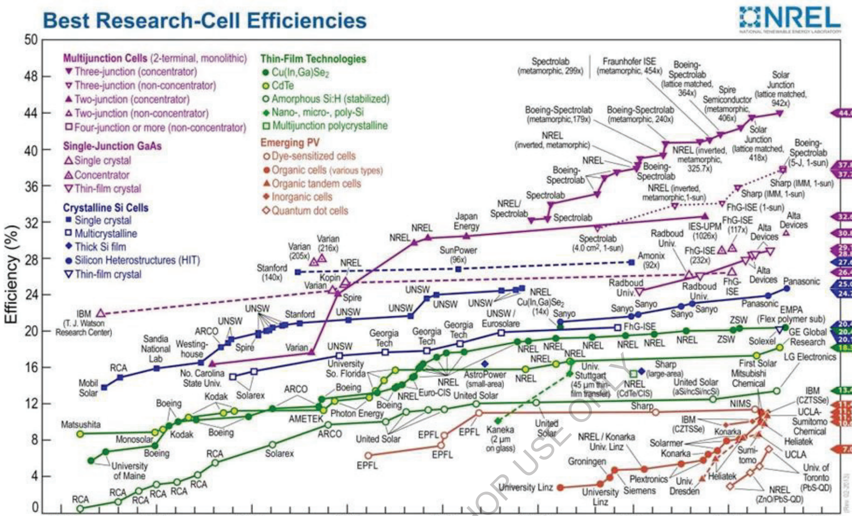
Conversion efficiencies (photon energy-to-electron energy) are approximately 10% to 15%. Silicon crystal is the most common semiconductor material, but lower in efficiency than gallium arsenide which is approximately 18% to 25%.

Silicon (Si) and gallium (Ga) photovoltaic crystalline wafers are grown in a furnace or crucible and sliced to a usable thickness to make the cell substrate.

Si cell is heated in a furnace to infuse n-silicon (negative-charge based) with p- impurities (positive-charge bases, or hole), making the n-p semiconductor available to produce a photoelectric current. Ga cells are doped (infused) with arsenic to make the n-p semiconductor for the gallium arsenide cells.

Cells are covered with a fused silica quartz glass (silicon dioxide) to reflect UV and IR energy. Back surfaces are coated with vaporized aluminum to enhance power. The cell is then attached to insulation and thermal coatings to help reduce cell temperature in space.

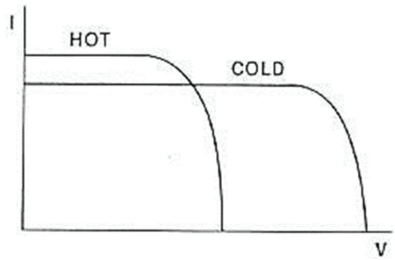
Newer amorphous or thin-film silicon and other substrates are more efficient, lower mass, and better adapted for spacecraft photovoltaic power systems (see comparative chart of photovoltaic technology below).



The largest PV arrays in space are the four modules on the International Space Station producing as much as 100 kW. The 262,400 cells that make up the 8 power channels for the ISS are 8 cm x 8 cm crystalline silicon cells folded for transport and extended in space (see image below). The ISS electrical power system is connected by about 13 kilometers of wire.

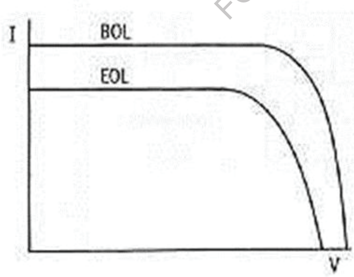
Performance & Degradation

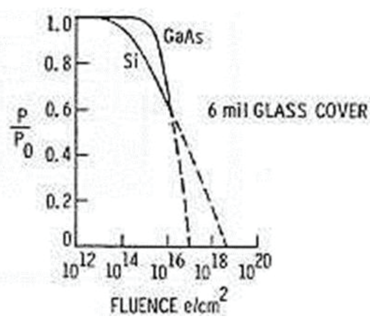
Higher energy photons (UV and X-ray) produce the same electron voltage in the cell, but deposit additional energy which heats and damages the molecular structure of the crystal, reducing its photon-to-electron conversion efficiency.



Lower energy photons (IR) do not produce mobile electrons (current) and only heat the substrate. Higher temperatures of the cell reduce power output and can reduce cell lifetime. Figure 3 shown below describes the typical decrease in power (although slight increase in current) due to increasing temperatures.

Proton and electron radiation in space reduces the lifetime of the cell. It also reduces the power output by approximately 5% per year, depending on the radiation level encountered by the PVA. This loss can be dramatic in high radiation regions, such as geostationary satellite orbits, with as much as 35% loss per year.





Voltage produced depends on the Si or Ga band-gap character (generally 0.6V per cell). Usable voltages of 10-150VDC (volts direct current) are produced in a PVA by wiring these cells in series (like batteries). Current is produced in the cell by the number of electrons made available from the number of photons striking the cell (intensity of light). An increase in current is made by wiring the cells in parallel (also like batteries). A cell's open circuit voltage (a significant load can dramatically reduce the voltage) is relatively constant and varies only slightly with temperature and aging effects.

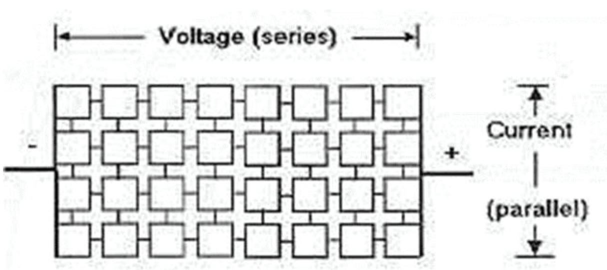
Current also varies slightly with temperature and age, but varies proportional to the light intensity. The 1/r2distance (from the Sun) decrease in available power, or energy, from the solar array is due to the decreasing number of photons which produce the current. The voltage of the array remains nearly constant, hence, the power from the array is reduced as the number of photons decrease as the PVA moves farther from the Sun.

Photovoltaic Array Voltage & Current

The voltage produced by a solar cell is typically 0.6 VDC, which represents the energy to move an electron from the potential well on the semiconductor surface. This means that if, as an example, an electrical power system requires a voltage supply of 100 V, and has 0.6 volt cells connected in series to produce a total voltage of 100 volts, it will need 100V/0.6V/cell = 167 cells connected in series.

Since the current supplied by a single solar photovoltaic cell is on the order of

0.01 amps or less, the cells must be connected in parallel to combine the electron flow (current) equivalent to the required current, in this example, 1. amps. The total number of cells in parallel would be 1.0 amp/0.01 amp/cell = 100 cells. The total array would then be 167 x 100 cells. This would develop 100 volts at 1.0 amp. This amounts to 1A x 100V = 100 Watts of power (1 Watt = 1 Volt x 1 Amp). Typical spacecraft power requirements are on the order of 1000 W to 2000 W.



Maximum power point is a design target for the greatest efficiency in a PVA. The maximum power point is shown as the point providing the largest rectangle that fits within the I-V curve as shown below. Since this is an I-V plot (current vs. voltage) the area under the curve is I times V or power. The greatest area means the most power.

The aging of a PVA will reduce the power output which must be planned for throughout the lifetime of the spacecraft mission. Thermal stress, radiation and coating degradation account for most of the aging effects. A conservative 10% per year degradation for Si cells does not account for the much greater reduction from the high radiation levels found in some Earth or Jovian orbits. The power output at the end of the mission (end-of-life, or EOL) must be used when designing the system for the minimum power level required. This is often considerably less than the beginning-of-life (BOL) value.

In addition, the temperature effects will slightly increase the PVA array size at Venus because of the higher temperature expected for the array surface (approximately 400 K). There would be a smaller array size decrease due to the lower operating temperature for the PVA at Jupiter although those temperatures are very low (100 K) and would have associated thermal problems. The actual array size will depend on the cell spacing and arrangement, the array stowage configuration, and the array deployment scheme.

Chemical Battery

The chemical battery is used in primary and secondary power systems with a number of limitations. The energy-to-mass ratio for batteries is low, the lifetimes of primary batteries are short, and the pressure & temperature limitations are usually restrictive.

However, the relatively low cost and rechargeable character of some batteries makes them useful on almost all spacecraft.

Primary supply - must have as high an energy density (high energy-to-mass ratio) as possible and cannot be rechargeable, since rechargeable batteries have a lower energy density than non-rechargeable batteries. These are used for short-term space flights, including launch vehicle subsystem power, missiles and suborbital flights.

Secondary supply - must be rechargeable and have as high an energy density as possible. Used for electrical power storage and spacecraft auxiliary power (pyrotechnic devices, space suits, etc.). For example the Extravehicular Mobility Unit (EMU) has a Li-Ion battery with 26.5 Amp hours and output voltage between 16-20 volts –sometimes referred as Long Life Battery (LLB) and it is comprised of 80-18650 lithium ion cells composed of five cells in series and sixteen cells in parallel.

The rechargeable battery is the only type of electrical energy (power) storage used on a spacecraft. The storage requirements and recharging requirements planned for the mission determine the type and size of battery used.

Battery Terminology

Total capacity - measured in amp-hours (AH). For example, 40 Amps for 1.5 hours = 60 AH

Energy storage - measured in watt-hours. This is equal to capacity times discharge voltage.

amp x volt = watt

energy = watt x second = Joule = watt-hours/3600

Specific energy density - the energy stored by a battery divided by its mass in watt-hr/kg.

Charge rate - rate at which battery can accept charge. A rule-of-thumb for charge rate is total capacity/15hr.

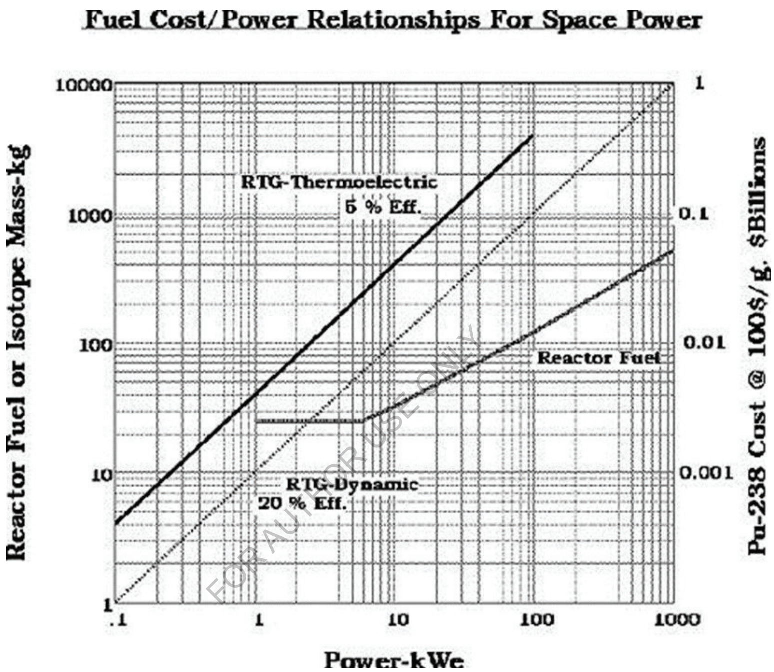
Depth of discharge (DOD) - percent of battery capacity used in discharge (75% DOD means that 25% battery capacity remains after discharge). DOD decreases with an increasing number of discharge cycles. Reasonable values for spacecraft applications are 60-80%.

