NAVIGATION

What Is Navigation?

- ***Navigation*** refers to the determination, at a given time, of the vehicle's location and velocity (the "state vector") as well as its attitude.
- A ***navigation sensor*** is a device used to measure a property from which the navigation system computes its navigation solution; examples include accelerometers, gyroscopes, and radio navigation receivers.

- ***Positioning*** is the determination of the position of a body, but not its velocity or attitude. Many navigation technologies, though strictly positioning systems, operate at a high enough rate for velocity to be derived from the rate of change of position.
- ***Tracking*** **or** ***surveillance*** differs from navigation in that the position and velocity information is obtained by a third party without necessarily using equipment on board the object tracked.

Most navigation techniques are based on either of two fundamental methods:

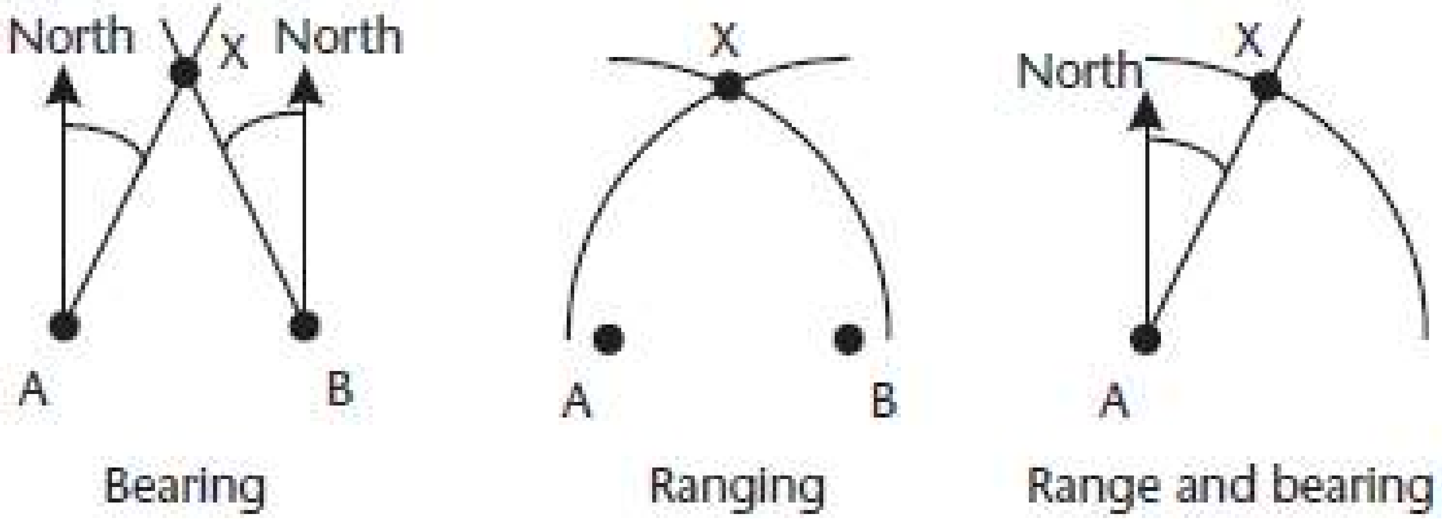
- **Position fixing**
- **Dead reckoning.**

1) Position Fixing

- There are a number of position-fixing methods.
- **Feature matching** compares features at the current location, such as landmarks, waypoints, or terrain height, with a map to determine the current position. This is generally easier for a human than for a machine.

- A two-dimensional position fix may be obtained by measuring the bearing to two known objects.
- A **bearing** is the angle between the line of sight to an object and north (either true or magnetic).
- The position fix may be extended to three dimensions by measuring the elevation angle to one of the reference objects, where the elevation is the angle between the line of sight to the object and a horizontal plane.

- If **range measurements** are taken from two known objects approximately in the same plane as the user, then the user position will lie on the intersection of two circles centered at the reference objects and with radii corresponding to the range measurements.
- However, there is generally a second intersection point. Often, prior information can be used to determine the correct position. Otherwise, a third range measurement is needed.

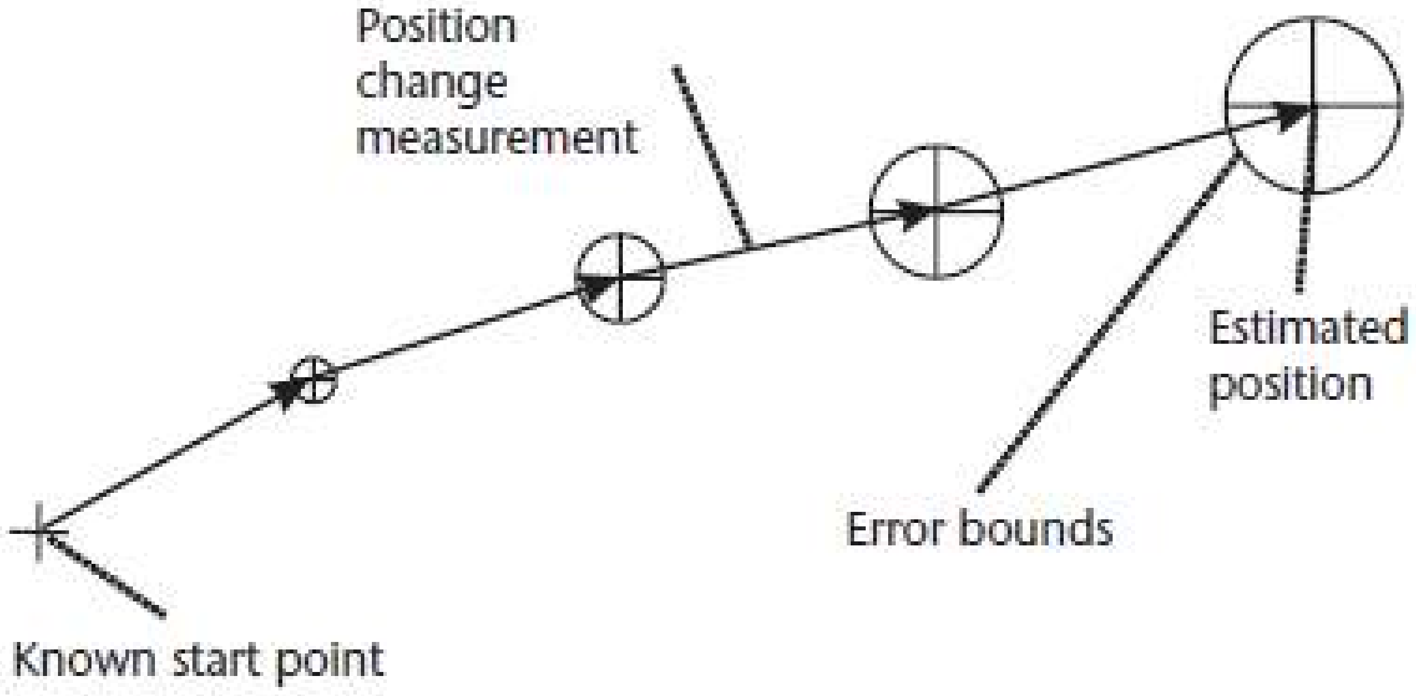


The bearing and ranging position fixing methods.

2) Dead Reckoning

- Dead reckoning (derived from ‘‘deduced reckoning’’) either measures the change in position or measures the velocity and integrates it.
- This is added to the previous position in order to obtain the current position.

Figure illustrates the concept of dead reckoning. Where the attitude is changing, the smaller the step size in the position calculation, the more accurate the navigation solution will be

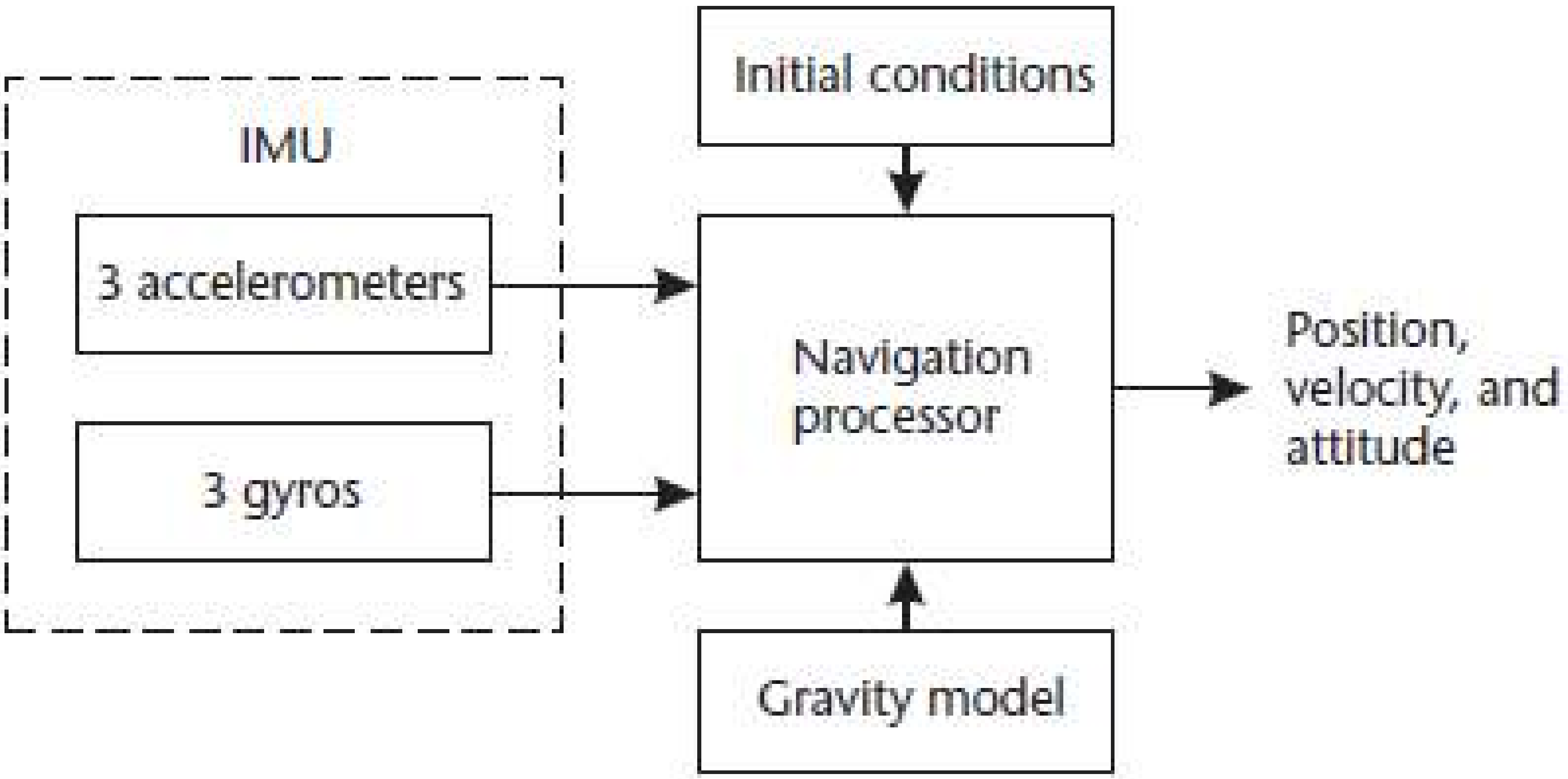


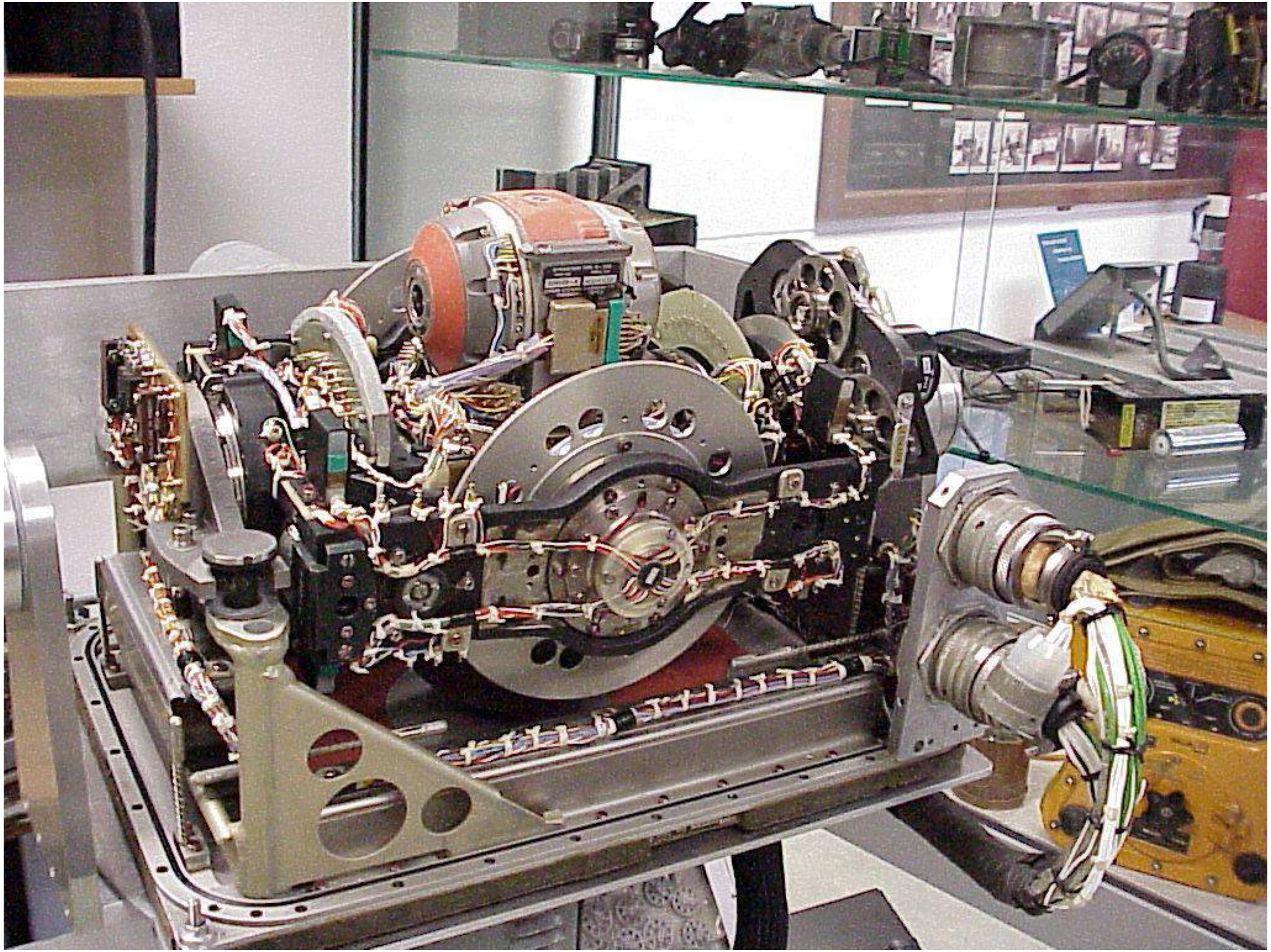
The dead reckoning method.

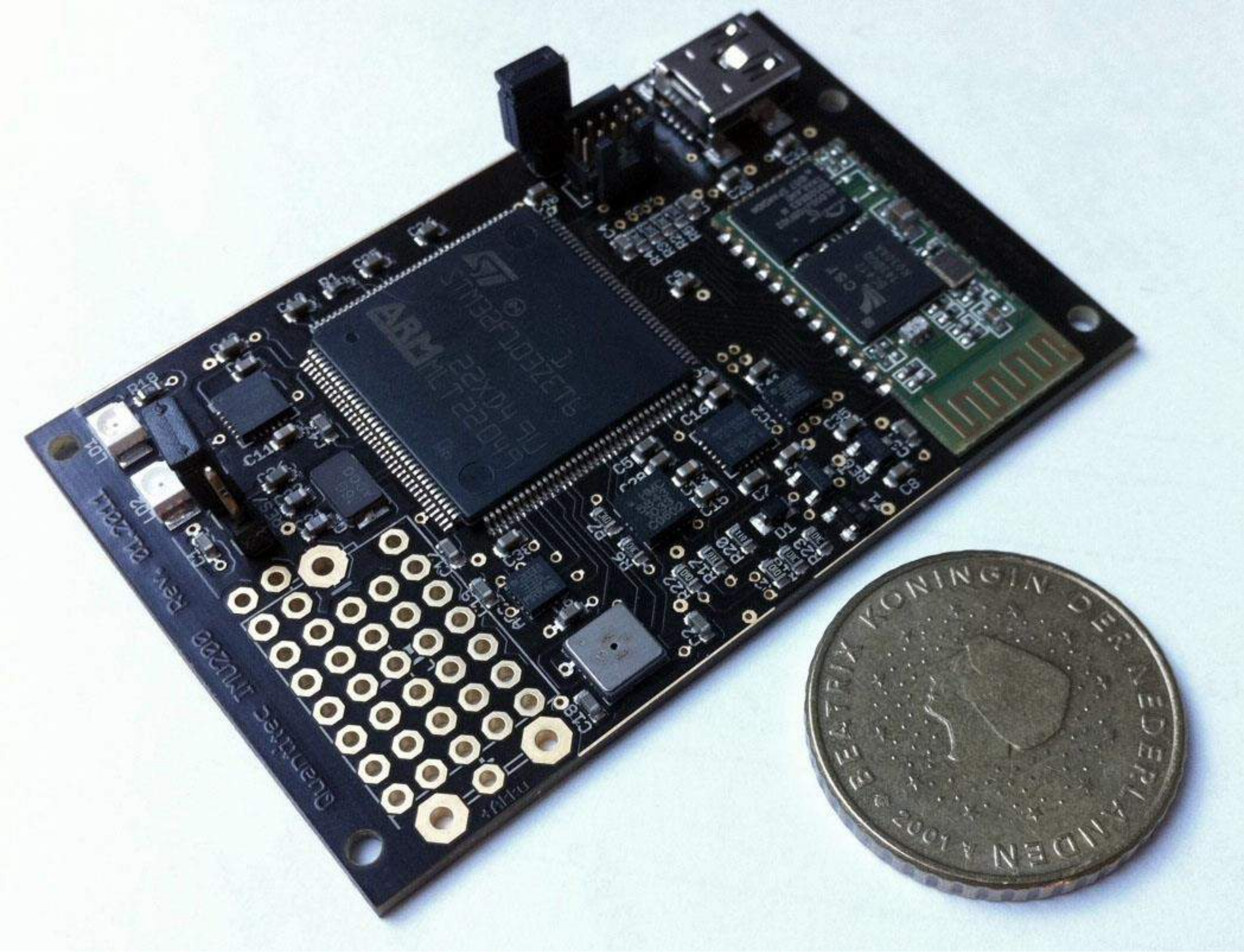
- The dead reckoning position solution is the sum of a series of relative position measurements.
- Each of these will be subject to an **error**; consequently, the error in the position solution will grow with time.
- Dead reckoning requires a known starting position, but after that will provide an uninterrupted navigation solution.
- A dead reckoning technique may be combined with one or more position fixing techniques in an ***integrated navigation system*** to get the benefits of both methods.

