Oles Honchar Dnipro National University
Department of Space Information Technology

Vladimir Belikov

Control Systems of Launch Vehicles and Spacecraft

Slides

Володимир Вікторович Бєліков
Системи керування ракетно-космічних літальних апаратів
Слайди

Dnipro, 2020

Анотація

Цей комплект слайдів є одночасно ілюстративним матеріалом і конспектом лекцій для навчальної дисципліни «Системи керування ракетно-космічними літальними апаратами» англійською мовою. Дисципліна призначена як ознайомча для іноземних студентів, які вивчають різні аспекти ракетно-космічної техніки, але не системи керування. Метою дисципліни є ознайомлення студентів із колом питань, теоретичних аспектів і технічних рішень, пов’язаних з системами управління ракет-носіїв і космічних апаратів різного призначення.

Ця версія слайдів була сформована під час стажування їх автора з 13.01.2020 р. по 13.02.2020 р. на основі слайдів, розроблених для цієї ж дисципліни у квітні-серпні 2019 року, з урахуванням досвіду викладання дисципліни англійською мовою іноземним студентам у першому семестрі 2019/20 навчального року, і містить усі виправлення і уточнення, потреба у яких була виявлена у ході викладання дисципліни і під час стажування автора.

Oles Honchar Dnipro National University
Dept. of Space Information Technologies

Control Systems of Launch Vehicles and Spacecraft

Системи керування
ракетно-космічних літальних апаратів

Lecturer: Vladimir Belikov

Dnipro, Ukraine, 2019

About this course

This is an optional introductory course of the bachelor’s degree level for undergraduate students mastering aerospace engineering, while not specializing in control systems, however, needing to have general systemic understanding of the main principles, design and operation of the control systems as parts of aerospace systems.

Depending on the curriculum, the course may include:

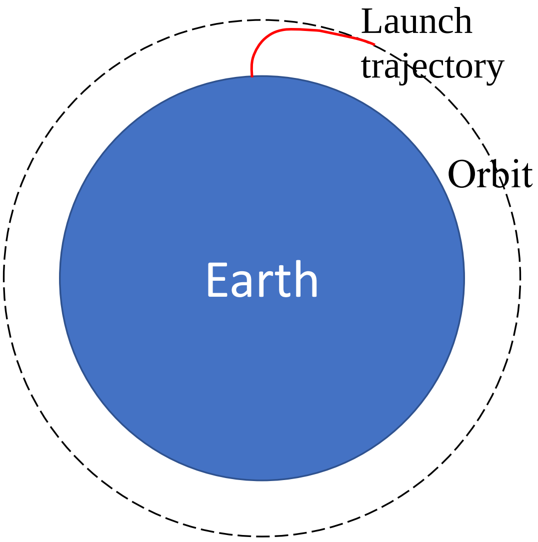
Lectures:	16 or 32 hours (1 or 2 lectures in 2 weeks)
Practice classes:	16 hours (1 class in two weeks)
Final control:	credit/no credit test

1

Introduction

Rockets, spacecraft and their control systems

Launch and insertion into an orbit



Most missions include orbiting the Earth. For satellites, orbiting is the main part of the mission. For interplanetary missions, it is a preparation to further flight.

The task of a launch is to bring a spacecraft from a particular point on Earth to a particular point in space in such a way that the spacecraft will come there at a required moment of time with a required vector of velocity.

This will ensure the spacecraft orbiting the Earth in the required orbit.

5

Control during the launch (1 of 3)

The main tasks of the control systems:

Keeping the rocket on the desired launch trajectory or close to it. This is to ensure the required parameters of the spacecraft at the end of the launch. This is the **flight control** task, a part of which is **guidance**.

Flight control includes dropping parts that are not necessary anymore – stages that have emptied the propellant tanks (**staging** or stage separation), the **fairing**, and, sometimes, some other parts. Finally, **deployment** of the spacecraft is performed (the spacecraft is separated from the final stage of the rocket).

Keeping to the necessary trajectory requires counteraction to **disturbances**, such as the wind and other irregularities of the atmosphere, vibration of the hull, oscillation of the fuel in tanks and others.

6

Control during the launch (2 of 3)

For this, the control system of the launcher needs:

To know the position of the rocket and parameters of its motion at any given moment of time relative to a particular reference system. This is the **navigation** task.

To know the **attitude** of the rocket, that is, angles between the axes of the reference system of the rocket and axes of the reference system, in which the launch is performed. This is the **orientation**, or **attitude determination** task.

To maintain the desired **attitude** of the rocket at any point of the trajectory. This is the **attitude control** task.

To maintain efficient operation of the thrusters by providing them with fuel and oxidant in optimal proportions. This is the **propellant flow control** task.

Control during the launch (3 of 3)

An additional task is collection of information about the state and motion of the rocket and transmission of it to the mission control center. This is the task of **telemetry**.

In some cases, receiving commands from the control center and their execution is necessary. This is rarely done during the launch, but may be necessary in some cases. Such commands may include certain corrections of the flight or even termination of the launch if it went completely wrong. This is the task of **telecommand**.

Sometimes, a rocket can collect scientific or commercial information during the launch and transmit it to the ground. For so-called ***sounding rockets*** (which are usually ***suborbital***, they do not reach any orbits), this is the main or the only purpose of the mission. This is the task of **data acquisition**.

Spacecraft missions

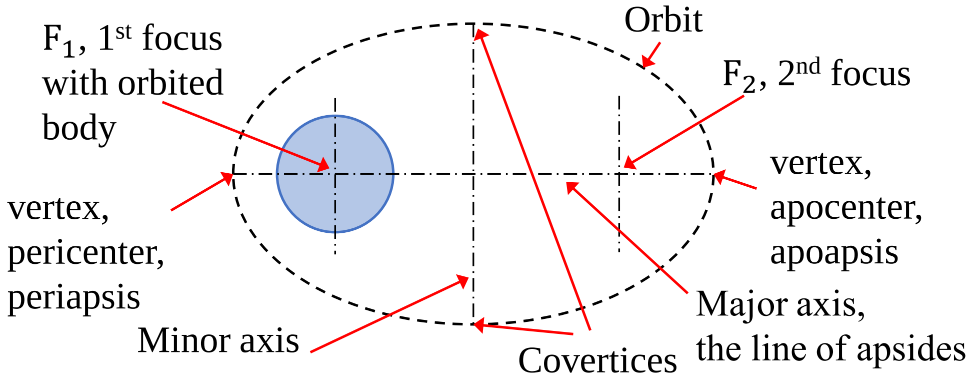
Missions of spacecraft vary substantially depending on the design and purpose of the spacecraft. They can be scientific, commercial or military, or combined. On some occasions, they may be educational (university satellites) or related to hobbies (amateur radio satellites). There are also utility missions (which can be commercial or free), like satellite navigation, maritime distress signaling, removal of space debris etc.

They can be **orbital, lunar, cislunar, or interplanetary**, or may involve orbiting a **Lagrange point** in the system “Sun-Earth” or “Earth-Moon”. Most common are missions involving orbiting the Earth.

Orbits around the Earth are called **geocentric orbits**. Other missions may involve orbiting other celestial bodies – the Moon, other planets, their moons, asteroids and comets, and the Sun.

Elliptical orbits

Orbits are typically elliptical. The orbited celestial body is in one of the **focuses** (or **foci**) of the ellipse. The highest point of an elliptical orbit is the **apogee** (around the Earth), **aphelion** (around the Sun), or **apocenter** (around other celestial bodies). The lowest point is **perigee**, **perihelion**, **pericenter**. Each of these two is also **apsis** (*pl.* **apsides).**



Orbit

F$_1$, 1st focus with orbited body

F$_2$, 2nd focus

vertex, pericenter, periapsis

vertex, apocenter, apoapsis

Minor axis

Covertices

Major axis, the line of apsides

Some Geocentric orbits (1)

The orbit of a satellite is chosen depending on its mission.

- **Low Earth orbits (LEO)** are those with the altitude of the apogee between 160 km and 2000 km above the sea level.
- **Medium Earth orbits (MEO)** are those with altitudes of the apogee from 2000 m to less than 35,786 km.
- **Geosynchronous orbits (GEO)** are those with the altitude 35,786 km. They are circular, and the orbital period of spacecraft in them is one **sidereal day**, 23 h, 56 m and 4 s approximately.
- **Geostationary orbit (GSO)** is a GEO in the plane of the Earth’s equator. A satellite stays above the same point on Earth.
- **High Earth orbits (HEO)** are orbits with the altitude of the apogee more than 35,786 km. Variants of HEO with the altitude of the perigee below 2000 km are **highly elliptical orbits**.

Some Geocentric orbits (2)

- **Polar orbits** are passing over the poles or close to them. They are used for mapping, reconnaissance (spying) and some weather and communication satellites.
- **Sun synchronous orbits** are close to polar orbits, but their altitudes are chosen such that satellites in them pass over the same latitudes at the same local time, making one turn in 2, 3, 4 or 6 hours. These are used for earth observation satellites and spy satellites.
- **Highly elliptical orbits** are used for satellites that should "hover" most of the time over a particular territory.
- **Transfer orbits** There are several kinds of orbits of this class. They are used for transfer of satellites between orbits.
- **Graveyard orbits** are for storing defunct satellites. A satellite from GEO is transferred into one of them at the end of its mission.

Spacecraft control

Certain control functions are common for all spacecraft missions, even though different satellites may implement these functions differently.

Navigation is determination of current location of the spacecraft in a particular reference system or several reference systems.

Attitude determination, control and stabilization ensures direction of sensors and antennas of the spacecraft as required for the mission.

Deployment of elements, like solar panels, antennas, etc. Operation of extendable and retractable elements and elements for docking.

Orbital maneuvers for correction of orbits, changing orbits, rendezvouses with other spacecraft, or for termination of the mission.

Reentry, including preparatory operations, reentry, and, for some spacecraft, landing.

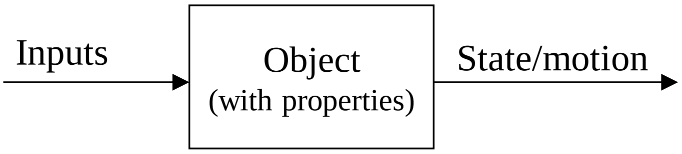
2

Basic notions

Elementary understanding of control systems

An object to be controlled

An object having “**properties**” performs “**motion**” under the influence of “**inputs**”. Inputs influence the motion depending on the properties of the object.



Inputs

Object

(with properties)

State/motion

Inputs are anything that influences the motion of the object – forces, voltages, currents, pressure, heat etc.

Properties are inertia of the object, determining how fast or slowly it reacts to inputs. They include mass, moments of inertia, electric capacity, electric inductance etc.

State of the object may include its location in space, velocity, voltage, current, temperature, pressure etc. And motion is the change of the state.

Inputs

These include **external inputs** and **control inputs**.

Examples of external inputs: gravity, aerodynamic force, voltage, amplitude, current, electromagnetic fields, pressure, flow, etc.

The source of some external inputs may be ***within the object***.

External inputs can be **known** and be either constants or functions of time and/or state and/or motion of the object (e.g., gravity). Or they can be **unknown**, affecting motion of the object in an unpredictable way. Such are called **disturbances** (e.g., wind).

C**ontrol inputs/influences** are created and applied to the object in order to make it move in a required way/pattern/direction, or make it maintain its current state.

Purpose of control

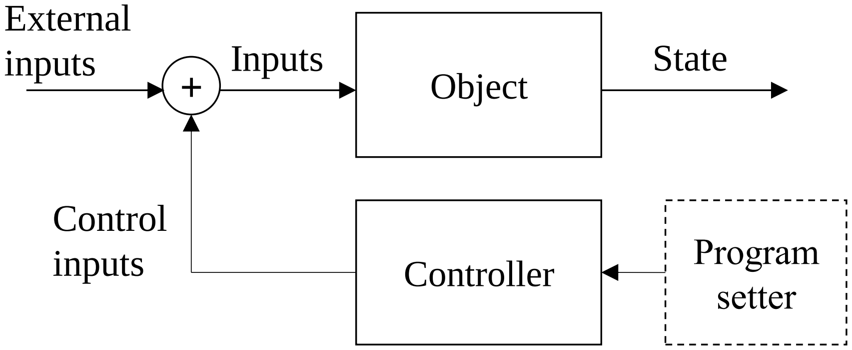
The object moves (changes its position or state) under the influence of external inputs. Not always this motion is desirable. Or, in some cases, it is not important how the object moves, but it is important that it should get to a particular position or state. There can be special requirements to achieving the desired state: either minimal time or minimal expense of energy.

The purpose of control is to ensure desired state/motion of the object given the known inputs and unpredictable disturbances.

For this, it is necessary to apply control inputs that can interact with or counteract external inputs in such a way, that the total of the external and control inputs will ensure the required state or motion of the object.

An open-loop control system

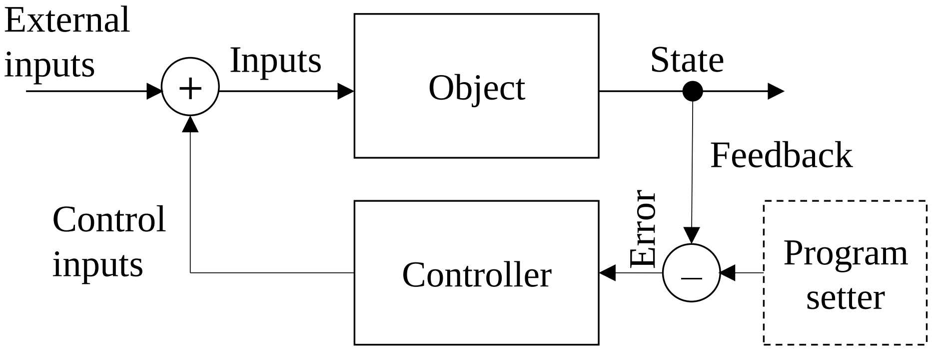
When external inputs are insignificant in comparison with control inputs or are totally predictable, it is possible to control the state and motion of the object by applying control inputs according to a predetermined program of required motion depending on time.



External inputs
+
Inputs
Object
State
Control inputs
Controller
Program setter

A feedback control system

Often, external inputs significantly and unpredictably affect motion. In this case, it is necessary to measure the deviation of the actual motion from the desired one and to produce control inputs that will ensure the motion to be close to the desired.



External inputs
+
Inputs
Object
State
Feedback
Control inputs
Error
–
Controller
Program setter

Sensors and actuators (1 of 2)

Usually inputs and motions are of different physical nature. Therefore, it is necessary to convert the physical nature of motion into the required physical nature of control inputs.

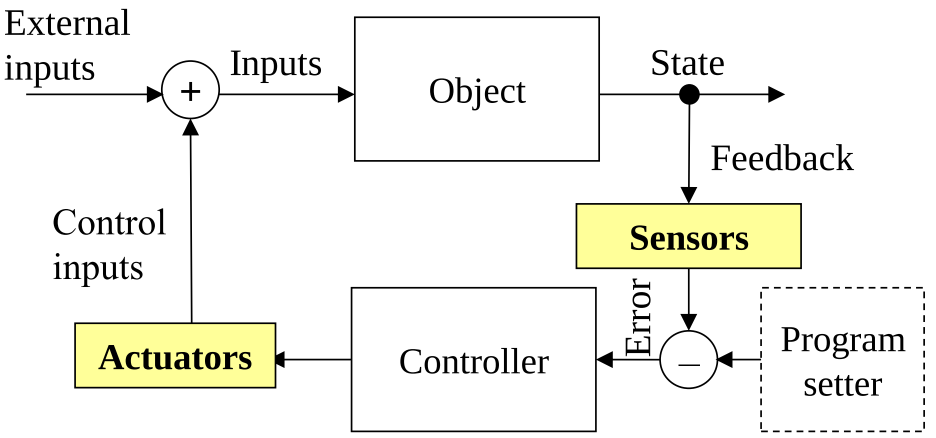
Moreover, physical nature of the controller is often different from that of both the inputs and motions of the controlled object.