Nuclear Power

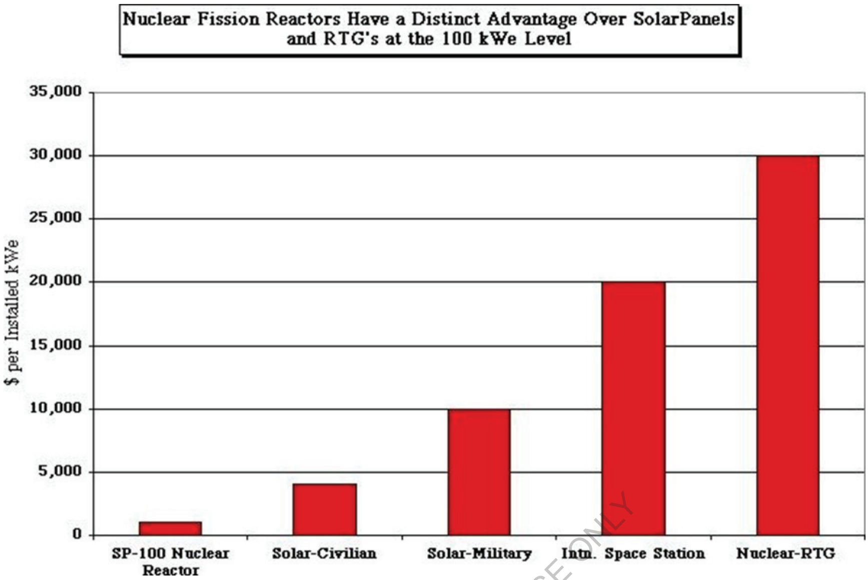
Energy released from the rearrangement in the nucleus of an atom is much greater than the chemical (electron shell) energy released in chemical reactions by a factor of 106 to 1010. The advantage of nuclear energy is obvious. Nuclear energy can come from simple energy level changes within the proton-neutron mix that results in the emission of X-rays or gamma-rays. Or, it can come from an instability in the compatible combinations of the nuclear particles which results in the emission of single particles, and/or multiple particle fragments, and X- and/or gamma-rays.

Elements that appear in the periodic table of elements are the stable varieties which retain their nucleus as long as the atom/nucleus is undisturbed by other high-energy nuclei. Should a neutron or proton, or multiple particle fragments, strike the stable atom's nucleus, the number of remaining protons and neutrons can change. That change creates an unstable nuclear combination, forming a radioactive isotope that will emit particles until it reaches a stable combination.

This fragmentation process is called fissioning. Because of the stability rules of the number of compatible neutrons and protons allowed in a single nucleus, and because of the binding energy of the collective nucleus, some radioisotopes emit only particles or fragments (usually alpha particles, which is a helium nucleus) at a statistically steady rate which reflects the level of instability of the nucleus. The change, or decay rate, is a constant, and is based on the nuclear mix of protons and neutrons. Other radioisotopes emit particles and/or fragments, but can also be excited by surrounding particles or fragments from nearby particles to ultimately emit similar particles. This type of fissioning is the key to the second type of nuclear activity that can generate far more energy per unit mass than the radioisotopes.



While radioisotopes decay at a constant rate, those radioisotopes can also be excited by other particles. If a radioisotope fragments into particles that do not initiate other nuclear reactions in nearby atoms/nuclei, the amount of energy released is a constant per emission/decay. Total energy released is the kinetic energy of the particle/fragment and gamma radiation. Thermal energy or heat is also created from the emitted particle(s) that are slowed by the "friction" in the surrounding material. These radioisotopes are useful for producing regulated heat that can be converted into electrical energy. This is the type of fuel used in the radioisotope thermoelectric generators (RTGs) that are employed in deep-space missions and in small science packages placed on the surface Moon and Mars.



Nuclear Radioisotope Power

Advantages of nuclear radioisotopes used for electrical power generation include its constant decay rate and generally long lifetime, although the decay lifetime depends on the specific radioisotope. Also, the relatively innocuous radioactive byproducts, generally low-energy alpha particles and low level gamma radiation, and the incapacity for runaway fission reactions make this power generation technique appealing for smaller space application. However, the amount of heat and the energy or heat produced per unit mass of material limit the radioisotope thermoelectric generator to roughly 10 kW. Larger nuclear power generators use fission reactors because of the inherently lower mass required. One serious disadvantage is that the radioisotope electric power generator is also an extremely expensive power source to produce.

An isotope's specific half-life (the time to decrease activity by 1/2) measures its activity per unit mass. A long-lived isotope (a half-life of 100 years or more) has a slower reaction rate, but requires more radioactive material for the same amount of heat as a higher activity isotope. A shorter half-life isotope produces more heat energy per kilogram of material but its thermal output decreases in a shorter time period. Ideal half- life values for spacecraft nuclear power applications range from approximately 10-100 years. More importantly, a radioisotope's fission product can also be selected for as low a physical and biological hazard as possible.

One relatively safe isotope to use for heat generation on spacecraft is 238Pu with a half- life of 87.7 years. Thus, this material will decrease its power output by a factor of 1- 0.5^(1/half-life) or about 0.7873%.

Nuclear power is typically used for missions beyond 2 AU because of the large surface area needed for a solar photovoltaic array at greater distances. Although expensive, the compact and long-lived nuclear power source also does not require a secondary electrical source like most photovoltaic systems. The relatively simple radioisotope power source can also be used for electrical power on robotic spacecraft operating on the Moon, Mars, Venus or even Mercury. The continual and constant power is useful for these surface planetary missions because the nights can be as long as 14 days on the Moon and 126 days on Venus. Without the need for a secondary system, a space-based electrical power system is reduced in mass, complexity, and launch costs.

Radioisotope Thermoelectric Generator (RTG)

The commercial nuclear thermoelectric power source which has been used on numerous spacecraft (Transit, Pioneer 10-11, Voyager II, Ulysses, Viking, Apollo 12-17, Galileo, Cassini, etc.) for more than 40 years has been produced by the Department of Energy for NASA.

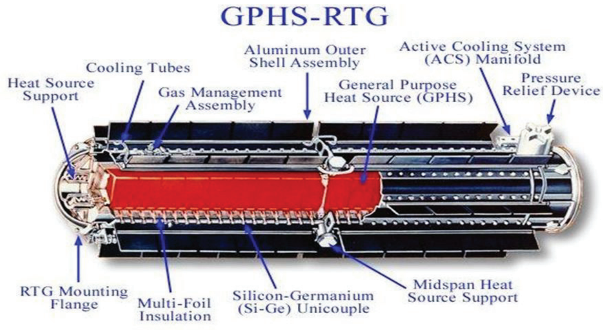
Voyager RTGs worked at about 67% by 2001 due the degradation of the bi- metallic thermocouples used to convert thermal energy into electrical energy, instead of the 83% expected, since it took about 23 years to produce them. PBOL= 470W.

Galileo RTGs decayed after sitting for about 4 years (1986-1989) because of Challenger accident.

RTGs are used in spacecraft when operating at very long distances from the Sun when sunlight and therefore the use of solar arrays is limited (beyond Mars distance, 1.52 AU)

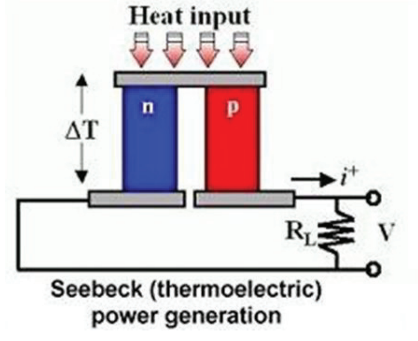
The basic units were first produced with less than 25 Watts (electric) output to as much as 500 Watts. Applications requiring more than the single unit power output of 1400 Watts employ a series of RTGs to supply sufficient power to the spacecraft's expected mission end. As an example, the Voyager I & II spacecraft had three 200 W units.

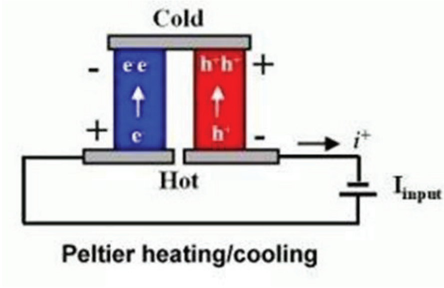
The energy conversion from heat into electricity is produced through either static or dynamic methods, with radioisotope generators producing electricity from thermoelectric or thermoionic elements. Thermoelectric conversion is used generally for lower temperatures, while thermoionic systems are useful for higher temperatures and the resulting temperature gradients.



Thermoelectric Conversion

Direct conversion of heat into electrical energy is possible with two dissimilar (bimetallic) conductors or semiconductors. Both methods utilize the Seebeck effect that certain materials produce electrical current with when exposed to a thermal gradient As with many other physical processes, the reverse is also possible; producing a hot & cold surface by applying electrical current, described by the Peltier effect. It is also possible to reverse hot and cold sides of the Peltier device by reversing current flow. As with other conversion techniques that are variations of the ideal Carnot cycle, including dynamic systems, the temperature difference between the hot and cold side of the converter will establish the maximum energy converted and the efficiency of conversion.



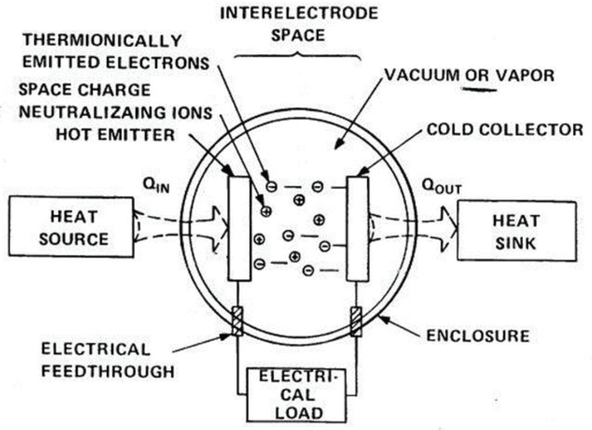


Semiconductors can be more efficient in producing electrical energy from modest temperature differences. Semiconductors used for heat-to-electrical power conversion include silicon-germanium, silver-antimony-germanium telluride, and lead telluride.

Although telluride materials have been used in most of the RTGs, they are limited in maximum operating temperature. As with other Carnot-cycle system, higher temperature differences are inherently more efficient. For the modest temperature sources used in RTGs, the thermoelectric power conversion elements are made of silicon-germanium (Angelo).

Thermoionic Conversion

Thermoionic converters are similar in function to vacuum tube diodes of the past in which electrons were "boiled" off of an emitter surface by resistive heaters then collected on a cooler conductive (collector) plate. Temperatures of 1,600-2,000 K at the emitter are needed for high currents to achieve a power conversion efficiency of 15-25%. Corresponding collector temperatures would be 800-1,000 K (Angelo).



Dynamics Thermal-to-Electric Conversion Techniques

Because the Stirling engine (25%-30% efficiency) is the simplest of the heat-to-rotary motion conversion mechanisms, and relatively efficient, it is attractive for moderate temperature electrical power systems using a sing-phase working fluid. The Sterling engine is currently being used in terrestrial power applications, but there have been no space power systems developed as yet.

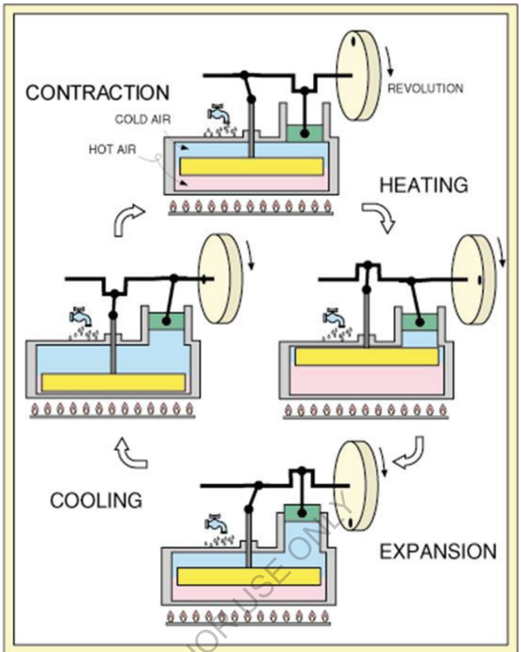
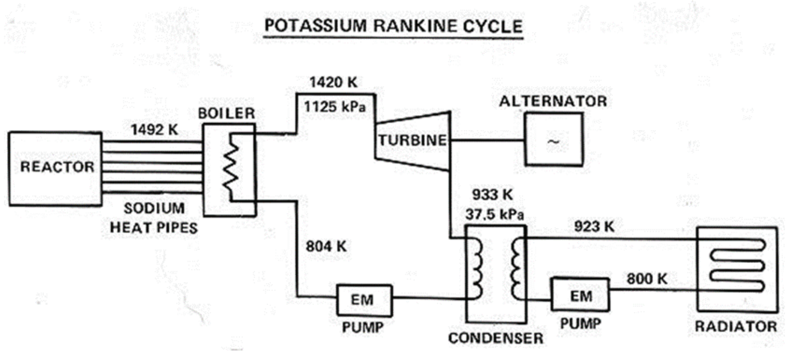


Diagram of the Sterling engine operation showing the expansion and contraction cycle within the primary chamber. The larger diffuser (yellow) is used to promote the chamber expansion/contraction by opening the chamber to heating from below or to cooling from above. The smaller piston and rod on the right drives the engine shaft in this 2-cycle operation (stirlingengine.com).

