SPACECRAFT THERMAL CONTROL SYSTEMS, MISSIONS AND NEEDS

SPACECRAFT THERMAL CONTROL 1
What is STCS, TC and STC 1
What makes STC different to thermal control on ground 4
Systems engineering 4
Systems design 5
Control and management 6
Quality assurance 6
Spacecraft missions and thermal problems 7
Missions phases 8
Missions types according to human life 8
Missions types according to payload 9
Missions types according to orbit 10
Orbit parameters 12
Orbit plane 13
Sun direction 15
Orbit period and eclipse fraction 15
Eclipse period 16
Missions attitudes 19
Some relevant cases 20
Thermal problems in space 23
The generic thermal balance 25
Power sources 29
Thermal radiation 30
Why is thermal radiation so important in space 31
Why is spacecraft thermal control required 32
Temperature levels, ranges and margins in spacecraft design 33
Temperature levels 33
Temperature ranges 34
Temperature margin 34

SPACECRAFT THERMAL CONTROL

WHAT IS STCS, TC AND STC

Nomenclature:

TC stands for Thermal Control.

STC stands for Spacecraft Thermal Control.

STCS stands for Spacecraft Thermal Control System (or Subsystem).

What is under Spacecraft Thermal Control System:

- System (interacting elements behaving as an entity): structural system, navigation system, power system, communication system, TCS…
- Control (to regulate, to command).
- Thermal (temperature and heat).

- Spacecraft (vehicle for a space missions, i.e. outside Earth's atmosphere): from nano-satellites (0.1 m, 1 kg, 10 W) to space stations (100 m, 10^5 kg, 100 kW).

The aim of STCS is to guarantee that all equipment and structures, during their whole life, are maintained within acceptable temperature margins, for the different thermal loads imposed, at minimum overall cost. Temperature restrictions can be imposed on extreme values (e.g. avoid T_{max}>50 ºC and T_{min}<0 ºC), and on spatial or temporal gradients (e.g. avoid $dT/dx>10^{-3}$ K/m to prevent optical misalignments, or avoid $dT/dt>10^{-3}$ K/s to prevent thermal shift in sensors).

The need arises because most active equipment can only work at room temperatures (e.g. from 0 ºC to 40 ºC for batteries, although a general rule of thumb is that life-span and reliability of any semiconductor device is inversely proportional to the junction temperature), whereas thermal expansion may deform structures. Thermal control is vital in any active system, and TC failure may be catastrophic (e.g. the Shuttle was designed to make an emergency re-entry in case of heat-radiator failure, because it was judged that after 3 hours without heat rejection the vehicle would be un-habitable).

The solution to the temperature control problem is a good thermal design (taking decisions to achieve the goal), in order to:

- Protect the equipment from damaging hot temperatures, either by proper heat insulation from external sources, or by proper heat removal from internal sources. Special needs are thermal protection (TPS) during ascent and descent through atmospheres, where surface temperatures well over 1000 K develop.
- Protect the equipment from damaging cold temperatures, by proper heat insulation from external sinks, by enhanced heat absorption from external sources, or by heat release from internal sources.

To guarantee that the goals of STCS will be met, an iterative procedure is followed, starting by an assumed full hardware specification and finding the corresponding temperature field, and changing the assumptions until an acceptable result is achieved. In other words, starting from the initial system requirements, the designer selects some preliminary TC solution from the state of the art in STCS, develops a mathematical model of the spacecraft and its environment, able to predict the thermal response under the variety of situations envisaged, performs a detailed physical test in one or a few representative situations, verifies that the prediction of the model is acceptable in this particular case, and extrapolates it to all the other foreseeable situations, usually in an iterative process of solution refining.

SCTS may refer just to the hardware used aboard for thermal control, or to all thermal aspects in spacecraft design. Tradition has coined 'thermal control' as synonymous of thermal engineering, covering a variety of sub-disciplines: technologies (state of the art), design (find a solution within the state of the art, or anew), control (sensors and actuators), analysis (worst cases), simulation (numerical), diagnostics (monitoring and testing)… STC does not end with ground design and verification, but continues with operational tele-monitoring and tele-control until the spacecraft end of life.

The design process is traditionally split in four phases, each one beginning with a statement of work (SOW) and ending with some readiness review:

1. Phase A. Feasibility study or conceptual design. It starts with a rough specification of objectives and requirements; then, the state of the art is analysed to pick out the most promising alternatives (it is no good to keep to 'the best' because, at that early stage in the design process, the specifications are not frozen); thence, the more demanding design tasks are pinpointed for further in-depth study, and finally, a preliminary design is established and documented. The only hardware development at this phase may be a mock-up to better illustrate the project (important for promotion and future team newcomers).
2. Phase B. Detailed design, with specification of interfaces, leaving no foreseen open problems. At this phase, contacts with possible equipment providers are established, and some special hardware may be developed for bread-boarding tests on difficult or key aspects of the design, to help engineering development. After this phase the design should be fixed, and a preliminary design review (PDR) formally passed to go on with the nest and most expensive phase: implementation (from paper drawings to the workshop).
3. Phase C. Manufacturing of components, assembly and verification at subsystem level. At this phase, non-conventional hardware items are manufactured (following the same procedures expected for final production) and subjected to qualification tests, to demonstrate that they withstand the most severe loads expected, i.e. for quality assurance of the method (these prototypes are not used for final assembling). Some times, a critical design review (CDR) is performed in the middle of this phase, after the manufacturing details are available but before the most expensive commitments.
4. Phase D. Delivery of a qualified product. At this phase, all hardware is subjected to an acceptance test (less severe than the qualification test) for quality assurance of the product. Many times, particularly for unique projects like spacecraft design, Phase C and D are combined in a Phase C/D envelop.

At each phase, not only the technical constraints must be clarified, but the financial constraints and other management endeavours (particularly the time-lining); i.e. the optimization of the three main characteristics of a project: performances, schedule, and cost, should be done from the three fronts (otherwise, pushing performances too high, as often done in space projects, tends to produce large increases in schedule or cost, or both).

For large or complex projects, like spacecraft design, the whole system is split into several subsystems, and clear interfaces between them are established by means of an interface requirement documents (IRD), in order to contain each subsystem design under known boundary conditions. For the STCS, the interface specifications are usually stated as an acceptable range in temperature value and temperature gradient (or heat flux).

What makes STC different to thermal control on ground

Every non-inert system must evacuate heat to the environment (to compensate entropy generation within), because thermal buffering is impracticable in the long term. Consequently, thermal control of a given system is basically driven by the available environment.

Besides being familiar with thermal control of habitable spaces (heating and air conditioning), we take for granted nowadays thermal control in vehicles too (from cars to aircraft), as well as thermal control in appliances (from insulators and thermostats in furnaces, to fans in computers). Thermal control of electronic equipment on ground has nowadays become one of the typical tasks of thermal engineers.

What makes STC different from thermal control on ground? The space environment: vacuum, and abrupt load changes at eclipses. We take for granted Earth's thermalising baths: the atmosphere and the oceans, and perhaps do not realise their contribution to thermal control on ground. For instance, typical day-night temperature variation in Madrid, with a continental climate, is about T_{mean}±10 ºC with an annual mean of $T_{annual,mean}$=15 ºC, whereas those values are ±100 ºC and 1 ºC on the Moon, and nearly the same on an artificial satellite, or an EVA suit, all being exposed to the same external environment.

Even the comfortable-looking shirt-leaves environment of habitable modules in space poses different thermal control problems than on ground, due to the absence of natural convection (there is always some artificially created forced air convection to help ventilation of persons and equipment).

Thence, the fields to cover in STCS may be grouped as:

- For a background in thermal control, a review of:
  - Thermodynamics
  - Heat Transfer
  - Control theory.
- For a background in spacecraft, a review of:
  - Missions and payloads
  - Thermal loads in space.
  - Technologies available for STC.
- For the actual spacecraft thermal control design:
  - Objectives and requirements of STC.
  - Numerical modelling, to predict temperature evolution for given geometry, materials, and interactions.
  - Physical testing, to accept and validate data and numerical modelling results.

SYSTEMS ENGINEERING

Leaving aside the ground infrastructure required (e.g. launch and ground stations constraints), the design of spacecraft, either robotic spacecraft (satellites and planetary probes), or spacecraft for human spaceflight (spaceships and space stations), must bring together knowledge from various disciplines, namely:

- Systems engineering for defining global goals (with priorities and reliabilities), distribute specialist tasks (payload, propulsion, navigation, electrical power, thermal control…), and keep clear interfaces in a top-down hierarchy.
- Project management for maintaining the design baseline under estimated budgets of time, cost, and risk, until the end product is delivered and/or operated. Clear hand-out procedures must be established to delimiter responsibilities of the different teams involved in the whole project: customers, developers, operators, end-users, sponsors...
- Astronautics (or Astrodynamics) for overall mission design (launcher, spacecraft platform (lately, the popular name for a space platform is bus), and ground coverage), and manoeuvres.
- Propulsion engineering for the design of the propulsion subsystem, which provides means of launching the spacecraft and transporting it from one orbit to another.
- Communications engineering for the design of the telemetry, tracking, and command (TTC) subsystem, which uses technologies and techniques of terrestrial radio and digital communications to communicate with the ground, and to perform tracking and ranging.
- Computer engineering for the design of the on-board data handling (OBDH) subsystem, which includes on-board computers and computer buses, and input/output devices.
- Software engineering for the on-board software which runs all the on-board applications, as well as low-level control software. This subsystem and the former one are very similar to terrestrial real-time and embedded software designs.
- Electrical engineering for the design of the power control subsystem (PCS), which generates, stores and distributes electrical power to all the on-board equipment. Up to now, all spacecraft power generators have no moving parts. Since the first use in 1958 of solar cells in spacecraft by both the USRR and the USA (Vanguard-1), most spacecraft get their electrical power from photovoltaic solar panels (wall-mounted or deployed). The exceptions are short-time manned vehicles (e.g. Apollo, Shuttle), which are powered by fuel cells, and deep space probes, which are powered by radioisotope thermoelectric generator (RTG).
- Control theory for the design of the attitude and orbit control (AOCS) subsystem, which points the spacecraft correctly, and maintains or changes the orbit according to the mission profile. Although the techniques in AOCS design are common with terrestrial methods, the hardware used for actuation and sensing in space is usually very specific to spacecraft.
- Thermal engineering for the design of the thermal control subsystem (TCS), which maintains environmental conditions compatible with operations of the spacecraft equipment. This subsystem has very space-specific technologies, since, in space, radiation and conduction usually dominate heat transfer, by opposition with Earth where convection is typically the main one.
- Mechanical engineering for the design of the spacecraft structures and mechanisms. These include beams, panels, and deployable appendages or separation devices (to separate from the launch vehicle).

Systems design

As for any design endeavour, spacecraft thermal control design must find a compromise solution (trade-off) fulfilling the requirements at the lowest cost (on power, mass…). Satellite thermal control design currently faces several challenges, such as higher payload dissipation, increasing heat transport distance, denser packing of the on-board electronics, longer life, shorter development time, increasing need for satellite radiator area. The TCS adds mass, cost, and complexity; approx. (3±2)% mass and (3±2)% cost of the spacecraft.

Once the mission and payload are specified, the usual STCS design methodology followed is:

- establish the thermal requirements
- establish the worst case for environmental heat loads and power dissipation
- elaborate the control means
- build mathematical models to simulate the satellite thermal behaviour
- analyse the design for worst environmental and dissipation heat loads
- verify the design against the requirements
- gather the budgets
- change the design if necessary
- verify the design by test and correlate the mathematical models.

Control and management

Project management must plan the design activities and establish procedures to control resources and guarantee success, by setting a well-structured timeline, with well-defined control milestones that allow a sure step-by-step progress, concluding the different phases in the project within budget and without compromising future phases.

In what concerns the thermal control design, a sequence of thermal analyses and tests are established as a minimum, but the design team should be ready to try many different configurations at the beginning of the project. If the numerical modelling correctly predicts the results of scarce experimental tests, confidence in its extrapolation to new cases is gained by the whole team and less development test are necessary.

Quality assurance

ESA has developed STEP-NRF (Network-model Results Format) and STEP-TAS (Thermal Analysis for Space) as open standards for product data exchange based on the ISO 10303 (better known by its informal name STEP, Standard for the Exchange of Product model data).

The thermal control project must be compliant with international standards of quality assurance from the methodology applied, to the analysis performed, and the test used for validation. Some relevant international standards from the European Cooperation for Space Standardization (ECSS, http://www.ecss.nl/) are:

- ECSS-E-30 Part 1[a]. Space engineering. Mechanical — Part 1: Thermal control
- ECSS-E-10-03A. Space engineering. Testing
- ECSS-E-10-04A. Space engineering. Space environment

SPACECRAFT MISSIONS AND THERMAL PROBLEMS

A mission is a set of activities to reach a goal. Spacecraft missions aim at taking advantage of outer space as a privileged place and unique way for:

- Observation of our Earth (globally and in detail), and the rest of the Universe (free from our atmospheric filter), for environmental monitoring.
- Communications (including navigation aids).
- Experimentation in physical and life sciences under microgravity, vacuum, radiations...
- Explorations, travelling to other worlds (by tele-presence, or in person).