Inertial Navigation

- An inertial navigation system (INS), sometimes known as an inertial navigation unit (INU), is a complete three-dimensional dead-reckoning navigation system.
- It comprises a set of **inertial sensors, known as an *inertial measurement unit* (IMU), together with a navigation processor**.
- The inertial sensors usually comprise **three mutually orthogonal accelerometers and three gyroscopes** aligned with the accelerometers.
- The navigation processor integrates the IMU outputs to give the position, velocity, and attitude.


IMU
3 accelerometers
3 gyros
Initial conditions
Navigation processor
Gravity model
Position, velocity, and attitude




BEATRIX KONINGIN DER NEDERLANDEN
2001

- The **gyros measure angular rate**, which is used by the navigation processor to maintain the INS’s attitude solution.
- The **accelerometers, measure specific force,** which is the acceleration due to all forces except for gravity.
- Integrating the acceleration produces the velocity solution, and integrating the velocity gives the position solution.

Gyroscope designs fall into:

- Spinning mass gyros
- Optical gyros
- Ring laser gyros (RLGs)
- Interferometric fiber optic gyros (IFOGs)
- Vibratory gyros.

- The **errors in an inertial navigation solution grow with time** as successive accelerometer and gyro errors are summed.

The principal **advantages** of inertial navigation are

- continuous operation,
- low short-term noise, and
- the provision of attitude, angular rate, and acceleration measurements, as well as position and velocity.
- The **main drawbacks** are the degradation in navigation accuracy with time and the cost.

Radio Navigation

- The first use of radio for navigation was in 1930 with the first navigation satellite launched in 1961.

There are seven basic radio navigation techniques:

- Marker beacons
- Direction finding,
- Bearing/elevation
- Passive ranging
- Two-way ranging
- Hyperbolic ranging and
- Doppler positioning.

- **Marker beacons** are the simplest technique—receiving a signal indicates that the user is in the vicinity of the transmitter.
- With the **direction finding**, or **angle of arrival (AOA),** technique, the reference station is a simple beacon and may be used for other purposes, such as broadcasting. A rotatable directional antenna is then used by the receiver to obtain a bearing.

- In the **bearing or elevation** technique, the reference station broadcasts a signal that varies with the direction of transmission, enabling the user to obtain a measurement of their bearing and/or elevation to the reference station without the need for a directional antenna.
- In a **passive ranging, or time of arrival (TOA),** system, the reference station broadcasts a timing signal from which the user can deduce their range from the transmitter.

- In a **hyperbolic ranging, or time difference of arrival (TDOA)**, system, the receiver measures the time difference (TD) in the signals broadcast by two transmitters, removing the need to synchronize the receiver clock.
- **Differential passive ranging,** which uses a second receiver at a known location to calibrate transmitter synchronization and other common-mode errors, can also be described as a TDOA technique.
- **Doppler positioning** relies on the transmitter moving along a prescribed trajectory. The receiver measures the Doppler shift of the received signal.

Satellite Navigation

- The world's first satellite navigation system was the U.S. Navy's Transit system. Development started in 1958, with the first experimental satellite launched in 1961 and system operational in 1964.

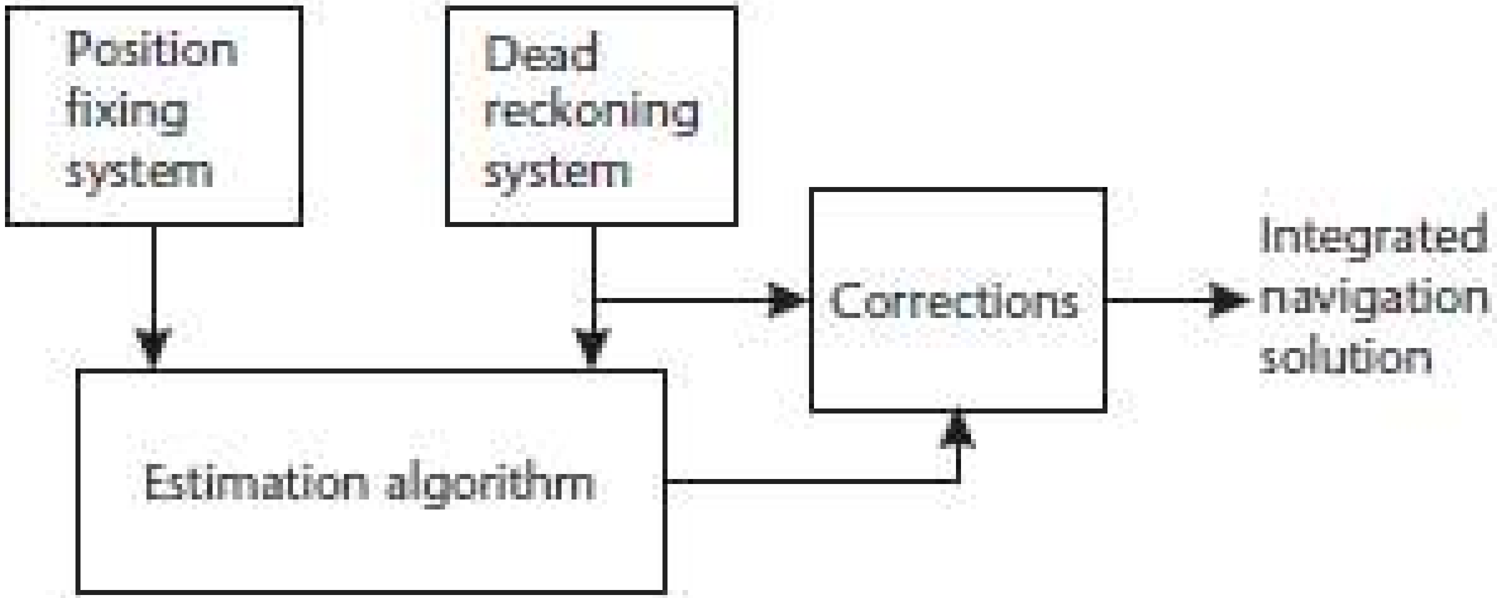
- Development of the **Global Positioning System (GPS)** started in 1973 when a number of U.S. military satellite navigation programs were merged. The first operational prototype satellite was launched in 1978 and initial operational capability (IOC) of the full GPS system was declared in 1993. Although developed as a military system, GPS is now used for a wide range of civil applications.
- **The Global Navigation Satellite System (GLONASS)** is operated by Russia and was developed in parallel to GPS, also as a military system. The first satellite was launched in 1982.
- A third satellite navigation system, **Galileo**, is under development by the European Union and other partners. Galileo is a civil system under civil control. The first satellite launch was at the end of 2005 .
- In addition, regional systems are being developed by China, India, and Japan, with proposals to expand the Chinese Compass system to global coverage.

- GPS, GLONASS, and Galileo are each designed to comprise a constellation of 24 or more satellites orbiting at a radius of between 25,000 and 30,000 km, ensuring that signals from at least four satellites are available at any location.
- The **main limitation** of satellite navigation is lack of signal continuity. GNSS signals are vulnerable to interference, both incidental and deliberate. They can also be blocked, attenuated, and reflected by buildings, terrain, and foliage.

The Complete Navigation System

- Different navigation applications have very different requirements in terms of accuracy, update rate, reliability, budget, size, and mass, and whether an attitude solution is required as well as position and velocity.
- Consequently, different combinations of navigation sensors are suitable for different applications.

- Position fixing and dead reckoning systems have very different error characteristics so, for many applications, a dead reckoning system such as INS is integrated with one or more position fixing systems, such as GNSS.



A typical integrated navigation architecture.

- The dead reckoning system provides the integrated navigation solution as it operates continuously, while measurements from the position-fixing system are used by an estimation algorithm to apply corrections to the dead reckoning system’s navigation solution.
- The estimation algorithm is usually based on the **Kalman filter.**

Inertial Sensors

- **Inertial sensors** comprise accelerometers and gyroscopes, commonly abbreviated to gyros.
- An ***accelerometer*** measures specific force and a *gyroscope* measures angular rate, both without an external reference.
- Devices that measure the velocity, acceleration, or angular rate of a body with respect to features in the environment are not inertial sensors.
- An ***inertial measurement unit*** **(IMU)** combines multiple accelerometers and gyros, usually three of each, to produce a three-dimensional measurement of specific force and angular rate.

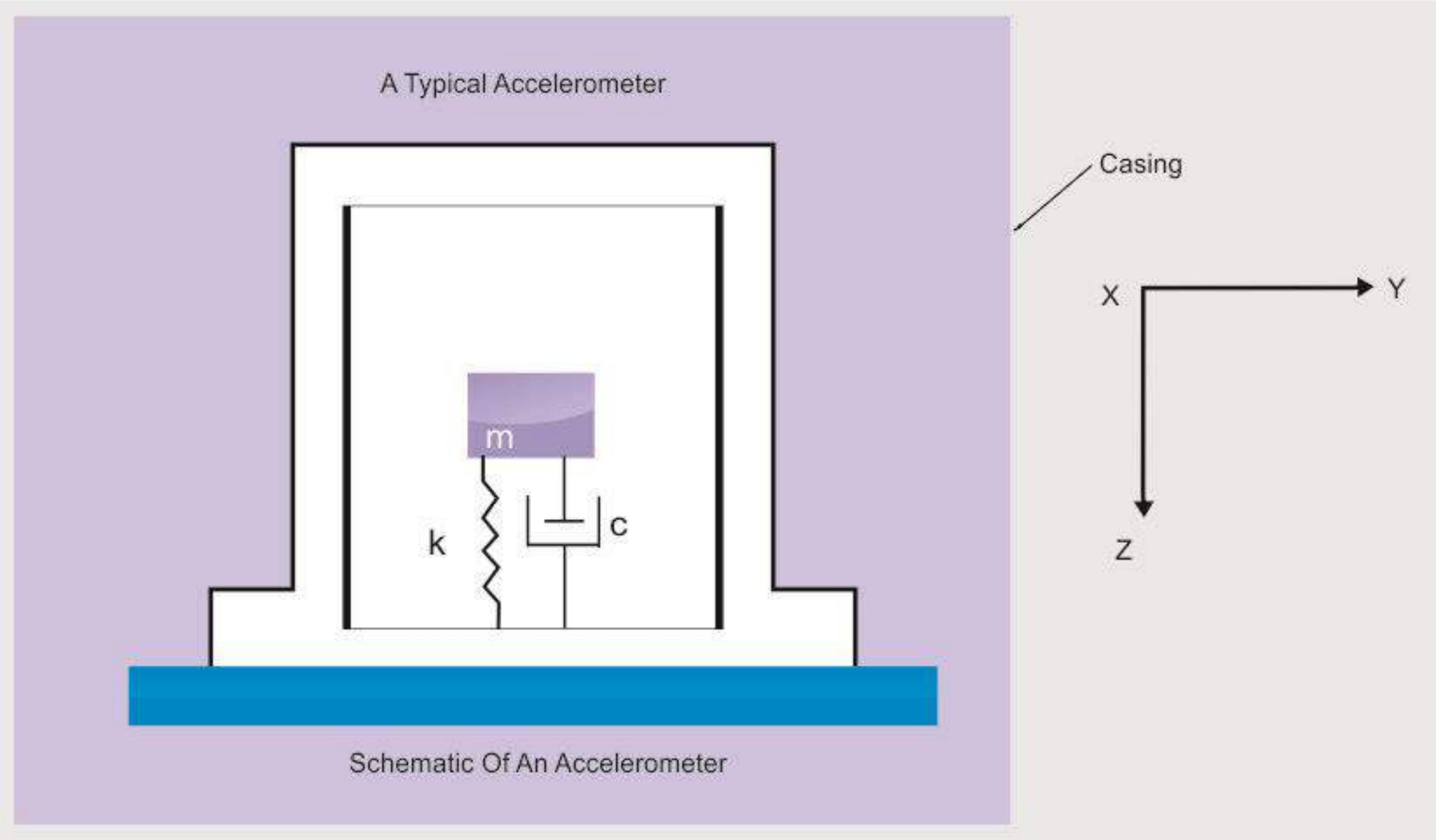
Accelerometers

- **An accelerometer** is a sensor that measures the physical acceleration experienced by an object due to inertial forces or due to mechanical excitation. It converts the mechanical motion into an electrical output.
- The ability of an accelerometer to sense acceleration can be put to use to measure a variety of things like tilt, vibration, rotation, collision, gravity, etc.

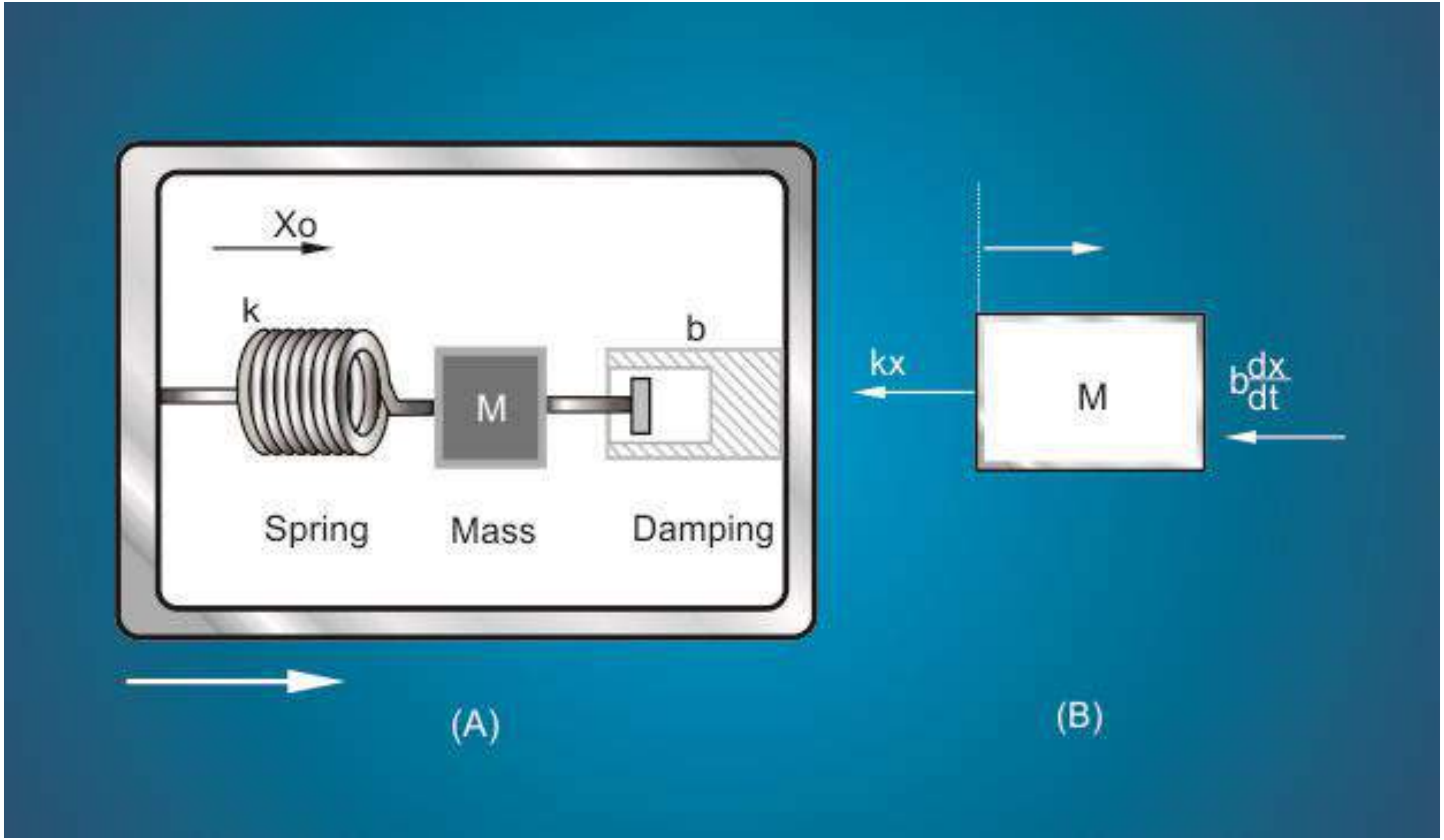
The sensing element is a proof mass (also known as seismic mass). The proof mass is attached to spring which in turn is connected to its casing.

In addition, a dashpot is also included in a system to provide desirable damping effect; otherwise system may oscillate at its natural frequency.

The dashpot is attached between the mass and the casing. The unit is rigidly mounted on the body whose acceleration is of interest.


A Typical Accelerometer
Casing
X
Y
Z
m
k
c
Schematic Of An Accelerometer

- When the system is subjected to linear acceleration, a force acts on the proof-mass. This causes it to deflect; the deflection is sensed by a suitable means and is converted into an equivalent electrical signal.