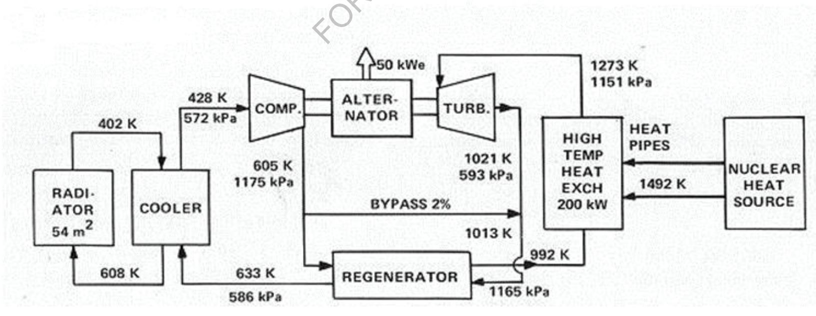
Rankine Engine Direct Heat to Rotary Motion Engine

The Rankine cycle engine (15%-20% efficiency) is a variation on the ideal Carnot cycle with a two phase liquid-gas expansion and compression, similar to the common steam engine. The expansion cycle absorbs heat, driving a turbine which, in turn, drives an electrical generator. The expanding gas cools and condenses for heat removal - both cycles being adiabatic. Water can be used for the two-phase liquid, but potassium and mercury are far more suitable at reactor temperatures. At lower temperatures, organic fluids may be a better choice.



Brayton Engine Direct Heat to Rotary Motion Engine

The Brayton cycle engine (20%-35% efficiency) is a gas cycle engine that is similar to the turbojet engine in the expansion and compression cycle and uses a single- compressible fluid system. The Brayton engine is another variation of the ideal Carnot cycle engine, but with constant pressure heating and cooling. Expanded gas drives a turbine and electrical generator, while a compressor is used to pressurize gas for greater efficiency.



Nuclear Reactor (fission power)

Nuclear reactors are controlled nuclear chain reaction devices containing fissionable radioisotopes. Their advantage is in the multiplication of energy release by neutrons, and the resulting induced fission from nearby nuclei, which, in turn, release more neutrons. While controlled fission can release 500,000 times more heat energy per mass than radioisotopes, the same chain reactions can be used for thermonuclear weapons at critically high reaction rates. Fissionable materials have extremely high heat energy content, but have accompanying high radiation hazards. Therefore, space applications of nuclear reactors are limited to large power sources with sufficient shielding to protect nearby equipment, and placed away from crew operations.

Fissionable materials contain excited, unstable nuclei that undergo spontaneous fission by releasing two particles or fragments, which include at least one neutron, but more often 2-3. Heavier nuclei that can split from the impact of slow or thermal neutrons produce a more-desirable fission process used for power reactors. These fissile isotopes include 233U,235U, and 238Pu (shorter half-time than plutonium 239 but more radioactive). Fissionable nuclei that require high-energy (fast) neutrons are better suited for fast neutron reactors or thermonuclear weapons but also power smaller reactors.

Reactor heat output is dependent on the mass and concentration of the fissile materials, as well as the neutron flow or flux. Moderating materials are used for slowing or absorbing the neutrons to regulate reaction rates, in combination with cooling the reactor core material. Although small fissionable cores may not need active thermal control, most reactors generate sufficient heat to raise internal temperatures to damaging or destructive levels. For efficiency and simplicity, temperature regulation, heat transfer, and neutron moderation may be combined by the choice of a single fluid.

Thermal Control & Neutron Moderation

As part of the reactor thermal and neutron control design, heat transport is included in the choice of heat and neutron control fluid (or gas, but that becomes difficult and more expensive for spacecraft applications). One common but potentially hazardous choice for heat transport is liquid metal coolant. These dense fluids have extremely high conductivity but retain radioactive activation by neutrons and other particles from the reactor core. Sodium, potassium mercury and lithium are attractive for their thermal conductivity, but particularly hazardous when exposed to the environment, especially if their radioactive isotopes are released accidentally. Water is a more common reactor coolant and moderator because of its modest conductivity, high neutron absorption capacity from the attached hydrogen atoms, and minimal hazards from leakage.

However, the large amount of water required and the potential for leakage makes it less attractive than some liquids, including metals, in space applications. Tradeoffs apply to coolant-moderator selection as in all other areas of reactor design.

An additional important design consideration for a power reactor is the minimization of the core mass. If the core neutrons are allowed to pass from the core to the external shielding they are lost to further core interactions. If the shielding reflects the escaping neutrons, those neutrons that would normally be lost by absorption could be returned to impact other nuclei and release more energy. Reflective properties of the shielding can be chosen from many inactive materials with high collisional cross-sectional areas and small neutron absorption cross sections. Typical materials with these thermal neutron reflective properties include graphite and beryllium. In addition, a reflective reactor core shell can be designed as an intermediate shield that can increase or decrease their exposed area by mechanically retracting or extending the cylindrical reflector covering the core. This technique is used as an effective, although not complete, thermal and neutron flux control technique in space power reactor designs.

Space Power Reactors

SNAP 2

The Space Nuclear Auxiliary Power reactor #2 project included both a development and an experimental unit. Each developed 50 kW of thermal power and 3 kW of electric power with a 9% conversion efficiency. Fuel for the reactor core was uranium-zirconium hydride (U-Zr-Hx), with a NaK liquid metal cooling system. The reactor's secondary cooling loop used mercury for heat transport through the power conversion engine.

Unshielded weight of the unit was 545 kg. Power conversion for SNAP 2 utilized a Rankine engine.

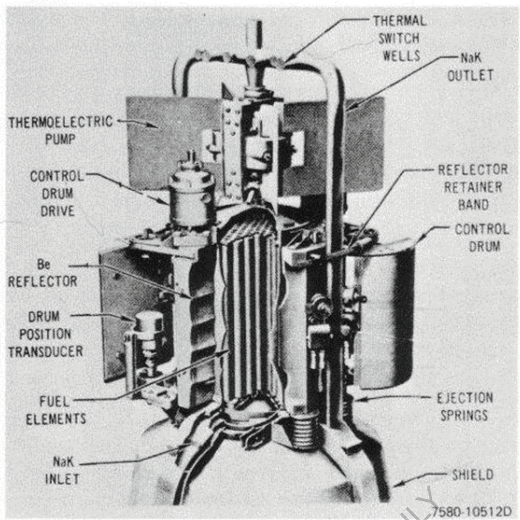
SNAP 8

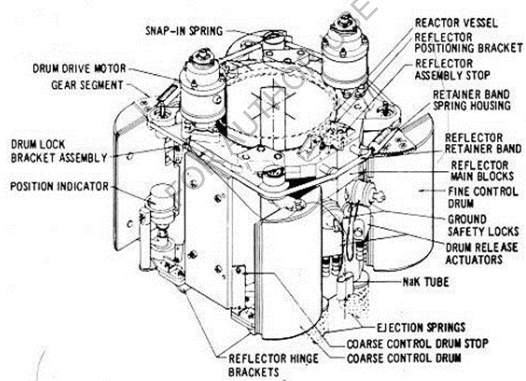
SNAP 8 was also a dual power reactor project that included both development and experiment units. Each developed 600 kW of thermal power during operation, and generated 35 kW electrical power at 8% efficiency. Like the SNAP 2 project, SNAP 8 used U-Zr-Hx fuel, NaK liquid metal cooling, a mercury secondary cooling loop, and a Rankine power conversion engine. Unshielded unit weight was 4,545 kg.

SNAP 10A

SNAP 10A was the only U. S. civil power reactor flown in space. The low-Earth orbit research tool was used to verify the fabricated design and operation parameters. SNAP 10A'sbasic design was based on the SNAP 2 technology, including 235U fissile material with zirconium hydride moderator for the basic core. Cooling and heat transport was accomplished with the same NaK liquid metal coolant loop. Power conversion for SNAP 10A differed from the SNAP 2 design, consisting of a SiGe thermoelectric array in contact with the circulating NaK cooling fluid.

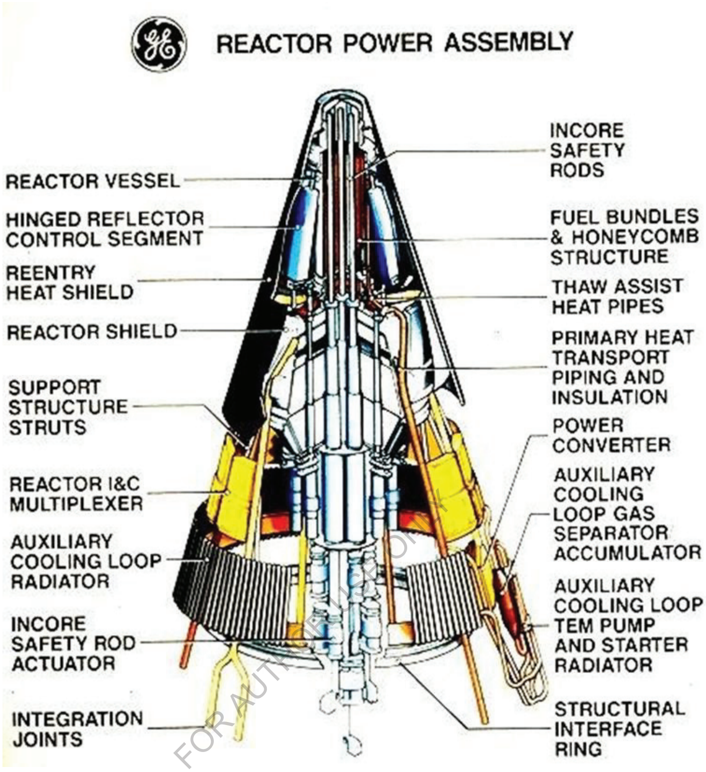
Snap 10A was successfully launched on an Atlas Agena booster on April 3, 1969 from Vandenberg, California into a near-circular 1,300 km orbit. Reactor activation consisted of subsystem starts and core reflector closure followed by critical neutron level reaching (criticality) and full power production within 12 hours. Due to an erroneous command, the SNAP 10A reactor was shut down prematurely after 43 days of continuous operation. The SNAP 10A core was mistakenly ejected into a higher orbit.





SNAP 100

Although never launched, SNAP 100 was a NASA initiative to develop powerful space reactors for future exploration missions and lunar/Mars outposts. The 100 kWe supply developed 2 MW thermal, using a lithium-cooled uranium reactor. The project which began in 1983 was terminated in 1993.



Electrodynamic Tethers

Tethers, essentially wires orbiting in space (see Figure 19), can produce electrical current by two methods. One is the generation of current by a loop passing through the Earth's magnetic field. The second is current generated by a wire floating in the Earth's electric field that uses surrounding ions in the Earth's upper atmosphere to complete the circuit. The first case of a wire loop passing through the Earth's magnetic field generates electrical energy at the expense of slowing the spacecraft. Electrical power is generated by the electromotive force as the wires are pushed through the Earth's magnetic field, like a current generator.

The force required to generate electrical power is equal to the decelerating force applied to the spacecraft in orbit. Hence, additional propulsion is needed to re-boost the spacecraft to the original orbit if the orbit is to be maintained. The second method has been tested in two flights on the Space Shuttle that confirmed electrical power can be generated in low-Earth orbit, although the method is not practical.

Fuel Cells

Fuel cells were first developed during the Gemini program because batteries could not provide the power needed since batteries had limited lifetime and could not provide the required power for a 14-day mission to the Moon. The fuel cell is a conversion process that generates electrical energy and heat from a chemical reaction from the fuels used. The type of fuel cell used on manned spacecraft is the hydrogen-oxygen fuel cell, which produces water, and heat. The reaction can occur at a very high rate with tremendous heat energy (the propulsion fuel for the SSME), or at a much lower rate using a catalyst to lower the reaction potential. The catalyst normally used is platinum. The temperature of this system is critical because of the temperature dependence of reaction rates and molecular diffusion.