Sensors convert parameters of motion of the controlled object into signals of an appropriate physical nature suitable for the controller.

Actuators convert control signals of the controller into control inputs of an appropriate physical nature suitable for affecting the controlled object.

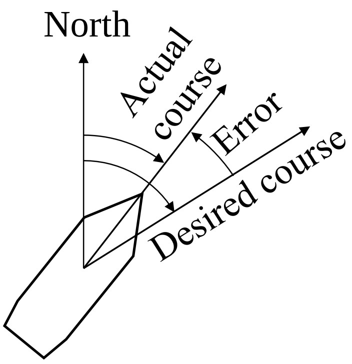
Sensors and actuators (2 of 2)

Sensors measure parameters of motion and provide input signals for the controller. Actuators convert output signals of the controller into the control inputs required for the controlled object.



A simple example of a control system (1/4)

Consider a ship that is to be kept for some time on a particular course (fixed azimuth). It can deviate from this course because of wind, streams in the sea and some other causes (external inputs). When it deviates, there is an error – the difference between the desired and actual courses, and this error has to be corrected.



A controller (a human or an automaton) measures the actual course (***motion*** or **state** of the controlled object), compares it with the desired course (***program***) and steers the ship by turning its rudder (***control inputs***) in order to bring the ship back to the required/desired course.

A simple example of a control system (2/4)

The controller should take into account that the ship does not respond to the turn of the rudder immediately. The ship starts and ends turning with a delay depending on its dynamic **properties**.

When the rudder is turned away from the neutral position, the water presses on the rudder and creates a **torque** (moment of force). The magnitude of the torque depends on what angle the rudder is turned, on the design of the rudder, the velocity of the ship and properties of the medium (water).

The torque creates **angular acceleration** of the ship around its center of mass. This causes increase of **angular velocity** of the ship, and the ship begins to turn. Also the ship gains **angular momentum** and **kinetic energy** of rotation – these depend on the angular velocity and the **moment of inertia** of the ship.

A simple example of a control system (3/4)

Now the ship is turning and gaining angular velocity, while coming closer to the desired course. **When should the controller return the rudder to the neutral position**?

When the rudder is back to the neutral position, the ship does not stop rotating immediately. It has angular velocity, and it has accumulated angular momentum and kinetic energy. Therefore, it keeps turning.

The momentum and the energy are gradually dissipated by the hydraulic drag of water; the ship slows down its rotation and finally stops turning.

Turn of the ship is not likely to stop exactly on the desired course. It may stop before or after this course, and then further controlled adjustment will be necessary.

A simple example of a control system (4/4)

Different ways of applying the control input.

- ► The rudder can be turned to the maximum angle, kept turned for a while and then returned to its neutral position.
- ► The rudder can be turned to the maximum angle, kept turned for a while, then turned in the opposite direction to create a torque for dissipating rotation, and then returned to its neutral position.
- ► The rudder can be turned proportionally to the error (deviation from the course).
- ► The rudder can be turned proportionally to the error and the angular velocity of the ship. The angular acceleration may also be taken into account.

The ship will move differently in each of these cases.

Steady state and transient process

A **steady state** is a state, in which an object remains in the same state (for instance, does not move, or moves with a constant velocity).

A **transient state** is a state, in which an object has not yet reached a steady state. The object can be in this state only temporarily.

In a transient state, the object moves, its state changes. This is a process in time, and it is called **transient process**.

When an object is in a steady state, and then its inputs change, its state ceases to be steady, and the object gets into a transient state. The transient process begins, which will end when the object reaches another state, which can be steady with new inputs. This takes time, and this time is called **transient time**.

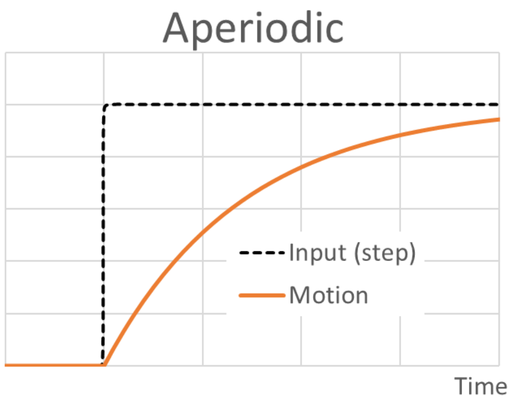
Transient process

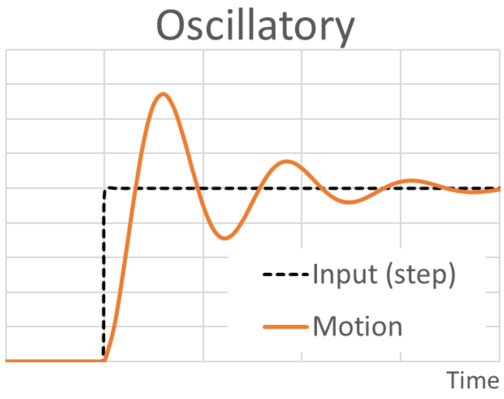
Because of **inertia** of various physical nature, the transient process does not happen in one moment, it takes some time – the transient time. In the transient process, the object gradually approaches the new steady state. How fast or slow this occurs, depends on the inertial properties of the object. It may even pass several times through the new steady state, not stopping in it.

The transient process is an important characteristics of a control system. The best way to analyze it is to apply the most adverse change of inputs and see what happens. And the most adverse change is a unitary **step change**, when the value of an input (or inputs) changes from 0 to 1 in no time at all. This is possible only theoretically, and therefore reactions to such changes are analyzed with the use of mathematical models rather than experiments.

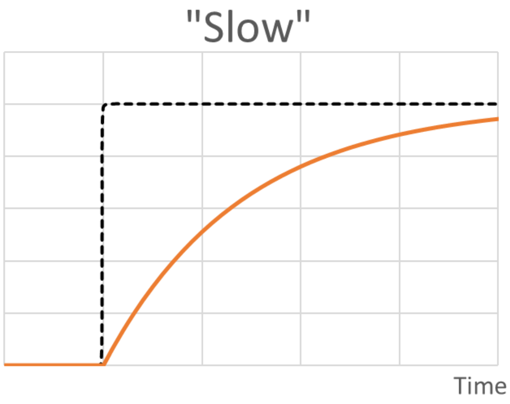
Aperiodic and oscillatory processes

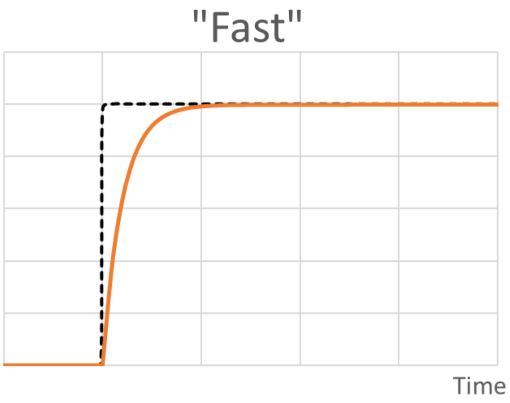
Approach of the object to the new steady state can be monotonous (aperiodic) or with oscillations, in which the object may pass through the steady state several times, before finally settling in it.





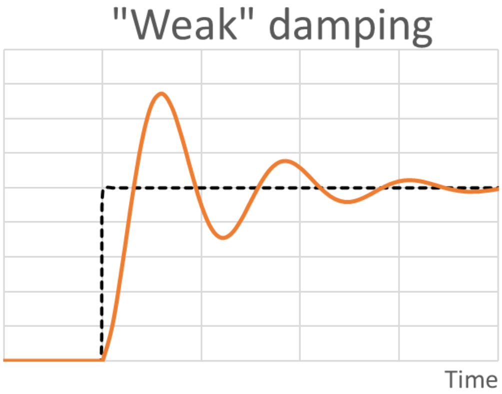
“Slow” and “fast” transient processes

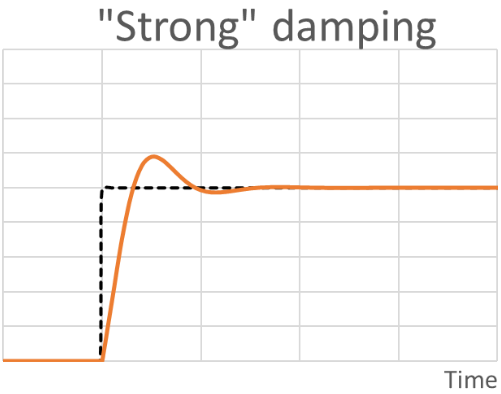




How slowly or quickly the transient process occurs depends on the inertia of the controlled system and on the power of the inputs, including the control inputs. By choosing properties of the controller, it is possible to make the transient process “fast enough”.

“Weak” and “strong” damping



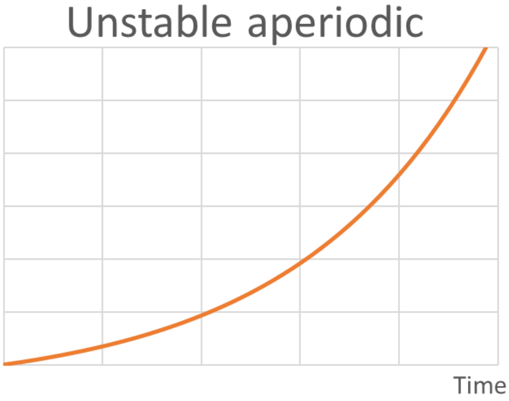


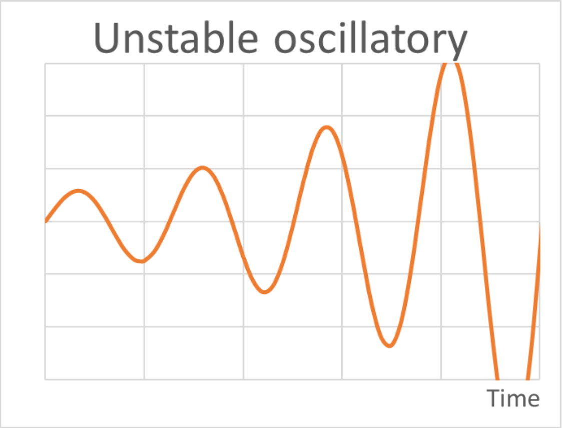
Oscillations depend on how much kinetic and potential energy the system accumulates and how fast one transits into the other. Damping occurs due to dissipation of that energy. The faster is the dissipation, the stronger is the damping. A controller can be designed for insuring required damping.

Stable and unstable systems

A **stable** system eventually arrives at a steady state. Previously shown diagrams are of stable systems.

An **unstable** system, once having started moving, never stops (this requires inflow of energy). It does not have a steady state.





Limits of instability

Theoretically, an unstable system, once having started moving, never stops. In reality, there are limits. The faster the system moves or accelerates, the stronger become the forces in it. At a certain point, these forces destroy the system, and its debris may appear to be quite stable.

Or an initially unstable system may become stable because dissipation of energy becomes equal to its inflow, or because of **nonlinearity** of its parameters.

Stabilization of an unstable system

A controlled object may be intrinsically unstable. An appropriate design of the controller can make it stable.

Transfer function

Motion of control systems is described by **differential equations**. When those equations are linear, they can be represented as **algebraic equations**. One way of doing this is the **Laplace transform**, as a result of which, operation of differentiation is replaced with the **Laplace operator *s***.

Transfer function (aka system function and, sometimes, gain function) is a theoretical expression of **the ratio** of the Laplace transforms of **the output to the input** of a control system or its element. It can be noted as W(s) or G(s).

$$W(s) = \frac{Laplace\ transform\ of\ the\ output}{Lalace\ transform\ of\ the\ input} = \frac{Y(s)}{X(s)}$$

Frequency response

Consider a unitary sinusoidal input, either external, or program. If the input was zero before the sinusoidal input is applied, the initial reaction of the system will be the **sum** of a **transient process** related to the properties of the system and the "**forced motion**" caused by the sinusoidal input.

After a while, the transient process ends, and the system (if it is stable) reaches a steady state, in which it moves sinusoidally, being forced by the sinusoidal input. However, the amplitude and phase of this "forced motion" may be different from those of the sinusoidal input.

Frequency characteristics show how the amplitude and/or phase of this motion depend on the frequency of the input, assuming that the input was applied "long ago", and the transient process has already stopped.

Bode plots – magnitude and phase

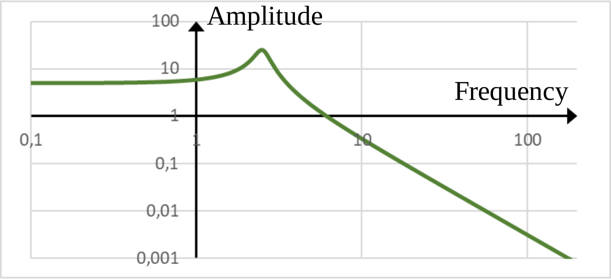
A Bode **magnitude plot** shows the magnitude of the response of a control system to a unitary sinusoidal input, depending on the frequency (usually, the angular frequency ω is used).

A Bode **phase plot** shows the phase of the response of the same.

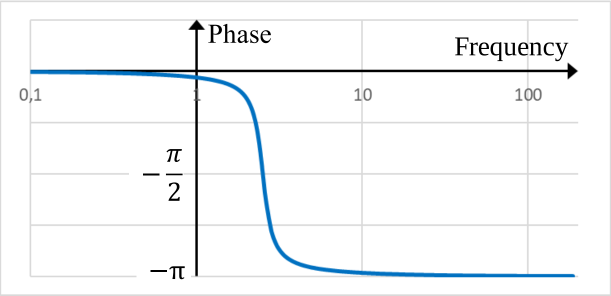
The horizontal axis of these plots represents frequency. And usually the scale of this axis is **logarithmic**; it shows the decimal logarithm of the frequency. The vertical axis of the magnitude plot is also logarithmic, while the vertical axis of the frequency plot is linear, in degrees or radians.

These two plots together are used for determining whether the system is stable or unstable.

An example of Bode plots



A Bode **magnitude plot**.
Both axes are logarithmic.
The horizontal axis represents frequency, and the vertical, magnitude of the output.



A Bode **phase plot**.
The horizontal axis is logarithmic and represents frequency. The vertical axis represents the phase shift of the output relative to the input.

The Nyquist plot (magnitude and phase)

A Nyquist plot shows both magnitude and phase of the response of an "**open-loop**" control system to a sinusoidal input. It is built on a complex plane for the angular imaginary frequency jω.