Specifications of a spacecraft mission usually starts by defining the payload and orbit (most of the time directly related), then the platform (or service module), then a suitable launcher, and finally the ground operations (system and payload, including end-users) and required ground segment

Spacecraft missions can be manned or robotic, the latter being classified according to payload objectives. Launchers and sounding rockets are not usually considered spacecraft, but the distinction becomes artificial when reusable vehicles are considered (from the Shuttle to space-planes). Even high altitude balloons (flying at around 40 km height) share most of the ‘space environment’ characteristics, at least in what concerns thermal control problems. Mission lifespan is another important parameter, since thermo-optical properties degrade over time.

Thermal loads are heavily dependent on spacecraft mission. The main thermal load is usually solar radiation, but it may be aerodynamic heating in planetary atmospheres, or even on-board power dissipation in heavy duty crafts. The final destination of all thermal loads in astronautics is an energy exit as thermal radiation towards the deep-space sink (at 2.7 K); in old spacecraft, radiators were just part of the external surfaces, but in heavy-duty-spacecraft there is not enough envelop area and deployable radiators are required (radiators in the ISS are second in size to solar panels).

Space vacuum makes thermal radiation dominate the energy balance at all times in a mission (e.g. in an Ariane 5 launch, pressure inside the fairings drops to one hundredth of sea level value 100 s after lift-off).

For the thermal analysis of a spacecraft the following data must be known:

- Spacecraft <u>geometry and materials</u> data for the body and surfaces (thermal capacities, conductivities, radiative properties, and desired operating temperature ranges), including expected <u>power dissipation</u> laws). If the spacecraft has moving parts (e.g. pointing solar panels and antennas, deployable camera shrouds…), the pointing or kinematics data must be given.
- Spacecraft <u>orbit and attitude</u> data around the planet (the central attracting body, in general, be it a planet, a moon, or the Sun). Orbit and attitude can be tracked passively by ground radars, actively by ground-station radio-navigation, or, in the case of LEO, by GPS navigation.
- <u>Sun orientation</u> data relative to planet or moon (for non-heliocentric orbits), to calculate solar radiation and albedo loads. Orbit, attitude and Sun orientation are usually considered together under spacecraft mission.

For the special case of spacecraft that descend and land on the surface of a celestial body (other than the Earth), the thermal problems are dependent on whether there is an atmosphere or not. Leaving aside the aerodynamic heating during the descent on planetary atmospheres (to be considered under the Thermal Protection heading in TCS Technologies), a spacecraft on the surface with an atmosphere may be subjected to direct solar radiation (if the atmosphere is transparent like in Mars, not in the case of Venus that is fully cloud-covered), scattered solar radiation, convection from the ambient gas, conduction to the ground, and radiation to the environment, which is no longer the background radiation at 2.7 K but the effective sky temperature, which depends on the atmosphere thickness and gas constitution (on the optical thickness of the atmosphere filter).

Missions phases

Spacecraft analysis is first and foremost centred on its steady (or periodic) cruise-phase configuration, but thermal control is not only required on this main phase other important phases are:

- Ground storage and stand-by at launch pad.
- Ascent within launcher coffin.
- Orbiting or cruise.
- Flyby.
- Re-entry (re-usable launchers, landers, and sample return vehicles).
- End of life disposal (usually by controlled disintegration at re-entry).

Missions types according to human life

- Unmanned missions (or robotic missions), like ordinary satellites, cargo vessels, deep probes. There are some 2500 satellites in Earth's orbit, among commercial, scientific, and military ones, from the smaller nano-satellites (say 0.3 m in size, 10 kg, 100 W), to the larger communication satellites (say 3 m in size, 10 tons, 10 kW). Pallets, attached to other spacecraft or free-flying like Eureka, lack for navigation capabilities. Cargo vessels, like Progress, ATV, H-II, are usually non-retrievable short-missions (from one week to one quarter), to upload materials and get rid of waste (the ATV carries a 8 tons payload with 20 tons total mass). Deep probes travel far from Earth's orbit, to the Moon, to other planets or their moons, to comets or towards the Sun, usually carrying landers and rovers (two sounding balloons were deployed in Venus atmosphere, too, and sounding aircraft have been investigated).
- Manned missions, comprising transfer vehicles like the Soyuz, Shuttle, or Orion (in 2014), orbiting stations like Mir and ISS, and flybys and landers to the Moon (to Mars in 2030s?). Manned missions have additional thermal needs for the habitable space, airlock, space suits, etc. The later poses one of the most delicate STCS problems, only second in importance to the air revitalisation system, what is usually managed together under the Environmental Control and Life Support System (ECLSS). Why manned missions? Ans.: Because we want to know by ourselves: to the first spacecraft on 4 October 1957, the robotic Sputnik-1, followed on 3 November 1957 Sputnik-2 already carried a living animal, the Laika dog (which died from overheating after few hours in orbit); on 12 April 1961, Vostok-1 carried the first person to space (Y. Gagarin orbited the Earth, and came back).

Missions types according to payload

The payload is that part of a cargo earning revenue, and all spacecraft expenses are justified in terms of the benefit (commercial, scientific, or other) the payload can provide, basically in the form of information (and in rare occasions by material sample return). The main types of non-military payloads are (se Table 1 for a summary):

- Earth observation and meteorology. Polar low Earth orbit (LEO) for global coverage, or geostationary orbit (GEO) for continuous coverage (Fig. 1). The first spacecraft, Sputnik I, had a radio beacon that was used for communication research (ionosphere transmittance, by signal analysis), atmospheric research (by orbit tracking), meteorite research (by temperature monitoring of the filled nitrogen gas), and thermal control research (temperature was measured at the surface and inside). GMES (Global Monitoring for Environment and Security), a joint initiative of the EU and ESA, with its Sentinel satellites, is the European Union contribution to the Global Earth Observation System (GEOS).
- Communication and navigation (point-to-point or broadcasted). The orbit may be GEO for low and middle latitudes coverage, Molniya and other high eccentricity orbits (HEO) for high latitudes, or a constellation of satellites in polar LEO or inclined middle Earth orbit (MEO) for global coverage. Galileo is becoming the European Union contribution to the Global Positioning System (GPS).
- Astronomy. The orbit may be LEO, MEO, HEO, GEO, Lagrange points… The Hubble Space Telescope (HST), with a 2.4 m in diameter primary mirror, was launched in 1990, is in a LEO at 590 km, and been serviced by the Shuttle several times, whereas its follow-on, the James Webb Space Telescope (JWST), expected to be launched in 2013, will have a 6.5 m in diameter primary mirror, and will be placed in a heliocentric orbit at 1 500 000 km (point L2, described below); the telescope will be thermally shielded from the Sun, and decoupled from the rest of the spacecraft systems, by a big sunshade made of five metallised sheets, maintaining the whole telescope, after a four months natural cooling down period, at some 40 K (further cooling down to 10 K for the middle infrared detector will be performed by a mechanical cryo-cooler).
- Human exploration and space stations. Low Earth orbit or deep probes to the Moon. After exploration comes scientific research and technology development.

Table 1. Typical missions and installed power.

Mission	Orbit	Attitude	Installed power [W]
Science (astronomy)	HEO	Sun, stars or planet pointing	200..1500
Telecommunication	GEO	Earth pointing	500..5000
Earth Observation & Meteorology	LEO-polar GEO	Earth pointing	500..5000
Global navigation	MEO, i=56º	Earth pointing	200..1500
Manned Vehicles	LEO+transfer	Earth pointing	1000..10 000
Manned Stations	LEO	Earth pointing	10 000..100 000

Missions types according to orbit

Main orbit characteristics are its centre of attraction (e.g. orbits around the Earth, around the Moon, the Sun...), and its size (or altitude over the attracting celestial body), besides other geometrical or temporal details. Main types of orbits are:

- LEO, i.e. low Earth orbits, are usually circular orbits at 300..900 km altitude (>250 km to avoid large drag, and <1000 km to avoid van Allen belts radiation; e.g. ISS at around 400 km, MetOp at 800 km). Orbit period is $T \approx 1.5$ h (90..100 min). The main orbit parameter relevant for TCS is inclination to the equatorial plane (e.g. 52º for ISS, 99º for MetOp), which governs possible eclipse periods. Low Earth orbits have many applications: as a destination for Earth observation, for space stations, for astronomy sensors without the atmospheric filter (e.g. Hubble telescope, at z=540 km and i=28º), for mobile communications (e.g. the Iridium constellation), or as a parking orbit for subsequent missions (e.g. Hohmann transfer orbit to GEO, Moon trips). The LEO Sun-synchronous orbit (SSO), where the orbit plane keeps an invariant position relative to Sun, is much used for Earth observation (e.g. Nimbus, NOAA, Landsat, Spot, Envisat, Spot, Goce, MetOp...). The SSO is a near-polar near-circular orbit usually at 600..900 km altitude, slightly retrograde ($i \approx$98..100º), that profits from the precession of non-equatorial orbits in oblate planets (a function of altitude and inclination), to match the 360 degrees per year Earth revolution around the Sun (≈1º/day eastward), so that the sub-satellite point passes over the same latitude at the same local time (e.g. at noon on the Equator, to minimise shadows, although seasonal variation of illumination cannot be avoided). Super-low orbits (SLEO) require continuous propulsion to balance air drag and avoid orbit decay: Goce (Gravity field and steady state Ocean Circulation Explorer) used ion thrusters to operate for almost 5 yr at 255 km altitude in a polar sun-synchronous orbit (at this altitude, deceleration is around $1.5\cdot10^{-5}$ m/s^2 due to air drag, and $6\cdot10^{-8}$ m/s^2 due to radiation pressure); Slats (Super Low Altitude Test Satellite) is expected to operate down to 220 km altitude with ion thrusters for months, and at 180 km altitude with the help of hydrazine thrusters for a week. Short-time missions can flight very low; e.g. the Space Shuttle orbited at z=245 km for 10 d (STS-9), and the Foton satellites at 280 km for 16 d. The cost to put a mass in LEO orbit is some 10 000 €/kg.
- MEO, i.e. midway Earth orbits, are usually circular orbits midway between LEO and GEO. Mainly used for navigation satellites, with orbital periods of 12..15 h (i.e. almost semi-synchronous orbit, two orbits per day) to ease tracking: there is a constellation of 31 GPS (54 have been launched) at 20 200 km altitude (i=55º), and 24 Glonass at 19 100 km (i=64.8º); 28 Galileo satellites (the first 2 launched in 2011), at 23 300 km altitude and i=56º, with a track repeat ratio of 5/3 (5 orbits in 3 days to return to the same sub-satellite point). Spacecraft at MEO and GEO are within the high-energy Van Allen radiation belts.
- GEO, i.e. geostationary Earth orbits, are equatorial circular orbits at 36 000 km altitude (a=42 164 km) going eastwards. GEO orbit period is precisely one sidereal day, T=86164 s (23.934 h), with null inclination, i=0 (i.e. equatorial), so that the ground track is a fix point on the Earth surface if the satellite moves in the same sense of the Earth's rotation). There are many examples: Meteosat (7 units), Intelsat (27 units), Astra (14 units), Telecom (5 units), Eutelsat (3 units), Inmarsat (11

units), Hispasat (3 units), Brazilsat (3 units), TDRS (5 units), EDRS (2 units),… The price to put a mass in GEO orbit is around 30 000 €/kg.

Fig. 1. Relative distances and viewing angles for LEO and GEO.

- Halo orbits around Lagrangian points. In the three-body's orbital mechanics, there are three axial points that rotate around the largest mass at the same rate as the small mass: L1 in between, L2 opposite to the small one, and L3 opposite the largest; besides, there are two other Lagrange points, L4 and L5, at 60º in the plane of the 3-body motion (Fig. 2). Two such systems are the Sun-Earth system (with points SEL1 and SEL2), and the Earth-Moon system (with points EML1 and EML2). A mass left at those points (or nearby) is marginally stable, and describe quasi-periodic orbits around those points, named halo orbits (Gr. $\eta\alpha\lambda o\sigma$, circle around, disc). SEL1 (at 0.99 ua) and SEL2 (at 1.01 ua) have great advantages because they have all bright sources (Sun, Earth, Moon) almost at fixed positions relative to the tree-body axis, and do not suffer eclipses; besides, the marginal instability around L or L2 (or L3) naturally removes uncontrolled debris (unlike at L4 and L5). SEL1 is the best to watch either at the Sun (with Earth at the back, like SOHO), or to watch the Earth fully lighted all the time (e.g. DISCOVR). SEL2 is the best to watch the rest of the universe; e.g. Planck, Herschel, Gaia, Spica, Euclid, JWST. To notice that SEL2 is slightly beyond the reach of Earth's umbra ($1.5\cdot10^9$ m against $1.4\cdot10^9$ m), so that solar radiation is not completely blocked (besides, halo orbits are designed to have large amplitude around the SEL2 point, so that solar panels are used to power the spacecraft, without eclipses). The Earth–Moon-L_2 point (61 500 km from the Moon) has been proposed as a location for a communication satellite covering the far side of the Moon. Sun-Mars Lagrange points are of little interest because Mars orbit is highly eccentric and perturbations would be high, and Mars' solar umbra extends about 30 000 km beyond Sun-Mars L2.

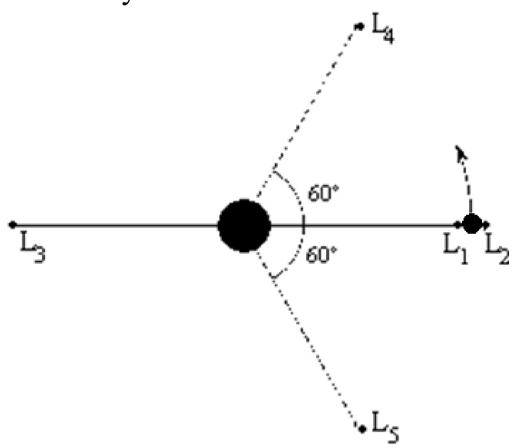

Fig. 2. The five Lagrangian points in a three-body system.