The combination of cryogenic hydrogen (H2) and Oxygen (O2) generates water (H2O), heat and power.

- This power is limited by the amount of cryogenic fuel available
- The water is removed by either dumping it over or sending it to the Life Support System
- The heat is dissipated using the thermal control system.

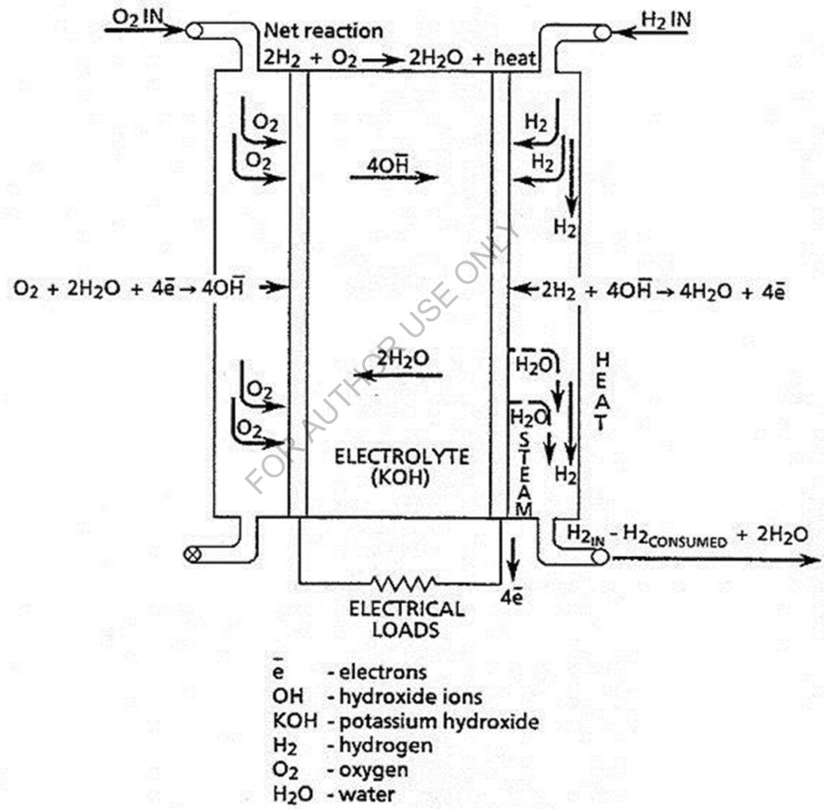
Space Shuttle Orbiter Fuel Cell Operation

Conceptual operation of the fuel cell begins as the reactants enter the fuel cells by flowing through a preheater where the cryogenic reactants are heated to 40°F or more. Heat for the preheater comes from the Freon cooling loop, as does the heat to increase the cryogenic oxygen temperature of the cabin breathing oxygen. Pressure of the cell reactants is reduced to approximately 135-150 psi after filtration. A second regulator drops the oxygen gas pressure to 60-62 psia and the hydrogen pressure to a value 4.5 to 6 psid (differential) below the oxygen pressure. Coolant pressure is also regulated with any decrease in reactant pressure triggering a drop in the coolant pressure to prevent deformation of the stack plates (NASA-EPS 2102).

Hydrogen gas is mixed with recirculated water vapor and exhaust from the cell stack. The hydrogen-water vapor mix enters a condenser where the saturated water vapor is cooled to remove water droplets which are then removed with the hydrogen pump/water separator. The hydrogen-water vapor gas is then routed to the cell plates to the anode interface (see Figure 20).

At the same time, oxygen gas enters the cell plates and is infused through the cathode to combine with water and returning electrons to produce hydroxyl ions (OH-). The hydroxyl ions then migrate through the conductive electrolyte (KOH and H2O) towards the anode.

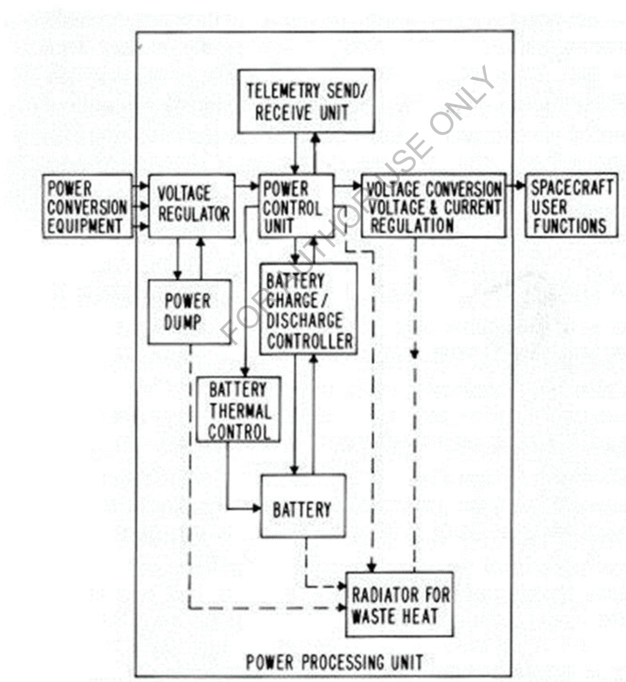
As the hydrogen gas and water infuse through the anode to the electrolyte, the hydrogen combined with the hydroxyl ions generate water plus electrons (see reaction equations in Figure 20 below) and heat.



Power Management: Regulation, Distribution and Control

The electrical power (5%-10% of the spacecraft dry mass is cabling) used by the spacecraft systems is produced by a primary system that has a variable output, and has a different operating current and voltage than is required by the spacecraft loads.

Regulation, control, and distribution circuitry are used to solve the problem of differing power source and load needs throughout the spacecraft. The regulation and control system must also satisfy the charging and discharging limitations of the secondary storage batteries, and overload protection for the loads, the primary supply, and the secondary storage.



In addition to the voltage and current being stepped up or down to satisfy the load requirements, alternating current is often needed for some of the components.

This DC to AC conversion is generally accomplished with an inverter through oscillators or solid state converters. The conversion loss can be large (40-50%) on terrestrial applications, but are generally less than 10% for spacecraft inverters.

DC-to-DC conversion is also required for some spacecraft components and the primary supply. This conversion must be efficient to keep heat loads small and reduce power consumption.

The bus voltage of spacecraft is usually DC since the spacecraft generates DC (Shuttle used to have three-phase 400 Hz AC). Low bus voltage (typically 20-50 V but some spacecraft are designed to operate in 100-120 V) require high currents and therefore high losses).

The bus voltage control can be:

Unregulated: The load bus voltage is the voltage of the batteries, used when there are significant variation of the load bus voltage.

Fully regulated: It is not efficient and will work on spacecraft with low power needs and highly regulated bus.

Quasi-regulated: It uses a battery charger in series with battery to regulate the bus voltage when the battery is charged (individually or parallel) but not when the battery is discharged.

Chapter 8: Guidance, Navigation & Control

Guidance, navigation and control (GN&C) systems include the onboard spacecraft and ground elements that are used to guide, orient, and change the spacecraft position according to the mission requirements. Orbital maintenance, orbit transfer, orbit precession, station keeping, pointing, midcourse corrections, tracking, and many other spacecraft operations require navigation and control systems. Those requirements may also include various errors in the trajectory or orbit during the mission. Since those errors or corrections could be made autonomously, or by ground controllers, most spacecraft have the capability of onboard control calculations and ground-based control. It is assumed in these discussions that the spacecraft or satellite is a simple rigid body.

Guidance - spacecraft trajectory control during the thrust phase

Navigation - determination of spacecraft position and velocity relative to a specified reference frame

Control - spacecraft attitude (orientation) and trajectory control

Orientation control - the adjustment or control of the spacecraft attitude or orientation

Position (thrust vector) control - the adjustment of a spacecraft orbit or trajectory. This requires onboard thrusters or atmospheric drag (aerobraking or the orbital decay in a planetary atmosphere).

Stabilization - a spacecraft requires a fixed orientation in the absence of controlling forces to keep its orientation from constantly drifting. The stabilization can be active or passive, and usually requires a spinning body (spacecraft spin or spinning gyros) which produces angular momentum. Gravitational gradient stabilization is an exception.

Attitude error - the "low frequency" component of spacecraft misalignment that is corrected with an accurate attitude control system.

Attitude jitter - the "high frequency" component of spacecraft misalignment that is not correctable by the attitude control system but can be reduced by proper design.

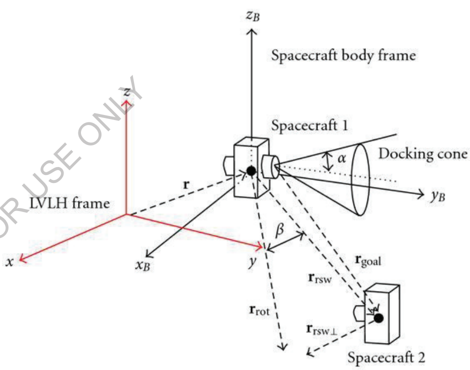
Absolute reference - position or motion with respect to a known coordinate reference

Relative reference - position or motion with respect to a previous measurement

Two basic areas of spacecraft guidance and control are attitude & trajectory determination and attitude & trajectory control. The subsystems and components that make up an autonomous spacecraft guidance and control system can be described as the attitude/trajectory determination and the attitude/trajectory control components. However, nearly all spacecraft missions are commanded or corrected by Earth-based control for changes in the mission program, or to overcome errors that may occur during the launch or flight phases. These ground-based spacecraft control elements provide the versatility and modifications necessary for the many variables that occur during a spacecraft's operational lifetime.

Attitude, Trajectory Determination

The determination of spacecraft attitude and/or position (within a specified tolerance) with respect to a specified reference frame. The reference frame is a coordinate system that could be a ground station position, the Earth's center, ecliptic (Earth orbit) coordinates, heliocentric coordinates, or others. The attitude/trajectory determination may also be in reference to other targets (Mars, for example, if traveling to that planet).



Requirements

The accuracy and precision of the attitude or position measurement and resulting position and/or orientation calculations must be sufficient for the attitude control element to meet the pointing (orientation) and path (orbit or trajectory) accuracy requirements.

Sensors

Sensors are used to measure spacecraft position, attitude, and/or motion. The sensors are actually detectors that measure force, electromagnetic radiation or magnetic fields. The sensor data can be converted into position and/or motion information by using a computer to calculate or compare spacecraft data with reference data (star positions, solar system coordinates, relative position, etc.).

Acceleration Sensor Examples

Accelerometer (inertial or linear motion) sensors

The force of motion (or gravity) is measured and converted, by simple time integration, into velocity and position by an onboard computer. Three accelerometers are required in order to calculate position and velocity in three dimensions. This type of inertial unit measures straight-line motion (the same as the spacecraft and aircraft inertial navigation system).

Gyroscopic rate (rotational motion) detector

A 3-axis acceleration sensor is similar to the accelerometer sensors, but uses a gyroscope for each rotational change. Acceleration is measured by the change between a fixed gyroscopic orientation and a moving (rotating) spacecraft. This unit measures rotational motion. Accelerometers can also be attached to measure acceleration in the inertial (fixed) frame.

Laser rate and motion detectors

The laser gyro uses an interference pattern produced by a split-beam laser and Doppler shifting to calculate either rotational or linear (straight-line) motion. Three of these laser gyros are used to provide very accurate 3-dimensional position and velocity information. Laser ring-gyros are also available that measure both linear and rotational motion.

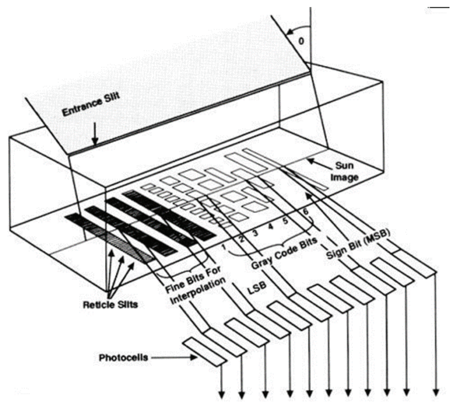
Optical Sensors

Star tracker (scanner)

A three-telescope unit (3 telescopes provide 3-dimensional data) is used for star or planet image identification, providing an accurate celestial reference. This requires a table of star and planet positions (ephemeris) and a computing system for star and/or planet identification and for position calculations. The system has high accuracy and solid state design but is costly and relatively heavy. It is also sensitive to the Sun, the Moon, the Earth and stray objects in the field of view.