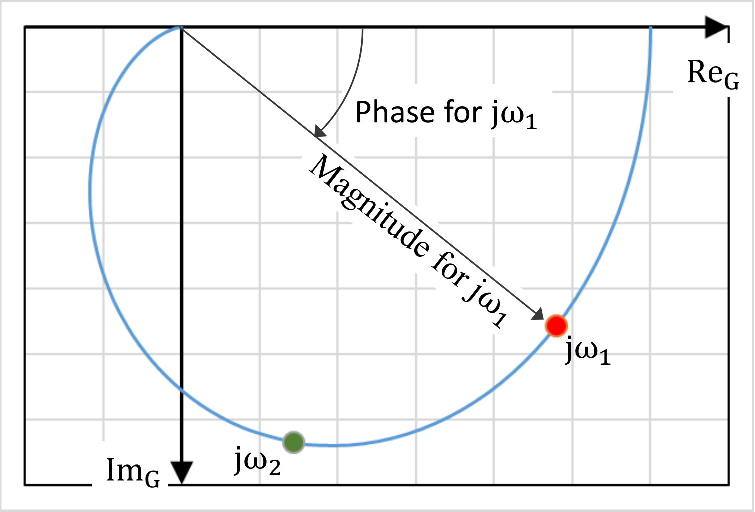
Each **point** on the curved line of the plot represent a particular **frequency**, the magnitude of the radius vector of the point represents the magnitude of the response, and the directional angle of the radius vector represents the phase of the response.

Basing on the Nyquist plot of an open-loop system, it is possible to conclude on its stability when the loop is closed, and the system operates as a feedback control system:

If the **Nyquist stability criterion** is satisfied, then the feedback system is stable.

An example of a Nyquist plot

This plot is of a simple element described by a second-order linear differential equation.



Marked points on the curve correspond to different frequencies:

$$\omega_1 < \omega_2$$

The curve comes to the 0 reference point at $\omega = \infty$

Linear and nonlinear systems

If a system consists of only **linear components**, it is linear.

A **linear component** is such that its output in the steady state is in a linear proportion to the input, or its integral, or its derivative. It can be described by a linear differential equation.
If a system contains at least one **nonlinear component**, it is non-linear.

A **nonlinear component** is such that its output in the steady state is not always or never linearly proportional to the input.

An ideally linear component is impossible. Some components are made as close to linear as possible within the range of their intended operation, but become nonlinear beyond that range. Others are made intentionally nonlinear. And yet others are intrinsically non-linear.

Basic linear elements

We are already familiar with the **aperiodic** and **oscillatory elements**. They are linear because they can be described by linear differential equations.

An **amplifying element** or **gain element** produces an output which is equal to the input times a constant gain.

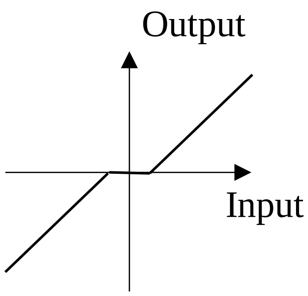
An **integrating element** or **integrator** produces an output which is equal to the integral of the input with respect to time plus the initial state of the integrator.

A **differentiating element** or **differentiator** produces an output which is equal to the derivative of the input with respect to time.

Not always accurate differentiation is possible. When the input changes too fast, its derivative is too big for implementing in a device.

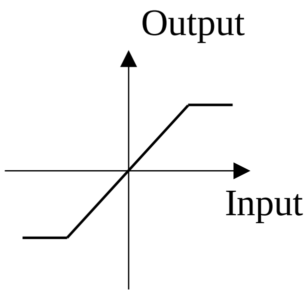
Some nonlinear elements (1 of 3)

A linear element with a **dead zone**.



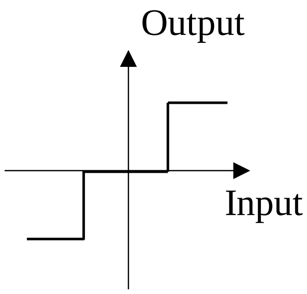
While the input is within a “dead zone”, the output remains zero.

A linear element with **saturation**.



When the input reaches a limit, the output stops changing.

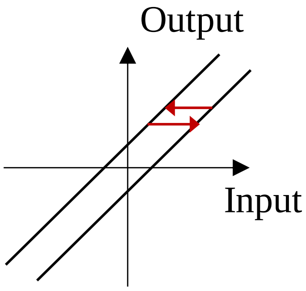
A **threshold** element.



The output is either zero or a particular level, depending on the input.

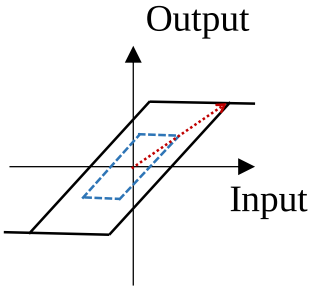
Some nonlinear elements (2 of 3)

A linear element with **backlash**.



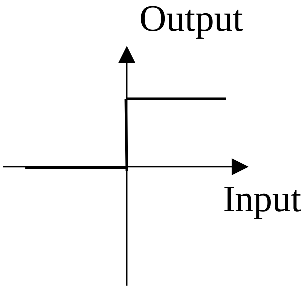
When the input, after growing, begins to decrease, the output does not follow immediately.

A linear element with **hysteresis**.



When the input is taken out of 0 along the red dotted line, it further moves along one of the inner loops (dashed line) or the outer loop, and never returns to 0.

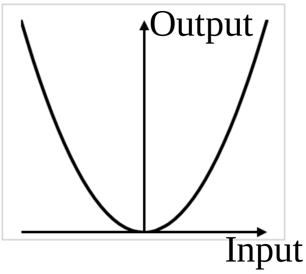
A **relay** element.



It is similar to a threshold element. It produces output in a certain range of input values, and produces 0 at others.

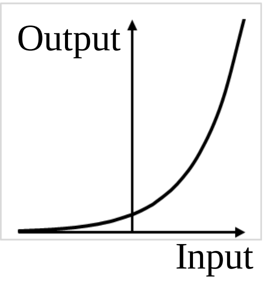
Some nonlinear elements (3 of 3)

A **quadratic** element.



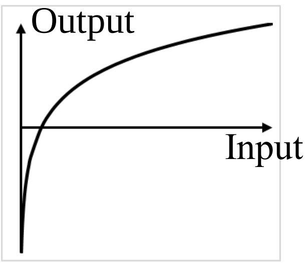
The output is proportional to the input squared.

An **exponential** element



The output is proportional to the exponent, a base raised to the power of the input.

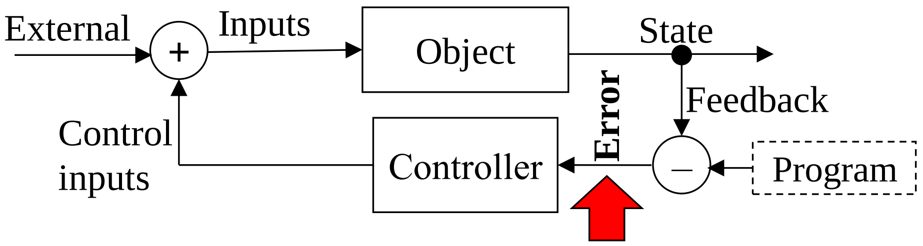
A **logarithmic** element.



The output is proportional to the logarithm of the input to a base.

Static and astatic control systems (1 of 2)

Consider the error of a feedback system.



A **static system** is a system, in which the error is not equal to zero, when it reaches a steady state. This residual error is a **static error**. That is, the steady state differs from the desired state set by the program.

The **static error** can be reduced by increasing the total gain of the controller, but it cannot be eliminated in a static system.

Static and astatic control systems (2 of 2)

The **static error** exists because, when the error supplied to the controller is small, the control inputs produced by the controller are not strong enough to counteract the external inputs.

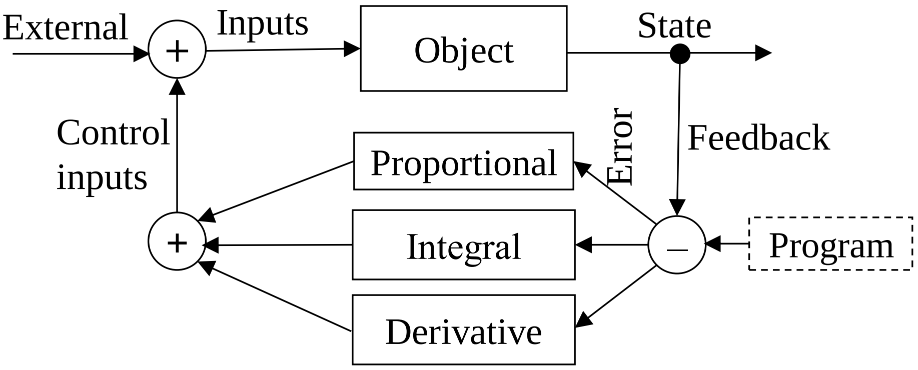
It is possible to increase the control inputs **by integrating the error** with respect to time. Then the control inputs stop changing only when the error is zero in the steady state.

An **astatic control system** is such that its error equals zero in the steady state.

Astatic control systems may have different **orders of astaticism**. The **first-order astaticism** (one integrator in the controller) ensures zero error in position. The **second-order astaticism** (2 integrators) eliminates the static error in speed. The **third-order astaticism** (3 integrators) eliminates the static error in acceleration.

The PID controller

PID stands for “**Proportional, Integral, Derivative**”. Its control input to the controlled object is formed as a sum of the proportionally amplified, integrated and differentiated error.



Correct combination of the 3 components provides fast reaching of the desired state with astaticism and strong damping.

3

Computers in control systems

Digital controllers of analog systems

A computer as a controller

In modern automation, often a digital computer is used as the controller or as its main part.

While controlling an object, a computer performs a control algorithm, which uses information on the current motion/state of the object and its desirable motion/state and computes required control inputs in order to bring the actual motion/state as close to the desired one as possible.

A single computer can act as several controllers of several objects by distributing its processing power and memory capacity for solving several control tasks either for several controlled objects or for several subsystems of the same controlled object.

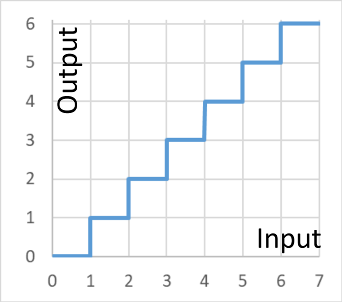
Convertors: ADC and DAC

Most digital computers operate in binary numbers. Their inputs and their outputs are **digital**. Moreover, they are **discrete**, because computers operate with a limited number of significant digits. Most parameters of motion, as well as external and control inputs of a controlled object are **analogous**. That is, they are **continuous** functions of time and other parameters.

Therefore, information about the motion of the controlled object should be presented as binary numbers. This is done with a device called **analog-to-digital convertor (ADC)**.

The output of the controlling algorithm is control signals in the form of binary numbers, which have to be converted into analog signals. This is done with a device called **digital-to-analog convertor (DAC)**.

Nonlinearity of the ADC

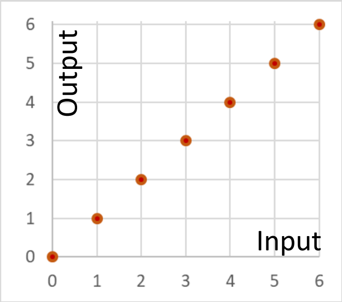


An ADC introduces a nonlinearity in the form of a step function, because its input is typically a continuous value, while its output is an integer binary number. Usually the number of steps is chosen such that this non-linearity can be regarded as almost neglectable.

Steps may be linearly proportional to the input, when each step corresponds to the same change of the input signal. Or they may be proportional to the logarithm of the input signal for increasing sensitivity to low values of the input signal.

The number of steps depends on the bit capacity (the number of bits) of the ADC. For instance, a 12-bit ADC produces 4096 steps.

Nonlinearity of the DAC



A DAC introduces a nonlinearity in the form of a dot function, because its input is an integer digital number, while its output is an intrinsically continuous value proportional to the input. With enough steps, this non-linearity can be regarded as almost neglectable.

Similarly to the ADC, output value may be linearly proportional to the steps of the input number, when a change of the level of the output corresponds to the same change of the input number. Or it may be proportional to the logarithm of the input number for increasing sensitivity to low values of the input number.

Latency

ADC samples the signal from a sensor once in an interval of time. Then it takes some time to convert it into a number, and only after that it becomes available to the controlling computer.

DAC also takes time to convert its input in digital form into an analog signal. It receives a number from the computer, and, while converting it into its output, cannot receive another number.

The computer reads the data from the ADC, processes them according to the control algorithm and sends data to the DAC. These processes also take some time.

All these delays together form the latency. If processes in the controlled object are so fast that during latency the state of the object changes significantly, then effective control becomes impossible. If this is the case, you need to make faster all these components.

Operating systems

An operating system (OS) provides **services to an applied program**, which include **input/output (I/O)**, access to data through the **file system**, **allocation of memory**, **allocation of the CPU time** and other resources to the applied program, testing and monitoring of the conditions of devices and the memory and other services.

A **multitasking OS** protects memory and other resources allocated to one applied program from access and modification by other simultaneously executed programs. Also it provides ways of communication between such programs. It maintains a discipline of access to resources necessary to many programs, which includes **prioritizing**, **queueing** and **spooling**.

The **OS requires resources** for its operation: CPU time, memory and others. This has to be considered in choosing the computer and OS. **Real time OS** usually provide better performance in control systems.

Program interrupts

Program interrupts are a way of ensuring **fast reaction** to events outside the computer or inside it.

An interrupt system of the computer monitors signals indicating **events requiring reaction**. When an event occurs, it sends an interrupt signal to the computer. For each event or type of an event there is a separate interrupt.

When an interrupt occurs, the CPU suspends execution of a current program and switches over to execute a program performing the necessary actions – an **interrupt handler**. After that, the CPU returns to executing the interrupted program.

Reaction to interrupts is different in different OSs. A **real-time OS** usually reacts to interrupts 10 to 20 times faster than other kinds of operating systems that can be run on the same computer.

Microcontrollers

A microcontroller is a **single-chip computer** with built-in memory, auxiliary devices, several I/O ports and, sometimes, ADC and DAC.

Microcontrollers are often used as **controllers of particular devices** of the controlled object, providing faster reaction to events occurring in it and reducing the computational load on the main computer.

The main computer can set required parameters of operation of each controller and receive information on the current state of the controlled device or subsystem. This gives the main computer the **overall control** over the controlled system as a whole, while leaving the tasks of immediate **local control** over certain subsystems and devices to microcontrollers.

Microcontrollers are often based on the **Harvard architecture**, while main computers are usually of the **von Neumann architecture**.

Ionizing radiation and electronics (1 of 2)

Ionizing radiation includes very **short electromagnetic waves** (**gamma rays** and **X-rays**) and **elementary particles** (mainly protons, neutrons and electrons).

This radiation is represented by gamma rays, X-rays and **cosmic rays** from stars and galaxies, and by the solar radiation and the **solar wind**.

The **cosmic rays and solar wind are flows of elementary particles**. The intensity of the solar wind is much higher than of the cosmic rays, but energy of some particles of the cosmic rays can be extremely high.

The **surface of the Earth is protected** from a large portion of this radiation by its atmosphere and magnetic field. But this protection is absent in space.

Ionizing radiation and electronics (2 of 2)

Electronics based on semiconductors (which includes almost all computers) is sensitive to this radiation because operation of semiconductor devices depends on flows of negatively charged electrons and positively charged "holes".