- HEO, i.e. high eccentricity orbits. On Earth, they are used for transfer orbits (e.g. Hohmann orbit from LEO to GEO), and for preferential high-latitude communications (Molniya orbits). They are also used for heliocentric deep probes to other celestial bodies, and for astronomical satellites like Solar Orbiter (with a perihelion at 0.28 AU and an aphelion at 1.3 AU) and the Parker Solar Probe (approaching the Sun at 0.05 AU, where solar irradiation is 700 kW/m^2). The first HEO was Molniya-1 satellite (1965, USSR) used for high latitude communications, having an altitude of 1500 km at perigee and 40 000 km at apogee, an inclination i=63.4º (to cancel the regression of apsides due to Earth oblateness), and a period T=12 h (semi-synchronous orbit), from which 11 h were over the North hemisphere.

- Swing-by orbits. They are nearly hyperbolic orbits used to increase the spacecraft momentum (gravity assist), currently used in all interplanetary probes. Notice that most of them lie close to the ecliptic plane; Ulysses spacecraft (launched in 1990) was the first spacecraft to orbit out of the ecliptic, to study the Sun at all latitudes.
- Landers and surface rovers. Their 'orbit period' and 'eclipse time' coincide with those of the planet-point they are located.

ORBIT PARAMETERS

Only a few parameters from orbital mechanics are relevant to spacecraft thermal control design, mainly Sun orientation, orbit period, and eclipse fraction, the latter being a key parameter to evaluate the importance of transient effects. However, the STC practitioner must understand orbit specifications (orbit plane position and ellipse parameters), and orbital parameters in general.

An inertial reference frame is always used, with an inertial reference plane containing the centre of the attracting body (the equatorial plane for planets and moons, or the ecliptic plane for heliocentric orbits), and an inertial reference direction (usually the vernal point, which is the point in the celestial sphere where the centre of the Sun crosses the equatorial plane of the Earth from south to north, i.e. at the spring equinox). Although this reference frame is not perfectly inertial (e.g. in the case of geocentric orbits, the origin has a centripetal acceleration $a_C=6\cdot10^{-3}$ m/s^2, and the equatorial plane oscillates and rotates with a 26 000 years period, making the vernal point to rotate 1.4º by century; in 500 b.C. the vernal point coincided with Aries constellation, but in 2000 a.D. it points towards Pisces, going towards Aquarius), the discrepancies are almost always neglected; the current epoch used to fix this slow-motion is Julian year 2000, 12:00 h of 1st January (labelled J2000).

The most used reference frames are:

- The heliocentric ecliptic reference frame. Sun centred (but not rotating with it), with one axis perpendicular to the ecliptic (not to its equatorial plane), and another axis pointing towards the vernal point. The acceleration of the solar system around its galactic centre is $0.22\cdot10^{-9}$ m/s^2, which can be always neglected. This is the frame used for deep probes. A less-used variation of the heliocentric frame is to choose an axis perpendicular to the Earth equator (i.e. pointing to the celestial North). Heliocentric coordinates are ecliptic latitude, ecliptic longitude and radial distance to Sun centre. Due to a common original formation, the eight planets in our solar system orbit the Sun in the same direction that the Sun is rotating, which is counter-clockwise when viewed from above the Sun's north pole, and six of them rotate about their axis in this same direction (except Venus and Uranus, which have retrograde rotation). Sun's rotation varies with latitude (e.g. at 16º latitude, the period is 25 Earth-days), and the rotation axis is tilted 7.25º to the ecliptic.

- The geocentric equatorial reference frame (also called the celestial sphere). Earth centred but fixed to the stars (not rotating with the planet), with an axis perpendicular to the planet equator (celestial North at the epoch, e.g. J2000) and another axis pointing towards the vernal point (at same epoch, J2000). This frame is used for orbital and attitude control of Earth satellites. Geocentric coordinates are declination (over the celestial equator, coinciding with geographic latitude of sub-sat point), right ascension (from vernal direction), and radial distance to the Earth's centre. Instead of the right ascension, α, the hour angle H can be used, which measures angles from the celestial meridian and westwards, instead of from the vernal point and eastwards. They are related by α+H=local_sideral_time of the observer (approx. solar time). Both, right ascension and hour angle are usually measured in hh:mm:ss. Right ascension and declination are like geographic longitude and latitude except that they are measured with respect to the celestial sphere, with the vernal direction as the origin instead of the prime meridian of Greenwich. Each fix star has a fixed set of celestial coordinates (declination, right ascension and distance; well, the change is too slow, e.g. for Polaris (the North star, with an equivalent temperature of 7200 K) declination is +89.3º, right ascension 02:31:49 h, and distance 430 light-years, with a 26 000 yr angular period and a 17 km/s expansion), but the Sun is so close to us that it changes its geocentric coordinates around a year period. For planets and moons, the coordinates at a certain time are called ephemerides.
- The geographic reference frame. Geocentric fixed equatorial reference frame (geographic, or terrestrial, or planet sphere). Fixed to the planet (rotating with it; not valid as inertial frame). Used for tracking records. Variables are geographical latitude, longitude, and altitude above the geoid.
- Local horizontal local vertical (LHLV) reference frame. Centred at the observer position (which must be specified), with a vertical axis (pointing to the zenith), and other in the horizontal plane pointing to the planet North. Not valid as inertial frame. The celestial sphere is seen tilted and rotating. Local coordinates are elevation (or altitude; vertical angle over the horizon), azimuth (horizontal angle from the North, i.e. compass angle), and distance. Not used for ranging; only used for raw angular tracking. This is the most ancient reference system (altitudes were not measured at the time).
- Body-centred (assuming a rigid body) pointing forwards in the direction of motion. Use in the inside the spacecraft and in its neighbourhood.

Orbit plane

The orbital plane of a satellite trapped in the gravitational field of a celestial body is an ellipse if only a central force is acting, i.e., if there is no propulsion, no air drag, no influence of other bodies (e.g. Sun and Moon), no radiation pressure, and the attracting body has a perfect spherical mass distribution. The celestial body is always in one of the focus of the ellipse. Two extreme cases for that ellipse are the circular orbit, and the up-and-down trajectory along the same direction. Non-trapped orbits are hyperbolas (parabolas in the limiting case) with the focus in the celestial body.

Orbit perturbations (i.e. the effect of planet geodesic asymmetry, attraction of other celestial bodies, radiation pressure), give rise to orbit and attitude perturbations, which may be advantageously used, as for

Sun-synchronous orbits, in which to slow rotation of the orbital plane (precesion) due to Earth's oblateness, helps to compensate the difference between solar day and sidereal day, to achieve passing the over the Equator at the same solar time in every orbit. A frozen orbit is one in which the sum of all perturbations is minimized, hence, a satellite in that orbit will stay there the longer, or it will require minimum orbit-keeping manoeuvres.

In the inertial reference frame described above (equatorial for planets and moons, or ecliptic for heliocentric orbits), and knowing that the orbit plane passes through the attraction centre, two additional angles are required to fix that plane (see Fig. 3 for the case of planetary orbits), namely:

- Inclination, i, with the reference plane. For orbits around planets, inclination is relative to planet equator, whereas for heliocentric orbits, inclination is relative to the ecliptic (Earth's orbital plane) and not to Sun equator. It is common practice to include motion sense into inclination data, by assigning the first quadrant to anticlockwise motion (as seen from the north) and the second quadrant to clockwise motion; i.e $0<i<\pi$ is defined as the angle between the angular momentum and the celestial North (ecliptic North, for heliocentric orbits). Thence, a geostationary orbit must have and inclination $i=0$ and not $i=180°$. Maximum attainable latitude is $\phi_m=\pm|i|$.
- Longitude of ascending node, Ω, which is the angle, in the reference plane, between the reference direction (vernal point) and the direction of the ascending node (when the satellite crosses the reference plane towards the positive pole).

Once the orbit plane is known, three additional parameters are needed to position the ellipse, usually taken as: semi-major axis a, eccentricity e, and angular position of the periapsis from the ascending node ω. Notice that the tern (Ω,i,ω) are the three Euler angles for coordinate rotations (ϕ,θ,ψ), respectively. Another additional parameter fixes satellite position in the orbit, its angular position from the periapsis, θ, called true anomaly. The periapsis radius, R_p, and apoapsis radius, R_a, are related to a and e by $a=(R_a+R_p)/2$ and $e=(R_a-R_p)/(R_a+R_p)$, or $R_p=a(1-e)$ and $R_p=a(1+e)$. It is often preferable to use periapsis and apoapsis altitudes (from the planet surface), H_p and H_a, instead of radii (from the planet centre), at least to avoid confusion with the mean radius of the planet, also named R_p.

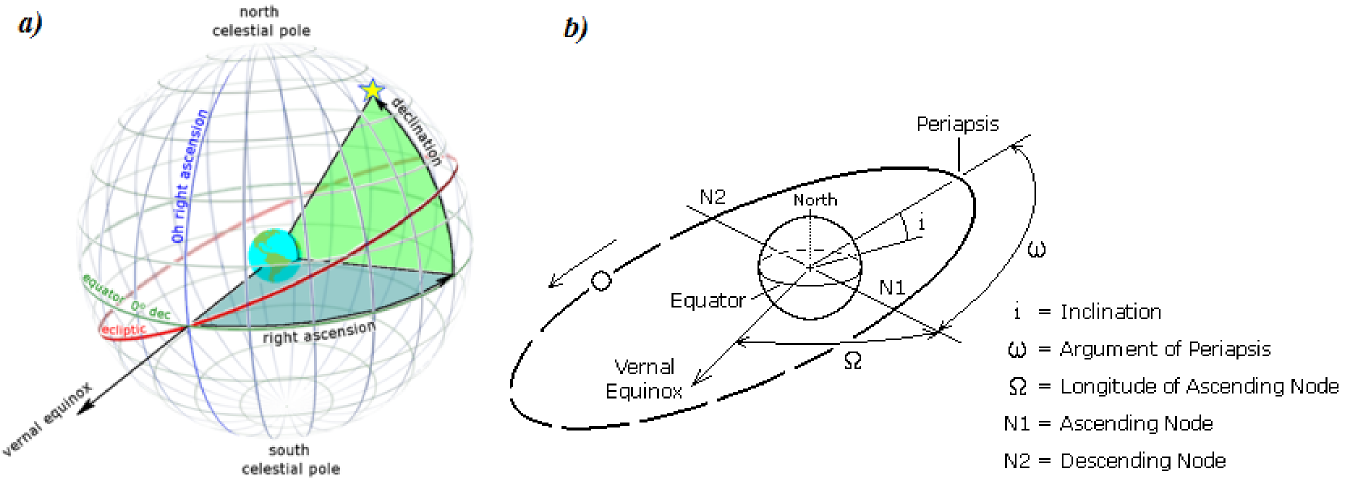

Fig. 3. a) Angular coordinates of a point (yellow star) in the celestial sphere: declination and right ascension (the Sun orbit is in red; the Earth should be spinning; http://en.wikipedia.org/wiki/Right_ascension). b) Planet centred orbit specification (http://www.braeunig.us/space/orbmech.htm)

Sun direction

Stars, far away, keep fixed positions in the celestial sphere, but the Sun, being so near (in comparison), moves around with a one year period. Planets and moons move faster in the celestial sphere (planet comes from Gr. αστήρ πλανήτης, wandering star), and their coordinates at a certain time are called ephemerides (Gr. ἐφημερίς, diary).

All astronomical objects look small from the Earth in angular size, so that solar radiation can be assumed collimated (parallel rays). Astronomical distances are difficult to compute; the Sun-Earth distance (R_{SE}) was not accurately measured until 1673 (Cassini), although the Earth-Moon distance (R_{EM}) was already accurately known since Antiquity: around 270 BCE, Aristarchus, in his book 'On the Sizes and Distances of the Sun and Moon', established that it was about sixty times the Earth radius, $R_{EM}=60\cdot R_E$, based on apparent angular diameter of Earth's shadow during a lunar eclipse (he measures $\delta_E=1.5°$) and Sun's angular diameter ($\delta_S=0.5°$); by geometry, $\arctan(2R_E/R_{EM})=\delta_E+\delta_S=1.5°+0.5°=2°=0.035$ rad, and hence $R_{EM}/R_E=2/0.035=57$. He also tried to find the distance to the Sun, but he arrived at $R_{SE}=20\cdot R_{EM}$, instead of the correct $R_{SE}=390\cdot R_{EM}$. Around the same time (say 250 BCE), Eratosthenes accurately computed the size of the Earth (R_E=6400 km), from knowledge of Egypt size.