Xo
k
b
M
Spring
Mass
Damping
(A)
kx
M
$b\frac{dx}{dt}$
(B)

Therefore, if m = proof mass of the body

x = relative movement of the proof-mass with respect to the frame

c = damping coefficient

k = spring stiffness

then

Summation of all forces on Proof mass = 0

$$m\,a + Fd + Fs = 0$$

$$m\,a = -\,Fd - Fs$$

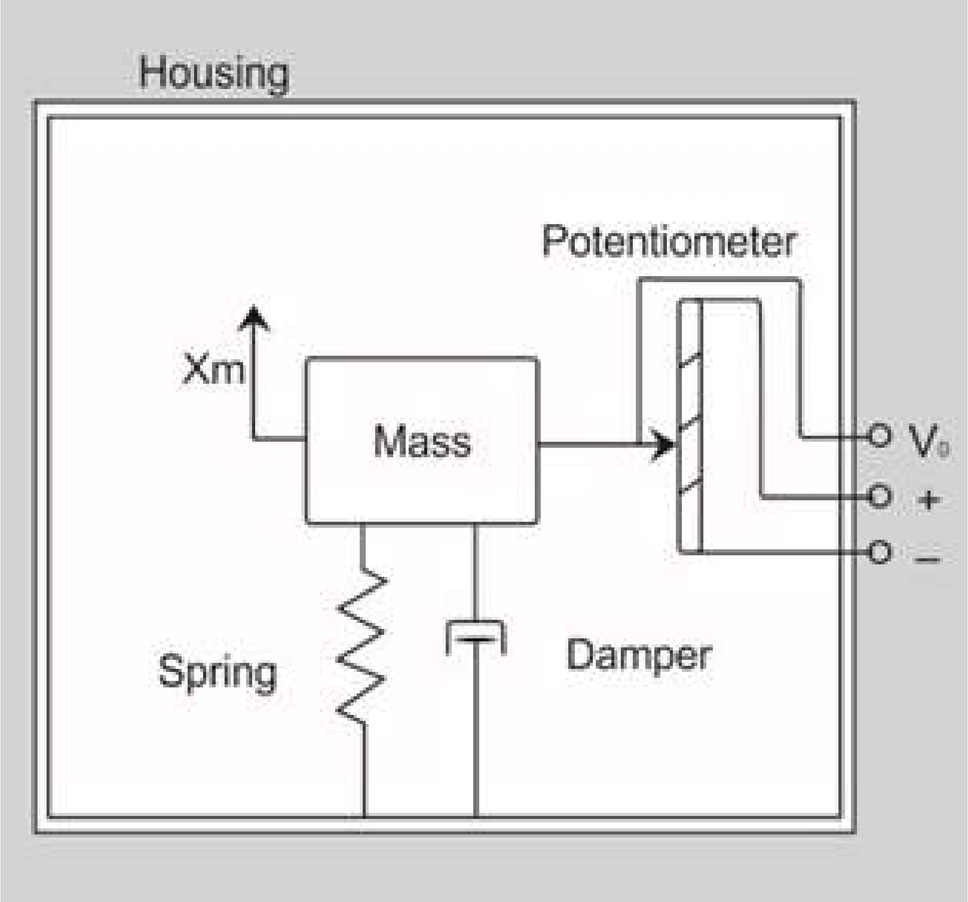
$$m\,a = -\,c\,\dot{x} - k\,x$$

$$a = -\left(\frac{c}{m}\right)\dot{x} - \left(\frac{k}{m}\right)x$$

TYPES OF ACCELEROMETERS

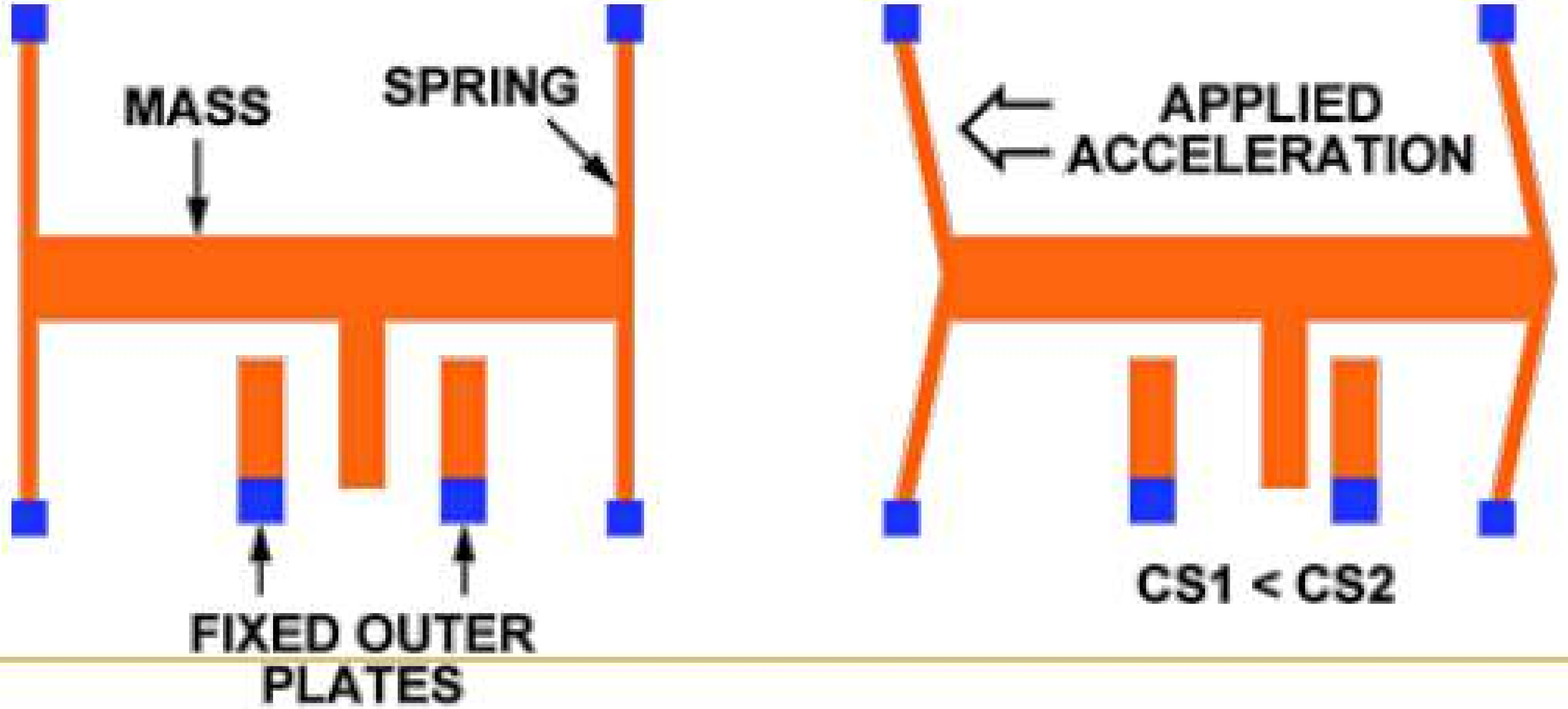
Potentiometric

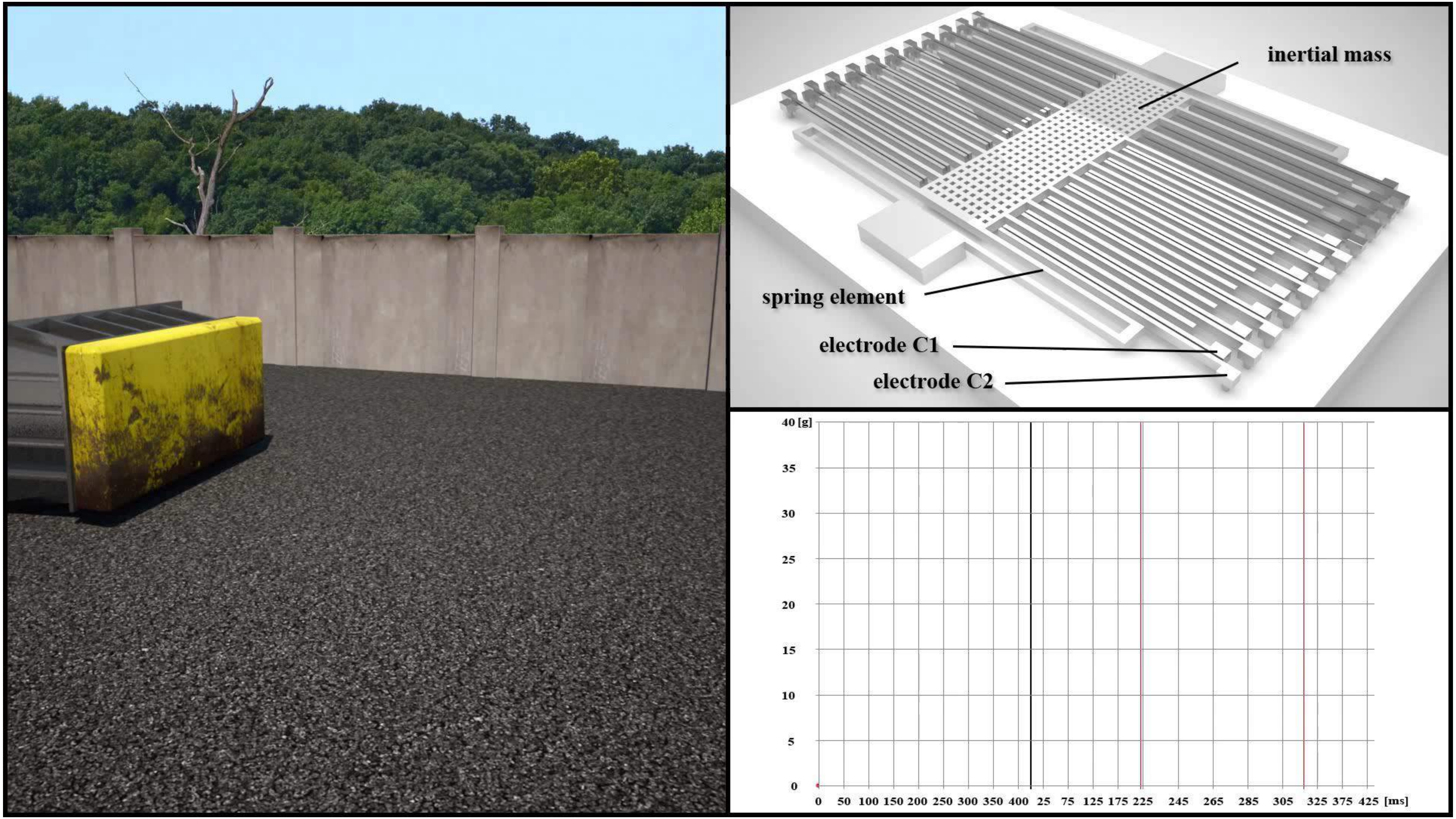
- One of the simplest accelerometer type - it measures motion of the proof mass motion by attaching the spring mass to the wiper arm of a potentiometer. Thus position of the mass and thereby, changing acceleration is translated to changing resistance.



Capacitive accelerometers

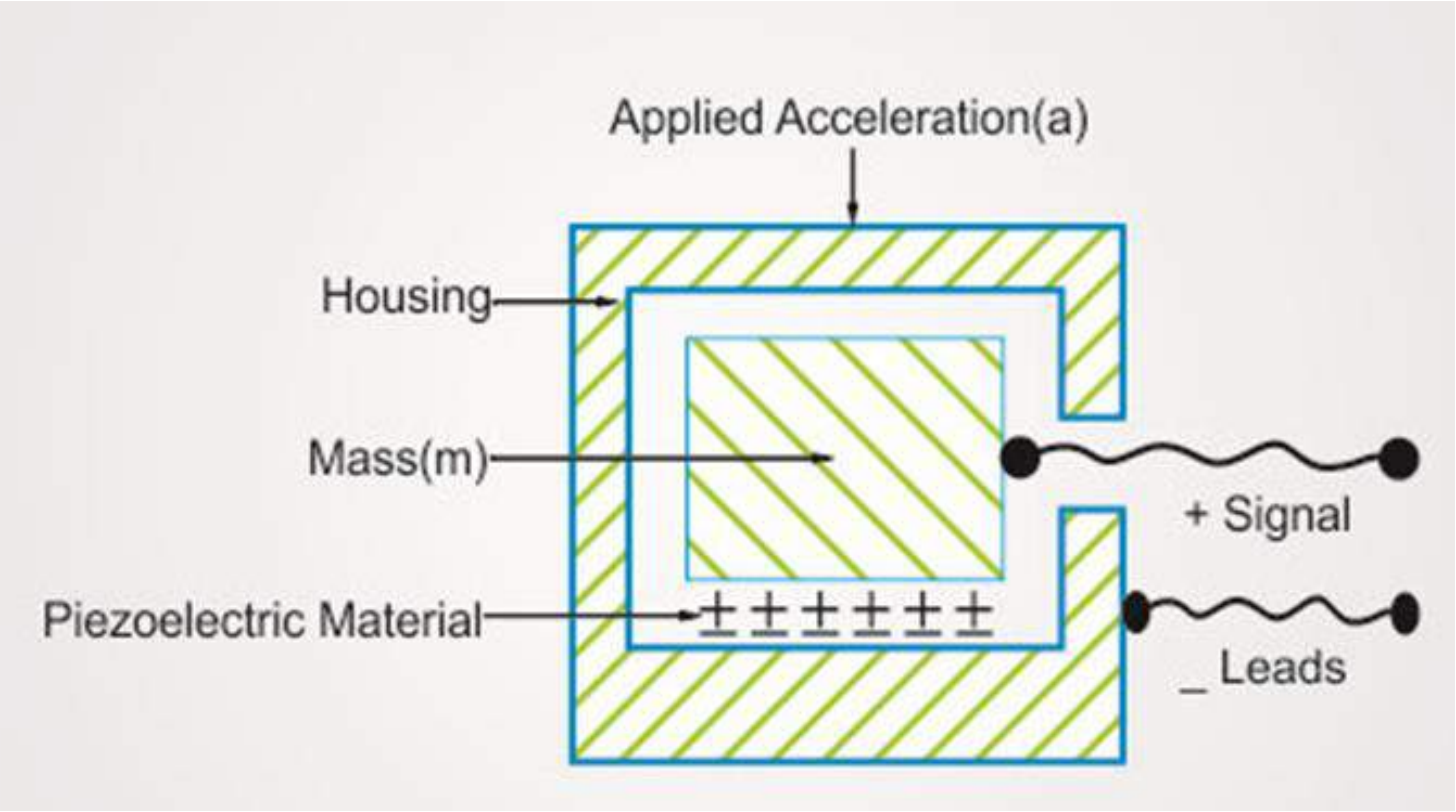
- Capacitive accelerometers sense a change in electrical capacitance, with respect to acceleration. Single capacitor or differential capacitors can be used; differential ones being more common.
- In these accelerometers, a diaphragm acting as a mass moves in the presence of acceleration. The diaphragm is sandwiched between the two fixed plates creating two capacitors; each with an individual fixed plate and each sharing the diaphragm as a movable plate. Movement of the diaphragm causes a capacitance shift by altering the distance between two parallel plates, the diaphragm itself being one of the plates.
- The two capacitors form the two arms of the bridge; the output of the bridge varies with the acceleration.


MASS
SPRING
FIXED OUTER PLATES
APPLIED ACCELERATION
CS1 < CS2


inertial mass
spring element
electrode C1
electrode C2
40 [g]
35
30
25
20
15
10
5
0
0 50 100 150 200 250 300 350 400 25 75 125 175 225 245 265 285 305 325 375 425 [ms]

Piezoelectric accelerometers

- Piezoelectric accelerometers employ piezoelectric effect. When piezoelectric materials are stressed, they are deformed and an electric charge is generated on the piezoelectric materials.
- In piezoelectric accelerometers, piezoelectric material is used as an active element. One side of the piezoelectric material is connected to rigid base. Seismic or proof mass is attached to the other side. When force (generated due to acceleration) is applied, piezoelectric material deforms to generate the charge. This charge is proportional to the applied force or in other words, proportional to acceleration (as mass is constant). The charge is converted to voltage using charge amplifiers and associated signal conditioning circuit.

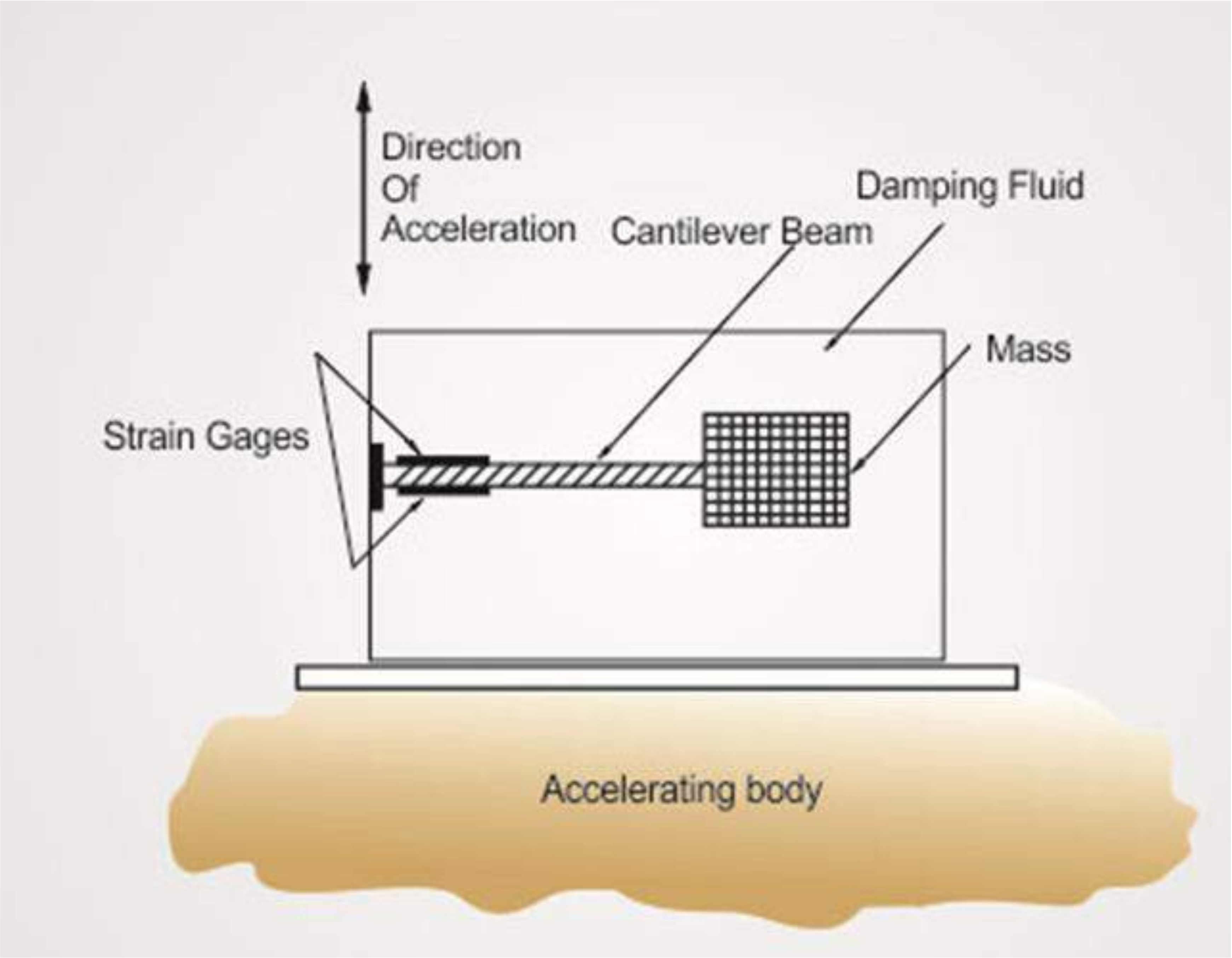

Applied Acceleration(a)
Housing
Mass(m)
Piezoelectric Material
+ Signal
_ Leads

Compared to other type of accelerometers, piezoelectric accelerometers offer unique advantages –

- Wide range
- Excellent linearity
- No wear and tear due to absence of moving parts
- No external power requirement