Two types of influence:

1. **Gradual deterioration** of properties of semiconductors.
2. **Single event effects** (SEE), which include:
   - Transient: single event **upset** (SEU);
   - Permanently damaging: single event **gate rupture** (SEGR), single-event **latchup** – a kind of a short circuit in a semiconductor device (SEL), and a single event **burnout** (SEB).

An SEE may change the value of a bit of data and cause software failure**.**

Ways of protection: software and hardware **watchdogs**, **hardware redundancy**, **data redundancy**, **shielding.**

Van Allen radiation belts

The **magnetic field** of the Earth captures many charged particles from the solar wind and redirects them to the poles, causing **auroras**: **Aurora Borealis** near the north magnetic pole and **Aurora Australis** near the south magnetic pole.

The areas of space where charged particles are captured are known as **Van Allen radiation belts**. They are characterized by high intensity of ionizing radiation consisting of charged particles.

The belts occupy altitudes from about **640 km** to about **58,000 km**. Usually there are two of them, the inner and the outer belt, and sometimes one more belt appears between them. Their intensity and size depend on the solar activity.

The metallic hull and plastic inner lining of spacecraft absorb most of the charged particles, however, particles cause **secondary X-ray radiation** when they hit the hull.

4

Launch vehicles as controlled objects

Inertial properties, forces, torques, disturbances

59

Forces acting on a rocket

The main forces acting on a launch vehicle during the launch are the **gravity**, the forces related to the **atmosphere** (mainly **aerodynamic**, but also **aerostatic** forces) and the **thrust**. Of these, usually only the thrust can be controlled. Sometimes, aerodynamic forces too.

When the vector of a force is not on the same line with the **center of mass** of the rocket, the force additionally creates a **torque**.

Gravity does not create noticeable torque; it can be brought to a resultant force applied to the center of mass. The resultant of aerodynamic forces is applied to a point not coinciding with the center of mass. The thrust may deviate from the longitudinal axis either intentionally (for the purpose of control) or because of design and manufacturing errors.

Disturbances are mainly caused by the atmosphere (the wind), by vibrations of the rocket and by sloshing of the propellant in the tanks.

60

Gravity (1 of 2)

Gravity is one of the four fundamental forces, and it is the force of mutual attraction of bodies having mass.

When the **distance between two objects is big**, much bigger than the size of the bigger of the two, **or** the objects are **perfectly spherical**, and masses are evenly distributed in them, we can assume that the attraction force is applied to the centers of mass of each object and use Newton's universal law of gravity formula with sufficient accuracy:

$$F = G\frac{m_1 m_2}{R^2}$$

Where $G \approx 6.67408 \times 10^{-11}\ \mathrm{m}^2 \cdot \mathrm{kg}^{-1} \cdot \mathrm{s}^{-2}$ is the universal gravitational constant, m_1 and m_2 are masses of two attracting objects and R is the radius between their centers of mass.

The Earth is not perfectly spherical, its actual shape is called "geoid".

Gravity (2 of 2)

The **polar radius** of the Earth is by ≈21 km shorter than the **equatorial radius**. Not much, taking into account that the average radius is 6371 km. Besides, masses are unevenly distributed inside the planet, and this unevenness is tiny in comparison with the size of the Earth.

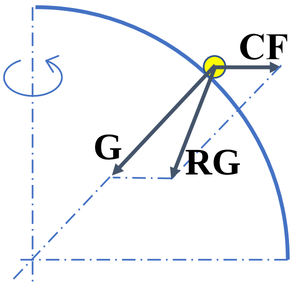
In many practical applications you can safely ignore these irregularities and assume that the Earth is a perfect ball with evenly distributed mass, but not when you deal with launching rockets and spacecraft.

For more accurate calculations, a **reference ellipsoid** model is used. The mass of this ellipsoid is equal to the mass of the Earth, 5.97237×10^{24} kg. And it is built in such a way that the total of the squared deviation of the points on the actual surface of the Earth from the respective points on the ellipsoid is minimal.

When even higher accuracy is needed, a **point cloud** model is used. In such models, Earth is viewed as a cloud of points with different masses.

Not only gravity

An object on the surface of Earth experiences **centripetal acceleration** due to the rotation of the planet. It creates a non-existing force acting on the object – the virtual **centrifugal force**, which, in fact, is a manifestation of inertia.

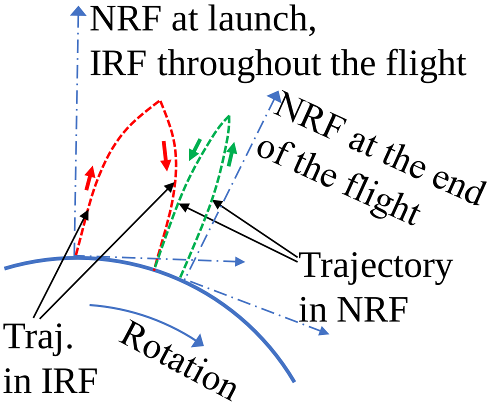


Here, **G** is the force of gravitational attraction, **CF** is the virtual centrifugal force, and **RG** is the resultant "gravitational" force, which is different both in magnitude and direction from G.
At the equator, CF is at maximum and opposite to G. At the poles, CF=0. Non-sphericity of Earth and the centrifugal force result in different free fall acceleration in different places.

The standard gravity $g_0 = 9.80665\ \mathrm{m \cdot s^{-2}}$, polar gravity, $g_P = 9.8321849\ \mathrm{m \cdot s^{-2}}$, equatorial gravity, $g_E = 9.78033\ \mathrm{m \cdot s^{-2}}$

Coriolis force

This is another virtual force, which, is a manifestation of inertia. It is observed in rotating non-inertial reference frame (**NRF**). Consider a vertical launch of a sounding rocket from the equator and neglect the atmosphere. We fix an inertial reference frame (**IRF**) at the moment of launch, while the NRF rotates with the Earth.



At launch, the rocket has horizontal velocity equal to that of the point of launch. It causes it move eastwards in IRF. However, the velocity of any point above this point is higher. This causes the rocket drift westwards in NRF, as the point of launch moves east.

Buoyant force

Launch vehicles are rather big. One of the smallest, Lambda 5S designed in Japan and capable of launching 5 kg to LEO, was 16.5 m. long and 0.74 m in diameter. Its volume is just 7 m^3. There are much bigger rockets, volume of some of them far exceeds 1000 m^3.

According to the **Archimedes' principle**, the rocket experiences the buoyant force equal to the weight of the displaced air. This is an **aerostatic force**.

The **density** of the air at the sea level is about $\mathbf{1.2\ \ kg}/\boldsymbol{m}^{\mathbf{3}}$. Then each cubic meter of displaced air creates the buoyant force of **11.77 N**. At the altitude of 5 km this force is reduced by almost half, and at 10 km, approximately by 4 times.

Even though this force is small in comparison with other forces, it still may need to be taken into account, especially at determining the mass of the rocket.

Bottom drag

When a rocket moves through the air at high speed (usually supersonic or hypersonic), and its engines are switched off, an **area of low pressure** (“vacuum”) forms behind the rear end of the rocket.

Normally, the rocket flies with its engines operating. However, this area forms for a short period of **“cold” staging**, when the empty stage is separated and pulled off the remaining stages, and only then, after some delay, the engines of the next stage are activated.

This low-pressure area tries to suck the rocket in, that is, creates an aerostatic force opposing the motion – **the bottom drag**.

If this happened close to the sea level, this force could be significant, around 80 to 90 kilonewtons per square meter, However, staging occurs at high altitudes, where pressure of the air is low, and staging takes several seconds, therefore, contribution of this force is rather insignificant (but not neglectable).

Aerodynamic forces

The geometry of launch vehicles is usually simple. In many cases, it is a cylinder with one end tapered and the other end blunt.

When the **incident flow** (which is the result of motion of the rocket through the air or motion of the air relative to the rocket) occurs along the longitudinal axis of the rocket, there is only the **longitudinal aerodynamic force**. When there is a non-zero angle between the incident flow and the longitudinal axis, there is also the **transversal aerodynamic force**.

The main causes of these forces are:

- Increase of the pressure where the incident flow meets the rocket. At supersonic speed, this increase becomes a shockwave.
- Friction of the air against the surface of the rocket.
- Turbulence of the air near the surface of the rocket.

Longitudinal force

It is usually assumed that the **angle of attack** of a rocket, that is, the angle between the longitudinal axis and the incident flow, is **small**. Then, its cosine can be assumed to be equal 1, and its sine, to be equal to the angle of attack in radians.

The longitudinal force is usually determined as follows:

$$X_1 = C_L \frac{\rho v^2}{2} S,$$

Where X_1 is the longitudinal aerodynamic force, ρ is the density of the air, v is the speed of the incident flow, S is the drag area (area of the largest cross-section, or the middle area), and C_L is the aerodynamic coefficient of the longitudinal force.

C_L is determined empirically and depends on the speed, shape and surface of the flying object. At small angles of attack, C_L is assumed not depending on that angle.

Transversal force

The transversal force exerted on a rocket is usually determined as follows:

$$Y_1 = C_T \frac{\rho v^2}{2} S,$$

Where Y_1 is the transversal aerodynamic force, ρ is the density of the air, v is the speed of the incident flow, S is the drag area (area of the largest cross-section, sometimes called the middle area), and C_T is the aerodynamic coefficient of the transversal force.

C_T is determined empirically. At small angles of attack, C_T is assumed to be linearly proportional to the angle of attack:

$$C_T = C_T^{\alpha} \alpha,$$

where α is the angle of attack, and C_T^{α} is an empirical coefficient equal to the average value of the derivative of C_T with respect to α.

69

Aerodynamic coefficients

Both C_L and C_T depend on the actual shape of the rocket, on the composition of the air and on the material and smoothness of its surface. These dependences are rather complicated, for this reason, the coefficients are measured in a **wind tunnel**, with a reduced-size model of the rocket. Thus determined coefficients are than scaled to the size of the real rocket.

Also, the coefficients depend on the speed of the incident flow relative to the speed of sound in the air. The ratio of these speeds is known as the **Mach number**.

The coefficients remain practically constant at Mach numbers from 0 to 0.6, then sharply increase with the growth of the Mach number, reach maximum at Mach=1.0, then reduce, and become constant again at Mach > 1.5. However, at Mach > 1.5 they are much larger than at Mach between 0 and 0.6.

70

Drag and lifting force

Longitudinal and transversal forces act in the reference frame tied to the rocket. In some cases, it is more convenient to consider aerodynamic forces in the reference system tied to the motion of the rocket. In this reference frame, the aerodynamic **drag** *X* is the component of the aerodynamic force projected onto the axis of the velocity of the rocket, and the **lifting force** *Y* is a projection perpendicular to this velocity.

For small angles of attack, the drag is

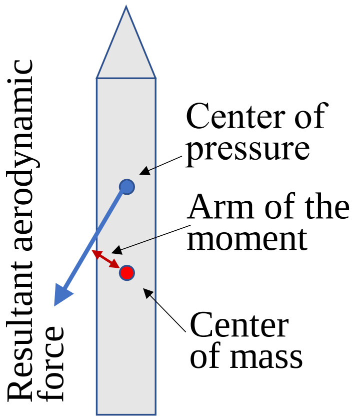
$$X = X_1 + Y_1\alpha$$

and the lifting force is

$$Y = -X_1\alpha + Y_1$$

Aerodynamic moment of force

All aerodynamic forces can be brought to a resultant force applied to a point inside the rocket, around which this resultant force does not create a moment of force. This point is called the **center of pressure**. In most cases, the center of pressure does not coincide with the center of mass.



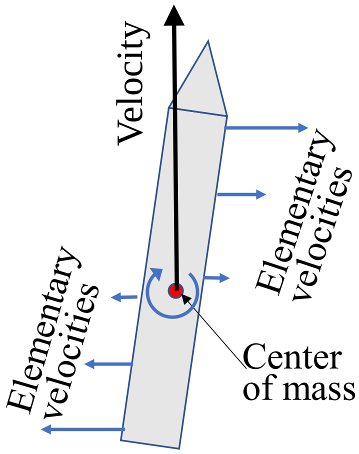
Whenever the resultant aerodynamic force is not on the line with the center of mass, there is an arm of application of the moment of force.

When the center of pressure is below the center of mass, the rocket is **dynamically stable**. When it is above the center of mass, the rocket is **dynamically unstable**.

As fuel in the rocket is consumed, the center of mass moves. Even if the rocket is initially stable, it may become unstable later.

Aerodynamic damping

When a rocket rotates around its center of mass, the elementary linear velocities of elements of the surface are oppositely directed on the opposite ends of the rocket.



Because of these elementary velocities, each element of the surface experiences an elementary drag directed oppositely to the velocities. The faster the rocket rotates, the stronger are those elementary drags.

Together, these elementary drags create a torque opposite to the rotation, and this torque dampens the rotation.

This damping effect is not strong, but it can dissipate rotation of a dynamically stable rocket.

73

Atmosphere (1 of 2)

Aerodynamic forces acting on a rocket depend on the atmospheric pressure, which changes with the altitude. At launch, these forces remain significant to altitudes of around 100 km. At reentry and descent, they are the strongest forces exerted on a spacecraft.

Consider the **ideal gas law**, formulated by Clapeyron:

$$PV = \frac{m}{M} RT,$$

where P, V and T are pressure, volume and absolute temperature, m is the mass of gas in the volume, M is its molar mass, and R is the ideal gas constant. $R \approx 8.3144598\ \mathrm{m^3 \cdot Pa \cdot K^{-1} \cdot mol^{-1}}$. The average molar mass of dry air is $M \approx 28.97\ \mathrm{g/mol} = 0.02897\ \mathrm{kg/mol}$.

By introducing density $\rho = m/V$ and specific gas constant $R_{\mathrm{SP}} = R/M$ we obtain:

$$P = \rho R_{\mathrm{SP}} T.$$

74

Atmosphere (2 of 2)

Therefore, we can estimate density at different altitudes by measuring pressure and absolute temperature:

$$\rho = \frac{P}{R_{\mathrm{SP}} T}$$

As pressure and temperature change all the time because of the weather, and also R_{SP} changes because of changing humidity, we cannot have the exact values of density at any given moment for any given altitude. Therefore, for determining parameters of controllability, typical and most adverse values are used.

They can be found in **Reference atmospheric models** (aka “**standard atmosphere**” models). There are several such models developed by different organizations. The most commonly used are the **US Standard Atmosphere** and the **International Standard Atmosphere**.

Wind (1 of 2)

Wind is probably the **most volatile disturbance**. Its strength can change in a second, and it adds to the incident flow of air. Since it is mostly horizontal, it creates the aerodynamic moment.

Instead of taking the current strength of the wind into account, estimations of controllability are done for the **maximum wind at different altitudes** (with a known probability that this maximum will not be exceeded). And its direction is assumed equally probable from any side.

The probability in this case is estimated over long-term observations. For instance, if the maximum speed of wind is indicated as 25 m/s at altitude 0 m with probability 0.997, it means that this speed can be exceeded. Indeed, in some hurricanes, the wind exceeded 57 m/s. When the wind is twice faster, the force it exerts is 4 times stronger.

Wind (2 of 2)

A model of wind developed in the USSR in the 1960th is given in the table below. Direction is assumed equally probable.

Altitude *h*, m	**Max. wind W, m/s, with the probability of 0.997**
0 to 10,000	$W= h \cdot 0.005 + 25$
10,000 to 15,000	$W = 75$
15,000 to 35,000	$W = h \cdot 0.00125 + 56.25$
>35,000	$W = 100$

Models of wind and atmosphere are developed for particular periods of time and particular territories. They have to be updated periodically, as the global and local climates change over time.