Horizon (limb) sensor

This imaging and computing system provides a relative position with respect to the Earth, the Moon, the Sun or a relatively large, nearby body.



Position identification resembles a cone with the thickness representing the error of measurement. A second independent system is required for locating (intersecting) the spacecraft on that cone. If no information is available on the approximate position of the spacecraft, a third system or sensor measurement is required to solve the position equation.

Sun sensor

The Sun's location can be measured from a series of slots and light sensors. The sensor exposure pattern (data) is then related to the Sun's position.

Magnetic field detector (magnetometer)

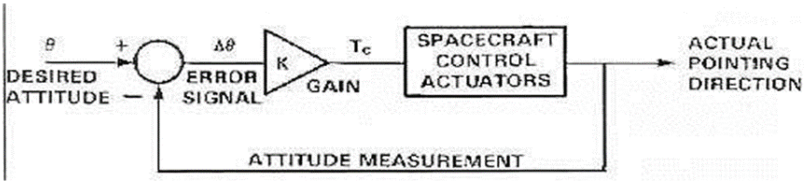
Planetary and solar magnetic field measurements by the spacecraft sensors can provide low to moderate accuracy position and/or attitude data if the magnetic field is well known.

Earlier manned space flight navigational systems beginning with the Mercury program and including Gemini and Apollo capsules, used optical terrain recognition instruments to supplement position information. Earth-reference and lunar- reference optical systems were advantageous to these capsules because of the landing reference required for alignment, relative position, and absolute reference.

This type of optical recognition device also on the Space Shuttle Orbiters, but later used for rendezvous rather than position reference because of the GPS addition to the Orbiter's GN&C system. The Global Positioning System provides very accurate position, velocity and acceleration data for spacecraft in LEO, including the ISS.

Attitude Trajectory Determination Output

The spacecraft position data from the attitude determination system is calculated and compared to spacecraft position data by an onboard computer. The difference in programmed and actual positions is used to calculate the necessary correction(s) to the position, if the difference is greater than allowed errors. This correction process is a feedback loop that requires defined precision in the sensors, positional calculations, control commands and correction estimates.



Attitude Trajectory Control

This is the process and control of the spacecraft position and/or attitude to provide spacecraft stabilization and pointing within a specified tolerance.

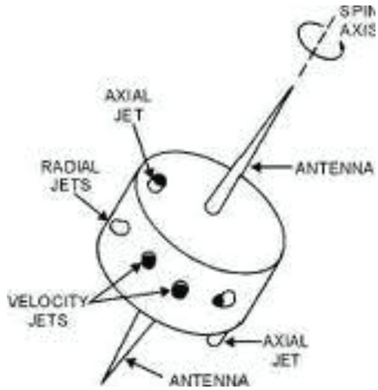
Torque or Force Control Subsystems

These provide the forces to adjust the spacecraft attitude or position. Since the control force is rarely in line with the center of mass, the control force produces a torque (force times perpendicular distance) which can rotate the spacecraft if not accounted for in the design and control program.

Spin Stabilization

This might also be considered to be a passive system, although the adjustment or control required for an accurate spacecraft spin alignment requires an active control system.

Simple spin



Spacecraft spin will provide a fixed axis of orientation for attitude stability. Spin rate for most spacecraft/satellites needs to be only a few rpm because of a lack of large external forces. Smaller (perturbing) forces, such as the non-spherical Earth or the Moon, can be, and generally are, a problem over a long period (weeks to years) and require correction. Most spin stabilized spacecraft require a nutation damper to reduce wobble (from unsymmetrical spacecraft) and torquers for precession control (from external torques or forces such as torque on a satellite due to the Earth's oblate shape, for example).

Dual-spun

Stabilization by the angular momentum produced from spinning is improved if part of the spacecraft is spun in the opposite direction. If both sections are spun at the same rate but in opposite directions, one section will be despun and provide a fixed and stable platform for sensors, antennas, etc. Mass symmetry along the long axis of the spacecraft is important. This configuration also requires a nutation damper and precession control.

Three-Axis Control

Spacecraft orientation and position controlled in all three dimensions by three-axis control systems that include thrusters and gyroscopic control. Other systems also have 3- dimensional control but are classified separately.

Thrusters

Three-axis cold gas or liquid fuel (mono- or bi-propellant) thruster system. These actually produce torque since the forces are not exactly along the three primary spacecraft axes. Accurate models of the resulting motion must be written into the computer code used for attitude control. Low thrust employed for most spacecraft.

Cold gas - simple, inexpensive, well developed and commonly used on smaller spacecraft & satellites.

Monopropellant - simple, efficient, relatively inexpensive. Used on moderate mass spacecraft.

Bipropellant - more complex, expensive and efficient (higher Isp than cold gas and monopropellant systems).

Gyroscopic Control

Reaction control wheels

Spacecraft control in this system is provided from slowly rotating gyroscopic wheels. The attitude changes result from the reaction force applied to spin up or slow down each of the 3 wheels. Redundance provided with fourth wheel, although this system is already heavy and requires high power.

Momentum wheels

This unit provides both 3-axis stabilization from three perpendicular gyroscopic wheels and attitude control from applied reaction forces. This advantage comes from the momentum wheels' rapid rotation, which produces significant angular momentum stability.

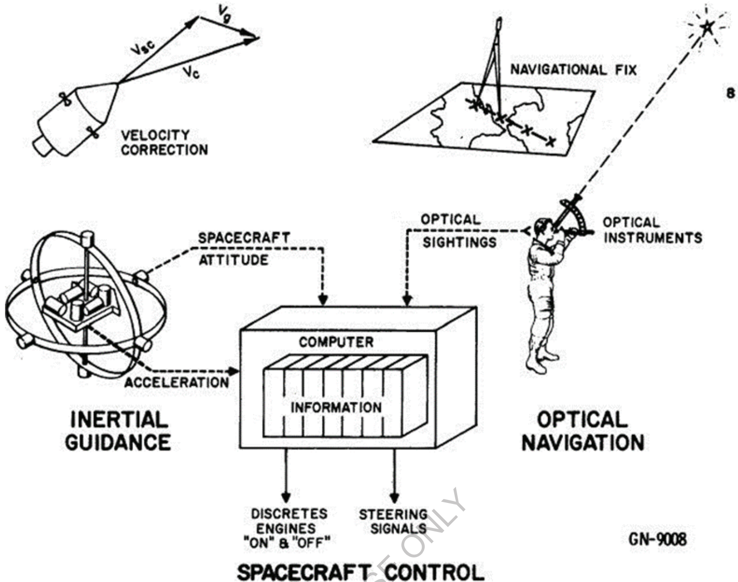
Control moment gyros

This consists of gimbaled momentum wheels that have an applied force to change orientation of the platform. This is heavier and less accurate than the momentum wheels, but has a much greater control torque on the spacecraft. Noise produced by system makes it more useful for larger spacecraft.

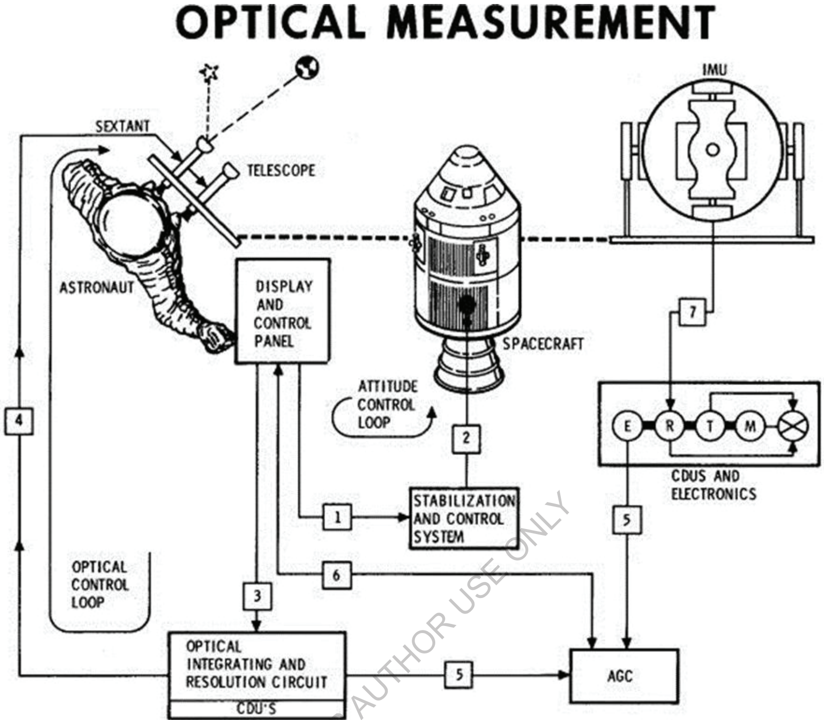
Apollo Mission Guidance, Navigation & Control Example

Manned spacecraft guidance and navigation evolved from simple onboard inertial measurements and radio guidance to much more sophisticated and complex systems, but still include the inertial measurement devices because of their inherent accuracy. Mercury, Gemini, Apollo, Skylab, STS, and ISS employ gyroscopic inertial measurements. However, today's rate gyros employ laser interferometers, or ring laser gyros, to offer much greater accuracy and less drift (accumulated) error.

Apollo spacecraft GN&C systems were well advanced beyond the Mercury and Gemini systems. Digital computers and communications, although not first flown on Apollo, were developed for the greater complexity of more distant navigation, as well as the more elaborate rendezvous and lunar landing requirements. Two primary types of instruments were used for Apollo's guidance and navigation; Optical and inertial. That produced absolute reference and accurate position changes respectively. A sketch of the multiple GN elements are shown below.



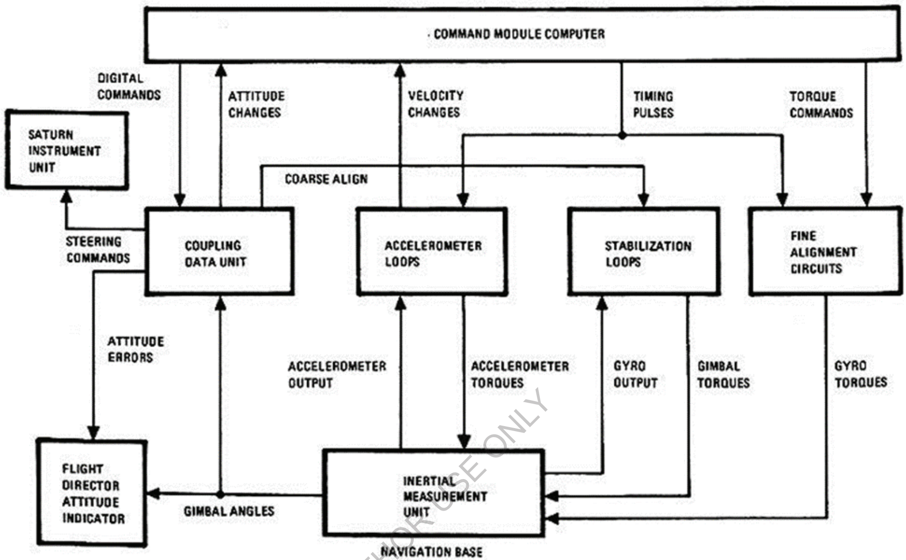
V_{sc}
V_g
V_c
VELOCITY
CORRECTION
NAVIGATIONAL FIX
8
SPACECRAFT
ATTITUDE
OPTICAL
SIGHTINGS
OPTICAL
INSTRUMENTS
COMPUTER
INFORMATION
ACCELERATION
INERTIAL
GUIDANCE
OPTICAL
NAVIGATION
DISCRETES
ENGINES
"ON" & "OFF"
STEERING
SIGNALS
GN-9008
SPACECRAFT CONTROL



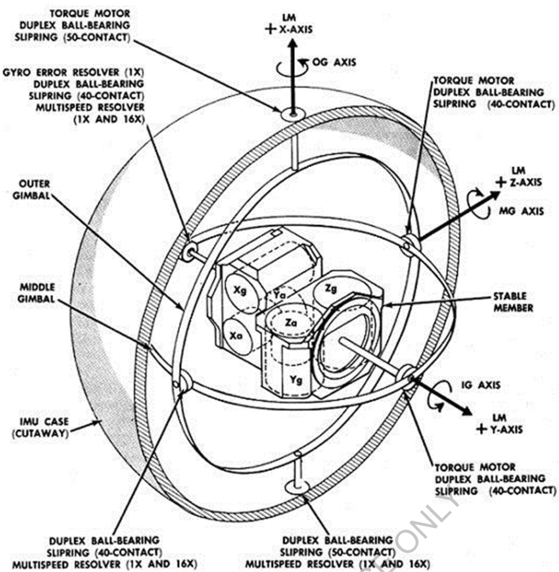
Apollo's optical instruments were employed in establishing absolute positional reference and included the telescope and star tracker, the sextant which was used for surface reference, and a horizon sensor (not shown). Absolute reference data was used for navigational reference and periodic alignment of the accurate inertial measuring units (IMUs).