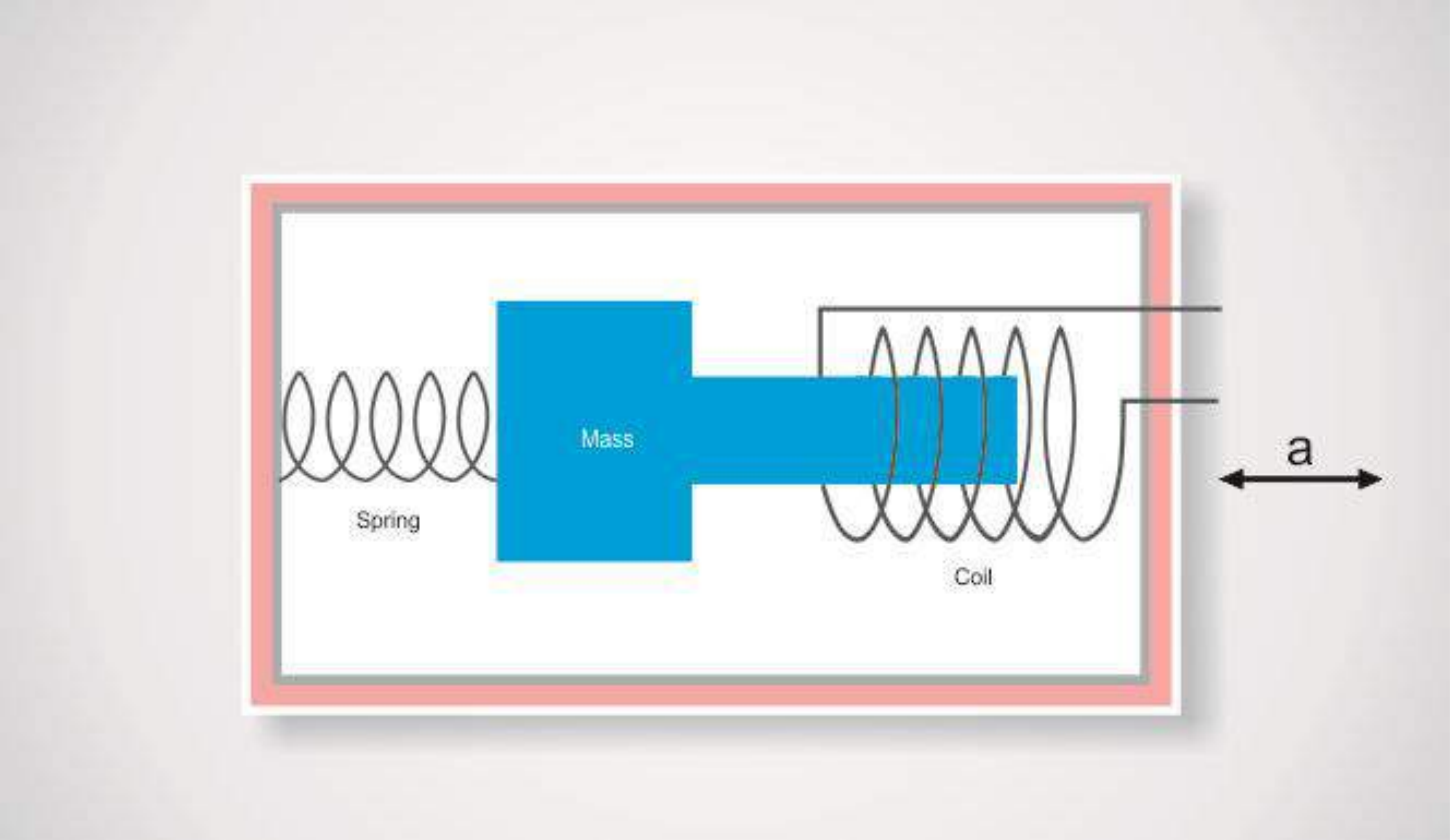
Piezo-resistive accelerometers

- Piezo-resistive accelerometers use piezo-resistive materials, i.e., strain gauges. On application of the force (due to acceleration), resistance of these strain gages changes. The change in resistance is monitored to measure the acceleration. Piezo-resistive elements are typically used in micro-machined structures. They can be designed to measure upto ±1000 g.


Direction
Of
Acceleration
Damping Fluid
Cantilever Beam
Mass
Strain Gages
Accelerating body

Variable inductance accelerometers

- In these accelerometers, proof mass is made of ferromagnetic materials. The proof mass is designed in the form of core which can move in or out of the coil.
- When the body is accelerated, the proof mass moves. In other words, portion of the core inside the coil changes and so the coil impedance.
- Thus, the coil impedance is a function of the applied acceleration.

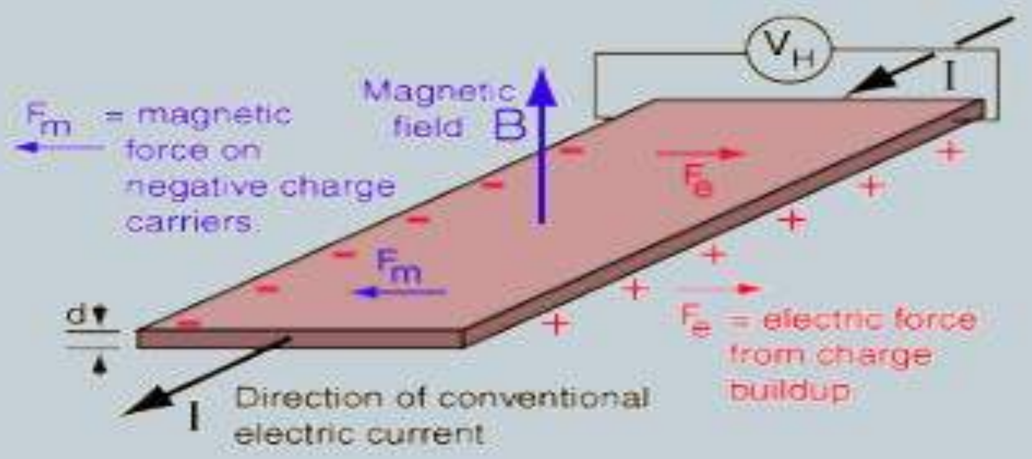


Mass
Spring
Coil
a

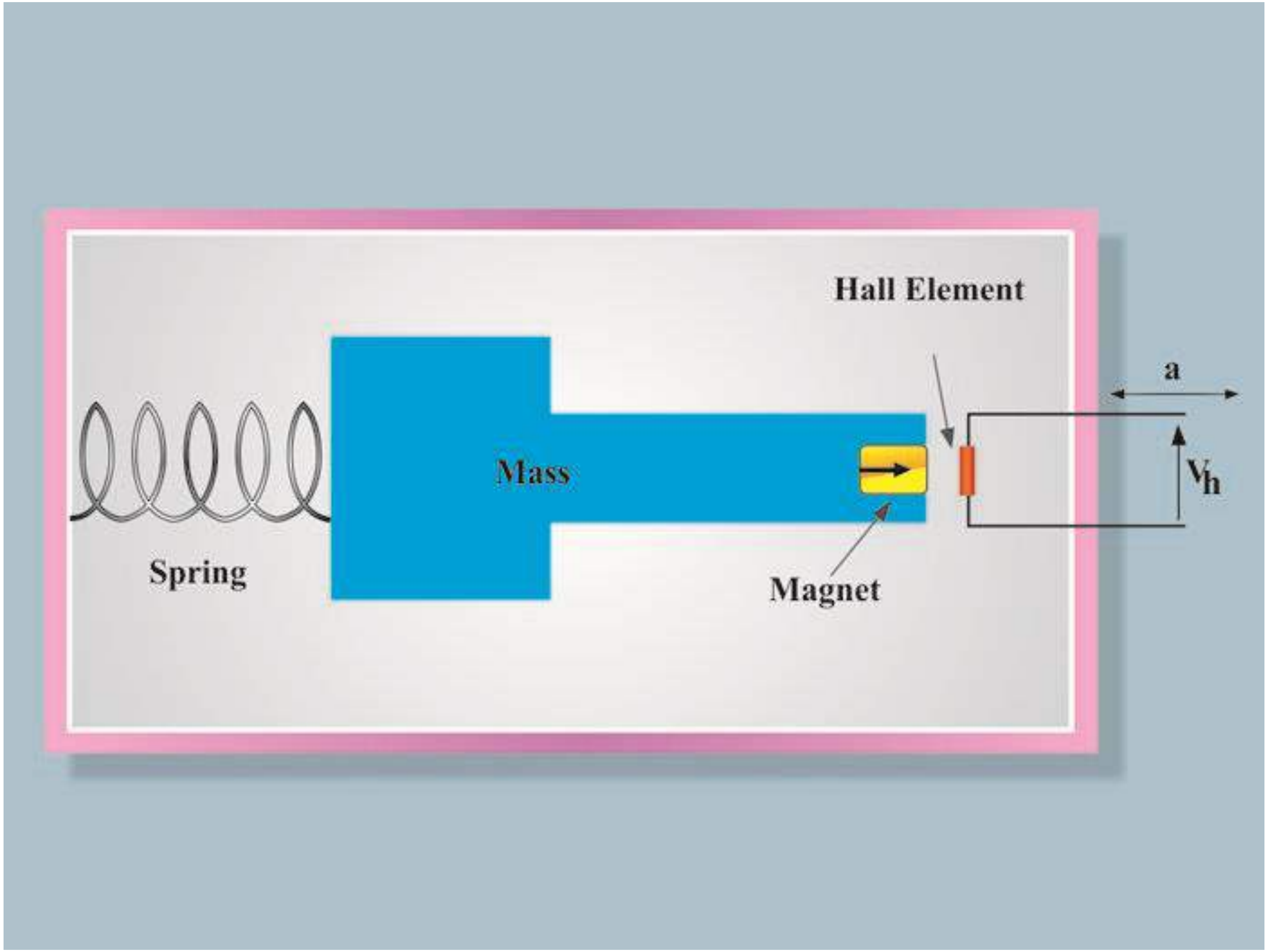
Hall Effect accelerometers

WHAT IS HALL EFFECT?

- The Hall effect is the production of a voltage difference (the Hall voltage) across a current carrying conductor (in presence of magnetic field), perpendicular to both current and the magnetic field.



- Hall Effect accelerometers measure voltage variations resulting from a change in the magnetic field.
- If a magnet is mounted/ integrated on a proof mass, the output of the hall element will vary according to the applied force due to the variation of the magnetic field sensed by the Hall element. Hall voltage is calibrated in terms of acceleration.


Hall Element
a
Mass
V_h
Spring
Magnet

Magnetoresistive accelerometers

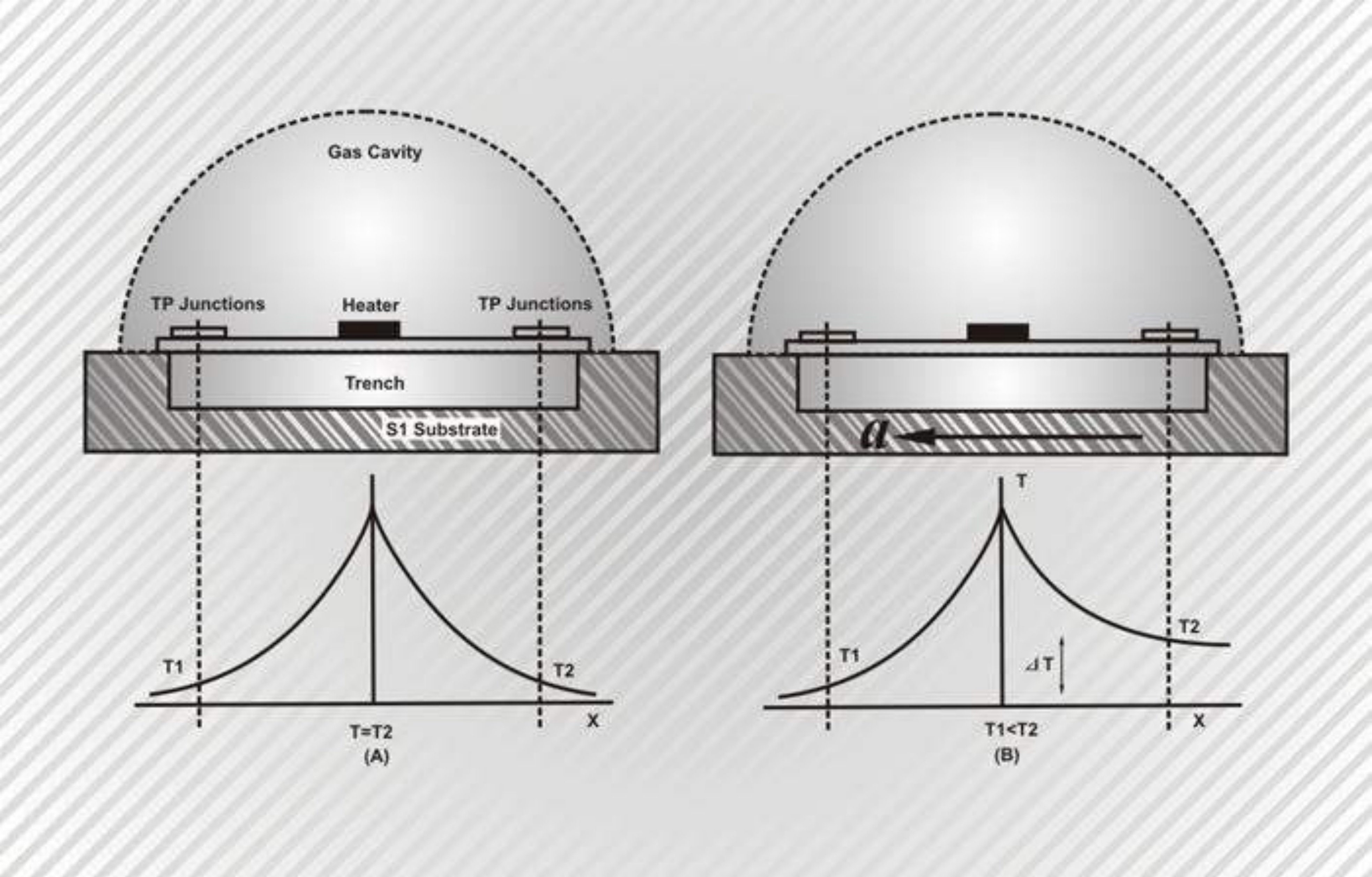
- Magnetoresistive accelerometers employ magnetoresistive effect. Resistance of magnetic materials changes when exposed to varying magnetic field. These accelerometers are similar to Hall Effect accelerometers; the only difference is the use of magnetoresistive material instead of Hall element. Hence, the change in resistance due to the applied acceleration is measured.

FBG Based accelerometers

- A Fiber Bragg grating (FBG) is a type of distributed Bragg reflector fabricated in a short section of optical fiber that reflects specific wavelengths of light and transmits all others.
- When a broad-spectrum light is transmitted through the fiber, and the transmitted beam impinges on the grating, a part of the signal is transmitted through, and another part is reflected off.
- The reflected signal is centered at Bragg wavelengths. Any change in the grating pitch of the fiber caused by strain or temperature results in a shift of Bragg wavelength. This is the property used for sensing of movement of mass in the accelerometers.
- In FBG sensor based accelerometers, the acceleration is coupled to a mechanical load on the FBG. Due to the strain experienced by the FBGs (as a result of applied acceleration), there is a shift in the reflected Bragg wavelengths. Shift in the wavelengths is then calibrated to the level of acceleration.

Heated Gas accelerometers

- Heated Gas accelerometers measure internal changes in heat transfer due to acceleration. These accelerometers use gas as a proof mass.
- Gas is enclosed in a cavity and a heat source is suspended at the center. Two (or more) thermistors are placed at equal distances from the suspended heat source.
- Under rest condition (or zero acceleration), the gas is heated to an equilibrium temperature, the heat gradient is symmetrical, and hence two thermistors are at same temperature. Under acceleration, the heat gradient become asymmetrical due to convective heat transfer, the gas shifts to the direction opposite the motion (the gas is the inertial mass) causing a temperature gradient. The temperature gradient is calibrated in terms of acceleration.

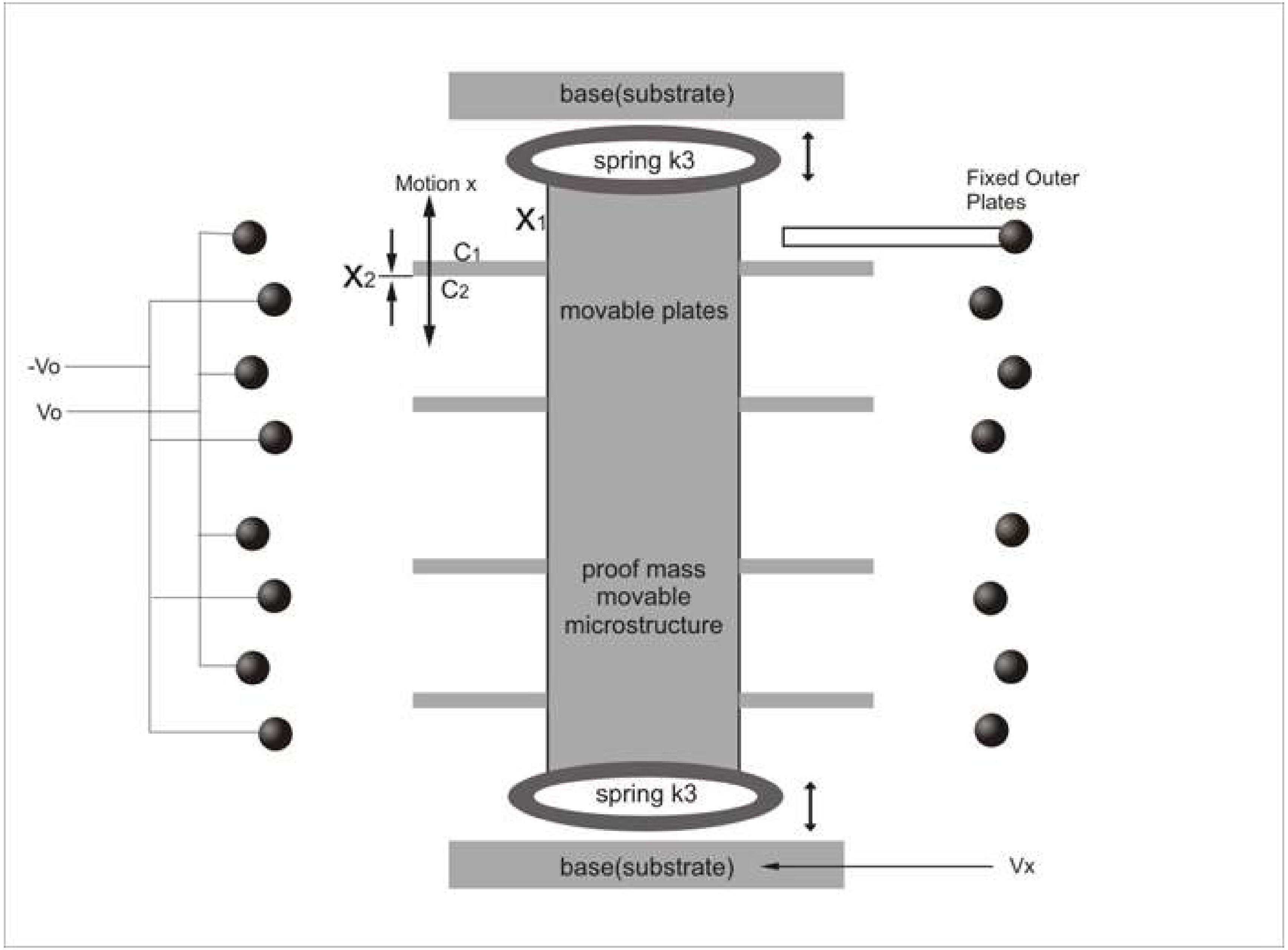


Gas Cavity
TP Junctions
Heater
TP Junctions
Trench
S1 Substrate
T1
T2
X
T=T2
(A)
a
T
T2
T1
⊿T
X
T1<T2
(B)

Amplitude linearity	Often referred to as amplitude non-linearity, amplitude linearity is a measure of how linear the output of an accelerometer is over its specified amplitude range. Amplitude linearity specifies the limits to how far the accelerometer's output will differ from the perfect linearity. Again, amplitude linearity is only valid at a (usually undisclosed) single frequency. It is specified as percentage of reading; sometimes expressed in a piecewise manner also.
Output polarity	Output polarity describes the direction of the accelerometer's output signal (whether it is positive or negative going), given a particular direction of the input acceleration.
Electronic Noise	This is the electronic noise generated by the amplifier circuit. Noise is specified as either "broadband", or "spectral". The broadband measurement is a measurement of the total noise energy over a specified bandwidth. Spectral noise is the noise measured at a specific frequency.
Size and Mass	Size and mass of an accelerometer can change the characteristics of the object being tested. The mass of the accelerometers should be significantly smaller than the mass of the system on which measurement is to be done.