Weight

There are several definitions of weight. According to some of them, it is a scalar value intrinsic to an object with mass, according to others, it is a vector.

We will use this definition: **Weight is a force exerted by an object on its support or suspension**. This force produces an equal and oppositely directed **reaction force** of the support or suspension. Weight is caused by any forces causing change in acceleration of an object, except the gravitational attraction (*and virtual forces of inertia*).

This definition implies **weightlessness**, which exists when no forces but gravity are exerted on a body. This state is known as **free fall**. In some other definitions, weightlessness does not exist.

We can speak of **standard weight**, a scalar value, the product of mass and the value of the standard gravity $g_0 = 9.80665\ \mathrm{m} \cdot \mathrm{s}^{-2}$.

Proper acceleration and g-force

Proper acceleration is acceleration of an object in a reference system freely falling in a gravity field. In other words, it is caused by all forces except gravity and virtual forces related to inertia.

A motionless object on the ground experiences proper acceleration equal to that of free fall and directed upwards. An object raised up and dropped, while falling freely (if we neglect the atmospheric drag) and accelerating towards the ground, experiences zero proper acceleration.

Note: Accelerometers measure only proper acceleration.

The **g-force** is not a force, it is the ratio of the proper acceleration to the standard acceleration of free fall. It is also called **g-load**. It is also the ratio of the current weight to the standard weight. It is expressed in units of g. For instance, 2g means that the proper acceleration is twice the acceleration of free fall, and the weight of an object is twice as big as of the same object motionless on the ground.

Thrust

Thrust is a force created by the engines of a rocket and the propulsion system in general.

Normally, the thrust is directed along the longitudinal axis of the rocket. However, in many rockets, its direction can be controlled to some extent in one way or another. Also, the magnitude of thrust can be controlled to some extent in many engines, especially liquid fuel engines.

Thrust of thrusters with bell-shaped nozzles (de Laval nozzle) **depends on the atmospheric pressure**. For this reason, nozzles of engines of the first stage are adjusted for the sea level, while nozzles of further stages are adjusted for higher altitudes or vacuum (nozzles of these are longer).

Uncertainties related to the thrust are: not absolutely accurate mounting of the engines on the rocket and/or deviation of the center of mass from the longitudinal axis; variation of the thrust of different engines of the same model and dependence on the changing atmospheric pressure.

Vibrations and sloshing

The rocket vibrates in flight. These vibrations cause periodical shift of the center of mass and slight bending of its hull and its longitudinal axis. As a result, thrust becomes periodically misaligned with the center of mass and creates periodically changing torque.

Fuel components of liquid-fuel rockets **slosh (oscillate)** in the tanks. These oscillations are not sinusoidal, and their periodicity is not constant. There are splashes, there are shifts of the components towards one side of the tanks when the rocket makes a turn.

Sensors of motion are located not in the same part of the rocket with the actuators of motion. Sensors measure these vibrations and oscillations and pass them to the control system.

These effects must be taken into account. And the transfer function of the control system should be designed such that it would dampen these effects.

81

Inertia (1 of 2)

Inertia of a rocket is represented by its **mass** and **moments of inertia**. Inertia accumulates **kinetic energy** and **kinetic momentum**, which needs taking into account in control systems.

Effects related to the **mass** of the rocket (linear acceleration, velocity and translational motion) can be brought to its center of mass.

Effects related to the **moments of inertia** of the rocket (angular acceleration and velocity, as well as rotational motion) can be brought to its main moments of inertia passing through the center of mass.

Rockets are designed in such ways that one of their **main axes of inertia** is as close as possible to their longitudinal axis, the other two design axes are chosen at the convenience of designers and are also close to the other two main axes of inertia.

82

Inertia (2 of 2)

The **moment of inertia around the longitudinal axis** of a rocket is much smaller than moments around the other two. Torques around this axis caused by disturbances are also small. This means that a small control torque around this axis should be enough.

The **moments of inertia around the other two axes** are much greater, and the torques caused by disturbances are also greater. This means that the rocket can accumulate substantial kinetic energy and rotational momentum in rotations around these axes. For dissipating them and stabilizing the rocket, much stronger control torques are needed. Besides, the rocket has to perform a program turn around one of those axes.

5

Spacecraft as controlled objects

Inertial properties, forces, torques, disturbances

External forces acting on spacecraft

These forces are usually much smaller than those acting on a rocket.

Gravity and inertial forces

Gravity is the smaller, the farther is the orbiter from the orbited body. For the Earth, at the altitude of 200 km, g ≈ 9.23 m/s^2, at 400 km, g ≈ 8.7 m/s^2, at 1,180 km, g ≈ 7 m/s^2, at 2,550 km, g ≈ 5 m/s^2, at 35,786 km (geostationary orbit), g ≈ 0.224 m/s^2, and so on.

The **centrifugal force** balances off the gravity in a circular orbit.

In a circular orbit, the linear velocity is the smaller, the higher is the altitude of the orbit. And in an elliptical orbit, the linear velocity is the highest at the periapsis and the lowest at the apoapsis.

The **Coriolis force** appears in reference systems tied to the Earth.

Residual atmosphere and drag

The upper limit of the atmosphere of Earth is usually assumed to be at the altitude 100 km. However, the atmosphere does not end there. Its noticeable influence is observed at altitudes up to 700 km. And even at altitudes from 700 to 10000 km there is still a part of the atmosphere (so-called **exosphere**).

The residual density of the air at such altitudes is so small that molecules of gases do not interact with each other. For this reason, they do not create aerodynamic effects, like lifting force, aerodynamic friction and aerodynamic drag. However, they create **atmospheric drag** and can cause a **torque** around the center of mass of the spacecraft.

This drag is a tiny force, but it acts permanently, over the long period of an orbital mission. It causes **orbital decay**, eventually resulting in re-entry of a satellite into the atmosphere. To counteract the decay, periodical orbital corrections are used.

Solar wind and light pressure

Solar wind is the flow of particles emitted by the Sun. Mainly it consists of protons and electrons, but also includes neutrons and nuclei of helium. The speed of these particles ranges from 250 to 800 km/s. It strength (density) varies considerably depending on the **solar activity** and becomes especially strong during **coronal mass ejections** directed towards the Earth.

Light pressure is caused by photons imparting their kinetic momentum when they are absorbed or reflected by a surface. This includes photons not only of visible light, but also of other spectrum bands of solar radiation. It can increase in a **solar flare**.

Both create a very weak force (and can create a weak torque), the effect of which accumulates over a considerable time interval and may contribute to the orbital decay.

87

Magnetic field

The Earth (and some other planets) have a magnetic field. If a spacecraft has a **permanent magnetization** or has **electromagnets** activated periodically, the magnetic field of the Earth interacts with the magnetic field of the spacecraft and creates a torque around the center of mass of the spacecraft, and, probably, a small force accelerating or decelerating the spacecraft.

If a spacecraft is **electrically charged**, which can be a result of interaction with charged particles in space or of operation of an ion thruster, it becomes a big charged particle moving in the magnetic field of the Earth. This creates the Lorentz force perpendicular both to the vector of velocity and the vector of the magnetic field.

88

Internal sources of forces and torques

Many spacecraft are equipped with a **propulsion system**. It can create thrust with the possibility of control of its vector, magnitude and time of application.

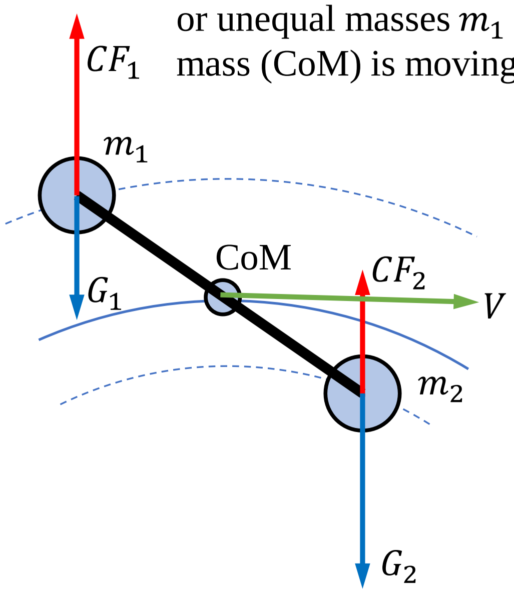
Moving parts of the spacecraft create forces and torques. This may include reaction wheels, solar panels, antennas, robotic arms, protecting lids of optic devices. In some cases, deployable parts, like solar panels and antennas, move only once, when they are deployed, in others, they can be moved after deployment. When they stop, they create an oppositely directed force, but not necessary exactly collinear with the initial force, therefore, the torques are not necessarily cancelled out.

Detachable parts create forces that are not cancelled out.

When the propulsion uses **liquid propellant**, it can move independently in the tank, creating forces and torques.

Tidal forces

Consider a spacecraft, which can be viewed as a dumb-bell with equal or unequal masses m_1 and m_2 connected by a rod. Its center of mass (CoM) is moving in an orbit with the velocity V.



Orbits of masses m_1 and m_2 do not coincide with the orbit of the CoM, therefore, gravity G_1, G_2 and centrifugal forces CF_1, CF_2 are different for m_1 and m_2.

The resultant forces acting on m_1 and m_2 are oppositely directed and create a torque around the CoM, which makes the spacecraft turn.

Eventually, rotation stops, and the spacecraft becomes oriented vertically.

Maneuvers of spacecraft

Maneuvering is achieved by change of the vector of velocity (direction and magnitude) of the center of mass. A spacecraft may require maneuvering for:

- correction of its orbit or moving into another orbit,
- creation of a grouping or a constellation with other satellites,
- a rendezvous and docking with another spacecraft,
- avoiding collisions with other spacecraft or space debris,
- re-entry or controlled orbital decay.

For maneuvering, the propulsion system is used. Normally, a spacecraft does not have too many thrusters, and the available thrusters are fixed or can change the direction of thrust in a limited range. Therefore, often the spacecraft needs to be properly oriented before firing the thrusters.

Certain maneuvers can be performed without thrusters, by controlling the atmospheric drag.

Attitude control of spacecraft

Depending on the mission, a spacecraft may need a particular orientation relative to stars, its orbit, the orbited celestial body, a particular place on that body or another spacecraft.

A particular attitude may be necessary for directing sensors to a particular area or its antenna to a ground communication center or another spacecraft, or for aligning its docking device with that of another spacecraft. For instance, Earth observation satellites often need their optic sensors directed to the nadir, but not always.

In some cases, the attitude should be changed quickly, in others, it should be kept constant for hours or days, with high accuracy.

The attitude is controlled by creating a torque with the use of actuators, like reaction wheels and magnetorquers. In some cases, propulsion can also be used for this purpose.

6

Coordinate systems

Some reference frames
used for different control tasks

Choice of reference frames

For solving different tasks of control, different reference frames are used. Particular reference frames are convenient for modeling and simulation of particular aspects of motion of rockets and spacecraft (the equation of motion are much simpler when are written in them). Some are used because sensors provide output of motion in those frames. Some are convenient for tracking and monitoring the flight from Earth.

There are reference frames related to the Earth, to the vehicle or its trajectory, to the orbit around the Earth. And, for interplanetary missions, reference frames related to the Solar System are necessary.

Coordinate systems can be Cartesian, polar, geographical, orbital and stellar. It is possible to convert coordinates in one reference frame into coordinates in another. For this, several elementary transformations are needed, which include rotations, translations (shifts) and other operations of vector and matrix algebra or quaternion algebra.

The launch coordinate system

This reference frame is connected to the point of launch. It is cartesian and it moves together with the point of launch, as the Earth rotates. For this reason, it is non-inertial.

The origin of this system coincides with the center of the rear cross-section of the rocket on the launchpad before the launch.

The **X** axis is directed towards the direction of the flight.

The **Y** axis is directed vertically up.

The **Z** axis completes the system making it a right-handed coordinate system.

The angle between the X axis and direction to the north is the **azimuth of the launch.**

This system is convenient for tracking the rocket during the launch.

The initial launch coordinate system

This reference frame is a copy of the launch system **at the moment of the liftoff**. After that, it does not accelerate in space. For this reason, it is inertial. Motion of the Earth around the Sun is neglected.

This system is convenient for modeling motion of the rocket during the launch, and it is modeled onboard the rocket for **inertial navigation systems**.

The launchpad coordinate system

This coordinate system is similar to the launch system, only its X axis is directed to the median azimuth of typical launches or to the north.

This system is used for positioning of the rocket on the launchpad and presetting the initial position of the rocket in the initial launch coordinate system.

The absolute geocentric system

This coordinate system is connected with the Earth. It becomes fixed at the moment of liftoff. For this reason, it is inertial.

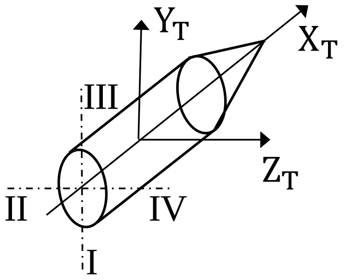
The zero reference point of this system coincides with the center of the gravitational attraction of the Earth. Its **X** axis is directed towards the north pole. Its **Y** axis is directed towards the crossing of the zero meridian with the equator. Its **Z** axis fulfils the system making it a right-handed coordinate system.

Geographical coordinate systems

These are various **map projections**. Coordinates in these systems are **latitude**, **longitude** and **altitude**. The most commonly used are variations of the **Mercator projection**. They are used for tracking the rocket during the launch and orbiting spacecraft during the mission, and for planning communication sessions with ground stations.

The tied coordinate system

This coordinate system is tied to the rocket or spacecraft. Its X axis is directed along the longitudinal axis and the other two axes are the designer axes of the vehicle making a right-handed system.



Its origin for a rocket is at the center of mass. It is convenient for considering forces, torques and attitude angles.

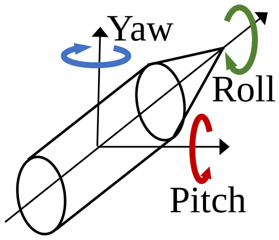
Planes I-III and II-IV are **stabilization planes**, and **planes of mass symmetry**.

The speed-related system

Its origin is at the center of mass, its X axis is along the vector of velocity, its Y axis is in the XY plane of the tied system.

Attitude angles

Three angles determine the attitude of a rocket or orbiting spacecraft.



Pitch is the angle between the X_T of the tied reference system and the horizontal plane. During the launch, this is usually the XZ plane of the initial launch system. For a satellite, it is the horizontal plane on the Earth beneath the spacecraft (the plane of **the local horizon**).

Yaw is the angle between the vertical plane passing through X_T and the vertical plane of the intended direction of flight.

Roll (or **bank**) is the angle between the vertical plane passing through X_T and the vertical plane of the intended direction of flight.

Pitch changes throughout the launch from 90° to the desired end value, Yaw is maintained close to 0 during the launch, but may change at the end of it. Roll changes after launch and then is maintained close to zero.

Transformations of reference frames (1 of 2)

It is often necessary to convert coordinates of a point or vector obtained or measured in one reference frame to its coordinates in another reference frame.