The two spherical coordinates used to position a point in the celestial sphere (be it a star, the Sun, a planet, a moon, or a spacecraft) are (Fig. 3a):

- Declination, δ, is the angular position over the celestial equator (coincides with geographic latitude of sub-body point). In the geocentric case (i.e. when the planet is the Earth), the Sun declination only depends on the day of the year, approximately in the form $\delta=-23.45°\cos(2\pi(N+10)/365)$, with N being calendar day from 1st January. Polaris declination is 89.3°.
- Right ascension, Ω, is the angular position along the celestial equator from the vernal point, being related to geographical longitude and planet rotation (notice that, for analogy with orbit parameters, the same symbol for longitude of the ascending node is used, Ω). It is also known as 'hour angle' when measured in hh:mm:ss, with 24 h corresponding to 360°). Right ascension of the Sun varies almost linearly with day of the year, N, with origin on the vernal pint (the 80th day of the year), i.e. $\Omega=(N-80\times360°/365$ (or $\Omega=(N-80\times24/365$ in hours). Polaris right ascension is 02:31:49 h (Polaris is 430 light-years away, separating at some 17 km/s; it has an equivalent temperature of 7200 K). Sirius (the brightest star seen from the north hemisphere), has as celestial coordinates (−16.43°, 06:45 h); α-Centauri (the closest star, at 4.38 light-years, and the brightest in the south hemisphere), has as celestial coordinates (−61°, 14:40 h).

Orbit period and eclipse fraction

The orbit period, T_o, for an elliptical orbit only depends on semi-major axis, a (half the distance between periapsis and apoaxis) in the way:

$$T_o=2\pi\sqrt{\frac{a^3}{GM}} \tag{1}$$

with G=6.7·10^{-11} m^3/(kg·s^2). For geocentric orbits, with M_{Earth}=6·10^{24} kg, orbit period increases from 5400 s (90 minutes) at 300 km LEO to 24 h at GEO (Fig. 4).

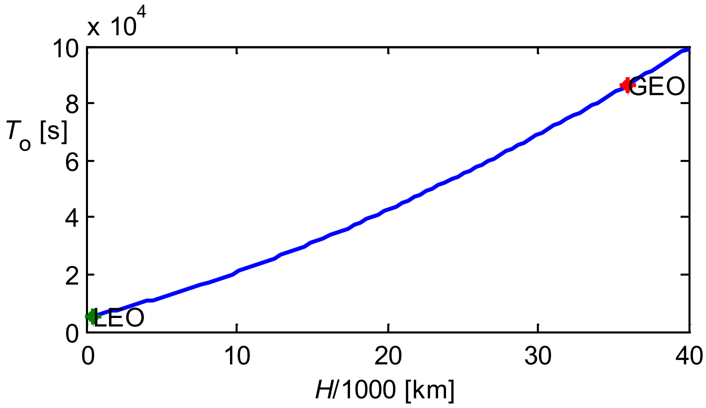

Fig. 4. Orbit period, T_o, for circular Earth orbits of altitude H.

Notice, by the way, that the high orbital speeds (e.g. at 300 km altitude $v_{LEO}=2\pi R/T_o=2\pi(6370+300)\cdot10^3/5400=7760$ m/s), imply large kinetic energies (e.g. $v^2/2$=30 MJ/kg, nearly the same as when burning a fuel), which must be dissipated by friction on re-entry (much higher for a probe coming from deep space). Notice also that orbital speed and energy decrease with altitude (e.g. from the 8 km/s and 30 MJ/kg at LEO, to 3 km/s and 4.5 MJ/kg at GEO). The Stardust probe successfully re-entered the Earth atmosphere in 2006 at 15 km/s, the present record (Apollo entered at 11 km/s).

Eclipse period

The eclipse period, T_e, depends on orbit period, T_o, orbit solar angle, β, relative altitude, $h\equiv H/R$, and orbit eccentricity. The orbit solar angle, β (or just 'beta angle'), is the angle from sunshine direction to the orbit plane, so that β=0 applies for any orbit passing through the subsolar point (i.e. when the Sun is in the orbit plane), and $\beta=\pi/2$ applies when the orbital plane is perpendicular to Sun rays; the beta angle is commonly given in degrees from −90º to 90º (β<0 for retrograde orbits).

Equatorial orbits around low-inclination planets or moons have small solar beta angles (their own orbits being almost in the ecliptic), which may often be neglected (e.g. the largest value for equatorial Earth satellites is β=23.5º at solstices; and β=1.5º for lunar equatorial orbits), whereas polar orbits may have solar angles from β=0 (coplanar orbit with the Sun) to $\beta=\pi/2$ (frontal orbit plane), depending on local launching time (β=0 for noon or midnight launching, and $\beta=\pi/2$ for dawn or dusk launching).

For circular orbits, the relative eclipse period, T_e/T_o, and the angle the eclipse starts, ϕ_{es} (measured from the orbit position closest to the Sun), is given, in terms of relative altitude, $h\equiv H/R$, and orbit solar angle, β, by:

$$\left.\begin{aligned}\frac{T_{\mathrm{e}}}{T_{\mathrm{o}}}&=\frac{1}{\pi}\arccos\left(\frac{\sqrt{2h+h^2}}{(1+h)\cos\beta}\right)\\ \phi_{\mathrm{es}}&=\pi\left(1-\frac{T_{\mathrm{e}}}{T_{\mathrm{o}}}\right)=\pi-\arccos\left(\frac{\sqrt{2h+h^2}}{(1+h)\cos\beta}\right)\end{aligned}\right\}\text{with}\quad \beta<\beta_{\max}=\frac{\pi}{2}-\arccos\frac{1}{1+h} \tag{2}$$

The relative eclipse period is plotted in Fig. 6 for some important circular orbits. Notice that, for a GEO eclipse (H=36 000 km), the maximum duration may be nearly double time than for a LEO eclipse, in spite of the fact that its duration relative to the orbital period is much shorter (a fraction of 0.05 vs. 0.4). In fact, for a given planet, and as a function of circular orbit altitude, maximum eclipse duration first falls from h=0 to h=0.215, and then monotonically increases as $\sqrt{h}$.

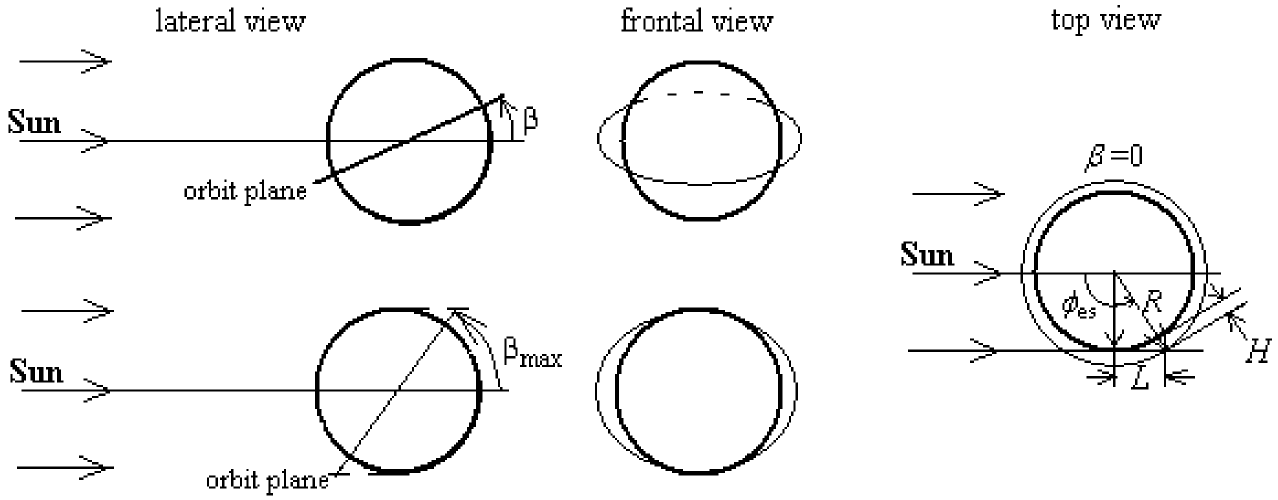

Fig. 5. a) Sketches to show the beta angle, β (inclination of the orbit plane relative to sunlit direction), and its maximum value for eclipses to occur, β_{max}. b) Eclipse start angle, ϕ_{es}, for β=0 (relative duration of eclipse is $T_e/T_o=\pi-2\phi_{es}/(2\pi)$).

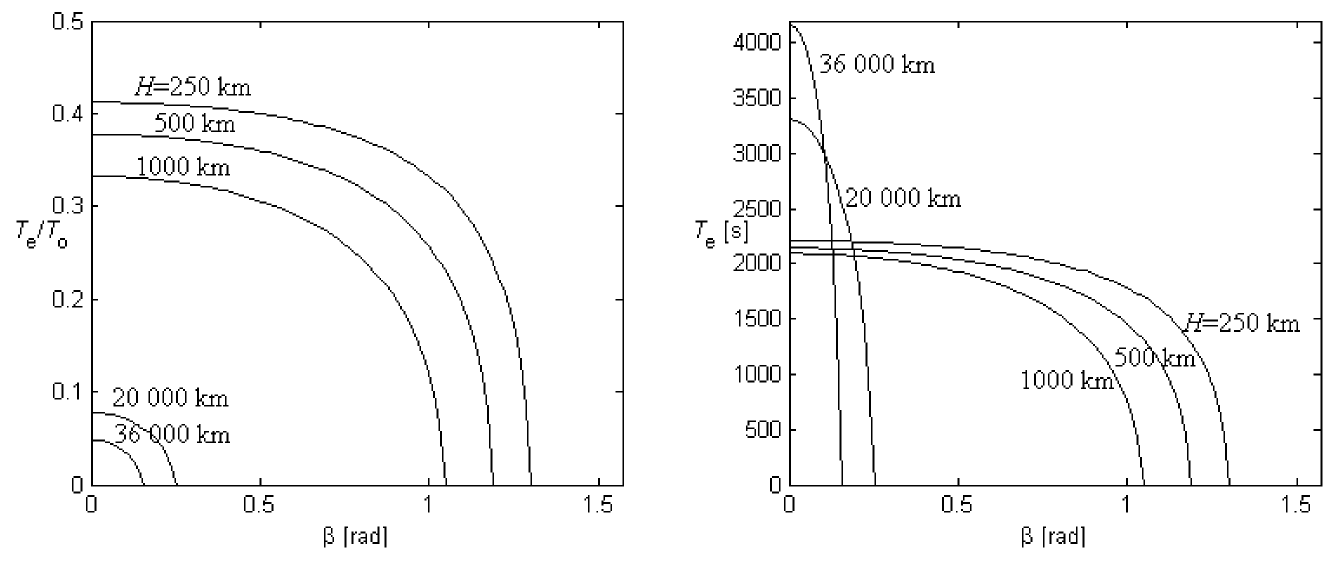

Fig. 6. a) Fraction of orbit period under eclipse, T_e/T_o, versus orbit solar angle β, for circular Earth orbits with altitude H; b) Eclipse duration.

For the case of the longest eclipse duration (i.e. for β=0, e.g. ecliptic orbit), equations in (2) can be recast to:

$$\left.\begin{aligned}&\frac{T_{\mathrm{e}}}{T_{\mathrm{o}}}\overset{\beta=0}{=}\frac{1}{\pi}\arcsin\left(\frac{1}{1+h}\right)\\&\phi_{\mathrm{es}}=\pi\left(1-\frac{T_{\mathrm{e}}}{T_{\mathrm{o}}}\right)\overset{\beta=0}{=}\frac{\pi}{2}+\arccos\left(\frac{1}{1+h}\right)=\pi-\arcsin\left(\frac{1}{1+h}\right)\overset{h<<1}{=}\frac{\pi}{2}+\sqrt{2h}\end{aligned}\right\}\qquad(3)$$

For instance, for a H=400 km orbit like that of the ISS, $h\equiv H/R$=0.063, maximum beta angle for eclipses to occur is $\beta_{\max}=\pi/2-\arccos(1/(1+h))=1.22=70.2^{\circ}$, therefore, any satellite at that altitude with an orbital inclination $i>\beta_{\max}-\delta_{\max}=70.2-23.5=46.7^{\circ}$ may potentially be sunlit during the whole orbit (the ISS has i=52º so that it has periods along the year without eclipse). For maximum eclipse duration (β=0), eclipse starts at $\phi_{\mathrm{es}}=\pi-\arcsin(1/(1+h))=1.92=110^{\circ}$, and ends at $\phi_{\mathrm{ee}}=2\pi-\phi_{\mathrm{es}}=4.4=250^{\circ}$.

By the way, the horizontal distance, L (see Fig. 5) of the satellite at the eclipse point, is $L=(R+H)\cos(\pi-\phi_{es})=\sqrt{2RH+H^2}\simeq\sqrt{2RH}$, so that a LEO spacecraft like the ISS at 400 km altitude is seen at $L=\sqrt{2RH}=\sqrt{2\cdot6370\cdot400}=2300\text{ km}$ distance when rising or setting in the horizon (a sounding balloon at 40 km can be seen from $L=\sqrt{2RH}=\sqrt{2\cdot6370\cdot40}=700\text{ km}$ distance, and an aircraft flying at 10 km can be seen from 350 km; always neglecting atmospheric and orographic effects (if only points above 10º over the horizon are considered, the range drastically reduces; e.g. drops to 60% for the ISS, i.e. $L_{10^{\circ}}$=1200 km).

Eclipses in high-eccentric-orbits (HEO, e.g. Hohmann transfer) may last up to 5 h, what is usually avoided by a proper choice of launch window. Besides eclipses by the planet, planet moons may cause eclipses, but because of their scarce frequency and short duration, they are not accounted for in thermal design.

Exercise 1. Deduce the eclipse duration in a equatorial circular LEO from orbital mechanics.

Sol.: There are many types of circular LEO orbits, which differ in orbit-plane position relative to the Sun-Earth axis, and altitude, H. The simplest analysis is made in the plane defined by the Sun-to-planet axis and the axis of the satellite orbit (lateral view in Fig. 5), because a single angle, β (angle between Sun-Earth axis and orbit plane), defines the orbit illumination. The general result is plotted in Fig. 6.