MEMS-Based Accelerometers

- First MEMS accelerometers used piezo resistors. However, piezo resistors are less sensitive than capacitive detection. Most of the MEMS accelerometer use capacitive sensing principle.
- Typical MEMS accelerometer is composed of movable proof mass with plates that is attached through a mechanical suspension system to a reference frame. Movable plates (part of the proof mass) and fixed outer plates form differential capacitor.
- Due to application of the force, proof mass deflects; the deflection is measured in terms of capacitance change.
-

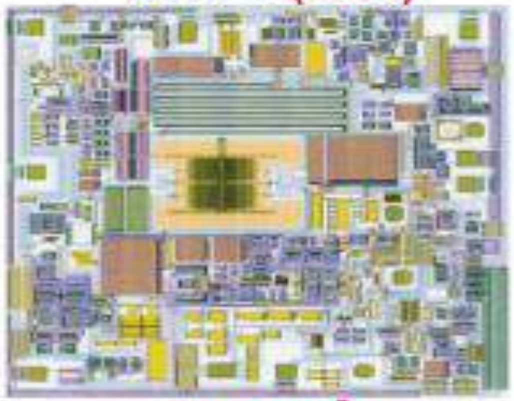


base(substrate)
spring k3
Motion x
Fixed Outer Plates
X1
X2
C1
C2
movable plates
-Vo
Vo
proof mass movable microstructure
spring k3
base(substrate)
Vx

Evolution of MEMS Accelerometers

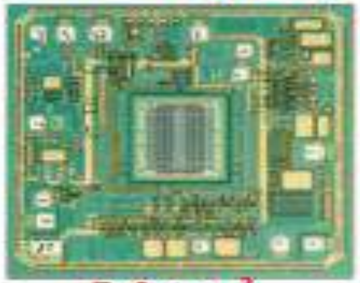
- Analog Devices Accelerometer (Automotive)

ADXL50 (1994)



10.8 mm²

ADXL76 (1996)



5.4 mm²

ADXL78 (2001)

2.7 mm²

ADXL180 (2006)

2.5 mm²

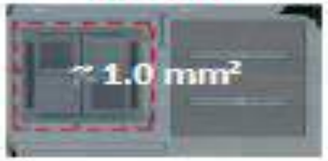
M. Judy, Proc. Solid-State Sensors, Actuators, and Microsystems Workshop, Hilton Head Island, SC, Jun. 2004

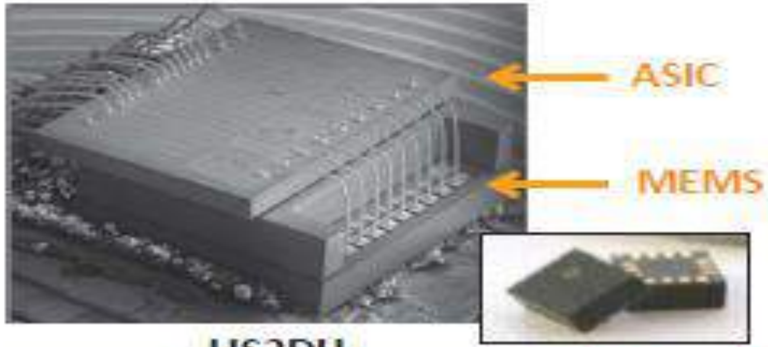
- STMicroelectronics Accelerometer (Consumer)

(2008)

(2010)

(2013)



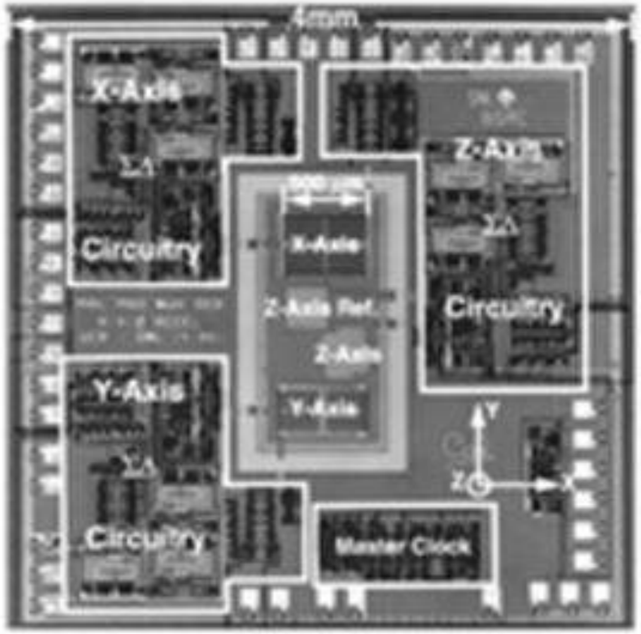


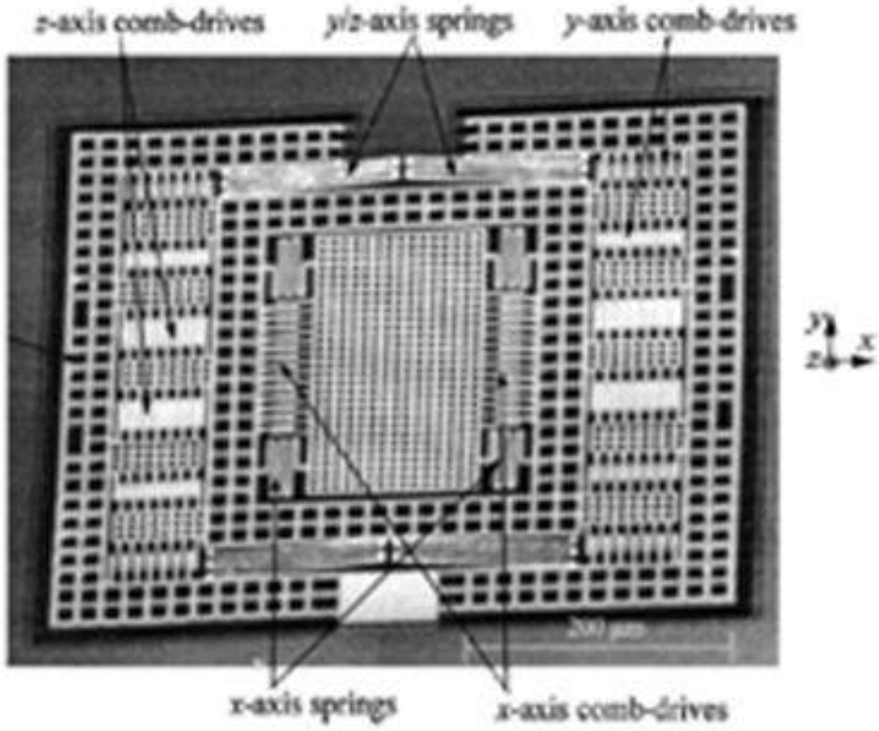
LIS2DH

2x2x0.9 mm³

Source: Yole Developpement, "MEMS Packaging sample report", 2012

• SEM photograph of MEMS 3D accelerometer is shown below





ACCELEROMETER SPECIFICATIONS

- Selection of a sensor requires proper understanding of the specifications. The specifications of an accelerometer include dynamic specifications, electrical specifications and mechanical specifications. Some of the important specifications of an accelerometer are as follows:

Sensitive Axis	Accelerometers are designed to detect inputs in reference to an axis; single-axis accelerometers can detect inputs only along one plane. Triaxial accelerometers can detect inputs in any plane.
Dynamic Range	Dynamic range refers to the maximum amplitude vibration that can be measured by an accelerometer before distortion occurs in the amplifier. It is normally specified in 'g's.
Sensitivity	Sensitivity refers to the ability of an accelerometer to detect motion. Sometimes referred to as the "scale factor" of the accelerometer, it is the ratio of the sensor's electrical output to mechanical input. It is typically specified in terms of mV/g and it is valid only at one frequency (usually 100 Hz) and at particular temperature (25° C). This indicates the voltage output per g of acceleration
Frequency Response	The frequency response specification shows the maximum deviation of sensitivity over a frequency range. More appropriately known as amplitude response, it is the sensitivity specified over the transducer's entire frequency range. The frequency response is specified over a tolerance band; they are specified in percentage and/or dBs, typical bands being ±10%, ±1 dB or ±3 dB. Upper frequency limit is typically governed primarily by the mechanical resonance of the sensor. Lower frequency limit appears because of "high pass" filtering used for reduction of the low frequency amplifier noise.
Mounted Resonance Frequency	This is the primary (largest) mechanical resonance of the sensor when mounted on the structure. At this frequency, accelerometer shows maximum sensitivity.
Transverse Sensitivity	Transverse sensitivity is the sensitivity of the accelerometer at 90 degrees to the sensitive axis of the sensor. Also referred to as cross-axis sensitivity, it is expressed as a percentage of the axial sensitivity. Ideally, it should be zero, but can be as much as 5%.

GYROSCOPES

WHAT IS A GYROSCOPE?

DEFINITION :

A **gyroscope** is a device for measuring or maintaining orientation, based on the principles of conservation of angular momentum.

PROPERTIES OF GYROSCOPES

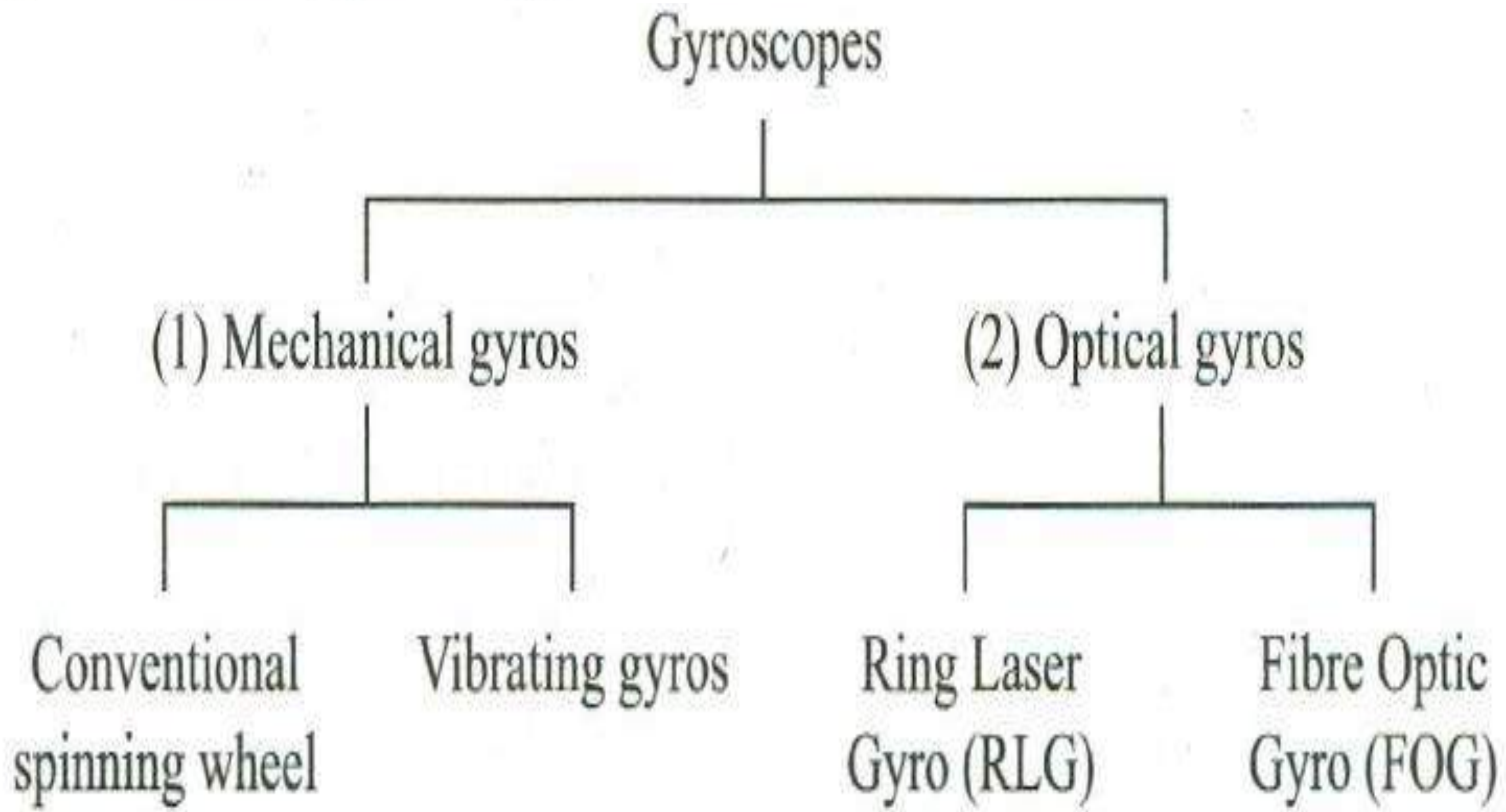
Gyroscopes have two basic properties:

Rigidity and Precession

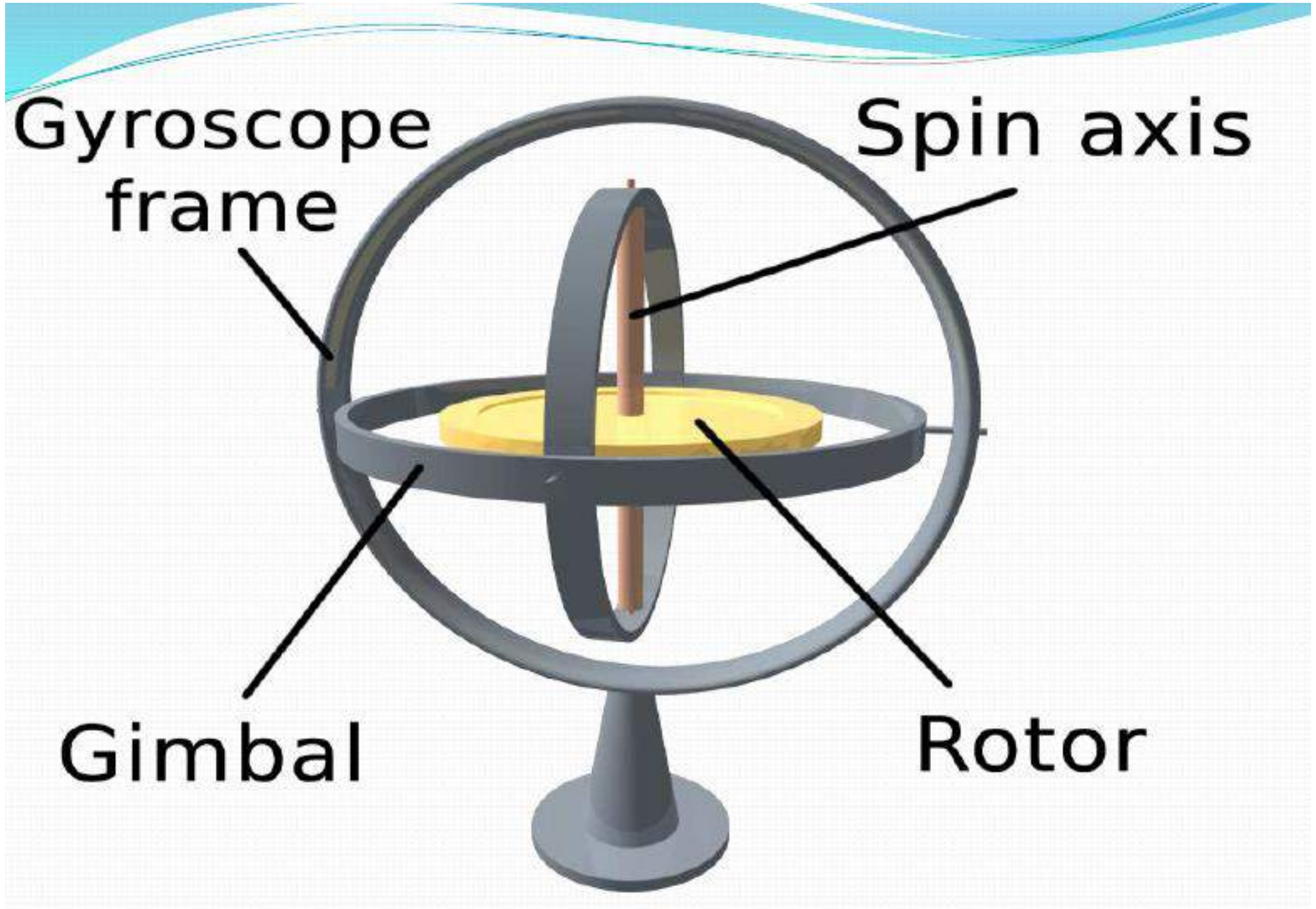
These properties are defined as follows:

1. RIGIDITY: The axis of rotation (spin axis) of the gyro wheel tends to remain in a fixed direction in space if no force is applied to it.
2. PRECESSION: The axis of rotation has a tendency to turn at a right angle to the direction of an applied force