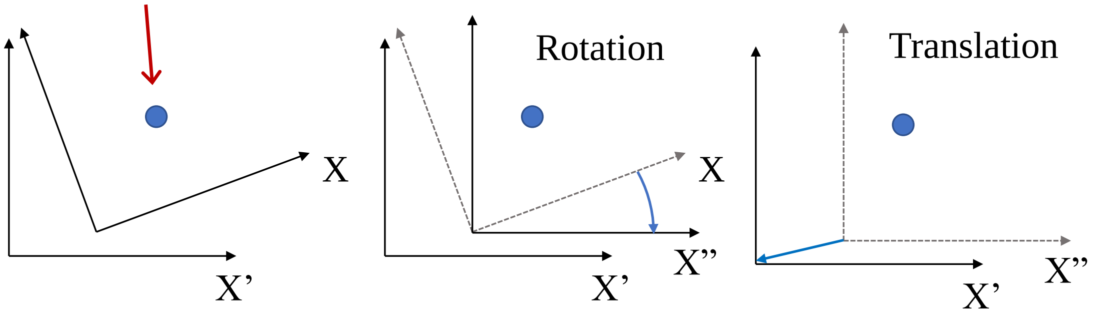
When both reference frames are cartesian, this transformation is done with operations of **rotation** and **translation**.

Rotation is turn of one frame around an axis in such a way that at least one of the axes of the rotating frame becomes parallel to the same axis of the fixed frame or to one of the reference planes of the fixed frame (after which, an additional rotation will align these axes). This is done with a **rotation matrix** multiplied by the vector of initial coordinates.

Translation brings the initial point of one of the system to the initial point of the other. For this, a **translation vector** is added to the vector of the transformed coordinates. Axes do not change their direction.

Transformations of reference frames (2 of 2)

Coordinates of the point known in reference frame X are to be converted into coordinates in frame X'.



In a 3D space, up to three rotations are necessary. Respectively, there should be three rotation matrixes. Their product is the **transformation matrix**. The order of rotations and, respectively, multiplications of the matrixes, is important, because these operations are **not commutative**.

Orbital elements (1 of 3)

One way of describing the position of a satellite in an orbit is:

1) to describe the orbit (this includes description of the ellipse and description of its location relative to the Earth or Sun),
2) and then to indicate the position of the satellite in it.

An ellipse can be described by **two parameters**, for instance:

- the semi-axes – **semi-major axis** a and **semi-minor axis** b, or
- the **eccentricity** $e = \sqrt{1 - b^2/a^2}$ and **semi-latus rectum** $\ell = b^2/a$. When $e = 0$, the orbit is a circle with radius ℓ; when $0 < e < 1$, it is an ellipse; when $e = 1$, it is a parabola, and when $e > 1$, it is a hyperbola.
- the **eccentricity** e and **semi-major axis** a.

Other combinations of two parameters, which describe an ellipse completely, may be used.

Orbital elements (2 of 3)

Two more orbital elements describe the orientation of the orbital plane.

1. The **inclination** i is the angle between a **reference plane** and the orbital plane. The reference plane for Earth orbits is the **equatorial plane** of the Earth. In interplanetary missions, it is the **plane of ecliptic**.

2. The **longitude of the ascending node** Ω is the angle between the **ascending node** and the **reference direction** in the reference plane.

- The **nodes** are points of crossing of the orbit with the **reference plane**. At the **ascending node**, an orbiter passes from “beneath” to “above” the reference plane. And “above” is where the north pole of Earth is. The reference plane is the plane of the equator for satellites and the ecliptic plane for interplanetary spacecraft.
- The reference direction is the direction to the **First Point of Aries** ( ), aka point of **vernal** (or March) **equinox**.

Note: This element does not exist for orbits in the reference plane.

Orbital elements (3 of 3)

One more orbital element describes the orientation of the orbit within the orbital plane:
Argument of periapsis ω is the angle between the line of the ascending node and the direction to the pericenter (**periapsis**) on the **line of apsides**. For an orbit in the reference plane, which does not have the ascending node, this angle has to be counted from the direction to the vernal equinox.

And the sixth orbital element describes the current position of the orbiter in the orbit:
True anomaly ν, θ, or f is the angle between the line to the periapsis and the position of the satellite in the orbit.

The true anomaly changes all the time, while other elements change more slowly and are assumed constant for an **epoch** – a time interval, at which they can be considered constant with sufficient accuracy.

7

Sensors

Measurements of accelerations, velocities, position, attitude, pressure, temperature

Accelerometers

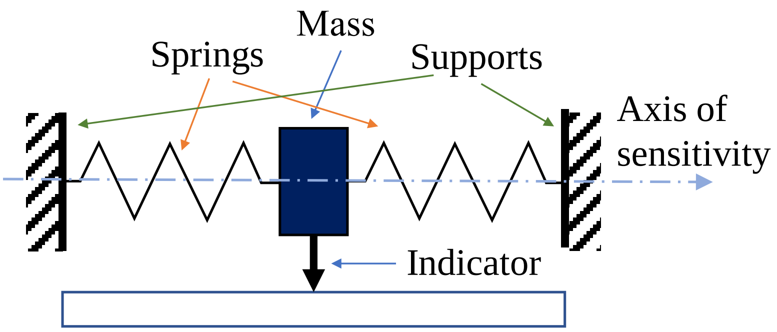
An accelerometer is a sensor that measures **proper acceleration** along its **sensitivity axis**. If the acceleration is not collinear with the axis, the accelerometer measures its projection onto the axis.

If we place an accelerometer on a motionless horizontal surface on Earth and direct its sensitivity axis strictly vertically, it will show an acceleration equal to that of free fall at this location and directed up.

Accelerometers usually produce electrical output linearly proportional to the projection of the acceleration on their sensitivity axes within the range of their sensitivity. Some accelerometers may have a dead zone and cannot measure small accelerations.

Concept of measuring acceleration

The **concept of an accelerometer**: In absence of acceleration, a massive object is balanced by two opposing forces in a neutral position. Acceleration creates an inertial force that shifts the mass from a neutral position. This shift is proportional to the magnitude of the acceleration along the axis of sensitivity. Then the shift can be read immediately or converted into voltage or current proportional to the acceleration.



This picture shows one of the possible implementations of this principle, but implementations used in practice are much more advanced .

Three-axis accelerometers

When an acceleration (a vector) is not collinear with the axis of sensitivity of an accelerometer, the latter shows the projection of this vector onto its axis of sensitivity. Three accelerometers with axes of sensitivity perpendicular to each other make a 3-axis accelerometer. Its output provides 3D projections of an arbitrarily directed vector of acceleration.

Some accelerometers are designed such that they have two axes of sensitivity (dual-axis accelerometers). Their output shows two coordinates of the projection of the acceleration vector onto the plane of their sensitivity. For measuring 3D, two such accelerometers are needed.

3D accelerometers are often available as ready-made devices with built-in microchips and produce digital output in the form of 3 projections of the acceleration vector.

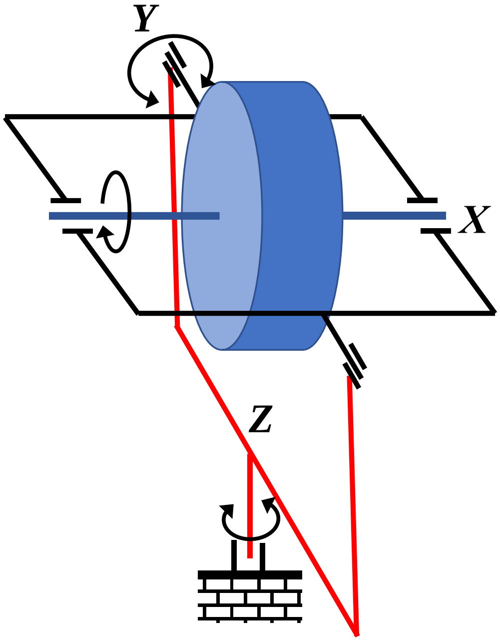
Velocity and position via acceleration

By integrating the vector of acceleration, we can obtain the vector of velocity. By integrating the vector of velocity, we can obtain the **path** of travel of a moving object, its **position** in the 3D space at any given moment and its **displacement** from the initial point.

This is called **inertial navigation**. But it is not a complete solution of the navigation problem. Acceleration caused by gravity is not measured and has to be taken into account by a correction according to a model of gravity to obtain the actual acceleration.

Inertial navigation systems quickly build up errors, therefore, they should be used for limited time intervals. If an accelerometer produces a constant error of $0.001\ \mathrm{m/s^2}$, and the motion is linear, in 1000 seconds (16 m 40 s) the error in measuring velocity will become $1\ \mathrm{m/s}$, while the error in travel, 500 meters.

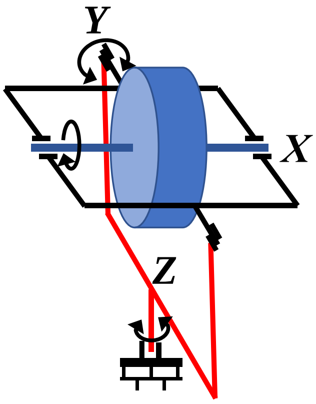
Gyroscopes (1 of 4)



A gyroscope is a flywheel usually suspended in a gimbal (a suspension with 3 degrees of freedom of rotation relatively to the support). The flywheel spins at high rate around axis ***X***.

When a torque is applied around axis ***Y***, the gyro turns around axis ***Z***. When a torque is applied around axis ***Z***, the gyro turns around axis ***Y***. This is precession, and its angular velocity is proportional to the torque and to the angular momentum of the flywheel. Also, a torque equal and opposite to the applied torque appears.

Gyroscopes (2 of 4)



When no torque is applied, but the support rotates around any axis or all axes together, the direction of axis ***X*** in space is preserved. However, small torques appear because of non-zero friction in bearings and eventually cause a drift of axis ***X***.

The higher is the rotation rate of the flywheel and the bigger is its moment of inertia around axis ***X***, the more accurate is the gyro.

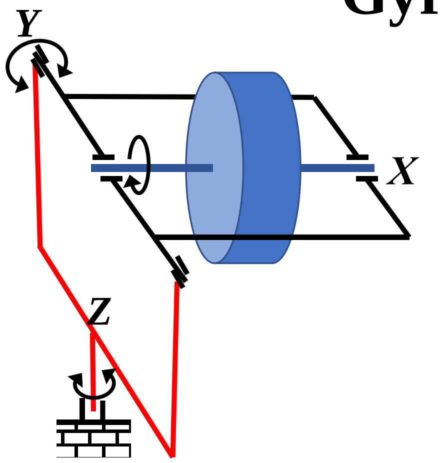
A torque around ***Z*** or rotation of the support around ***Y*** may cause the inner frame of the gimbal come to the plane of the outer frame. Then the gyro folds and loses its property. Rotation rate of the flywheel has to be maintained constant by external supply of energy.

Gyroscopes (3 of 4)

If we measure angles of turn around axes ***Z*** and ***Y***, we will know the change of attitude around two axes. One more gyro with its axis ***X*** perpendicular to that of the first one enables determination of the attitude around three axes. However, this possibility is rarely used.

A gyro with its axis ***Z*** rigidly tied to the support becomes a gyro with 2 degrees of freedom. It becomes **sensitive to rotation** of the support **around *Z***, because such rotation causes a torque around ***Z***, and the gyro begins to precess around ***Y***. To prevent frames from folding, this precession is balanced with two opposite springs. As a result, the angle of precession of this gyro becomes proportional to the angular rate of rotation of the support around ***Z***. A damper is added to prevent oscillations of the inner frame on the springs. Such gyros are used as **angular rate sensors**.

Gyroscopes (4 of 4)



If the gimbal axis ***Y*** does not pass through the center of mass of the flywheel, then a proper acceleration along axis ***Z*** will cause a torque around axis ***Y***, and the gyro will precess around axis ***Z***. The angular velocity of this precession will be proportional to the acceleration, and the total angle of turn will be proportional to the proper velocity. This is the principle of operation of a **gyro integrator.**

There are many other navigation and attitude determination instruments and sensors based on gyroscopes.

113

Inertial platform (1 of 2)

An **inertial platform** (aka **stabilized platform** or **gyroscopic platform**) is a device maintaining a constant attitude towards axes of an inertial reference system, typically, the initial launch reference system. It is suspended in a 3-axial gimbal onboard of a rocket.

Three gyroscopes on it with their axes ***X*** directed perpendicularly to each other represent the axes of the required inertial reference system. Each axis of the gimbal is equipped with a sensor of the angle of turn and a torque motor.

As soon as any or all gyroscopes begin to even slightly precess, signals of their precession are amplified and sent to the torque motors, which create torques around respective axes of the gimbal of the platform, until the gyros return into their initial positions.

114

Inertial platform (2 of 2)

On the inertial platform, accelerometers, angular rate sensors and gyro integrators are mounted.

The axes of sensitivity of the accelerometers and angular rate sensors are directed in parallel to the axes of the inertial reference system maintained by the platform.

The axes of sensitivity of the gyro integrators are directed along the so-called “ballistic directions”. Some rockets have two such directions, others only one.

Inertial platforms ***are not used now***. They were convenient before appearance of onboard computers. They still can be found in some rockets of old design. Their main use was for **inertial navigation systems** of ICBMs (intercontinental ballistic missiles), but they were also used in space rockets as supplements to other methods of navigation.

115

Strapdown system

A **strapdown inertial navigation system** is called so because navigation and attitude determination instruments (sensors of motion) in it are strapped down (rigidly attached) to the vehicle.

Such systems came in use when onboard computers achieved substantial computing power and became able to perform conversions of coordinates between different reference systems fast enough. This opened a possibility to maintain (in the memory of the computer) an abstract model of an inertial reference system using measurements of accelerations and rotations taken in the non-inertial reference system tied to the vehicle.

Having built this model, the computer can calculate projections of the current position, velocity and acceleration of the vehicle as well as its attitude in an inertial reference system by performing transformations of coordinates, numeric integration and other necessary calculations.

116

Laser gyroscopes

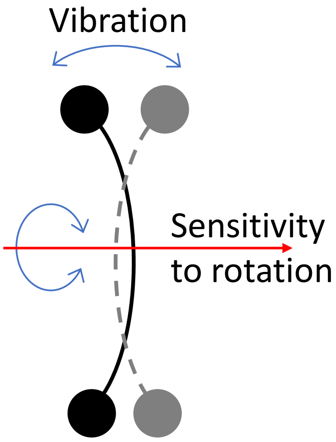
A **ring laser gyroscope** consists of a closed path of four perpendicular tubes forming a rectangle in a plane. At the angles where the tubes join, mirrors are places at the angle of 45° to each of the joining tubes.

A beam of a laser is split into two beams and injected into the path such that one of the beams travels clockwise, and the other, counterclockwise. At one of the angles, both beams are let out of the path and projected on a screen, where they form an interference picture.

When the rectangular path rotates around the axis perpendicular to its plane, the interference pattern changes due to the **Sagnac effect**. By measuring the change of the interference pattern, it is possible to calculate the angular rate of rotation around this axis.

The same principle is used in **fiber optic gyroscopes**, where the path is made as a coil of optic fiber up to 5 km long.

Vibrating structure gyroscope



A **vibrating structure gyroscope** (aka **Coriolis vibratory gyroscope**) consists of two flexibly connected masses. The connecting rod and masses are made vibrate in one plane.

This structure is sensitive to rotations around the axis in the plane of vibration that is perpendicular to the rod connecting the masses. When a rotation of the plane occurs around this axis, its position in space begins to change, while the masses continue vibrating in the previous plane due to the Coriolis effect.