In an equatorial circular LEO, the angle between the orbit plane and Sun direction, β, coincides with Sun declination, δ, i.e. $\beta=\delta$, with $\beta_{\max}=\delta_{\max}=23.45^{\circ}=0.41$ rad (i.e. $|\beta|$=[0,0.41]), and its influence in eclipse fraction, T_e/T_o, can be neglecte, as seen in Fig. 6a for low altitudes.

In the special case of equatorial circular LEO at equinox (δ=0, β=0, i.e. an ecliptic orbit, also applicable to a polar orbit coplanar with the Sun), the eclipse fraction is simply the circumference fraction, $T_e/T_o=1-\phi_{es}/\pi$, with $\phi_{es}=\pi-\mathrm{arsin}(R/(R+H))$; see Fig. 5. For instance, for H=500 km, $\phi_{es}=\pi-\arcsin(R/(R+H))=\pi-\arcsin(6378/(6378+500))=1.95$ rad (i.e. 112º), and $T_e/T_o=1-\phi_{es}/\pi=1-1.95/3.14=0.38$, i.e. the satellite is 38% of its orbit period under eclipse.

From orbital mechanics, the acceleration balance is $\omega^2(R+H)=GM/(R+H)^2$, and thus $T_o=2\pi((R+H)^3/(GM))^{1/2}\approx 2\pi((R+H)^3/(gR^2))^{1/2}=2\pi((6378\cdot 10^3+500\cdot 10^3)^3/(9.8\cdot(6378\cdot 10^3)^2))^{1/2}=$ 5650 s (i.e. 94 min), what finally yields a value for the maximum eclipse time of $T_e=0.38\cdot 5650=2150$ s (36 minutes).

Exercise 2. Deduce the duration of GEO eclipses from orbital mechanics.

Sol.: The geostationary orbit is an equatorial orbit with a period of one sidereal day (86164 s, and, with (1), a=42164 km), with the satellite going East (as the Earth itself). GEO eclipses can only occur for very small solar declination, when $|\delta|<R_E/R_{GEO}=6378/42164=0.15=8.7^\circ$ (see Fig. 1 and Fig. 5). With the usual approximation $\delta=-23.45^\circ\cos(2\pi(N+10)/365)$, with N being day number from the 1st of January, the result is that eclipses can only occur from 27th February to 11th April and from 28th August to 11th October; with maximum duration at equinoxes, starting in this case at an orbit angle of $\phi_{es}=\pi-\arcsin(R_E/R_{GEO})=\pi-\arcsin(6378/42164)=2.99$ rad (171º). The eclipse fraction is then $T_e/T_o=1-\phi_{es}/\pi=1-2.99/3.14=0.048$, i.e. $T_e=86164\cdot 0{,}048=4140$ s (69 min; 73 minutes including the penumbra, which lasts 2.1 minutes at equinox). In terms of orbit period, which for GEO is 24 h, the eclipse fraction is 0.05, as shown in Fig. 6 for the last curve (H=36 000 km) at β=0.

Missions attitudes

The orientation of the spacecraft in relation to the Sun is of paramount importance to thermal control, as well as, to a minor extent, the orientation towards a planet. Basically, two type of orientation (attitude) can be considered (Fig. 7):

- Most spacecraft are three-axis stabilised to point back to Earth for communications or imaging (i.e. with one side of the spacecraft always facing nadir), with deployed solar panels tracking the Sun (they must counter-rotate slowly to compensate for the slow spin of the main body tracking the Earth).
- Spin stabilised spacecraft (10..1000 rpm; e.g. Meteosat at 100 rpm) are simple to operate and have simplified TCS (solar panels and antenna are usually fixed to the main body, although some large antenna may be de-spun, like in Intelsat).

Spacecraft attitude is particularly important for photovoltaic power. Solar panels must obviously be in in the outside of a spacecraft, and may be mounted as:

- Mounted on the walls of the spacecraft main body (as used in small satellites and spinners).
- Deployed (once in orbit) as attachments of the main body (as used in most satellites). The attachment can be:
    - Fix (through hinges) and moving as a solid body with the rest of the spacecraft.
    - Rotating panels (most often around one axis only), to maximise solar input.

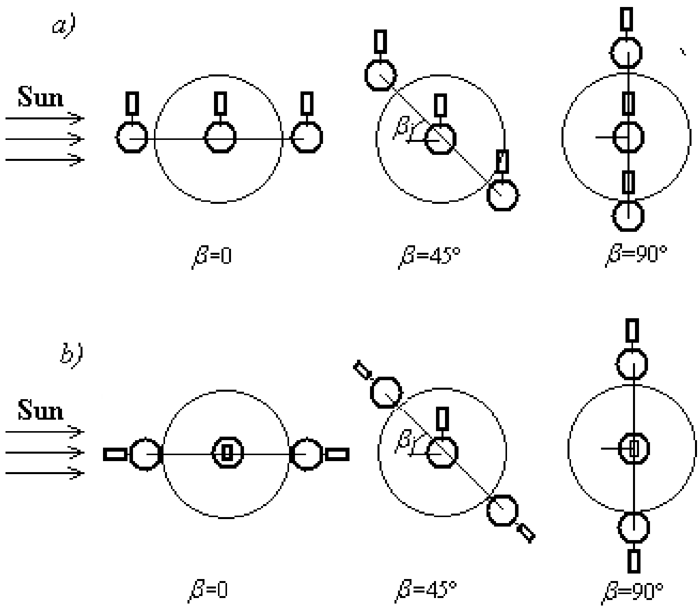

Fig. 7. Side view of satellite orientation relative to the Sun (β is the angle between the Sun and the orbit plane), for the two most used orbiting attitudes: a) inertial (spinning or three-axis fixed), b) nadir facing (i.e. facing the attracting body).

Some relevant cases

This is a simple compilation of mission data relevant to thermal control, for some different type of spacecraft of historical interest.

<u>Sputnik-1 data</u>

- The first spacecraft was launched by the USSR on 4th October 1957. It had a radio beacon that was used for communication research (ionosphere transmittance, by signal analysis), atmospheric research (by orbit tracking), meteorite research (by pressure and temperature monitoring of the filled nitrogen gas), and thermal control research (temperature was measured at the surface and inside).
- It was a hollow sphere of 0.58 m in diameter and 2 mm thickness, 84 kg in total, with two 2.5 m omni-directional antennas bent at their middle (looking as 4 whiskers pointing to one side).
- Its orbit had perigee at 215 km, apogee at 940 km, and 65º of inclination. It operated for 22 days until the 3 silver-zinc batteries went off, and fell after 3 months.
- Sputnik-1 thermal control used a highly polished 1 mm-thick heat shield made of aluminium-magnesium-titanium alloy, with temperature and pressure transducers encoded in the duration of radio beeps (if outside the 0..50 ºC nominal range). The downlink telemetry included data on temperatures inside and on the surface of the sphere. A temperature regulation system contained a fan, a dual thermal switch, and a control thermal switch. If the temperature inside the satellite exceeded 36 °C the fan was turned on and when it fell below 20 °C the fan was turned off by the dual thermal switch. If the temperature exceeded 50 °C or fell below 0 °C, another control thermal switch was activated, changing the duration of the of radio signal pulses.
- The sphere was filled with dry nitrogen, pressurized to 130 kPa. For the pressure control the satellite had a barometric switch, activated when the pressure inside the satellite fell below 35 kPa, changing the duration of radio signal impulse.

Shuttle data.

- Space Shuttle may refer just to the Orbiter, or to the whole Space Transportation System (STS) of the USA (a similar system was developed by the USSR, but it did not get operational). Its development started in the early 1970s, the first flight was on 12th April 1981, and it is to be decommissioned in 2010; the substitute program, Orion, expected to fly in 2014, was cancelled in 2009 in favour of commercial development of launchers). Six shuttles have been built; the first orbiter, Enterprise, was not built for actual space flight, and was used only for testing purposes. Five space-worthy orbiters were built: Columbia, Challenger, Discovery, Atlantis, and Endeavour. Challenger disintegrated 73 seconds after launch in 1986 (O-ring leak in the boosters, due to cold weather), and Endeavour was built as a replacement. Columbia broke apart during re-entry in 2003 (at launch, a frosty foam piece, detached from the external tank thermal insulation, damaged the Shuttle wing, causing lost of control and disintegration on re-enetry).
- The orbiter is 8.7 m in diameter and 37 m long (the payload bay is 4.6 m by 18 m long), with a mass of 69 000 kg (plus 25 000 kg maximum payload). It can put 24 400 kg into LEO, or 3800 kg into GEO. Before complete depletion of propellant, as running dry would destroy the engines, the main engines are shut down, stopping the oxygen supply first, since liquid oxygen reacts violently with hot engine metal. The two solid boosters are recovered by parachute, but the big fuel tank is disposed off during re-entry (the insulating foams burns away, and the heating causes a pressure build-up in the remaining liquid oxygen and hydrogen until the tank explodes).
- Shuttle thermal control. Once in orbit, there are eight radiator modules located in the back of the cargo bay doors, arranged to form two independent pumped freon loops (of R21, dichlorofluoromethane), each able to cope with 70% of the orbiter nominal cooling power, for partial redundancy. Each module consists of a curved aluminium honeycomb structure, 4.6×3.2 m^2 in size, some 2 cm thick, with an external aluminium face 0.28 mm thick, to which 26 parallel coolant tubes are bonded on the inner side, and a silver-teflon thermal control tape 0.13 mm thick is bonded on the outside. In 1999, additional strips of aluminium, 0.5 mm thick and 10 mm wide (5 m long), were bonded along the freon tubes to act as thermal doublers and, most important, to prevent tube piercing by space debris (in 2009 STS-128 got a debris impact that, where not for those doublers, would have impose a mission-abort and immediate return to Earth). When the bay doors are closed during ascent and re-entry, water-spray boilers provide the cooling (with ammonia-spray boilers activated during the descent when atmospheric pressure becomes greater than water vapour pressure).
- On re-entry, a large air drag is produced by sailing with 40º nose-up attitude around 120 km altitude at 8.2 km/s (Mach 25). Descent is un-powered, with lift to drag ratio varying considerably with speed, ranging from 1:1 at hypersonic speeds, 2:1 at supersonic speeds and reaching 4.5:1 at subsonic speeds during approach and landing. After landing, the vehicle stands on the runway for several minutes to permit the fumes from poisonous hydrazine and ammonia (N_2H_4 is used as a propellant for attitude control, and NH_3 for evaporative cooling), to dissipate, and for the shuttle fuselage to cool before the astronauts disembark.
- The STS cost breakdown is roughly 20% each: solid boosters, ground ops., flight ops., orbiter, and external tank (filled).

ISS data

- The ISS is an orbiting complex (about 400 km LEO) devoted to the study of 'living in space' (i.e. the effects on humans of the space environment), ancillary research on physical sciences, and stepping stone to further human exploration. In-orbit assembly began in 1998, and has been continuously inhabited since the first resident crew entered the station on 2nd November 2000, with a crew of six since 29 May 2009 (and more than double during crew exchange visits).
- Size. The truss length is 108 m, solar array span is 73 m, mass is 470 000 kg, 840 m^3 pressurised volume. Most data here down refers to the US-led part of the ISS; there is a smaller part contributed by Russia, with some peculiarities; e.g. each Russian module has its own solar panels, they use 28 V DC instead of 124 V DC, their thermal control system is independent and differs in working liquids, etc.
- Power. The ISS has 4 main solar array wings (double foldable blankets with solar cells on both sides, with central telescopic mat), each 34×12 m^2 in size, producing up to 32 kW at 160 V-DC (with DC-DC to 124 V end-use), plus nickel-hydrogen rechargeable batteries for the night (up to 35 minutes of the 90 minute orbit). The solar arrays normally track the Sun, with the "alpha gimbal" used as the primary rotation to follow the Sun as the space station moves around the Earth, and the "beta gimbal" used to adjust for the angle of the space station's orbit to the ecliptic.
- Thermal control. Habitable modules are wrapped in blankets (around 0.1 m tick) which are bullet-proof barriers for micrometeorite, ionizing radiation, and thermal transfer. Waste heat from inside the modules is evacuated by internal water loops to external heat exchangers; heat from these exchangers, plus heat from outside-mounted equipment (through cold plates), is transported by an anhydrous-ammonia loop (two independent circuits pressurised to about 2 MPa, going along the main truss, each capable of rejecting up to 35 kW), to radiator panels which emit to the cold empty space. Radiators are the second largest panels, after the solar wings. There are two main triple-radiators devoted to the main habitable modules, plus four single radiators each one devoted to cool the power conditioner at each solar wing. Each radiator consist of 8 panels of 4×3 m^2 deployed by a scissor mechanism to 25 m full length, has a mass of 1000 kg, and is made of bonded aluminium honeycomb panels coated with a white ceramic thermal paint (Z-93, with α=0.15 and ε=0.91) on 0.25 mm thin aluminium sheet, with the ammonia piping (2 mm internal diameter) embedded. All are orbital replaceable units (ORU), connected via flex hoses and quick-disconnect valves, and there are several spare units always attached to the truss (e.g. a radiator ORU is a box of 4×4×0.5 m^3). Radiator wings are usually edge to the Sun (to enhance cooling), and can be rotated ±115º to enhance heat release (edge to the Sun) or to prevent NH_3 freezing (at −77 ºC) by facing Earth during the eclipse phase (although the ammonia piping can tolerate freezing, with its 10% volume contraction; the worst cold case design is −93 ºC).
- Orbit. Mean altitude is around 390 km, with i=51.6º inclination to the Equator; solar beta angle changes some 4.5º/day, with bounds at ±(23.5+51.6)º=±75º. The ISS looses 80 m/day in altitude; it would take 1..5 yr to go down from the maximum height of 425 km (to be reached by

Soyuz) to the minimum of 300 km. The crew follows the Earth's circadian rhythm, waking up at 7:00 UTC (except when the Shuttle was docked, when Shuttle-MET was followed).