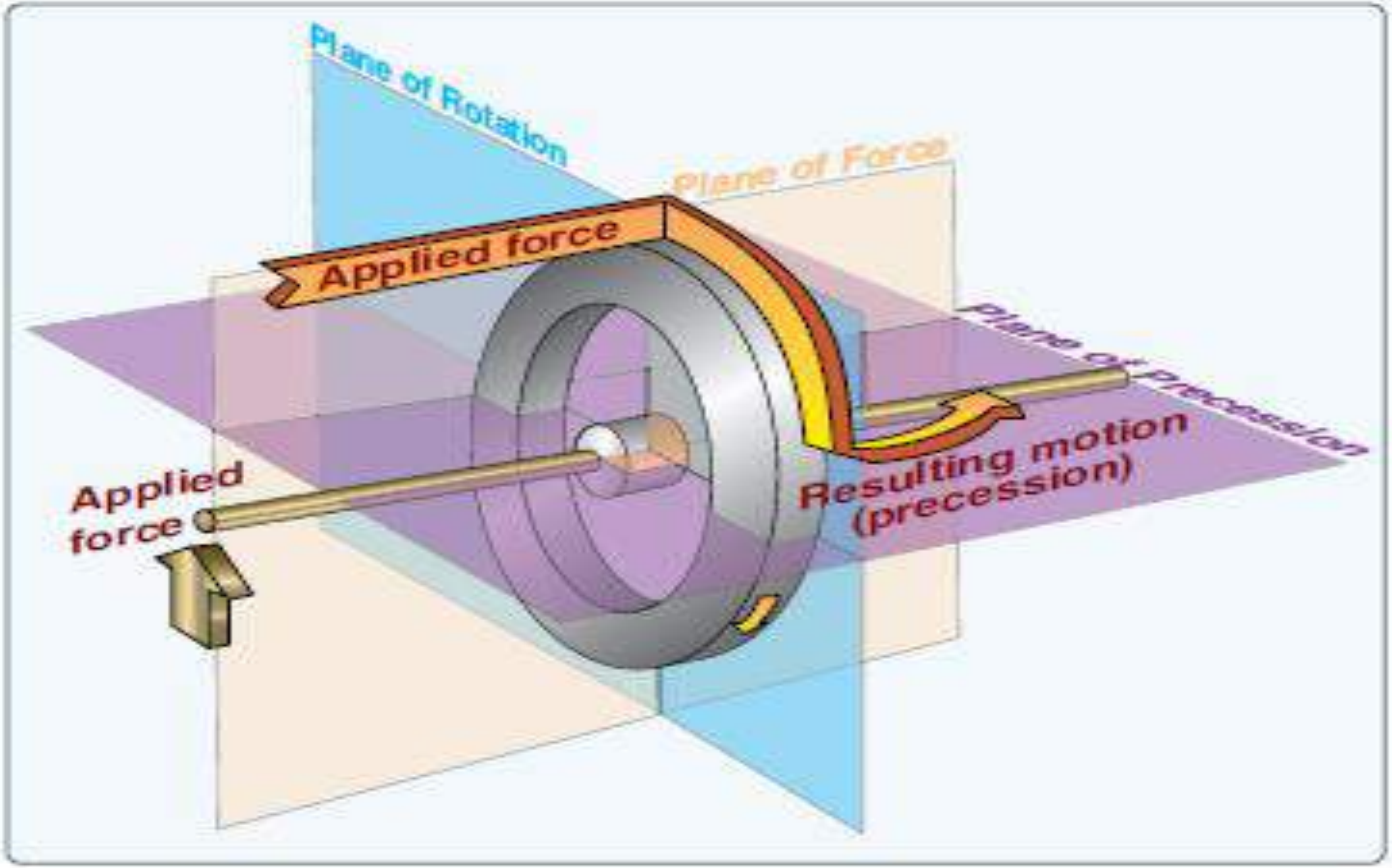
There are basically two main types of gyro



Spinning Wheel Gyros



Gyro precession

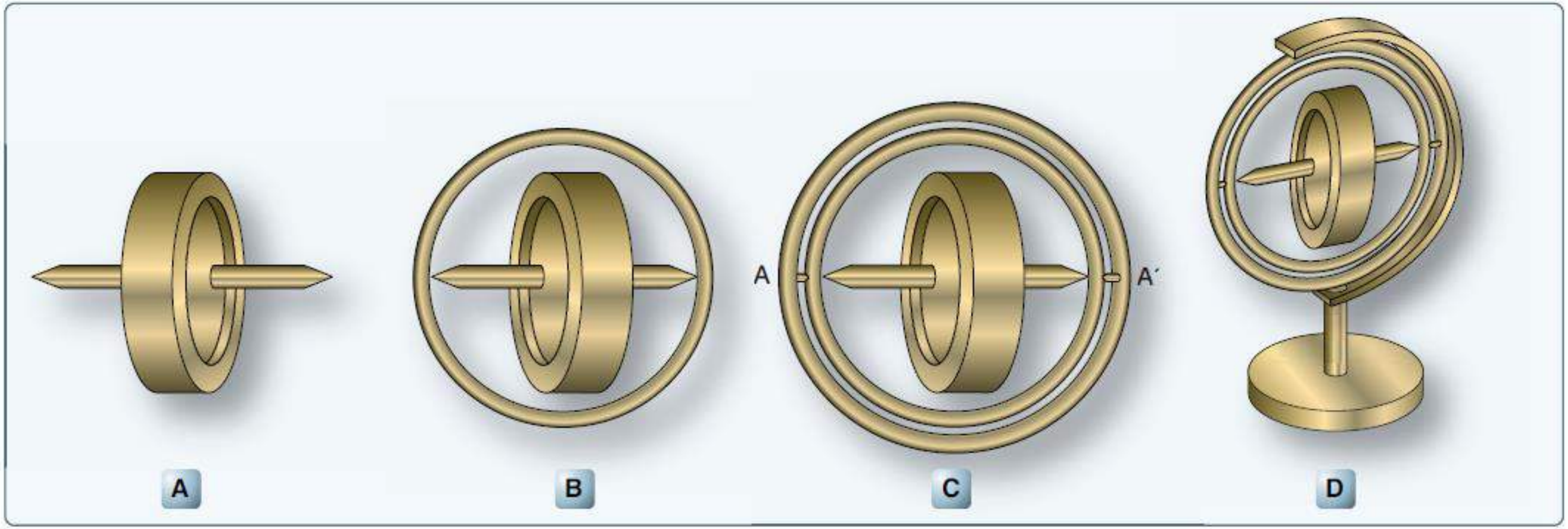


The Cubli

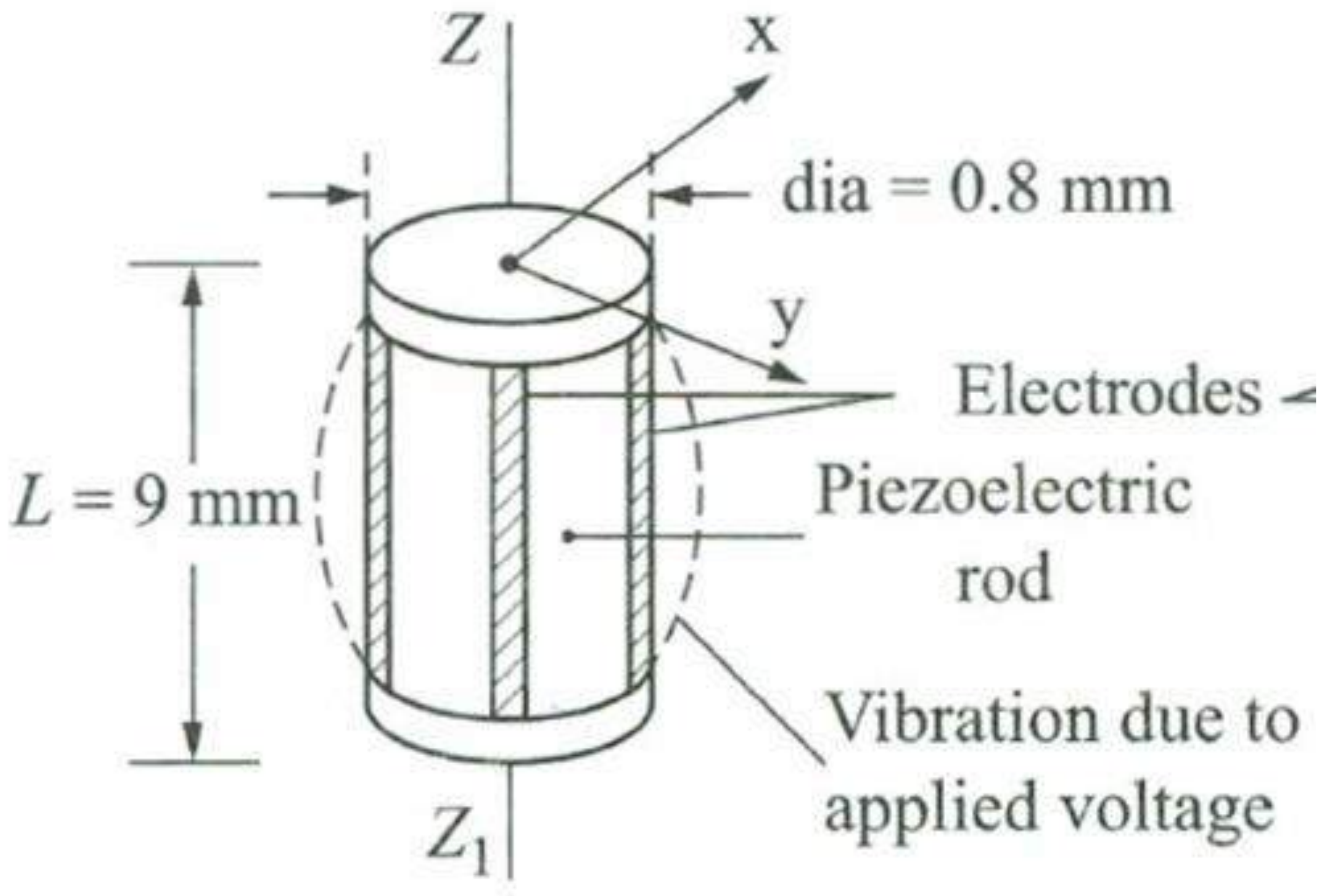
Building a cube that can jump up and balance

IDSC

ETH zürich


A
A´
A
B
C
D

Vibrating Gyros



- ***Micro Electro-Mechanical Systems (MEMS) Technology Gyros***
- These gyros exploit the effects of the Coriolis forces which are experienced when a vibrating mass is subjected to a rate of rotation about an axis in the plane of vibration. There are two basic configurations which are being exploited; **a tuning fork configuration and a vibrating cylinder configuration.**
- Both configurations can exploit micro-machining technology and integrated circuit manufacturing methods enabling miniature, extremely robust sensors to be constructed which have no wearing parts in their mechanism and with MTBFs in excess of 100,000 hours, and, very importantly, they can be of relatively low cost

Optical Gyroscopes

- Optical gyroscopes such as the ring laser gyro and the fibre optic gyro measure angular rate of rotation by sensing the resulting difference in the transit times for laser light waves travelling around a closed path in opposite directions. This time difference is proportional to the input rotation rate and the effect is known as the ‘Sagnac effect’ after the French physicist G. Sagnac who, in fact, demonstrated that rotation rate could be sensed optically with the Sagnac interferometer as long ago as 1913.

The Sagnac effect time difference, ΔT, between the clockwise (cw) and anti-clockwise (acw) paths is given by

$$\Delta T = \frac{4A}{c^2}\dot{\theta} \tag{5.1}$$

where A is the area enclosed by the closed path, c the velocity of light and $\dot{\theta}$ the angular rate of rotation about an axis normal to the plane of the closed path.

The difference in optical path length ΔL, where L is the perimeter of the path, is given by

$$\begin{aligned} \Delta L &= c\Delta T \\ \Delta L &= \frac{4A}{c}\dot{\theta} \end{aligned} \tag{5.2}$$

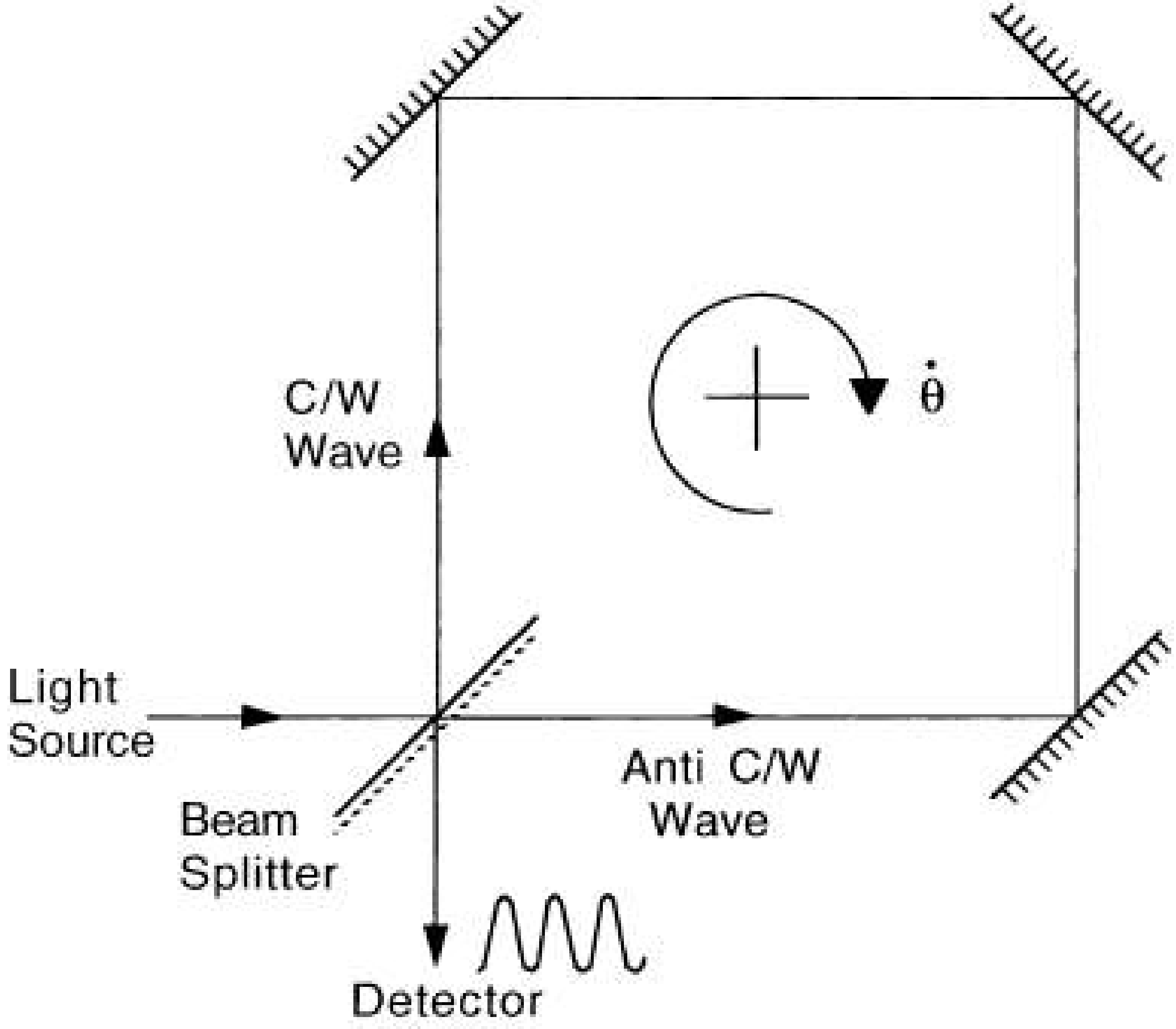
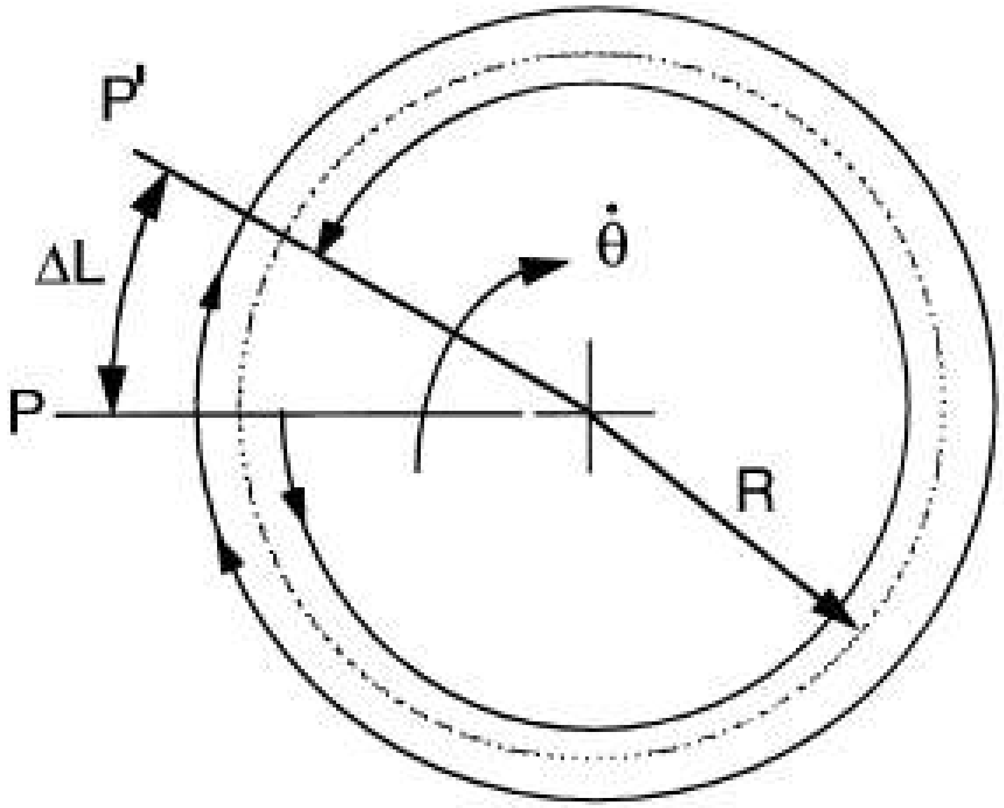


Fig. 5.4 Sagnac interferometer. The clockwise and anti-clockwise waves interfere to produce a finge pattern which shifts when the interferometer is subjected to input rate, $\dot{\theta}$.

Referring to Figure , consider a photon of light starting from P and travelling round the perimeter in a cw direction and a photon starting from P travelling in the acw direction.

In the absence of an input rate, the transit times of the two photons will be identical and equal to $2\pi R/c = T$. Now consider the path rotating at a rate $\dot{\theta}$.



Sagnac effect.

In time T, P has moved to P^1 and the path length for the cw photon is equal to $(2\pi R + R\dot{\theta}T)$ and the path length for the acw photon is equal to $(2\pi R - R\dot{\theta}T)$.