The deviation of the rod with masses from the rotating plane of vibration is proportional to the angular rate of rotation around the axis of sensitivity, frequency of the vibration, masses and rigidity of the rod.

MEMS

MEMS stands for “**Microelectromechanical system**”. In Europe this can be called **MST** (microsystems technology). Usually they include a tiny mass that can move on an elastic suspension and sensors detecting the motion of the mass. Dimensions of MEMS sensors are usually in the range of several millimeters or less than 1 mm. They can be combined into units containing several sensors, a microcontroller for converting signals of sensors into outcomes of measurement and a connector for one of the standard interfaces.

Some **inertial measurement units** (IMU) are bases on MEMS accelerometers and MEMS vibrating structure gyroscopes. They can measure acceleration and rotation. Such devices are widely used in automobiles, laptop computers, hand-held mobile devices, photo and video cameras, robots, etc. Specially designed MEMS are used in rockets and spacecraft.

Star tracker

A **star tracker** is a device for attitude determination of spacecraft. It consists of an **image sensor**, an optical system (a lens) projecting image of the sky onto the sensor and a microcomputer processing the images.

The computer of a star tracker holds a complete map of the sky with stellar coordinates of all sufficiently bright stars. By analyzing what part of the sky is seen by the tracker and at what angle of turn around the optic axis of the lens, the computer produces output of three attitude angles or a **quaternion** of attitude in a desired reference system. The algorithm includes elimination of moving objects (like planets, comets and asteroids), corrections for the drift of the **First Point of Aries**, for **parallax** of the nearer stars and some others.

A star tracker does not work when directed to the Sun or the Earth. Location of a tracker on the spacecraft should be chosen carefully. More than one tracker may be needed.

Satellite navigation

Instead of (or in addition to) using inertial sensors for navigation (determining position of a vehicle in space), satellite navigation is commonly used. A small satellite navigation microchip with an antenna produces an output containing three coordinates.

Commonly available satnav receivers operate when the vehicle moves no faster than 250 m/s. This is not enough for rockets and spacecraft. Special receivers operating at much higher speeds are needed.

Satnav receivers rely upon radio signals from navigation satellites. These signals can be jammed by a strong geomagnetic storm or with an evil purpose, or they can be switched off. Besides, the signals are unavailable beyond the orbits of the navigation satellites. Inertial navigation can be used as a backup in such situations for a reasonably short time, as errors of inertial navigation build up rather quickly.

Telemetry navigation

Position of a vehicle in space can be measured from an observation station on the ground. For this, a radar or lidar system can be used. The **Doppler effect** can be used for very accurate determination of projection of the velocity of the vehicle onto the line connecting it with the station. With three stations, it is possible to determine the entire vector of velocity and to increase the accuracy of determination of the position.

The obtained navigation parameters (coordinates and velocity) can be converted into coordinates in a desirable reference system and then transmitted as telemetric signals to the vehicle or used for issuing telecommands for correcting its position and velocity and transmitting them to the vehicle.

Since this way of navigation relies upon radio signals, it is susceptible to natural or intentional disruptions of radio communications.

Radar and lidar navigation

A spacecraft can have an onboard radar or lidar system for the purpose of navigation.

Such navigation can be used for determining position and velocity of a spacecraft relative to other spacecraft, especially for the purpose of docking.

By installing **retroreflectors** with precisely known coordinates on the ground, it is possible to determine coordinates and velocity of a spacecraft relative to the Earth. Three or more retroreflectors can be used for **triangulation** of the position. Retroreflectors can also be mounted on one or both of the docking spacecraft.

A radar system with a **phased array** can be used for quick scanning of a big part of the sky to detect approach of other spacecraft or objects of space debris.

Rotary encoder and tachometer

A **rotary encoder** converts the angle of turn of a shaft into an electric signal (or a digital code) proportional to the angle.

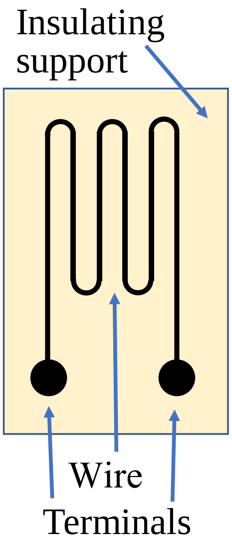
Such encoders can be **absolute** (the angle between a reference position and a current position of the shaft) or **incremental** (measuring the angle of turn from the previous position).

Rotary encoders can be conductive (rarely used now), optical, magnetic and electromagnetic.

Rotary encoders can be a part of other sensors (measuring the angle of precession of gyros, turn of a stabilized platform) or sensors for feedback from any rotating mechanism of a vehicle.

A **tachometer** is a device measuring rotation rate (for instance, rotations per minute) of the shaft of a motor. Tachometers are usually based on electromagnetic inductance.

Strain gauge



A strain gage consists of a long wire (made of certain alloys) firmly attached to a flexible insulating support. When the support bends, the wire stretches and changes its electrical resistance. Being attached to an object by an adhesive, it can measure strain in the object.

The strain gauge is connected to one of the arms of a **Wheatstone bridge**. The bridge is balanced when there is no strain. As soon as a strain is applied, the resistance of the gauge changes, the bridge becomes unbalanced, and a voltage between measuring nodes of the bridge appears. This voltage is proportional to the strain.

Strain gauge and load cell

The wire or foil of the gauge is of high specific resistance; its resistance is much higher than that of the wires connected to the terminals.

When the Wheatstone bridge becomes unbalanced by the sensor, its output voltage usually is tiny. Initially it was measured by a galvanometer. Now it can be measured by a digital voltmeter and converted into a digital format.

A **load cell** is a sensor for measuring weight or force. A common type is a rod with one end fixed in a support and the other end accepting the weight. To the cylindrical side of the rod, one, two or four strain gauges are attached. The strain in the rod caused by the weight is converted into the weight.

Strain gauges and load cells require **calibration** and **setting of the zero point**.

Pressure gauge

A **pressure gage** (or a **pressure meter**) is a device for measuring pressure of liquids and gases. It can measure either **absolute pressure** (comparative to vacuum), or **differential pressure** (relative to the ambient pressure or pressure in another vessel).

Pressure gages differ in designs and physical principle of operation, which are completely different for very high and very low pressures. For instance, **metal bellows** convert pressure into mechanical displacement, which then can be converted into an electric signal. A strain gauge or a **piezoresistive sensor** can convert pressure directly into electric signals.

Pressure gages are used for measuring pressure in propellant tanks, propellant feeding pipelines, hydraulic actuators, pressurized gas cylinders, combustion chambers, etc. A pressure gauge measuring atmospheric pressure outside a rocket can be an additional source of information on the altitude up to 10-15 km above the sea level.

Piezoelectric sensors

These sensors are based on the **piezoelectric effect**. Certain kinds of crystals, and also some kinds of ceramics, being mechanically stressed, produce electric potential on their opposite sides. The voltage of this potential is proportional to the stress. This effect can be used for making **strain gauges** and **pressure gauges**.

Besides, when the stress changes very quickly, so does the voltage. Therefore, such sensors can be used for sensing high frequency vibrations in the sonic and ultrasonic ranges, from 0 Hz to far more than 100 kHz, depending on the material.

The effect is reversible: by applying voltage it is possible to make a piezoelectric element shorter or longer. Thus, the effect can be used for certain **actuators** or for sources of sound or ultrasound of controlled frequency.

Liquid level gauge

A **liquid level gauge** is a device for measuring level of liquids, that is, distance from the bottom of a tank (or from a given reference point) to the surface of liquid. They can be used in tanks of launch vehicles, but not in tanks with liquid propellant components of spacecraft.

A level gauge may consist of a guiding rail, rod or pipe and a float sliding along the guide. A transducer (electromagnetic or capacitive) converts the position of the float on the guide into an electric signal, which then can be converted into a digital code and used by the control system.

The level of liquid can be also measured with a sonar or a laser ruler. However, such measures need averaging out because liquid in tanks oscillates in flight.

Flow meter

A **flow meter** is a device for measuring flow – the volume of liquid or gas passing through the meter in a unit of time.

Flow meters are based on different principles of operation.
► A small turbine mounted in the flow – the rotation rate of the turbine is proportional to the flow.
► A plate with an orifice in a pipe; by measuring the drop of pressure it is possible to evaluate the flow.
►An ultrasonic signal sent along the flow, and, by the Doppler shift of its frequency, the speed can be measured and converted into the flow rate.
► And so on. There are 8 principles and dozens of variations of their implementations for measuring flow.

Measuring flow is important for controlling rocket engines and thrusters, and, in some cases, some other components of spacecraft.

Measurement of temperature (1 of 2)

Temperature at different places of a vehicle can be measured for the purpose of telemetry or control, or both. Normal operation of many components is possible in a limited range of temperatures. For a manned spacecraft, the acceptable range of temperature in the crew cabin is even narrower.

Thermocouples are widely used for measuring temperature. A thermocouple is a device consisting of two pieces of wire of different metals (or alloys). An end of one wire is connected to an end of the other wire. On their opposite ends, a voltage proportional to the temperature forms.

Typically, thermocouples measure temperatures from –100° to 1300° Celsius, however, each is designed for a narrower range, depending on the metals in the couple. Tungsten/rhenium-alloy thermocouples can measure temperatures up to 2760°C.

Measurement of temperature (2 of 2)

R**esistance temperature detectors** (**RTD**), aka **resistance thermometers**, provide higher accuracy than thermocouples. They are based on wire or film of a metal (platinum, copper, nickel) or carbon.

Typically, RTDs measure temperatures from –100° to 500° Celsius, however, each is designed for a narrower range. For instance, copper RTDs are not used at above 300°C, because copper oxidizes quickly at such temperatures. Some RTDs can measure temperatures up to 900°C and, specially designed carbon RTDs, down to –273°C.

Thermistors *are not used* for measuring temperature, however, they are used for stabilizing parameters of electronic devices at changing temperatures or for overload or overheat protection of devices.

Temperature can be measured at a distance by light emission or heat radiation of a hot object with pyrometers and infrared thermometers.

Measurement of time (1 of 4)

Time is measured by a process with known speed – motion of the Sun in the sky, spilling of water or sand from a reservoir, radioactive decay, etc. By measuring current progress of the process, it is possible to conclude on the time having passed since the known beginning of the process.

If this process is periodical (like swings of a **pendulum**, or any other), the number of periods multiplied by the known time of one period gives the time.

The first mechanical clocks were based on the property of a pendulum to swing with an approximately constant period, which depends on the length of the pendulum, gravity, drag caused by air and friction. The next development was a clock based on a **balance wheel** swiveling back and forth with an approximately constant period, which depends on the moment of inertia of the wheel, a spiral spring and drag.

Measurement of time (2 of 4)

Quartz clocks are based on piezoelectric properties of quartz crystals. A quartz crystal of a particular size, determining the period of its free oscillations, is connected to a resonance circuit of an electronic generator and ensures stability of the frequency of the generated signal. Time is measured by the number of periods of oscillations of the signal. The frequency can be rather high, which enables measuring of rather short periods of time.

Atomic clocks are based on the property of atoms to emit electromagnetic waves at certain conditions. For instance, an electron of an atom can be excited by application of external energy, and its energy in the atom increases. Then, this electron returns to its stable energy level by emitting a quantum of visible light or other electromagnetic wave of a particular frequency. Atomic clocks are currently the most accurate clocks.

Measurement of time (3 of 4)

Accurate measurement o time is necessary for different purposes onboard of a rocket or spacecraft and for synchronization of operation of spacecraft and ground stations. In most cases, quartz clocks provide sufficient accuracy, but sometimes, for certain purposes atomic clocks are needed. For instance, each satellite of a **satellite navigation system** has an atomic clock onboard.

Accurate time together with observations over celestial bodies can be used for navigation on the surface of the Earth. It was the primary way of navigation, especially at sea, since the invention of sufficiently accurate clocks called **chronometers** and until invention of satellite navigation, and is still used sometimes.

Accurate time is used in satnav systems for determining the location, and it can be obtained as a byproduct from a satnav receiver.

Measurement of time (4 of 4)

In long-term spaceflights, when accuracy of time is of primary importance, clocks are adjusted for the effects described in the **special and general theories of relativity**. Onboard a spacecraft moving at high speed, time slightly slows down (the special theory), but it also slightly speeds up because the strength of gravity is less then on Earth (the general theory). These two effects, though, acting oppositely to each other, do not cancel out. The difference between them depends on the altitude of the satellite, because the strength of the gravitational field and the speed of the satellite also depend on the altitude and are different at different altitudes. Atomic clocks of navigation satellites have this correction.

Clocks require periodic synchronization with the **standard time**. In a satellite navigation system, clocks of all satellites are periodically synchronized with the main clock of the system on the ground.

Electric measurements

Vehicles contain many electric and electronic circuits, electrical parameters of which need to be measured for the purpose of control and/or telemetry. Output of many sensors is also in the form of electric signals. Typically, these measurements are required in digital format.

This is done by comparing a signal with a “standard” reference signal and counting the number of “steps” between the measured and the reference signals. For instance, voltage can be measured by a linear increase of voltage from 0 to the maximum value with simultaneous counting of ticks of time. Counting stops when two voltages become equal. The result of counting is proportional to the measured voltage.

These measurements include voltage, current, power, frequency, phase, resistance, inductance, capacitance, electric charge, attenuation, signal-to-noise ratio (SNR) etc.

Liftoff sensor

The **liftoff sensor**, or **liftoff indicator**, is installed at the bottom of a rocket, and its purpose is to inform the control system of the moment when the rocket lifts off the **launch pad**.

Many processes in the rocket and its control system should begin exactly at the moment of liftoff. Also, this moment is the zero moment of the time of flight.

This sensor is typically an electromechanical “button” of a special design. It is “depressed” before the liftoff, being squeezed between the bottom of the rocket and the launchpad. At the moment when it is released, it locks or breaks an electric circuit. Its mechanism designed such that it becomes released when pressure on it becomes equal to zero.

There can be 3 or 4 liftoff sensors, and the flight is deemed as having begun when two of them indicate liftoff.

8

Actuators

Implementation of control by applying control inputs

Propulsion system

The purpose of the **propulsion system** is **propulsion**, that is, pushing of a rocket or spacecraft. The nature of propulsion is the **reaction force** explained by Newton’s third law of mechanical motion. A **propellant** is ejected from a **rocket engine** with force at high speed, and this creates oppositely directed reaction force called **thrust**.

A propulsion system consists of one or more rocket engines, storage for the propellant, a system for feeding propellant into the engines and a control system. Often it includes actuators for controlling the vector of thrust – its direction and magnitude. The control system of the propulsion system is integrated into the control system of the vehicle.

Each stage of a rocket has its own propulsion system. Typically, a stage or a spacecraft has only one propulsion system. Some spacecraft may have two such systems with different propellants for different purposes.

Rocket engine

A rocket engine is a device used for propulsion by accelerating and ejecting stored **propellant** as **reaction mass** at high speed in the form of a directed propulsive **jet**. As a result, a reaction force appears directed oppositely to the jet.

Acceleration of the propellant for the jet requires energy. Most rocket engines use **combustion** – burning as a result of a chemical reaction of the propellant.

Small thrusters of spacecraft may use compressed gas. Expansion of the gas in the thruster provides energy for creating a small propulsive force. **Ion thrusters** are also used in modern spacecraft.