Although the IMUs were accurate, slight friction in the gyro assemblies generated continual drift in the reference data.

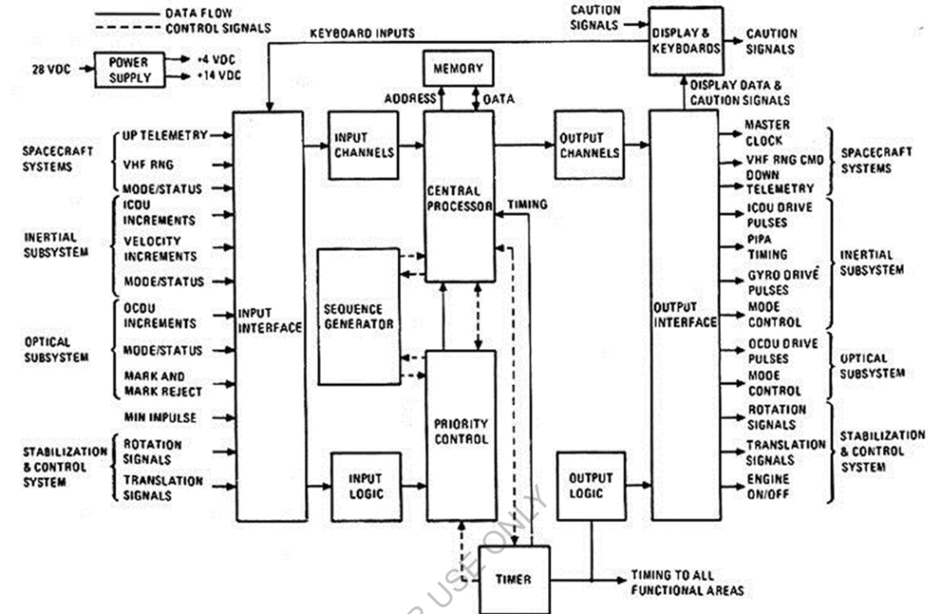
A diagram of the Apollo Command Module command and data flow for the GN&C subsystems is shown below. This unit includes the 3-dimensional gyroscopic inertial measuring unit (3-axis gimbaled accelerometers).



At the heart of the guidance inputs for the Apollo GN&C is the inertial measuring unit, the IMU. The IMU inertial, or fixed in space by a gimbal assembly, contained three accelerometers for each of the three dimensional coordinates. With the fixed (inertial) reference, the accelerometer data is provided in constant X, Y and Z values. Three-axis accelerometers directly attached to the spacecraft would also provide acceleration data in the three axes, but would be varying in reference coordinates with the spacecraft rotation. The three-axis gimbal assembly shown below includes dual accelerometers for each axis, and torque or rotation motors to align the reference orientation of the accelerometer platform to a suitable reference. As an illustration, a very useful coordinate reference for the Apollo missions was the landing site coordinates for the Lunar Module landings.

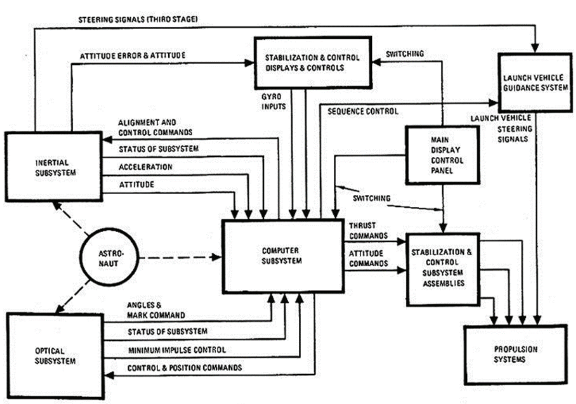


Central to the Apollo CM GN&C is the onboard digital computer. Its functions and processing power, minuscule by today's standards, managed the entire mission, allowing control inputs from programmed data, from astronauts, and from ground control instructions. A diagram of the Apollo Command Module computer and its functional interface is shown below.

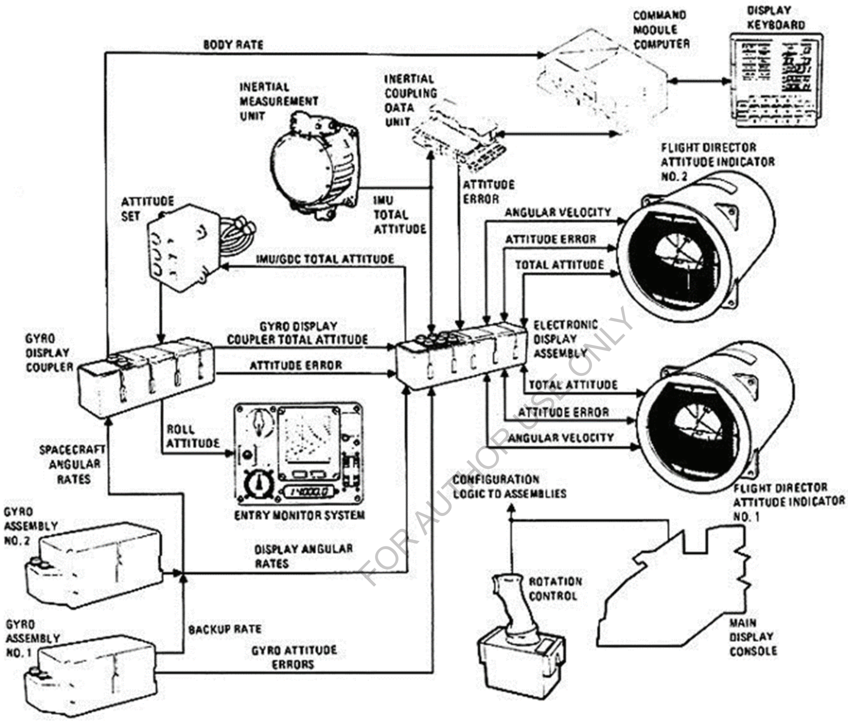


P-264

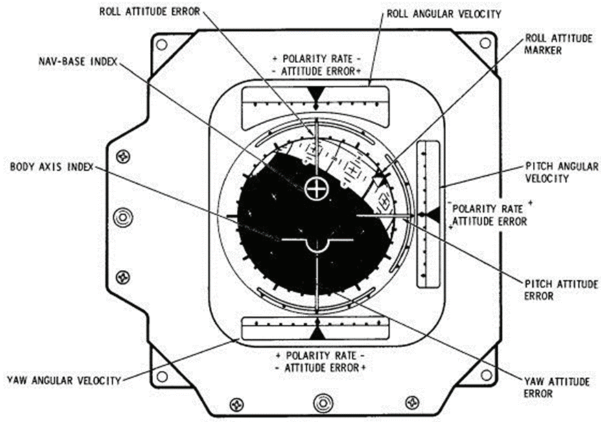
Command Module control and stabilization included inputs not only from the astronauts, ground control commands, and computed data, but also incorporated optical sighting data input for navigation. A schematic of the data flow in the Apollo CM GN&C command system is shown below.



Integrated elements of the Apollo CM GN&C system include determination and control components and reference and display units for astronaut flight operations. An aircraft-type attitude indicator/flight director is included as part of the display output, as shown below.

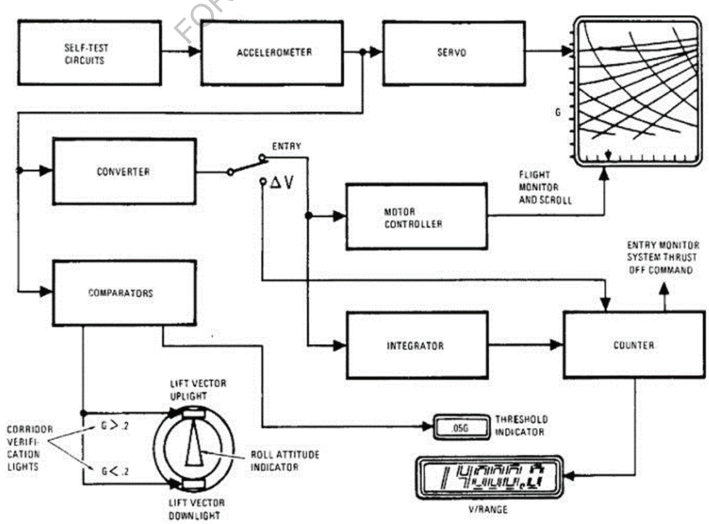


FLIGHT DIRECTOR ATTITUDE INDICATOR



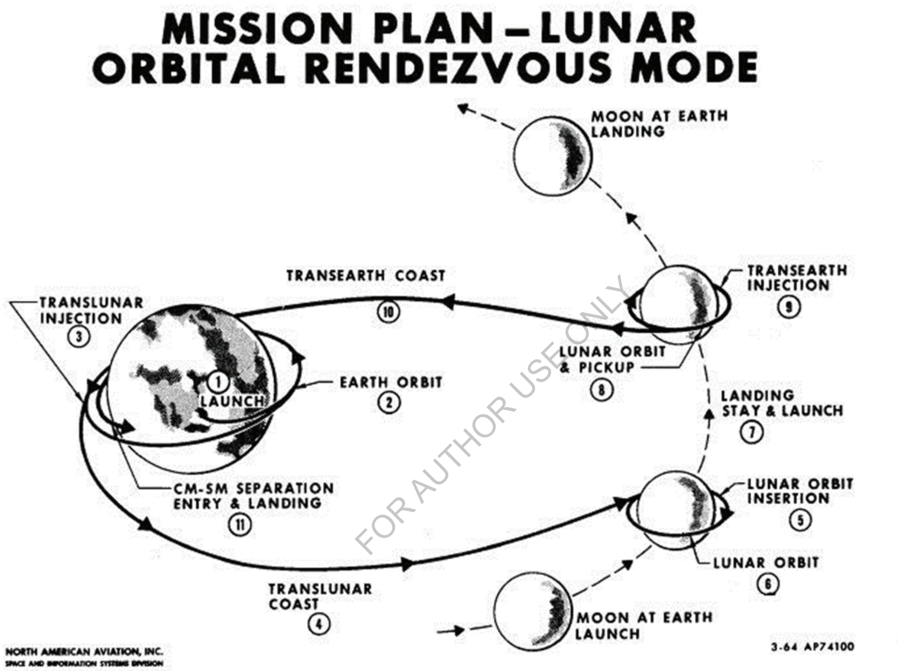
NOTES:
1. ATTITUDE ERROR - ATTITUDE DESIRED - ACTUAL ATTITUDE.
2. THE BALL IS OF THE INSIDE-OUT CONVENTION.
3. EULER ANGLE CONVENTION IS PITCH, YAW, ROLL.
4. THE BALL ATTITUDE SHOWN IS PITCH 345°, YAW 335°, AND ROLL 300°, WITH RESPECT TO THE NAVIGATION BASE INDEX.