The difference in transit time,

$$\Delta T = \frac{(2\pi R + R\dot{\theta}T)}{c} - \frac{(2\pi R - R\dot{\theta}T)}{c}$$

substituting $T = 2\pi R/c$ yields

$$\Delta T = \frac{4\pi R^2}{c^2}\dot{\theta} \tag{5.3}$$

and

$$\Delta L = \frac{4\pi R^2}{c}\dot{\theta} \tag{5.4}$$

$A = \pi R^2$ hence formulae (5.3) and (5.4) are identical with (5.1) and (5.2).

It can be shown that these formulae are unchanged when the optical path comprises a medium of refractive index n such as when an optical fibre provides the closed path.

The Sagnac effect is very small for low rates of rotation. Michelson and Gale in 1925 used a Sagnac interferometer with a rectangular cavity 600 m × 330 m to measure the Earth's rotation rate and measured a path difference of only 1/4 fringe.

Ring Laser Gyroscope (RLG)

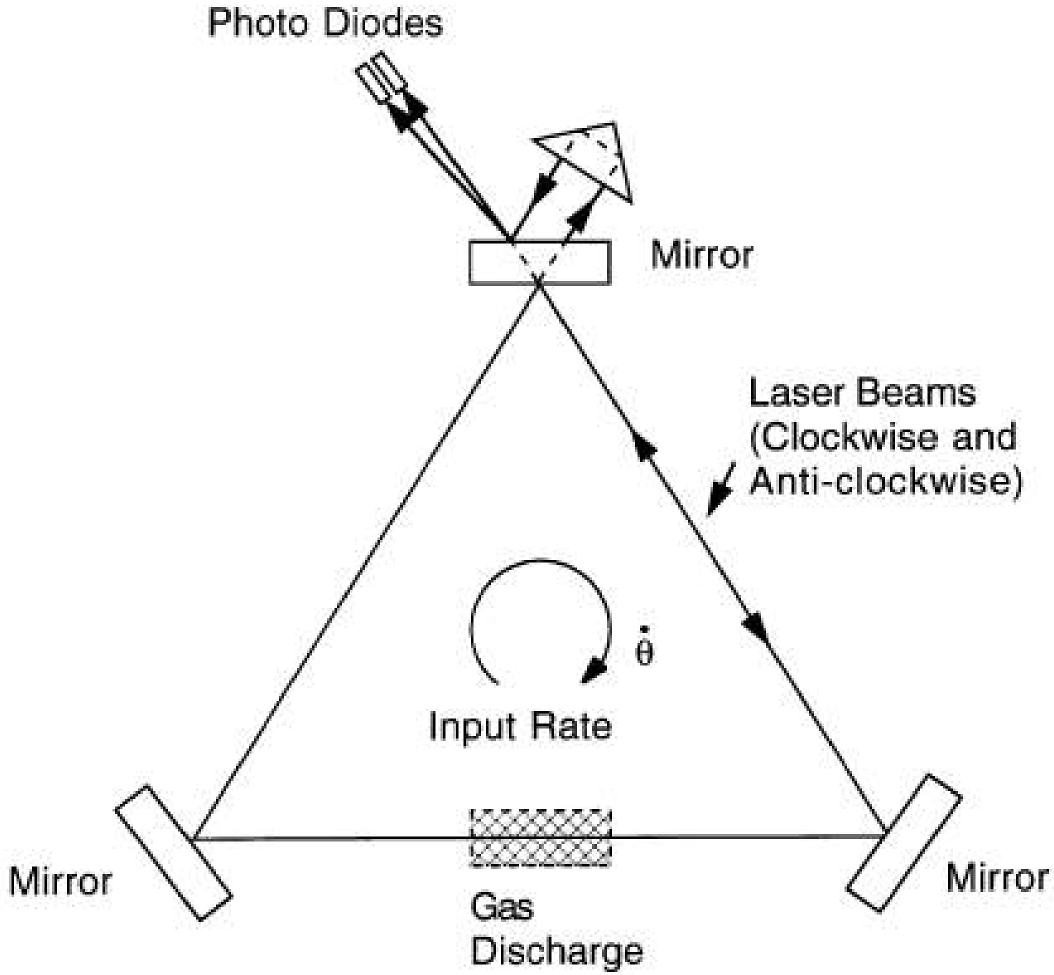


Fig. 5.6 Laser gyro schematic.

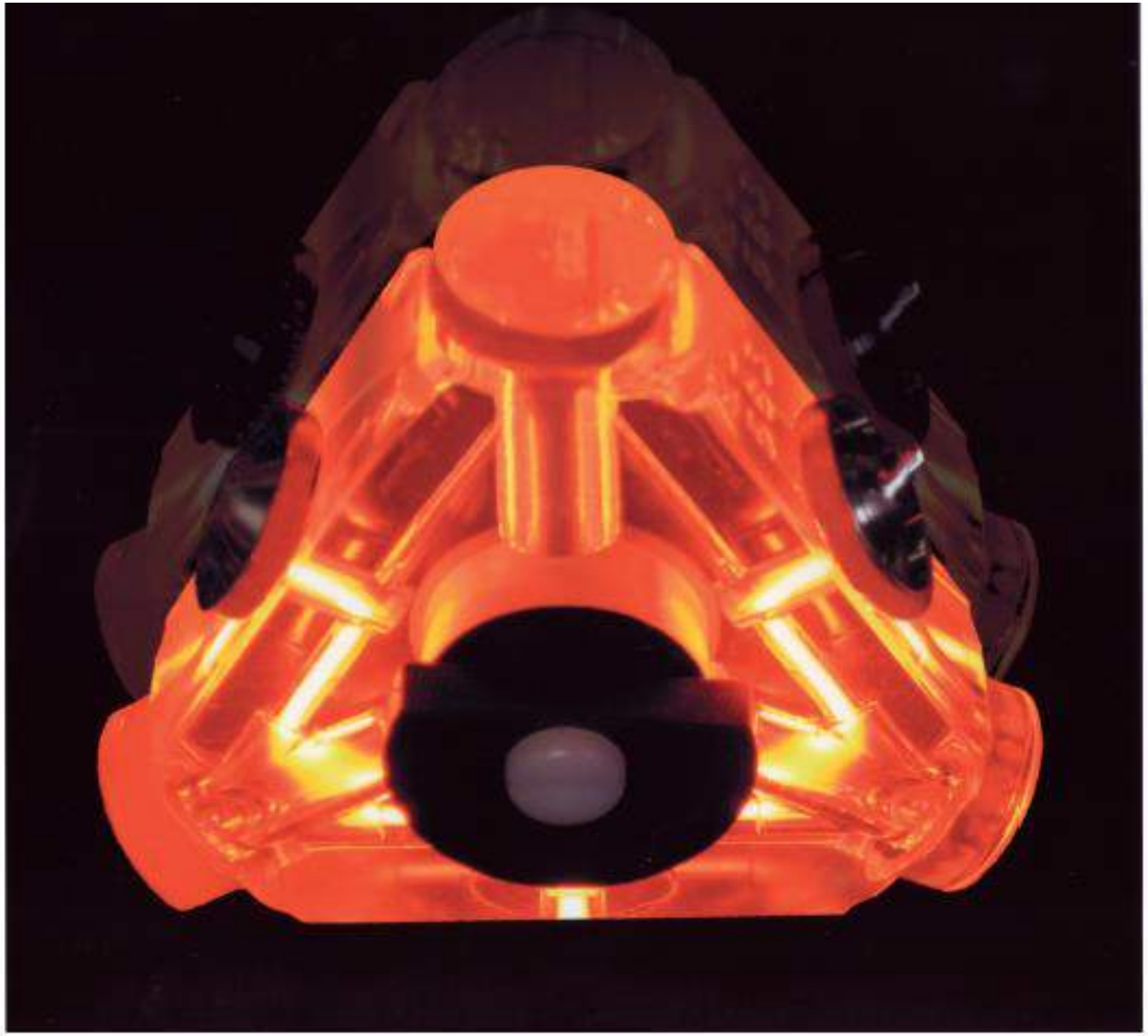


Fig. 5.10 Ring laser gyro (courtesy of Honeywell).

The basic elements of the RLG are shown schematically in Figure The two counter rotating laser beams are generated from the lasing action of a helium–neon gas discharge within the optical cavity, the triangular closed path being formed by reflecting mirrors at each corner of the triangle. This closed path forms the resonant cavity and the longitudinal mode frequency, f, is determined by the cavity optical path length, L, being given by $f = nc/L$ where n is an integer and c the velocity of light. At zero input rotation rate, the cw and acw path lengths are equal and there is zero difference between the frequencies of the cw and acw waves. When the RLG is rotated about an axis normal to the plane of the closed path there is a difference in the path length of the cw and acw travelling waves, as shown earlier, which causes a frequency difference between the two waves. This frequency difference is measured by allowing a small percentage of the two laser beams to be transmitted through one of the mirrors. A corner prism is generally used to reflect one of the beams so that it can be combined with the other to generate a fringe pattern at the read-out detector photo-diodes. An input rotation rate causes the fringe pattern to move relative to the read-out photo-diodes at a rate and in a direction proportional to the frequency difference (positive or negative). A sinusoidal output signal is generated by each fringe as it passes by the photo-diodes. These are spaced so that there is a 90° phase difference between their outputs so that the direction of rotation can be determined from which photo-diode output is leading. The photo-diode outputs are then converted into positive (or negative) pulses by suitable pulse triggering and direction logic circuits.

The frequency difference, Δf, resulting from a difference in optical path length, ΔL, is given by

$$\frac{\Delta L}{L} = \frac{\Delta f}{f}$$

From equation (5.2)

$$\Delta L = \frac{4A}{c}\dot{\theta}$$

whence

$$\Delta f = \frac{4Af}{cL}\dot{\theta} \tag{5.5}$$

The wavelength of the laser transition $\lambda = c/f$, hence

$$\Delta f = \frac{4A}{\lambda L}\dot{\theta} \tag{5.6}$$

i.e.

$$\Delta f = K_0\dot{\theta} \tag{5.7}$$

where K_0 is the gyro scale factor $= 4A/\lambda L$.

The gyro thus behaves as an *integrating rate gyro*

$$\int_0^T \Delta f\, dt = K_0 \int_0^\theta d\theta \tag{5.8}$$

The angle turned through about the gyro input axis in the time period, T, is equal to the net number of positive (or negative) pulses counted in that period. The RLG

- *High accuracy* – The RLG meets the dynamic range for a pure IN system of being able to measure angular rates from 0.01°/hour to 400°/s to the required accuracy – a dynamic range of 10^8:1.
- *Insensitivity to acceleration* – The RLG has no acceleration sensitive bias errors, as it is based on optical effects rather than inertial effects.
- *Very high rate range* – This is limited only by the noise/bandwidth characteristics of the read out electronics: ±1,000°/s is no problem.
- *Very high scale factor accuracy* – Errors are in the 5 to 10 ppm bracket.
- *Negligible warm up time* – Full gyro operation from the instant of turn-on.
- *Excellent turn-on to turn-on performance* – Performance capabilities can be maintained over several years without calibration.

- *Random noise uncertainty* – This is measured in 'degrees per $\sqrt{\text{hour}}$', and is one of the RLG's most important error characteristics. The error is significantly higher than experienced with angular momentum gyros. It affects the system heading determination in the gyro compassing phase during the initial alignment process. This is because it extends the time required to filter the Earth's rate signal from the gyro noise in order to determine the initial heading.
- *Very high reliability* – MTBFs in excess of 60,000 hours are being demonstrated in large scale service. 100,000 hours MTBF is being approached.
- *Life* – The laser beams ultimately destroy the mirrors but this is over a very long period. A life equivalent to over 100 years usage is being claimed.
- *Volume* – This is determined by the path length, typically 20 cm for inertial performance.

The Interferometric Fibre Optic Gyro

The implementation of the interferometric type of FOG is best explained in a series of stages starting with the simple basic system shown in Figure

Light from the laser diode source is passed through a first beam splitter and a single optical mode is selected. The light passes through a second beam splitter and propagates in both directions around the fibre coil. In the absence of rotation, as already explained, the transit times are identical so that when the light arrives back at the second beam splitter, perfect constructive interference occurs with accompanying fringe pattern. The gyro output signal is obtained by directing the returning light via the first beam splitter to a photo-detector. As explained, an input rotation rate about an axis normal to the plane of the coil results in a difference in the transit times between the clockwise and anti-clockwise beams as given by equation (5.1), viz.

$$\Delta T = \frac{4A}{c^2}\dot{\theta}$$

If the fibre coil has N turns, $A = \pi R^2$ N where R is the mean radius of the coil, and N can be expressed in terms of the length of the coil, L and the mean radius R, i.e.

$$N = \frac{L}{2\pi R}$$

hence

$$\Delta T = \frac{LD}{c^2}\dot{\theta} \qquad (5.9)$$

where D is the coil diameter $= 2R$.

This transit time difference results in the Sagnac phase shift, given by

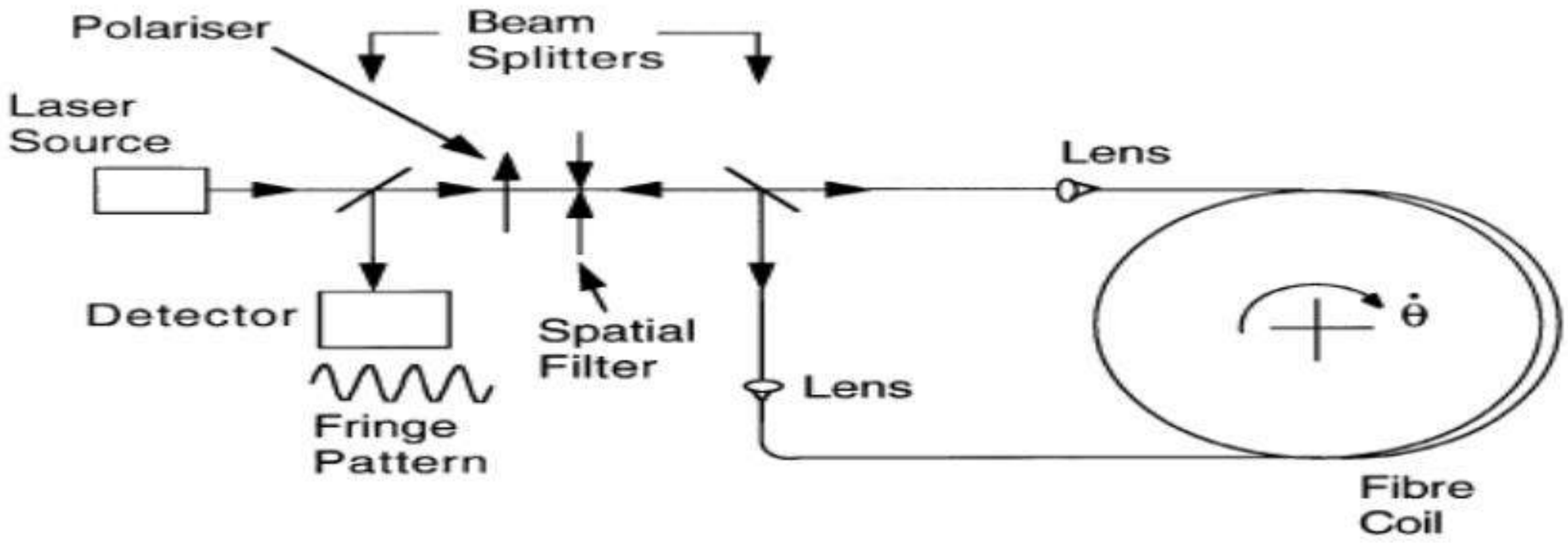
$$\Phi_s = \omega \Delta T$$

where ω is the angular frequency of the laser source $= 2\pi c/\lambda$, and λ is the laser wavelength.

Hence

$$\Phi_s = \frac{2\pi LD}{\lambda c}\dot{\theta} \qquad (5.10)$$

This phase shift between the clockwise and anti-clockwise travelling light waves results in a reduction in the intensity of the light at the detector. This change in


Polariser
Beam
Splitters
Laser
Source
Lens
Detector
Spatial
Filter
Lens
Fringe
Pattern
Fibre
Coil
$\dot{\theta}$