There are attempts to develop engines using nuclear energy. Among such attempts is the **nuclear thermal rocket**, which was tested but never flew, and **nuclear pulse propulsion**, which so far exists as a theoretical concept.

Specific impulse

The specific impulse is a key parameter of efficiency of a rocket engine or thruster. It indicates how much thrust an engine creates while spending one kilogram of propellant in one second.

In other words, the specific impulse is a ratio of thrust to the **flow rate** of the propellant: $I_{SP} = F_P / FR$,
where I_{SP} is the specific impulse, F_P is the propulsion force and FR is the flow rate, in units of mass or units of weight per unit of time.

If F_P and FR are kilograms-force, then the unit of I_{SP} is second (of time). If F_P is in newtons and FR is kilograms mass, then the unit of I_{SP} is $\mathrm{N \cdot s/kg}$. For the specific impulse, $1\mathrm{s} \approx 9.80665\ \mathrm{N \cdot s/kg}$.

The **specific fuel consumption** is the inverse of the specific impulse.

I_{SP} in $\mathrm{N \cdot s/kg}$ is equal to the **exhaust velocity** of the propellant in m/s.

Propellants (1 of 3)

Combustion rocket engines may use liquid or solid propellants.
Liquid propellants are usually **bipropellants** (two-component propellants) – **fuel** and **oxidizer** (or **oxidant**) stored in separate tanks and mixed in the combustion chamber of the engine, where they burn releasing the energy of oxidization.

Common components of liquid propellants (oxidizer and fuel) are:
1) **Cryogenic**:
- **LOX** (liquid oxygen) and **RP-1** (highly refined kerosene) ;
- LOX and **Liquid hydrogen**;
- LOX and **Liquid methane**.

2) **Storable** in reasonable range of temperatures, also **hypergolic**:
- **dinitrogen tetroxide** (N_2O_4) and **Hydrazine** or **monomethyl-hydrazine** (MMH) or **unsymmetrical dimethylhydrazine** (DMH).

143

Propellants (2 of 3)

The mentioned bipropellants provide specific impulse around 300 to 320 seconds. The highest I_{SP} is that of LOX and liquid hydrogen, 400 s.

The highest $I_{SP} = 542$ s ever achieved with combustion propellants was with a **tripropellant** of lithium, fluorine and hydrogen. But it is not used because fluorine is toxic, highly reactive and may cause explosions. Lithium and fluorine are also very expensive. Besides, components must be stored in three separate tanks, and the fuel feeding system has to be more complicated.

In some cases, especially for spacecraft, monopropellants are used. Their specific impulse is smaller than that of bipropellants, which makes the feeding system simpler. Engines with monopropellants can start and stop many times, operating briefly or longer, with small intervals. For instance, hydrogen peroxide decomposing in presence of a catalyst releases sufficient energy for thrust.

144

Propellants (3 of 3)

Solid propellants are made of small granules of oxidizer, powder or flakes of fuel and a compound holding them together. The oxidizer is a chemical compound releasing oxygen when heated. The fuel is an explosive material.

They are used for main engines in some rockets, rocket boosters and small engines for pulling some detachable parts away from the rocket. Engines with them are simpler and do not require feeding systems. However, control over them is limited – once the propellant is ignited, it cannot be stopped until it burns down completely. Also, the magnitude of thrust cannot be controlled. Their specific impulse is 260 to 280 seconds.

A propellant for an ion thruster is usually an inert gas stored in a high pressure gas cylinder. This propellant does not produce thrust of its own.

145

145

Propellant feeding

Two main ways of feeding liquid propellant components are **pressure feed** (also, **displacement feed)** and **turbopump feed**.

Pressure feed is implemented by creating controlled excessive pressure in the tanks above the level of liquid components. This pressure can be created by controlled supply of pressurized gas from gas cylinders or by burning small amounts of propellant components and injecting their combustion product into the space above the components in the tanks.

In turbopump feed, a **turbopump** is used, which is a gas turbine and two pumps on the same axis – one for the fuel and the other for oxidant. Gas for rotating the turbine is usually made by burning fuel and oxidizer. This gas is injected into the turbine at high speed and pressure and rotates it. The exhaust of the turbine is fed into a nozzle similar to the nozzles of the main engine, only smaller, for creating small additional thrust.

146

146

Control of thrust (1 of 4)

Control of thrust of a rocket is achieved by magnitude and direction of the vector of the reaction force. Control of thrust of spacecraft can also be achieved by the impulse the thruster conveys to the spacecraft.

Control of the magnitude of thrust (**throttling**) it is possible with liquid propellant engines. It is achieved by controlling flow rate of propellant components. In traditional engines with **de Laval nozzle**, the magnitude can be controlled in a limited range – from maximum to a safe minimum, at which burning of components remains stable. In **aerospike engines**, the magnitude can be controlled in a wider range.

When an even number of six or more engines is used, the magnitude can be reduced by shutting down a pair or more opposite engines.

Magnitude of thrust of solid propellant engines cannot be controlled.

Control of thrust (2 of 4)

The magnitude of thrust is controlled for several reasons – to achieve desirable acceleration, to counteract disturbances, to reduce acceleration of a rocket carrying a manned spacecraft when a considerable part of the propellant has been already spent, to reduce the uncontrollable **residual impulse** of the engine after the **shutdown** (**thrust termination**) prior to staging or separation of the payload.

Control of the direction of thrust. When the resultant force of all thrusters does not pass through the center of mass of the vehicle, a torque appears. This torque can be controlled by changing the direction of the vector of thrust in order to achieve desirable rotation of the vehicle around its axes of inertia.

Jet rudders. Four blades (or vanes) made of heat-resistant material, usually, graphite, are installed right in the exhaust jet at the outlet of the nozzle. The direction of thrust is changed by swinging the blades.

Control of thrust (3 of 4)

By turning jet rudders in the plane I-III (see picture in slide 97) by the same angle to the left, we create torque around the Z axis of the vehicle and make it turn to the left. We can turn the rocket to the right by turning these blades to the right. This implements control of yaw. Similarly, we can control pitch by turning blades in the plane II-IV up or down. By turning any or both pairs of blades in the same plane in opposite directions we can control roll.

Jet rudders were used in early rockets.

Steering thrusters. Four small engines are installed at the bottom of the rocket near the main engine. Each is mounted in a suspension providing one-axis rotational degree of freedom. The direction of thrust is changed by swinging the thrusters similarly to swinging jet rudders. The swinging is done by a **steering machine**, individual for each thruster.

149

Control of thrust (4 of 4)

Swinging of the nozzle of the main engine. Either the nozzle or the whole engine is swung being suspended in a biaxial gimbal for control of pitch and yaw. With only one main engine, for controlling roll, four small thrusters can be used. Pairs of such thrusters are placed in the same plane (for instance, I-III) on the opposite sides of the bottom in fixed position perpendicular to the X axis. Two opposite thrusters of each pair are fired to control the roll.

With four or more main engines, it is possible to control pitch, yaw and roll by swinging the engines similarly to the steering thrusters.

Air rudders were used in early rockets. Four blades were mounted at the bottom outside the hull. By turning them similarly to turning the jet rudders, it is possible to control all three angles of attitude. Instead of changing the direction of thrust, they create **aerodynamic force** perpendicular to the X axis and thus create the control torque.

150

Ion thrusters (1 of 2)

An ion thruster generates thrust by ionizing a propellant and accelerating ions. For acceleration, electrostatic field (Coulomb force) or electromagnetic field (Lorentz force) can be used. In the course of acceleration, ions are focused into a narrow jet.

The exhaust velocity of ion thrusters reaches 15 to 20 km/s, and theoretically can be much higher. The specific impulse is typically 1500 s to 2000 s, and up to 10,000 s in experimental models. Thus, they need much less propellant for conveying the same impulse to a spacecraft than other types of propulsion.

However, their thrust is small, typically in the range of 20 mN to 250 mN, and up to 2 N in some models. They require a lot of electric power for ionizing propellant and accelerating ions, from 400 W for a 20 mN thruster to ten or more kilowatts for more powerful thrusters.

Ion thrusters (2 of 2)

In most ion thrusters, xenon is used as propellant. Sometimes, argon. In experimental thrusters, liquid cesium, which is vaporized during ionization, and some other substances.

In ionization, molecules of the propellant are turned into **cations** (positively charged ions). Then, a jet of cations is ejected, while excessive electrons remain with the spacecraft. In order to avoid build-up of negative charge on the spacecraft to billions of volts, **electron guns** are used, similar to those that were used in CRT screens of old TV sets and computer displays. Electrons are ejected into the jet of ions, where they randomly recombine with ions.

Ion thrusters can operate only in vacuum. They can work in short intervals or continuously over long time periods. They are used for attitude control and maneuvering, and, sometimes, as the main propulsion for accelerating interplanetary robotic spacecraft.

Hall-effect and plasma thrusters

A **Hall-effect thruster** (or just Hall thruster) is a type of an ion thruster that is based on the Hall effect, appearance of voltage on a conductor in magnetic field in transversal direction to the field. It uses strong magnetic field perpendicular to strong electric field to create a “virtual cathode” for accelerating ions, and a real cathode called neutralizer emits electrons into the stream. As a result, electrons and ions are ejected together, like in plasma thrusters.

A **plasma thruster** uses plasma (a mix of ions and free electrons) as a propellant. Instead of using high voltage for accelerating ions, they induce currents and voltages in plasma itself, and this results in acceleration of plasma as a whole rather than of just ions. The exhaust velocity of the propellant and specific impulse in these thrusters are somewhat lower than in ion thrusters, but they have their advantages and are used in spacecraft for the same purposes as ion thrusters.

Reaction wheel (1 of 2)

A **reaction wheel** is an electric motor attached to a flywheel and capable of rotating it in either direction. Its stator can be partially placed inside the flywheel. When the motor accelerates or decelerates the rotor, this creates a torque to the stator around the axis of rotation of the rotor in the direction opposite to the acceleration/deceleration.

Bearings of reaction wheels produce only tiny friction, and the whole device operates in vacuum to exclude aerodynamic drag.

Three reaction wheels mounted on a spacecraft coaxially with its main axes of inertia provide a possibility of attitude control around all three axes without using thrusters and propellant or any external forces for creating torque. If we need the spacecraft to turn clockwise around an axis, we rotate the respective wheel counterclockwise, and then, when we need to stop rotation of the spacecraft, we decelerate the wheel to a halt.

Reaction wheel (2 of 2)

Reaction wheels act almost immediately, within milliseconds, and provide fast and accurate attitude control.

Reaction wheels can accumulate considerable angular momentum. Their maximum rotation rate is typically 30,000 to 60,000 rpm. However, when they reach the maximum angular velocity, they become saturated and cannot provide more torque in the same direction. Then they have to be “unloaded” from the excessive momentum with other actuators, for instance, thrusters.

A reaction wheel can be used as a **momentum wheel**. If we preliminary impart a rotational momentum to the wheel, it becomes like a gyro with the spacecraft being its gimbal. A disturbing torque around an axis perpendicular to the axis of spin will be counteracted by the gyro, however, a precession around the third perpendicular axis will begin.

Magnetorquer

A **magnetorquer** is an electromagnet installed on a satellite for creating a controlled torque by interaction of the magnetic field created by the electromagnet with the magnetic field of the Earth.

Three magnetorquers positioned perpendicularly to each other on a spacecraft can create a resultant magnetic dipole of any spatial direction and of required magnitude, thus providing torque around any desirable axis. With addition of a model of the magnetic field of Earth and known current position of the satellite in space, this enables attitude control.

Magnetic field of Earth is weak, therefore, the control is slow. The model of the field is not precise, besides, the field changes being influenced by processes inside the Earth and by solar activity, therefore, the control is not accurate. However, magnetorquers is a cheap method for unloading reaction wheels. Influence of magnetorquers on other equipment of the spacecraft should be also considered.

Servomotor

A **servomotor** is a circular or linear electric motor with built-in feedback sensors reporting its current position and angular velocity (or only the position). Different types of electric motors can be used as servomotors. It is important that they should be easily reversible.

A servomotor is often incorporated with a local closed-loop control system operating as a PID controller (see slide 45). Then a larger control system using such servomotor acts as a “program setter” to the local control system of the motor (see slide 18) – it can set the required angular or linear position and/or velocity to the local control system, and the latter will operate the motor accordingly using the local feedback loop.

Modern servomotors may include a digital controller as a part of their local control system. Then the required position and velocity can be supplied to it in a digital format.

157

Solenoid actuator

A **solenoid actuator** consists of a **solenoid** (a coil of wire wound on an empty core), a **plunger** made of a ferromagnetic material and a source of force returning the plunger into its initial position (usually a steel spring). When electric current is applied to the coil, the latter creates a magnetic field that pulls the plunger into the coil attempting to center it across the length of the coil. When the current is cut off, the spring returns the plunger to its initial position.

This actuator converts electric signals into linear “pull” motion. By attaching to the plunger a non-magnetic rod passing through the spring (or by using a strong spring), the actuator may be used for producing linear “push” motion.

158

Valves

Valves are used for controlling flow of liquids and gases through pipes. They can have two states – fully open or completely shut (“bang-bang” valves) or can be set in any intermediate state between fully open and completely shut, including the terminal states.

Valves with two states can be controlled by solenoid actuators, and valves with any intermediate state, by servomotors. Often, for the purpose of control, pressure sensors before and after the valve may be needed, and, sometimes, a flow meter either before or after the valve.

Valves of different sizes and different purposes, substances, pressures and flows are used in vehicles. Some of the uses: control of feeding of propellant components, control of pressurization in pressure feed systems, in fueling pipelines, etc. In manned spacecraft, there are many more systems requiring control with valves.

Hydraulic and pneumatic actuators

With **hydraulic actuators**, it is possible to control a strong force by applying a small force. This possibility is based on the property of liquids to distribute applied pressure across their volume. It is also possible to control high power with much smaller power. This is because control of flow requires less power than maintaining of flow. However, a source of high hydraulic power is necessary in this case.

For instance, swinging of steering thrusters or nozzles of the main engine of a rocket requires strong force, while not too much power. Steering machines performing this swinging are often hydraulic.

With **pneumatic actuators**, it is possible to quickly release energy (to provide high power over short intervals of time), which was previously stored or is slowly accumulated in the form of compressed air or gas.

For instance, small attitude control thrusters of spacecraft using compressed gas as propellant are pneumatic actuators.

Single-action actuators

Such actuators are used for operations performed only once during a mission. For instance, this may include staging of the rocket, separation of the payload from the last stage, dropping of certain components, like fairing, deployment of deployable elements of spacecraft (like solar panels and antennas), opening of the parachute at reentry, etc.

These include latches and clips, which are initially locked and then are opened when necessary by a solenoid, small solid propellant thrusters for pulling off empty stages from the remaining rocket, etc.

When strong forces are needed to hold a separable element before separation, **pyrogenic fasteners** are used. Of them, **explosive bolts** are common. An explosive bolt is a bolt with a small charge of an explosive material and an electrically activated primer. When electric current is applied to the primer, the charge explodes and shears the bolt, releasing the separable element.

The end

This is the end of the course, but not the end of the story about control systems of launch vehicles and spacecraft.

A lot of things were left beyond this course: algorithms of control and models of controlled motion, aspects of implementation of the control systems, many devices and engineering solutions specific to particular missions, ground control of space missions and so on. These are topics for separate courses for future professionals in control systems.

However, this course provides a sufficient framework for further study of the subject and for awareness of the control systems to those who design and operate vehicles, but not their control systems.