- Attitude. Three axis stabilisation of the main body. Normal attitude is local-vertical local-horizontal mode (LVLH), meaning that one side is constantly facing down Earth, another side is facing the speed vector, and a third one (the truss) is perpendicular; Columbus is at the front (to the right of Node 2, Harmony, where the Shuttle docks). It is maintained by control moment gyroscopes (CMG, each 98 kg spinning at 6600 rpm); when the CMG saturates, thrusters are fired. Solar wings tilt along their long axis, to point more directly to the Sun during daytime, and are on glide mode in night-time. The angle between the orbit plane of ISS and solar direction varies between +75° and −75°.
- The cost of payload uploading to the ISS is around 20 000 €/kg (the cost to send payload to the Moon is estimated to be five times that). Columbus module joined the ISS (starboard of Node 2, Harmony) on 11th February 2008; maximum waste heat is 20 kW.

Planck data

- ESA's Planck spacecraft, launched in May 2009 and reaching its working position at the Earth/Sun's L_2 Lagrangian point in July, is designed to observe the anisotropies of the cosmic microwave background over the entire sky. It has a cylindrical body 4.2 m long and 4.2 m in diameter, spinning at 1 rpm with its axis pointing along the Earth and Sun line, with a mass of 1800 kg. The circular base facing the Sun is covered with solar cells; there is sunlit not only because the L_2-point (at 1.5×10^9 m from Earth) is beyond Earth's umbra cone vertex (at 1.4×10^9 m), but because it follows a Lissajous orbit with 0.4×10^9 m radius. A common service module was designed and built by Thales Alenia Space in Turin, for both Planck and Herschel missions, holding all the electronics for communications, power, orbit and attitude control, coolers...
- Passive thermal control is performed by three conical shields and a telescope baffle, able to keep the sensors at 50 K from the 300 K of the service module (some 175 K at the first cone-shield, 125 K at the second, and 75 K at the third one). The two main sensors, however, require active cooling, one down to 20 K, and the other down to 0.1 K in three stages: at 20 K with a H_2 closed-cycle sorption cooler, at 4 K with a Joule-Thomson closed-cycle system with ^{4}He, and at 0.1 K with an expandable dilution cooler based on the endothermic ^{3}He-^{4}He mixing, what limits Planck lifespan to 15 months).

THERMAL PROBLEMS IN SPACE

Let us now change from spacecraft missions to thermal control. The aim of STCS is to solve the following thermal problems:

- The operational temperature of electronic systems has a narrow range of acceptable values, outside which, the equipment is disabled or permanently damaged. And the space environment is harsh on thermal loads, with wide load-span and sudden changes.
- Thermal stresses may be high, as when lightweight deployed parts suffer some 100 ºC temperature jumps within few minutes at the entrance or exit of eclipses.
- Thermal expansion due to temperature gradients may cause unwanted optical deflections and structural deformations. Some fine instruments, like the capacitive accelerometers in GCCE

gradiometer, demand temperature stability of the order of millikelvins (this is achieved by fine thermal control of outer shells and thermal insulation of the inner core).

- Some spacecraft must survive from aerodynamic heating in planetary descent (notably for human re-entry).
- Some equipment must be kept at cryogenic temperatures, like cryostats and infrared detectors (to increase the signal-to-noise ratio).

However wide-in-scope the above objectives may appear, we are just focusing on traditional thermal control systems in space; thermal problems in space are even more varied. A possible classification may be:

- Traditional STCS, i.e. delicate heating and cooling to keep a temperature margin, including thermal protection systems, refrigeration systems, thermal energy storage on phase change materials, etc.
- Thermally-driven electrical power generation (e.g. thermoelectric radioisotope devices, helium-closed Brayton cycle, organic Rankine cycles).
- Other thermally-driven machines, as an absorption refrigerator, a heat pump to raise radiator emission, etc.

To solve those problems, a general background on thermal engineering is required. First, a clear understanding of the thermodynamic terms:

- Temperature is 'the level' of thermal energy, such that, if two systems with different levels are brought in contact, thermal energy will flow as heat from the high level to the low level. It suffices here to know that there are many indirect means of measuring temperature by calibration of the mechanical or electrical response of a small probe (the thermometer).
- Thermal energy is 'the amount' of energy stored in the microscopic motion of the molecules within the body, such that the total energy of an isolated system cannot change, and for non-isolated systems an energy balance can be applied to compute thermal energy changes, ΔE. For the simplest model of a perfect substance of mass m and thermal capacity c, the energetic equation of state is $\Delta E=mc\Delta T$, and, the energy balance $\Delta E=Q+W$.
- Heat, Q, is 'the flow of energy' due to a temperature difference between two systems. The science of Heat Transfer studies heat flow 'rates', $\dot{Q} \equiv \mathrm{d}Q/\mathrm{d}t$, usually under the continuum model of local thermodynamic equilibrium.

Some preliminary questions

- What and how a thermometer reads? Ans.: It reads the temperature corresponding to the thermal balance between the probe and its environment (usually not at equilibrium), and not directly (by comparison with a unit), but indirectly by a calibration function in terms of a related magnitude (thermal expansion, thermoelectric coupling, electrical resistance…).
- What would be the reading of a mercury-in-glass thermometer inside an evacuated thermos, and in space? Ans.: Around 300 K, i.e. typical room temperatures. In the case of the evacuated thermos, the 'floating thermometer' would come into radiative contact with the walls and would finally equilibrate. In space, the 'floating thermometer', assuming it is far from a planet, would

come into radiative contact with the Sun and deep space, the same as for a planet or a moon; if the thermometer were under the shadow of a another body (a planet or a spacecraft), then the temperature would be much lower, but it would exchange radiant energy with this third body still.

- What does it mean then, that at 300 km altitude on Earth the temperature is around 1000 K? Ans.: It is a computed value, from kinetic energies of the particles, $E_k=(1/2)Mv^2=(3/2)RT$, where M is the molar mass if R=8.3 J/(mol·K) is used (in gas kinetic theory the mass of a molecule and Boltzmann's constant are more common) e.g. at room conditions, the root-mean-square molecular speed in air (M=0.029 kg/mol) is some 500 m/s, what gives a kinetic temperature $T=Mv^2/(3R)=0.03\cdot500^2/(3\cdot8.3)=300$ K. A sizeable object in this environment will take a very long time to equilibrate with this 'gas' if isolated, but radiative coupling with other celestial bodies prevent that.
- What temperature ranges are usually encountered in STC? Ans.: The thermal balance is driven by heat transfer (heat balance), since thermometers are rigid enclosures (and work exchange is minimal). In outer space, there are only very-hot light-emitting objects (>3000 K, only the Sun because other stars are too far away), and 'cold objects without own light' (<1000 K), the hottest being Venus surface (with 700 K) and the coldest deep space (at 2.7 K).

The generic thermal balance

The traditional formulation of thermal balance equation in STC follows the lumped approach of Thermodynamics, where subsystems are considered at uniform temperature (isothermal), instead of other discretization approaches (as the finite element model), or the partial differential heat equation in a continuum.

From the three types of thermodynamic systems: open, close, and isolated, the latter is only of theoretical interest (perfect isolation cannot exist), and open systems (i.e. those with mass exchange) are usually restricted to fluid flow in piping systems, so that the energy balance for a close system is the basic equation in STC, which, in power terms reads:

$$\frac{\mathrm{d}E}{\mathrm{d}t}=\dot{W}+\dot{Q} \quad \rightarrow \quad \dot{Q}=KA\Delta T \tag{4}$$

where $\dot{W}$ is the electrical power dissipated within the unit (or nuclear power, or chemical transformation, if any), and $\dot{Q}$ is the neat heat flow-rate, which is due to the temperature difference ΔT (it does not implies it is linearly dependent, or that it only depends on ΔT and not on T), is proportional to the flow area A, and K is an overall conductance coefficient, dependent on materials properties, geometry and temperature.

It may help to correlate the thermal energy balance terms with the more general balance equation (valid for mass, momentum, energy, entropy…): Accumulation = Production + Flux (the latter might be Diffusion fluxes and Convective fluxes). The Thermodynamics summary may be:

1. First law: there is no production/consumption term in the total-energy balance (although, when talking about the 'thermal energy balance', other energy contributions are considered sources or sinks).
2. Second law: there is a tendency towards equilibrium within an isolated system, or towards a steady state if the boundary conditions are steady (towards a periodic state in any case). Do not confuse equilibrium with steady state, although the term 'equilibrium temperature' usually refers to 'local steady-state temperature').

Exercise 3. Deduce the global thermal capacity of a small electronic box which, when thermally isolated, heats up at a rate of 72 ºC/hour with a power consumption of 2 W.

Sol.: The energy balance is $mc\,\mathrm{d}T/\mathrm{d}t=\dot{W}$, and thence the global thermal capacity is $mc=\dot{W}/(\mathrm{d}T/\mathrm{d}t)=2/(72/3600)=100\ \mathrm{J/K}$.

Heat transfer modes

There are basically two heat transfer modes: by contact (conduction and convection), and without contact (thermal radiation). In microgravity there is no natural convection: crew comfort and avionics cooling are achieved by ventilation with some forced-air, but there is no natural convection in weightlessness (halogen lamps cannot work in space).

Thermal balance in STC

In spacecraft thermal control, it is practice to rename the dissipated work within an element as internal heat generation, $\dot{Q}_{\mathrm{gen}}$ (negative for electrical production in solar cells), and split the heat flux in several terms: heat transfer with the space environment, $\dot{Q}_{\mathrm{env}}$ (accounting for heat received from the Sun and the orbiting planet), and heat transfer with other parts of the spacecraft by conduction, $\dot{Q}_{\mathrm{cond}}$, convection, $\dot{Q}_{\mathrm{conv}}$ (only applicable in special cases in space), and radiation, $\dot{Q}_{\mathrm{rad}}$. The energy balance (for a close system) then is laid out in the form:

$$\frac{\mathrm{d}E}{\mathrm{d}t}=\dot{W}+\dot{Q}\ \xrightarrow{\mathrm{STC}}\ mc\frac{\mathrm{d}T}{\mathrm{d}t}+\frac{dE_{\mathrm{ele}}}{\mathrm{d}t}=\dot{W}_{\mathrm{net}}+\dot{Q}_{\mathrm{net}}=\dot{W}_{\mathrm{in}}-\dot{W}_{\mathrm{out}}+\dot{Q}_{\mathrm{cond,net}}+\dot{Q}_{\mathrm{conv,net}}+\dot{Q}_{\mathrm{rad,net}} \qquad (5)$$

In spacecraft thermal control, thermal radiation is of paramount importance and is covered in detail aside when studying the Space environment and the Thermal modelling. For the time being, and with the purpose of advancing some STC calculations to be later developed in detail, it suffices here to recall that a body at temperature T (>0 K) emits electromagnetic (EM) radiation, which can be modelled by Stefan's law, $\dot{Q}_{\mathrm{rad}}=A\varepsilon\sigma T^4$, where A is the emitting area, ε is the emissivity (ε=1 for the ideal emitter known as black-body, to be here assumed), and σ is the universal Stefan-Boltzmann's constant, σ=5.67·10^{-8} W/(m^2·K^4). For instance a black-body at 300 K emits $\dot{Q}_{\mathrm{rad}}/A=\sigma T^4=5.67\cdot10^{-8}\cdot300^4=460\ \mathrm{W/m^2}$. The term "black body" was introduced by Gustav Kirchhoff in 1860.

An un-powered spherical object in space, exposed only to solar radiation of irradiance E (E=1360 W/m^2 at one astronomical unit distance from the Sun, 1 AU=150·10^9 m), and to deep-space background radiation, which is at 2.7 K and can be assumed to be at 0 K for calculations, the thermal balance at the steady state is simply $0=\dot{Q}_{\text{solar}}-\dot{Q}_{\text{space}}=E\pi R^2-4\pi R^2\varepsilon\sigma T^4$, showing that the steady temperature does not depend on object size, and it is simply $T=(E/(4\varepsilon\sigma))^{1/4}$. Notice that solar radiation is absorbed is proportional to frontal area, πR^2, whereas the body emits in the whole surface area, $4\pi R^2$. The dependence of solar irradiance E with Sun distance (decaying with the square of Sun-distance, d^2) and of steady black-body temperature (decaying with the square-root of Sun-distance, $d^{1/2}$), are plotted in Fig. 8, showing the relative position of solar planets, although it must be mentioned here that real planetary temperatures depart from this simple model, as explained under Thermal characteristics of planetary missions.