In addition to the flight navigation systems used to get to and from the Moon, the Apollo CM guidance and navigation system included an atmosphere entry monitor that displayed the computed and actual trajectory of the entry. This monitor system allowed manual operation of the attitude and thrust control system in case of a guidance malfunction.



Apollo's Major Guidance, Navigation, & Control Maneuvers

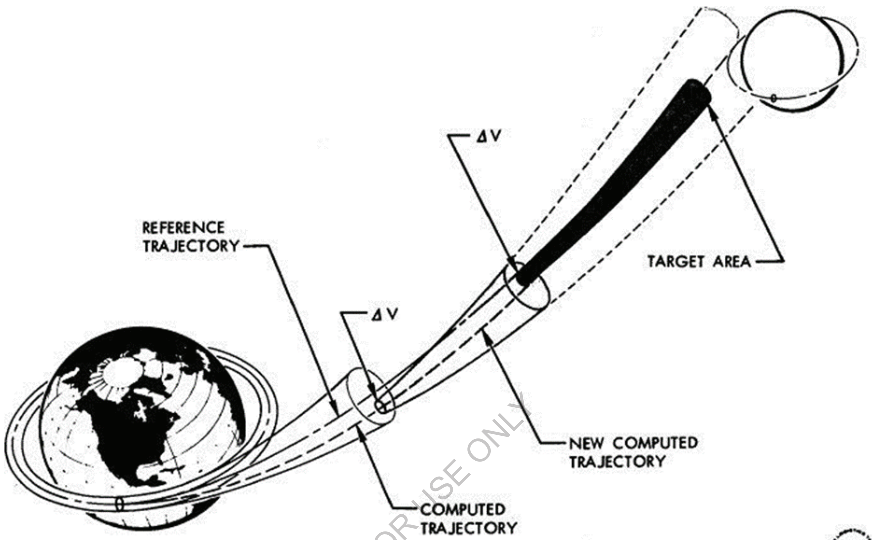
The Apollo lunar flight missions required extraordinary accuracy for the technology and time, and required a number of complex and challenging maneuvers. These included the tanslunar boost and lunar intercept, lunar orbit, lunar descent and ascent, transearth return boost and Earth intercept, and the critical atmospheric entry. A graphic summary of the flight operations is shown below.

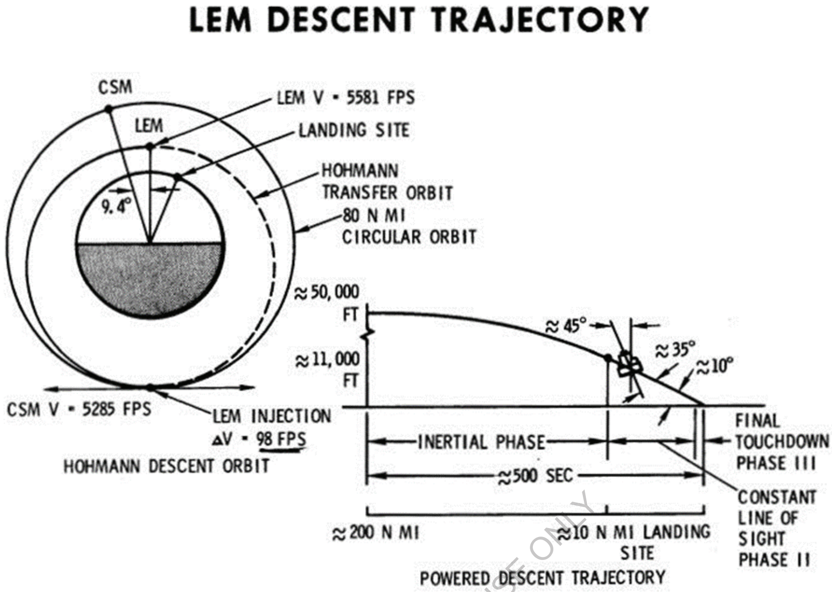


After low-Earth orbit entry and preparation for departure, the translunar boost was set up for the lunar rendezvous. Accuracy was not sufficient to arrive at a precise position for lunar orbit on the initial boost, so midcourse trajectory corrections were included in the flight operations.

Inertial reference within the GN system was switched to space reference from Earth reference in the departure flight. A diagram of the flight elements of the lunar rendezvous, landing and return is shown below.

MIDCOURSE POSITION & TRAJECTORY DETERMINATION

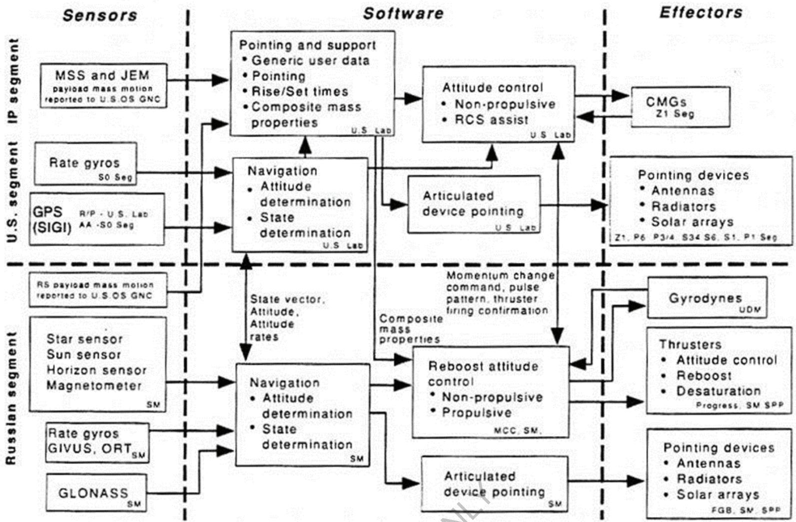




International Space Station

The International Space Station illustrates a useful example of a GN&C system because of its size and complexity. The ISS GN&C combines the attitude & trajectory determination components of both the U.S. and the Russian elements, providing greater functional utility and increased accuracy. Pointing capability is provided for the U.S. and Russian requirements that drive the antennas, PVAs, radiators, and instruments, as well as for antennae and experiments for the international partners including Japan and Canada.

A diagram of the GN&C determination and control components in the figure below shows a broad interrelation of both U.S. and Russian determination sensors in providing input to the navigational and pointing software that drives both U.S. and Russian control components.



Spacecraft Torques and Disturbances

Numerous influences on an orbital spacecraft generate orbital trajectories that are not precisely repeated. For example, the Earth's non spherical shape, an oblate spheroid, produces a precessional force on the orbit plane that depends on the inclination and altitude. Lunar and solar gravitational influences also change orbital positions and alignment over both long and short periods, making orbital maintenance more practice than theory. Other external forces on spacecraft produce torques and instabilities that influence spacecraft trajectory and attitude over short and long periods. These forces may be small or negligible, and some may require active control systems to provide a stable platform. A list of the more common torques is given below.

Chapter 9: Spacecraft Remote Sensing

Introduction to Spacecraft Remote Sensing

Spacecraft Remote Sensing is the practice of capturing, monitoring, and recognizing (or sensing) distant or remote things or occurrences is known as remote sensing. In a more narrow sense, the term "remote sensing" refers to the practice and technology of gathering data using instruments onboard aerial (aircraft, balloons), spaceborne (satellites, space shuttles), or both platforms, about the Earth's surface (land and ocean), and atmosphere. According on its application, remote sensing can be divided into:

- When satellite systems are employed, satellite remote sensing
- When images are taken to record visible light, this is known as photography and photogrammetry.
- Using the thermal infrared section of the spectrum for remote thermal sensing
- When using electromagnetic frequencies, radar remote sensing
- When optical signals are sent toward the ground and the range between the sensor and the ground is calculated using the pulses' return times, this process is known as lidar remote sensing.
- Today, remotely sensed data is combined with other cutting-edge geographic information like Google Earth, mobile mapping, and geographic information systems.

The functionality of Remote Sensing

Our five senses help us to understand the environment around us. Our sensory organs must come into contact with things to use sensations (including touch and taste). Nevertheless, receiving a significant amount of data about our environment through our senses of hearing and sight does not necessitate tight interaction between our internal organs and the outside world. In this way, we constantly engage in remote sensing. In general, the term "remote sensing" describes actions that involve capturing, monitoring, and perceiving (sensing) objects or happenings in distant (remote) locations. The sensors (for instance, specialized kinds of cameras and digital scanners) can be mounted in aircraft, satellites, or space shuttles to take images of the objects or events on the Earth's surface.

Various aspects of Digital Imaging using Spacecraft Remote Sensing

The Earth's surface characteristics or the environment both produce energy "signals," which are detected and recorded by all sensors (sensing systems). Video recorders and aerial cameras are two common examples of remote sensing devices. Examples include electronic scanning, linear/area arrays, laser scanning systems, and other more complicated sensing devices. These remote sensing devices can record data in conventional or digital representations, such as a matrix of "intensity value" corresponding to the mean illumination (energy) observed within an image pixel, or they can record data in analog format, such as hardcopy aerial photography or video footage. Digital remote-sensing photos may be used immediately in computers, and analog data can be scanned into digital form.

Regardless of whether they are analog or digital, everyday cameras are distinct from optoelectronic sensors. The Earth's surface may be seen from above using remote sensors carried by aircraft, satellites, space shuttles, and other platforms. This synoptic perspective allows us to concurrently analyze and comprehend items or events over a wide region and ascertain their spatial connections. On the other hand, regular cameras can only show an item or event from the side or the ground. In actuality, the majority of remote sensing satellites are outfitted with sensors that are Earth-facing. As they circle the planet in regular orbits, they serve as the "eyes in the sky" and continuously monitor it.

Such natural phenomena as growth, climate, wild fires, volcanoes, and others change continuously over time. In pollution monitoring, the revisiting time is an essential factor. For instance, vegetation changes are best detected using anniversary or close-to-anniversary photos, reflecting the seasonal and yearly phenological cycles that govern plant growth. The re-visit time of many weather detectors is substantially shorter: The NOAA AVHRR (National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer) local area coverage is every half-day, the Geostationary Operational Environmental Satellite (GOES) is every 30 minutes, and Meteosat Second Generation is every 15 minutes.

The basic principles of the Electromagnetic Radiation Spectra

Solar energy, sometimes called electromagnetic waves, is the primary energy source for remotely sensed data. The fundamental component of electromagnetic radiation, the photon, may travel both as a particle and as a wave with various frequencies and wavelengths. electromagnetic radiation is reflected, transmitted, or absorbed as it reaches the Earth's surface and makes contact. The amount of solar radiation that is reflected, absorbed, or transmitted changes with wavelength for any particular substance or surface. This crucial characteristic of substance is the foundation of remote sensing technology and allows for identifying and separating various compounds found on the Earth's surface. A kind of energy with wave-like characteristics, electromagnetic radiation mostly comes from the sun. The electromagnetic spectrum is the collection of wavelengths of solar radiation moving at the speed of light (C, or 3×10^8 ms^{-1}). The waves oscillate in all directions, transverse to their path of passage, and spread across time and space like sea waves.

Any substance on the surface of our planet will reflect or emit the varying numbers of photons at specific wavelengths when it interacts with external electromagnetic energy or is activated by internal processes. A remote sensor's photon energy is often expressed in power units like Watts per square meter per wavelength unit. The plot of power variation vs wavelength produces a distinct pattern that may be used to identify the substance being detected.

Wavelength and frequency are two parameters used to describe electromagnetic waves. The distance between each wave's crest is measured in wavelength. The frequency () is the quantity of oscillations that are finished each second. In order to more accurately characterize electromagnetic waves, other words like crest, trough, amplitude, and period are also required. The maximum height of a wave is its crest, while its lowest point is its dip. The distance between a wave's midway and crest or trough is referred to as the magnitude, which is another crucial element in the wave definition. Period is the time it takes two crests to pass each other at a still point.

Various Aspects of Remote Sensing Data Reduction

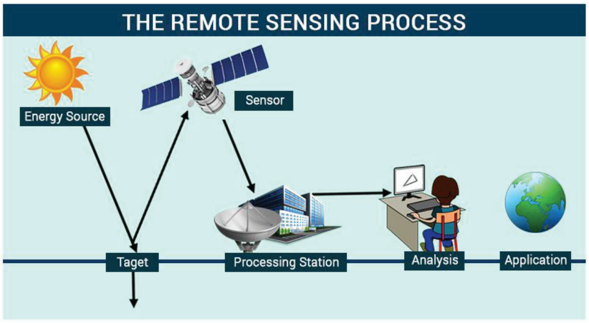
We must educate ourselves on the several types of sensors used to collect this data and their purposes and capabilities to comprehend the properties of remotely sensed data. In this chapter, we divide remote sensors into two major categories: analog and digital. As an illustration of an analog image, aerial images clearly have the benefit of capturing excellent spatial information. Digital imaging like satellite imagery typically has improved spectral, radiometric, and temporal resolution. Improved resolution of each type has been the trend over the past 40 years. A lot of work has gone into designing and building sensors that have the highest resolution possible for the jobs they are meant to do.

Greater spatial resolution makes it possible to identify ever-smaller individual targets—less than 1 m in size. Since individual entities' spectral fingerprints may be more clearly differentiated, as with hyperspectral sensors, identification of individual entities (features, objects) is improved with higher spectral resolution. Sharper pictures produced by higher radiometric resolution enable greater separation of various target materials from their backgrounds. Finally, increased temporal resolution has made it possible to use temporal analysis to observe the movement of ships and other objects in addition to monitoring ecosystems and natural hazards.

There are often apparent trade-offs among various types of resolutions. For instance, in conventional photographic emulsions, advances in spatial resolution are based on smaller film grains, which result in corresponding reductions in radiometric precision, or the ability to depict a broader range of brightness values. A lower IFOV is needed for higher spatial resolution in multi-spectral scanning systems, which results in less energy reaching the sensor. In order to counteract this impact, one can either lower wavelength range by enlarging the spectral window to pass more energy or decrease radiometric precision by splitting the energy into fewer light conditions. According to reports, NASA and the European Space Agency have suggested developing remote sensing systems that can deliver data incorporating high spatial resolution and revisit capabilities, among other things, to resolve the contradiction between spatial and temporal sensitivities.

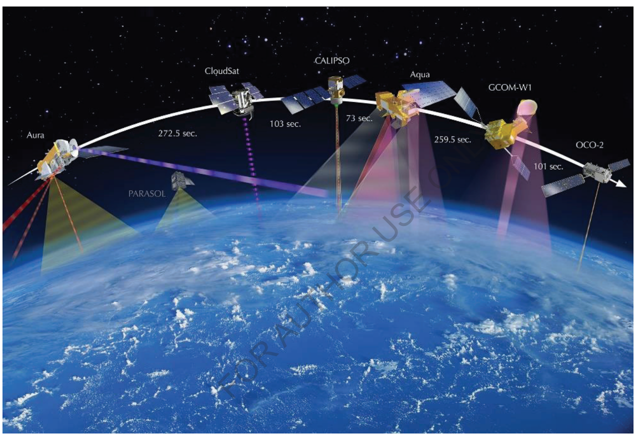
Remote Sensors

Remote sensors are most frequently divided into the passive and the active sensor groups depending on how reflected or emitted energy is used and measured during sensing. According to how the data is recorded, passive sensors are further divided into three categories: across-track, along-track, and thermal infrared scanners. Active sensors are further divided into radar and Lidar, which have different data gathering techniques and geometrical qualities in the pictures produced. We looked at each sort of sensor's mechanism and method for acquiring digital images in this chapter. This information is necessary for both the assessment and interpretation of images and for selecting the best kind of sensor for a particular project.



Earth Observation Satellites Characteristics

Earth asset satellites and environmental satellites are the two main categories under which Earth monitoring spacecraft may be divided. The Earth resources satellites are intended for observing, evaluating, and charting environmental conditions, ecosystems, and natural resources. The Earth resource satellites generally take pictures with intermediate spatial resolution (10 to 100 m), swath widths of under 200 km, and revisit intervals of two to one months. Environmental satellites encompass meteorological and oceanographic satellites, which collect photographs with much finer quality and far broader swaths of tens of thousands of kilometers. They frequently cover the entire Earth daily or hourly and have a substantial temporal resolution.



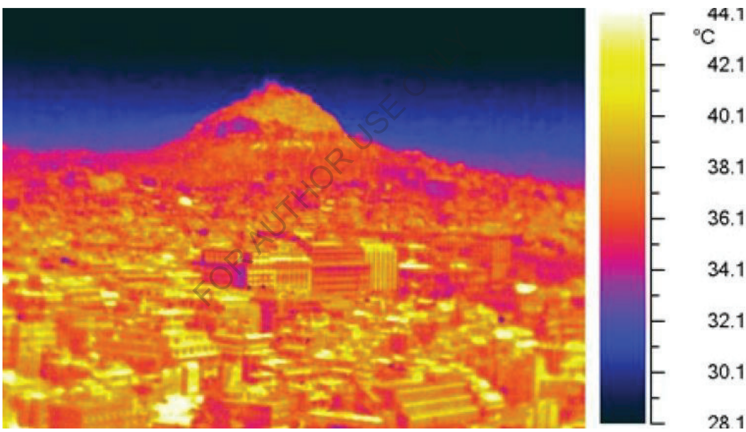
Earth resource and environmental satellites have enhanced spatial data research and technology over the past few decades and added to our understanding of the planet Earth. Local and regional studies frequently utilize images from Earth component satellites, whereas regional and global studies frequently use pictures from atmospheric satellites. The latter becomes more crucial when worries about climate change are addressed. Commercial satellites are a new class of spacecraft that have been made accessible for public usage since 1999. These satellites are capable of stereo imaging and offer images with extremely high spatial resolution, frequently a few meters or even sub-meters.

Commercial satellites may also provide very accurate position data, therefore their photos are frequently employed for urban, mapping, and military purposes. New concerns about personal privacy and spatial intelligence arise as these photographs are often made available online and through telecommunications.

Modern spaceborne remote sensing is evolving quickly. The United States and a select few other countries have created "due use" satellite systems for use by the authorities, the military, and commercial businesses. Many people are interested in satellite constellations because they frequently include a sensor web made up of other communications satellites, in-situ sensors, and ground control systems.

Thermal Remote Sensing Applications

If a substance's absolute temperature exceeds the freezing temperature, it will radiate electromagnetic energy into space. A material's absorption coefficient and kinetic temperature are the two factors that determine how much radiant energy it emits. In addition, thermal conductivity, capacity, and inertia all play significant roles in determining a material's temperature range as well as how that temperature changes over time. Thermal inertia, closely related to an object's thickness, primarily regulates a particular material's diurnal temperature change.



Thermal-IR pictures gauge the materials' radiant temperatures. Only specific areas of the thermal-IR spectrum, defined as atmospheric apertures, may be employed for thermal satellite data because carbon dioxide, ozone, and water vapor strongly absorb thermal-IR radiation. Several environmental elements need to be taken into consideration because they may have varying degrees of effect on thermal patterns, even if thermal pictures can offer identifiable signatures or indirect indicators of qualities of materials that are diagnostic. Thermal-IR sensors frequently have to choose between temperature accuracy and spatial resolution. Depending on the objective of a particular project, precise ground temperature computation may or may not be required during thermal picture assessment and interpretation.

In addition to identifying soils, minerals, and rocks, calculating soil water and evaporation and transpiration from vegetation, tracking active volcanoes and forest fires, identifying thermal plumes in lakes and rivers, and researching the urban heat island effect, thermal-IR photographs obtained from space and by aircraft have been used in a variety of purposes.

\

Figure Sources & References

https://www.mathworks.com/company/user_stories/ohb-develops-satellite-guidance-navigation-and-control-software-for-autonomous-formation-flying.html

https://mars.nasa.gov/mro/mission/spacecraft/parts/gnc/

https://www.eetimes.com/taking-care-of-power-systems-for-space-missions/

https://aerospace.honeywell.com/us/en/products-and-services/product/hardware-and-systems/space/small-satellite-specific-bus-products/telemetry-tracking-command

https://en.wikipedia.org/wiki/Solar_electric_propulsion

https://medium.com/teamindus/structural-evolution-of-the-teamindus-spacecraft-that-will-land-on-the-moon-b5aa6bc73ccc

https://astronomy.swin.edu.au/cosmos/e/Electromagnetic+Radiation

https://www.altenergymag.com/article/2017/12/can-we-harvest-energy-from-space/27549

https://beyondpenguins.ehe.osu.edu/issue/energy-and-the-polar-environment/solar-energy-albedo-and-the-polar-regions

https://www.nasa.gov/mission_pages/soho/15-years.html

https://www.vox.com/the-highlight/2019/7/16/17690740/cosmic-rays-universe-theory-science

https://www.nasa.gov/feature/goddard/2021/five-questions-about-space-weather-and-its-effects-on-earth-answered

https://www.theweek.in/news/sci-tech/2019/04/01/New-insight-into-how-Suns-powerful-magnetic-field-effects-Earth.html

https://www.livescience.com/21668-why-earth-magnetic-field-wonky.html

https://www.nasa.gov/mission_pages/sunearth/news/gallery/20130228-radiationbelts.htmlhttps://www.iso.org/news/ref4321.html

https://www.innovationnewsnetwork.com/studying-comets-understand-evolution-solar-system/10902/

https://ai-solutions.com/_freeflyeruniversityguide/patched_conics_transfer.htm

https://www.ucf.edu/news/ucf-researchers-develop-groundbreaking-new-rocket-propulsion-system/

https://www.renesas.com/tw/en/application/key-technology/radiation-hardened-solutions/satellite-telemetry-and-command

https://www.google.com/url?sa=i&url=https%3A%2F%2Faerospace.org%2Farticle%2Fpropelling-electric-propulsion-forward&psig=AOvVaw3lAe1CXZkBOw-

WU2CHMFpa&ust=1665754588605000&source=images&cd=vfe&ved=0CAgQjRxqFwoTCLDx_Map3foCFQAAAAAdAAAAABAD

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.spacetec.partners%2Fservices%2Fstrategy%2Fhyprogeo-hybrid-propulsion-module-for-transfer-to-geo%2F&psig=AOvVaw0JRn3danJCX3DEHbULJL66&ust=1665754847136000&source=images&cd=vfe&ved=0CAgQjRxqFwoTCMCRvMKq3foCFQAAAAAdAAAAABAD

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww1.grc.nasa.gov%2Fresearch-and-engineering%2Fnuclear-thermal-propulsion-systems%2F&psig=AOvVaw0isfJFW8fhoG9Mqe67fPC7&ust=1665755191622000&source=images&cd=vfe&ved=0CAgQjRxqFwoTCLCmmumr3foCFQAAAAAdAAAAABAD

https://www.google.com/url?sa=i&url=https%3A%2F%2Fmars.nasa.gov%2Fmro%2Fmission%2Fspacecraft%2Fparts%2Fantennas%2F&psig=AOvVaw3nqdupL9UqAfiX36PjUWcQ&ust=1665756000870000&source=images&cd=vfe&ved=0CAgQjRxqFwoTCPiS0emu3foCFQAAAAAdAAAAABAD

https://cdn1.byjus.com/wp-content/uploads/2018/11/physics/wp-content/uploads/2016/06/23.png

https://upload.wikimedia.org/wikipedia/commons/7/77/A-Train_w-Time2013_Web.jpg

https://www.gim-international.com/cache/0/5/6/b/9/056b986afd500bb7820ae1304d2a58a38b0a1974.jpeg

Brown, C. D., Spacecraft Mission Design, 1992, AIAA, Wash., D.C.

Brown, C. D., Spacecraft Design, 1981, University of Colorado.

Isakowitz, International Reference Guide to Launch Vehicles, 1994, AIAA, Washington D.C.

Larson & Wertz (ed.), Space Mission Analysis and Design, 2nd edition, 1992, Kluwer Academic

Publishers, Boston.

Glenn http://facilities.grc.nasa.gov/epl/index.html

Sutton - Rocket Propulsion Elements, George Sutton & Oscar Boblarz, Wiley Interscience m, 7th edition.

Gunn - Development of nuclear rocket engine technology, GUNN, STANLEY V., Rockwell International,,

AIAA-1989-2386.

KS OmniScriptum Publishing
Brivibas gatve 197
LV-1039 Riga, Latvia
Telefax: +371 686 204 55

info@omniscriptum.com
www.omniscriptum.com