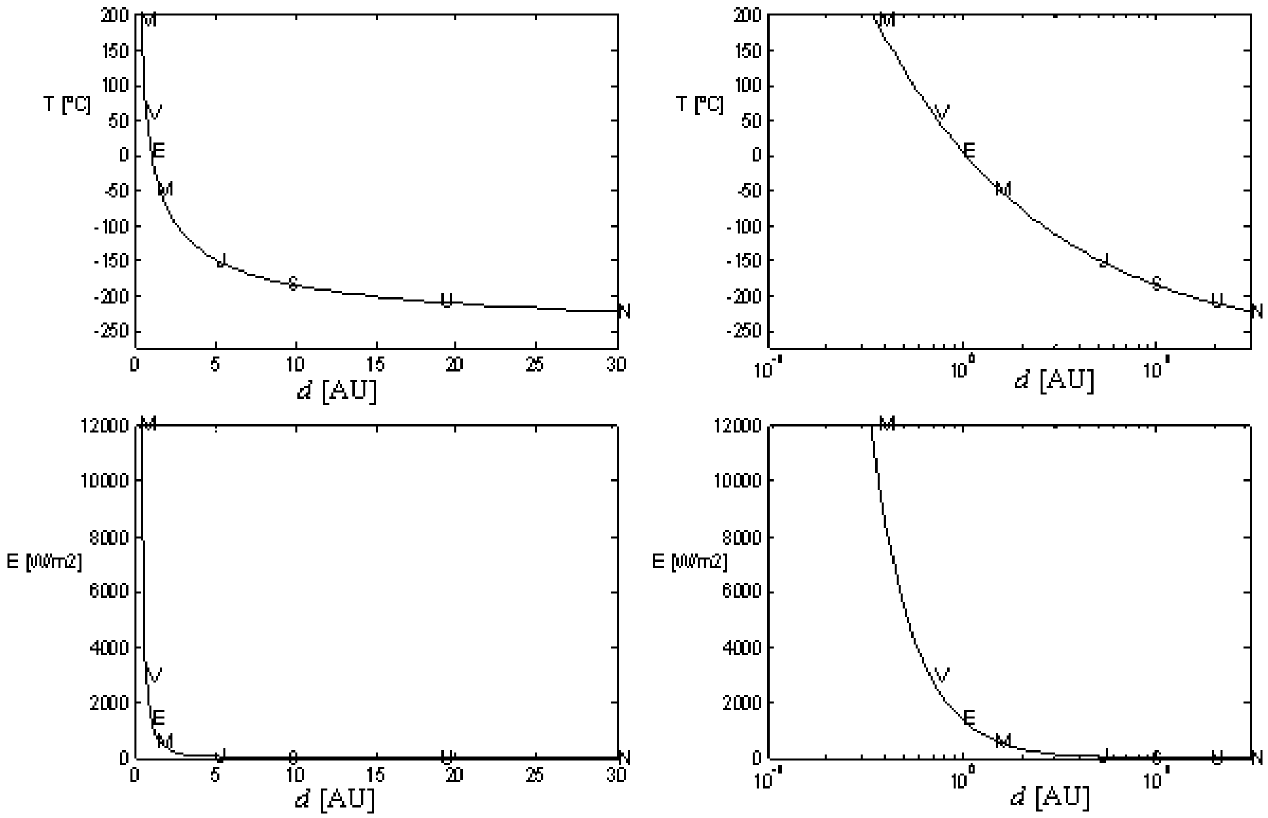

Fig. 8. Effect of distance to the Sun, on steady-state temperature for a passive black-body, T, and Sun irradiance, E, both in linear scale and semi-logarithmic scale. Solar planet positions are shown, but real surface planet temperatures are different because of their non-blackbody and non-passive character.

Exercise 4. Find the solar irradiance on Mercury, and on the Solar Orbiter spacecraft.

Sol.: We have just seen that at 1 ua, E=1360 W/m^2 (the 'solar constant'), and irradiation decreases with distance squared, $E=E_1(R_{SE}/R_{Sp})^2$, where $R_{SE}=150\cdot10^9$ m is the Sun-Earth mean distance, and R_{Sp} the Sun-probe distance. Mercury has an elliptic orbit, with $R_{SM,peri}=46\cdot10^9$ m (0.31 ua) at perihelion, hence $E_{M,peri}=E_1(1/0.31)^2=1360\cdot10.4=14$ kW/m^2, and $R_{SM,aphe}=70\cdot10^9$ m (0.47 ua) at aphelion, hence $E_{M,aphe}=E_1(1/0.47)^2=1360\cdot4.5=6$ kW/m^2. Solar Orbiter is expected to be launched in 2019, with a perihelion at 0.28 ua and an aphelion at 0.9 ua, so that it will be exposed to a maximum solar irradiance of $E_{SO,peri}=E_1(1/0.28)^2=1360\cdot13=17$ kW/m^2. In spite of the Sun proximity, solar radiation can still be considered as parallel rays (from 0.28 ua, the Sun disc is seen with a half-angle of 1º; 0.27º from Earth).

Exercise 5. Find the steady temperature of the Earth as if it were a black-body.

Sol.: The general expression for the steady temperature of an spherical black-body is $T=(E/(4\sigma))^{1/4}=(1360/(4\cdot5.67\cdot10^{-8}))^{1/4}=279$ K. The real average surface temperature on Earth surface is 288 K; the reason it is hotter than a perfect absorber (in spite of having a solar absorptance of just 70% of a black-body) is that its emissivity is still lower, 60% that of a black-body.

Exercise 6. Consider a small appendage of a GEO satellite at equinox, consisting basically on an aluminium sphere 0.2 m in diameter and 1 mm thick. Find the temperature evolution, neglecting thermal interactions with the rest of the satellite.

Sol:. We take as thermooptical properties of anodized aluminium α=0.15 and ε=0.80. The geostationary orbit (GEO) is so far (its radius is 6.6 times that of the Earth) that we can neglect thermal inputs from the planet (it can be shown that the view factor from a small sphere at GEO to the Earth is F_{sp}=0.006, the absorbed IR flux is $\dot{Q}_p$=0.14 W, and maximum albedo absorption is $\dot{Q}_{a0}$=0.04 W, decreasing with orbit angle and being zero during eclipse). Hence, from the steady energy balance, $\dot{Q}_{in}=\dot{Q}_{out}$, $\dot{Q}_s=\dot{Q}_\infty$, $\pi R^2\alpha E=4\pi R^2\varepsilon\sigma T^4$, we obtain the steady temperature, $T=(\alpha E/(4\varepsilon\sigma))^{1/4}=(0.15\cdot1360/(4\cdot0.80\cdot5.67\cdot10^{-8}))^{1/4}=183$ K (−90 ºC); a blackbody would reach 278 K (5 ºC).

Taking orbit angle ϕ from the subsolar point, we found in Exercise 2 that the eclipse period starts at 2.99 rad and ends at 2π−2.99=3.29 rad (i.e. it is in shadow from 171º until 189º). The geometry of GEO and the appendage results in a constant solar absorption except during the eclipse period, i.e. $\dot{Q}_{in}=\dot{Q}_s F_s$, with $\dot{Q}_s=\alpha A_{\text{frontal}}E$=0.15·0.03·1360=6.4 W, and the lighting factor of Fig. E6.1.

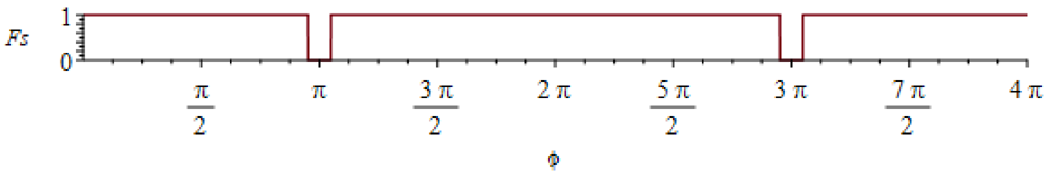

Fig. E6.1. Factor to account for eclipses.

The temperature evolution is obtained from the energy balance, $mc\mathrm{d}T/\mathrm{d}t=\dot{Q}_s F_s-\dot{Q}_\infty$ with $mc=A\delta c$=0.13·0.001·900=0.12 J/K, $\dot{Q}_\infty=\varepsilon A\sigma T^4$=0.8·0.13·5.67·10^{-8}·$T^4$ and any initial conditions, since $T(t)$ will become periodic after a while. Changing to angular variables with $\phi=2\pi t/T_o$, with GEO period T_o=86164 s (a sidereal day), the numerical simulation during two orbits, starting with an initial value $T(0)$=300 K to see its effect, is shown in Fig. E6.2.

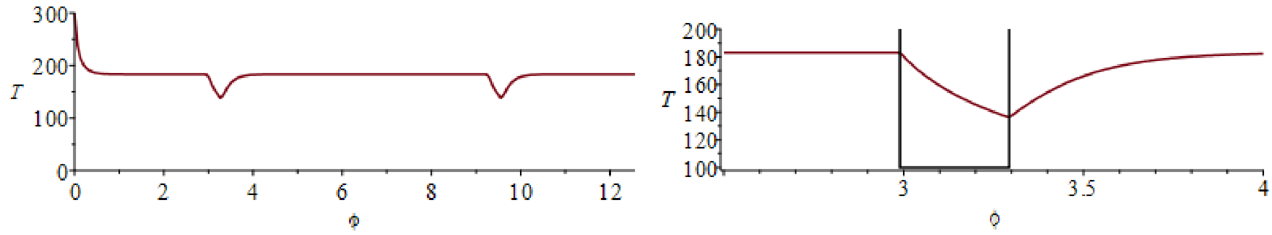

Fig. E6.2. Temperature T [K] vs. orbit angle ϕ [rad]. Numerical simulation for two orbits after a bizarre initial value (soon faded away). b) Detail of the cooling during eclipse, and the warming up.

The cooling when entering eclipse is governed by $4\pi R^2\delta\rho c\mathrm{d}T/\mathrm{d}t=-4\pi R^2\varepsilon\sigma T^4$, which can be integrated to $T(t)=T_{st}[1/(1+t/\tau)^{1/3}]$, where the relaxation time is $\tau=\delta\rho c/(3\varepsilon\sigma T_{st}^3)$ $=0.001\cdot2700\cdot900/(3\cdot0.8\cdot5.67\cdot10^{-8}\cdot183^3=2900$ s (49 minutes); i.e., after 49 min the initial temperature of 183 K would have fallen to $183\cdot[1/(1+1)^{1/3}]=145$ K, so that by the end of the eclipse (70 min, found in Exercise 2) the temperature would be at a minimum of 136 K, when then the Sun bathing would bring the temperature back to 183 K after a transient phase that can be analyzed similarly (the warming up is now governed by $4\pi R^2\delta\rho c\mathrm{d}T/\mathrm{d}t=\pi R^2\alpha E-4\pi R^2\varepsilon\sigma T^4$).

Power sources

Spacecraft need electrical energy to power the various spacecraft subsystems, and all power systems dissipate heat at the generator and at all consumers. Most spacecraft are powered by photovoltaic solar panels systems (SPS, deployable or body-mounted, with rechargeable batteries to provide electrical power during periods), because, in spite of their high initial cost, they are the most efficient in power/mass ratio, reliable, and safe. The exceptions are deep space probes, which are powered by radioisotope power systems (RPS), and short-time manned vehicles (e.g. Apollo, Shuttle), which are powered by fuel cells systems (FCS).

Solar panels are sets of solar cells that generate electricity by the photovoltaic effect, with an efficiency of η=15..25% for space-qualified cells, defined as maximum electrical power (depends on the load) divided by incident solar radiation (1360 W/m^2 at mean Sun-Earth distance). Roughly, half of incident solar radiation on a solar cell is absorbed and dissipated as thermal energy, with the other half split between the photovoltaic conversion, and the radiation reflected back at the surface. Efficiency degrades (some 2%/yr in power output) by ionizing radiation, outgassing, micrometeoroids, and thermal cycling. The record of solar powered distance was 2.6 AU by Rosetta spacecraft. On the contrary, for missions close to the Sun, solar cells will be so severely degraded by heating, that a thermopile like those used in RTG may be a better choice.

Radioisotope power systems do not get its energy from the environment but from radioisotope heater units (RHU) carried aboard, based on the natural decay of radioactive nuclei. This nuclear disintegration generates ionizing radiation (usually alpha-particles) which is readily absorbed by matter. Besides using RHU as direct heaters to keep sensitive deep-probe equipment warm, RHU are used to generate electricity in a so called radioisotope thermal generator (RTG), based on the thermoelectric process (Seebeck effect), with a typical efficiency of η=10%, or by thermoionic processes (based on the Edison effect, less efficient, but larger power range). The most used radioisotope is Pu-238 (in PuO_2 ceramic material form), a non-fissile isotope of plutonium with a half-life of 87.7 years and an initial heat release of 560 W per 1 kg of pure ^{238}Pu metal, which reduces to 6 W/kg of electricity when the whole RTG unit is considered (i.e. RHU, thermopile, casing, and heat sink). NASA is developing a Stirling Radioisotope Generator (SRG) producing 55 W electricity from 150 W RHU (i.e. with an efficiency of η=37% for an operating temperature of 900 K).

Thermal radiation

Several different meanings can be found in the literature, for thermal radiation:

- Thermal radiation is any EM-radiation which causes thermal effects (i.e. that may cause temperature changes): non-ionizing-UV, visible, IR, microwaves (on polar molecules), and any other radiation that is absorbed by matter, because it changes the internal energy of matter.
- Thermal radiation is the EM-radiation emitted by bodies by the effect of being hot (in absolute terms, i.e. depending on its temperature). Temperature is a marking of the level of microscopic agitation of atoms and molecules, and those microscopic motions create an EM-field that radiate energy outwards. We can see the radiation emitted by hot objects, as hot iron (1000 K), incandescent filaments (3000 K), and the Sun (6000 K); the average visual threshold is 800 K. By the way, halogen lamps (quartz bulbs) do not work well under microgravity, because they require natural convection for the regeneration of the tungsten filament. The thermal control of halogen lamps is very interesting, being able to change luminous efficiency from the typical 5% of the traditional glass bulb, to the typical 15% of the normal quartz bulb, up to the 35% when the bulb has an internal IR-reflecting coating.
- Thermal radiation is synonymous of infrared radiation (i.e. from visible to microwaves, 0.7 μm to 1000 μm, although there is little interest for λ>30 μm). The British astronomer Sir William Herschel, the discoverer of Uranus, found infrared radiation in 1800 when noticing that a thermometer placed below the red line in a prism was heated (he called it calorific rays); the effect is enhanced by the fact that the infrared rays are little dispersed (the refractive index of glass at 2 μm, glass is opaque to long infrared radiation, is n=1.50), while visible rays spread more ((the refractive index of glass at 0.4 μm is n=1.53). Herschel's experiment is easily reproducible if a glass prism is available.

Among the many characteristics of EM-radiation, direction and speed of propagation, collimation and coherence of the beam, intensity, frequency and phase of the wave, polarization, momentum…, the two important ones in thermal radiation (besides directionality) are:

- Power intensity. Thermal emission power can be modelled by Stefan-Boltzmann's law, $M=\varepsilon\sigma T^4$, with ε=1 for the ideal emitter (one that emits the most radiative power for a given temperature, corresponding to a small aperture in a large isothermal cavity of any kind of material or shape, or approximated by a black-painted surface).
- Spectrum. In thermodynamic equilibrium, radiation within an isothermal cavity (black-body radiation) cannot have all its photons with the same energy; because the distribution of maximum entropy is not uniform, $u=h\nu_{mean}=hc/\lambda_{mean}$, but that given by Planck's law:

$$u_\lambda = \frac{8\pi hc}{\lambda^5\left[\exp\left(\frac{hc}{k\lambda T}\right)-1\right]} \tag{6}$$

This is similar to the case of an ideal gas at equilibrium, whose molecules cannot all have the same average kinetic energy $E_k=(3/2)k_BT$, and the distribution of maximum entropy is given by Maxwell-Boltzmann's law of molecular speed distribution.

As for radiation-matter interaction, we are familiar with materials behaviour in the visible range (opacity and transparency, reflected colours, and so on), but we lack a feeling about the response to IR radiation. Most materials are opaque to infrared radiation, but there are some special materials that are nearly transparent to IR-radiation (e.g. germanium, sapphire, several salt crystals, and some thin films as the plastic film shown in Fig. 9). Infrared absorption in opaque materials may occur within the first micrometres in metals, within the first millimetre in non-metal solids and liquids, or require large absorption paths in semi-transparent gases liquids and solids. Care must be paid when taking about radiation matter interaction, to distinguish between interface properties and bulk properties; e.g. the high extinction rate in metals (high bulk absorption) should not be confused with their surface absorptance, which can be extremely low for polished surfaces, reflecting more than 90% of the incoming infrared radiation.

Fig. 9. Some apparent black bodies are not black-bodies: visible vs. infrared opacity. (http://en.wikipedia.org/).

Why is thermal radiation so important in space

Thermal radiation is of paramount importance in space because of the vacuum environment (and to a minor extent of the lack of natural convection within pressurised enclosures, although there is always some forced convection in habitable spaces). If different surfaces are exposed to solar radiation in the open air (on ground), we know that black surfaces get hotter than white surfaces by some degrees, but, if the experiment is performed under vacuum (as suggested in Fig. 10), differences in temperature might reach more than a hundred degrees.

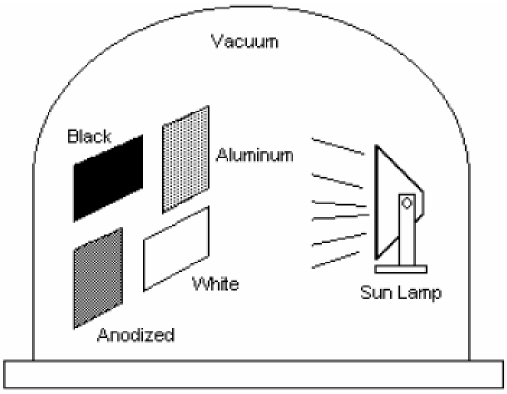

Fig. 10. A big change in radiative effects takes place from atmospheric to vacuum environment.

Why is spacecraft thermal control required

One may think that thermal control is only important in high power devices, but spacecraft are not powerful vehicles (the complete ISS has 100 kW peak, like a small car, whereas a ship or aircraft may be 1000 times more powerful). On the other hand, one may think that thermal control is only important in very hot and very cold environments. Of course, there are spacecraft missions going close to the Sun (e.g. Solar Power+ is foreseen to approach at 9.5 Sun radii, getting 510 times more solar heat than at Earth distances), or going far away from the Sun, feeling cryogenic temperatures, but most spacecraft operate in Earth orbit.

The main reason why thermal control is required in space, even when orbiting near the Earth, is the vacuum environment, as just said, which impinges on several aspects of spacecraft behaviour; basically:

- Spacecraft operate in a harsh environment that inflicts large thermal spans, up to a hundred degrees when entering or leaving an eclipse, even more when deep probes go near or far from the Sun. There are special cases with really severe heat loads, as for re-entry probes, air-breaking (the cheapest way to capture a probe around a planet with atmosphere), ascent…
- Spacecraft operate in a harsh environment that inflicts large thermal gradients by lack of a thermal bath, up to a hundred degrees between the sunlit and shadowed sides of the vehicle (rubber tyres in moon rovers may span from −50 ºC to 100 ºC from one side to the other, and rubber gets brittle below −60 ºC).
- Spacecraft systems are as delicate as living matter. Some human temperature limits may be worth recalling: inside the body 310±0.5 K is normal, whereas 310±5 K is a life risk (hospital); ambient air at 294±3 K means comfort, 294±10 K demands special clothing, and 294±30 K requires air heating or refrigeration (HVAC). Most delicate components for TCS are: sensors, batteries, optics, liquid reservoirs, joints and bearings. Why so narrow T-margins?: living matter, and non-living active components (electronics, but also mechanisms) cannot withstand large temperature excursions; thermal expansion may cause optical misalignment and structural failure (shell deflection, tile detachment), temperature shift may cause loose of calibration (bolometers must be kept at T_{calibr}±0.1 K). There is no safe threshold; electric-charge diffusion doubles every 10 ºC increase.

Basically, the TCS must keep all the components of a spacecraft within their respective temperature limits (very narrow), during all mission life, against the hostile thermal environment: no convection, direct solar radiation, cryogenic space, on board energy dissipation… Briefly, Spacecraft thermal design aims at:

- Avoiding overheating (damage, not recoverable) by proper heat rejection (only for interior planet missions).
- Avoiding overcooling (dormant, usually recoverable) with heaters (for some SC parts in interior-planet missions, and for the whole SC in exterior-planet missions).

What is actually controlled, temperature or heat?

Temperature is what we want to be controlled, and the way to do it is by controlling the heat fluxes that govern it. What is needed in STC, heating or cooling? In orbit, a spacecraft must operate around 300 K because, not only living matter, but electronics and other active components, have this operating range.

The task of thermal control at around 300 K may seem simple when realising that, in outer space, we have at our disposal a 5800 K heat source (the Sun), and a 2.7 K heat sink (the background space), but it is not so easy (think of mixing boiling water with icy water to take a shower; and sometimes the hot sources fades out, as in eclipses).

Several solutions have been worked out (to be detailed aside) to control heat fluxes and then body temperature; they are traditionally grouped as:

- Passive means (i.e. without moving parts and without external power), just with an appropriate design of geometries and materials, including PCM (phase change materials, not pulse-code-modulation) and heat pipes.
- Active means, (i.e. with moving parts or additional power), like heaters, thermostated louvers, pumped fluid loops, refrigerators, electro-optical coatings, etc.

In summary, the objective of STC is to adequately manage the different heat flow-rates, to control the temperature of each component (but recall that other means of temperature change might work in other instances, e.g. gas compression/expansion).

TEMPERATURE LEVELS, RANGES AND MARGINS IN SPACECRAFT DESIGN

Nomenclature:

- Level refers to location in the temperature scale, for the system or the environment.
- Range refers to values within which the temperature of the system is specified (i.e. maximum and minimum for given behaviour).
- Margin refers to difference between one range and the following, either on the maximum values or for the minima.

Temperature levels

What temperature values may be faced in STC? Ans.: In between the Sun surface 5800 K (i.e. the photosphere; the external Sun corona, at $(1..10)\cdot10^6$ K, is nearly transparent), and the deep-space background radiation at 2.7 K. The steady temperature of an isothermal spherical blackbody, is a good indicator of how hot or cold the space environment is, as a function of Sun distance (e.g. 180 ºC at Mercury, −222 ºC at Neptune). Special equipment like research furnaces, radioisotope heaters and generators, cryostats and so on, may pose special requirements. It is customary to divide the temperature scale in three levels:

- Cryogenic temperatures, if T<100 K (or T<200 K since there are few applications in between; ECSS-30 sets the limit at 120 K). The range T<200 K is of interest for image detectors and associated optics, for long-term sample preservation, for ground tests in vacuum, and for deep probes (further than Mars, e.g. 95 K for Europa or Titan). To reach temperatures of 200 K, reverse Turbo-Brayton coolers can be used; for 70 K, Stirling cryo-coolers; for 20 K, two-stage Stirling cryo-coolers; for 4 K, He-JT coolers, for 1 K, He3-sorption pump cooler; and for <0.1 K, adiabatic demagnetisation is used.
- Ordinary temperatures, if 200..500 K (i.e. including freezers and heaters). Normal working temperatures for most equipment on ground and in space is in the range 300±30 K. Sometimes

there may be short periods of extreme temperatures; e.g. a Mars lander usually works in the 200..300 K range, but during descent it may be exposed to T>2000 K (and the Mars Poles environment may be at 150 K).
- High temperature, if T>500 K (ECSS-30 sets the limit at 420 K). Usually for TPS, but also for Venus landing (735 K), and further approaching the Sun (Solar Orbiter perihelion is at 60 solar radii, 0.28 AU, and will get 17 kW/m^2, but passing at high speed in a highly elliptic orbit). During descent in planetary atmospheres, parts of the spacecraft must bear T>2000 K for a while; Shuttle tiles worked at 2000 K during many re-entries. Radioisotope heaters work at some 600..800 K.

Temperature ranges

Several temperature ranges are defined for thermal control:
- Optimum (i.e. desired for best performance and reliability).
- Operational (i.e. demanded to be within calibration tolerances). This is the basic design driver.
- Storage (i.e. not damaging when un-powered).
- Survival (i.e. low damage or partial failure).
- Isolated damage (i.e. confined damage, non-invasive to neighbour components).

Some temperature range data for operational state are:
- Batteries, −5..25 ºC while charging, −10..50 ºC while discharging. Minimum ∇T between cells.
- Star sensors, −5..25 ºC operational, −20..50 ºC in storage. Minimum ∇T in optical instruments.
- Propellants, −10..50 ºC (for safety). Hydrazine fuel must be kept above its freezing point (275 K).
- Electronics, −20..70 ºC.
- Electric motors, −40..80 ºC.
- Antennas, −100..100 ºC.
- Pyrotechnical devices, −100..100 ºC.
- Solar arrays, −190..110 ºC. The hottest solar cells developed (for work at 0.25 AU) withstand 115 ºC, with the cover glass layer at 117 ºC, a kapton rear layer at 114 ºC, and a copper base layer at 100 ºC.
- Multi-layer insulation (MLI) blankets, −150..230 ºC.
- External mechanisms, −200..130 ºC.

Temperature margin

Different temperature margins are sketched Fig. 11. Example data for an electronic box:
- Hottest qualification test: 70 ºC.
- Hottest acceptance test: 60 ºC. Qualification margin, 70−60=10 ºC.
- Hottest prediction, including foreseeable uncertainties: 50 ºC. Acceptance margin, 60−50=10 ºC.
- Hottest normal prediction: 40 ºC. Uncertainty margin of model, 50−40=10 ºC.
- Coldest normal prediction: 10 ºC.
- Coldest prediction, including foreseeable uncertainties: 0 ºC. Uncertainty margin of model, 10−0=10 ºC.
- Coldest acceptance test: −10 ºC. Acceptance margin, 0−(−10)=10 ºC.
- Coldest qualification test: −20 ºC. Qualification margin, −10−(−20)=10 ºC.

A synopsis of the relations between temperature ranges and margins is presented in Fig. 8. For instance, with the example values given before, if the manufacturer of some equipment establishes a safe operating range for it between −20..70 ºC, then the thermal designer must propose a TCS solution expected to keep the equipment in the 0..50 ºC, such that in the final acceptance test for the flight hardware it demonstrates to operate well in the −10..60 ºC range (only spare parts will be tested at the limit −20..70 ºC.

However, to for the TCS designer to demonstrate that the proposal is valid, an appropriate mathematical model must be run and show to yield predictions within a shorter range, 10..40 ºC for the normal cold and hot simulations.

Qualification margin
Acceptance margin
Uncertainty margin
Qualification test range | Acceptance test range | Design target range | Model prediction range | Worst hot case, Worst cold case
Uncertainty margin
Acceptance margin
Qualification margin

Fig. 11. Temperature ranges and margins in spacecraft thermal control design.

Back to Spacecraft Thermal Control