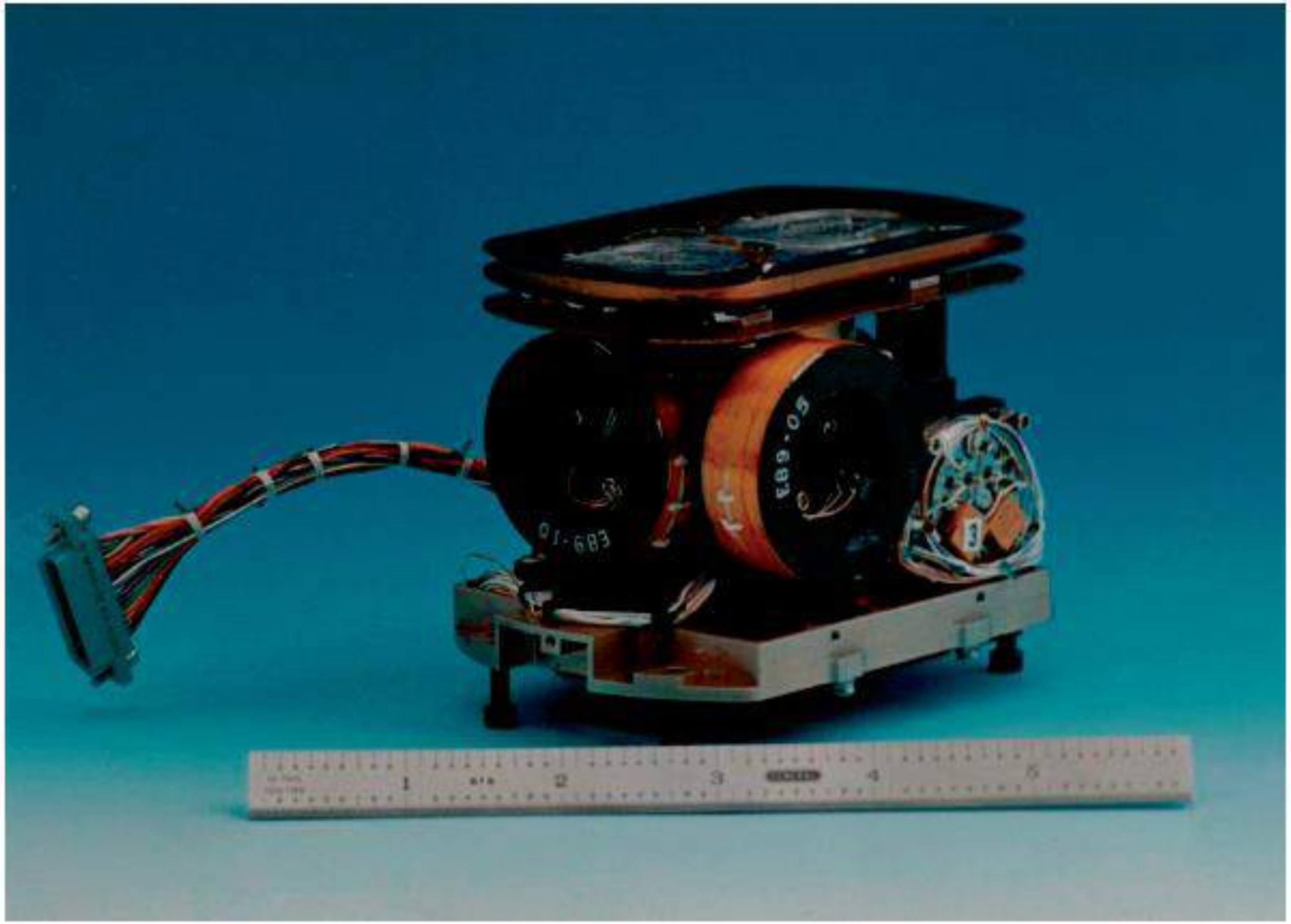
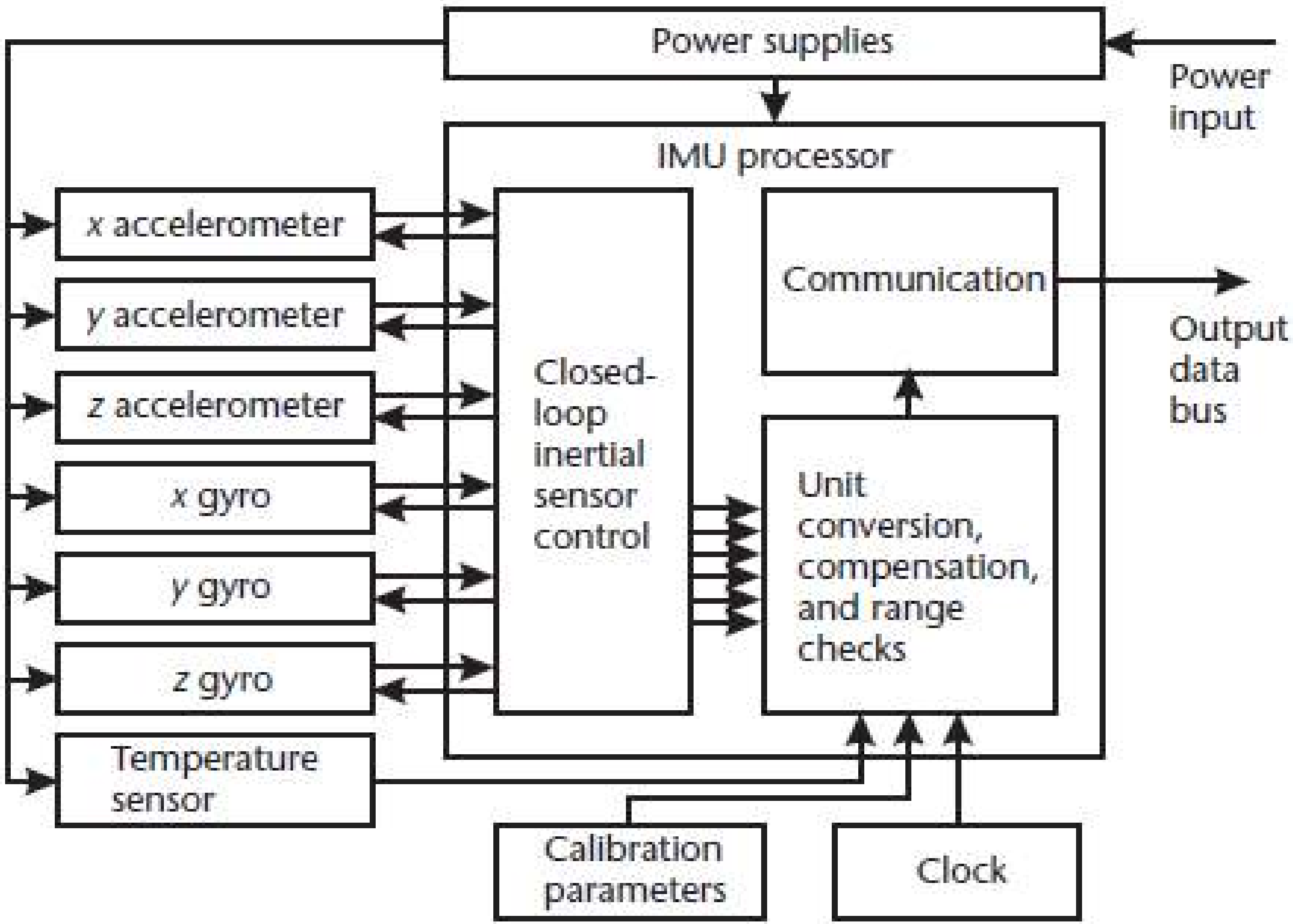


Fig. 5.14 Fibre optic gyro inertial measuring unit (courtesy of Smiths Industries).

Inertial Navigation

Inertial Measurement Unit

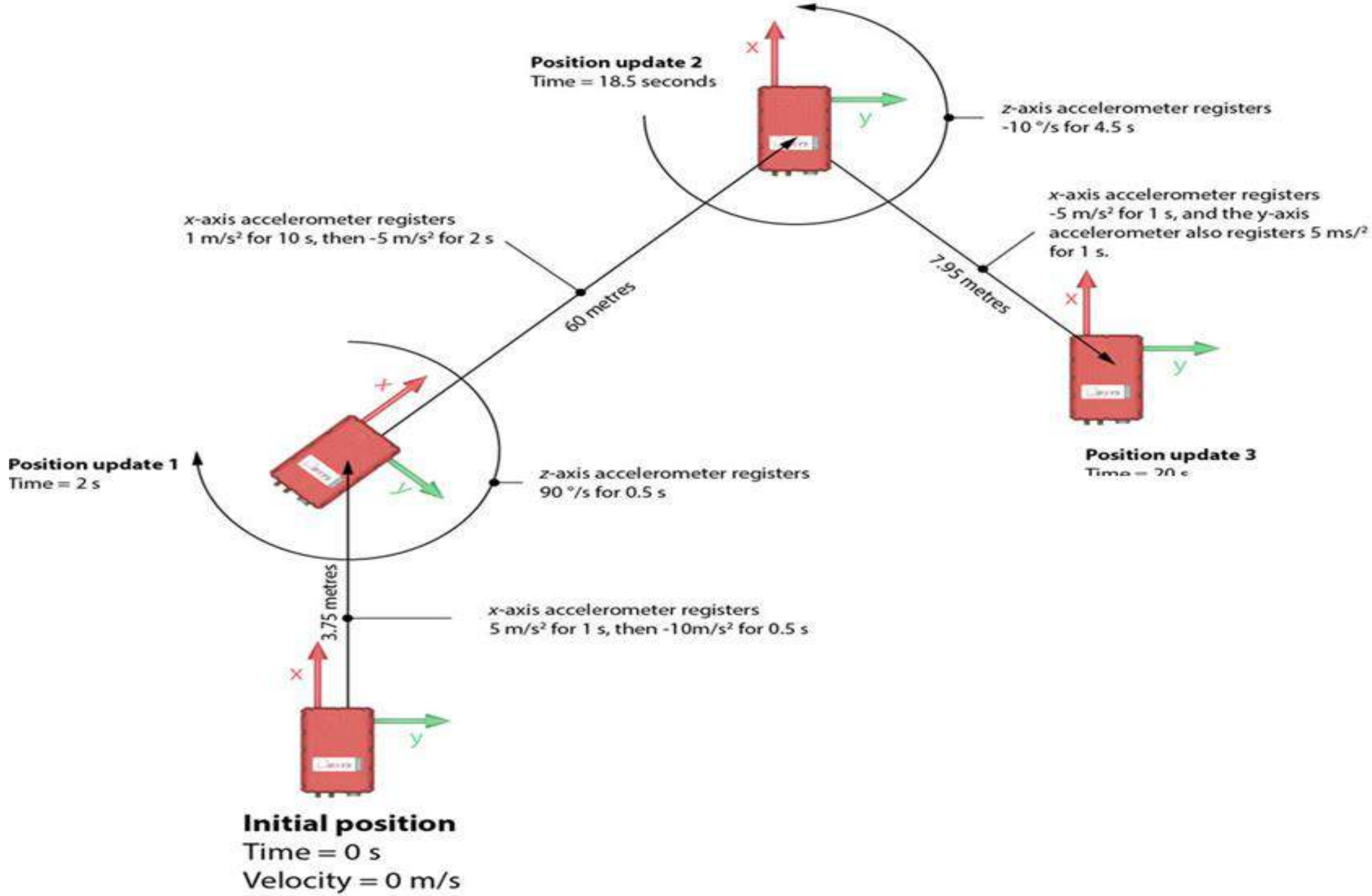


Schematic of an inertial measurement unit.

- The IMU processor performs unit conversion on the inertial sensor outputs, provides compensation for the known errors of the inertial sensors, and performs range checks to detect sensor failure.
- Unit conversion transforms the inertial sensor outputs from potential difference, current, or pulses into units of specific force and angular rate.

- Using the measurements taken from the three accelerometers and three gyros, the inertial navigation system keeps track of where it is in three-dimensional space.
- It does this using a process called **dead reckoning.**
- The actual process of dead reckoning is : you take information from some source (gyros and accelerometers in this case) and turn them into a movement that can be added to your last known position to see where you are now.

Dead reckoning

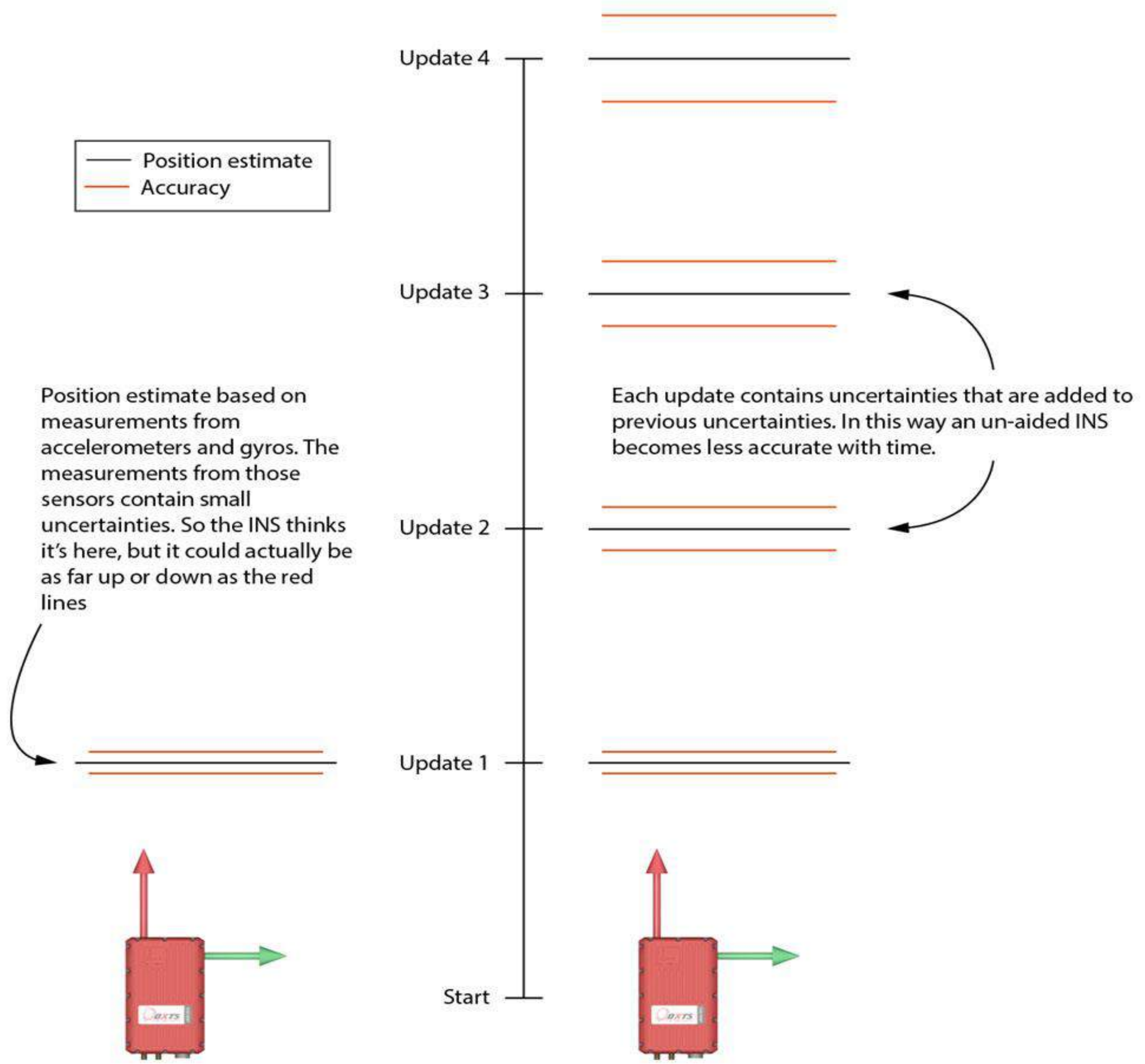


- You can see that initially the INS is stationary and aligned with its x-axis pointing straight up. The fig then shows three other positions and the information recorded by the sensors between them. Of course in reality the INS would update its position tens or hundreds of times per second, but in this example position updates are only shown when key changes take place for ease of understanding.
- So at time zero, the INS is stationary (and does not know where it is). It then sees an acceleration of 5 m/s^2 on the x-axis accelerometer for 1 second, which gives it a velocity of 5 m/s (or 18 km/h). It then immediately comes to a complete stop—detecting an acceleration of -10 m/s^2 for 0.5 seconds. As no other measurements were registered on the other sensors, the navigator can easily work out that it has moved 3.75 metres in the direction of the x-axis. Again, at this point, the INS doesn’t know where it is as we haven’t given it any position information to begin with.

- As soon as the INS stops at position update 1, the z-axis gyro detects a value of 90 °/s for 0.5 seconds; so it knows that it has just turned 45° in a clockwise direction. Again, as soon as that movement is complete the INS again sees acceleration on the x-axis accelerometer. This time it's 1 m/s^2 for 10 seconds followed by -5 m/s^2 for 2 seconds. Using the same techniques as before, the INS can work out that it has now moved 60 metres further on at a 45° angle from where it was at position update 1.

- **Drift** is one of the problems faced by **unaided INS.**
- Un-aided means systems that only use accelerometer and gyro measurements to calculate their position.
- Drift is the term used to describe the accumulation of small errors in the accelerometer and gyro measurements, which gradually cause the INS position estimate to become more and more inaccurate.
- Each time an accelerometer or gyro is read, there is a miniscule error in the reading. If we were just taking a single reading to work out how fast we were accelerating or turning, this wouldn’t be a problem. But because the navigation computer is adding up each measurement to work out how it has moved on from the previous position estimate, the miniscule error grows with time.

How drift can accumulate in an un-aided INS



- We can **solve the problem of drift** by combining an INS with GPS to create a GPS-aided INS (also written as GPS+INS), and also solve the problems that affect GPS too.

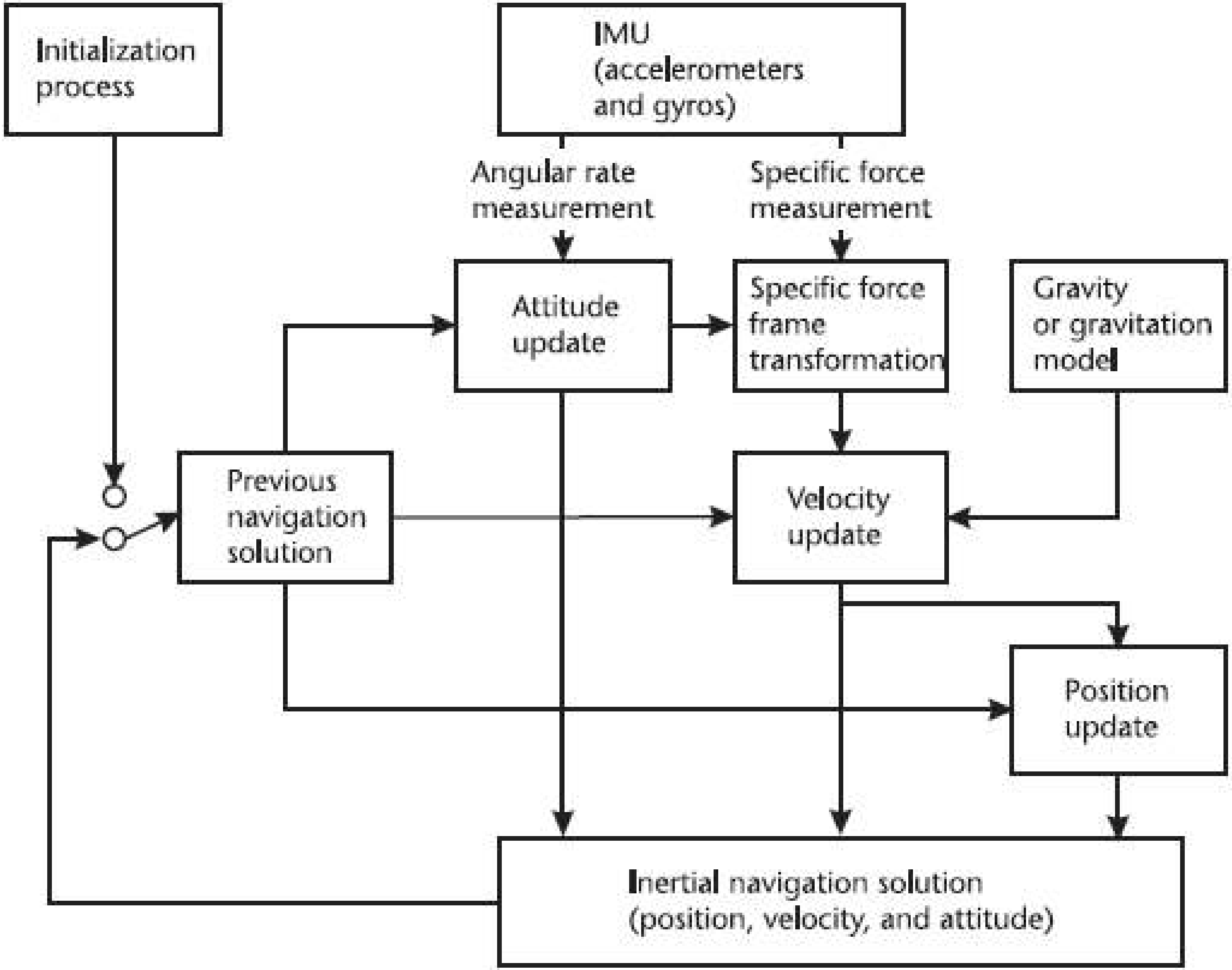
GPS and INS are wholly complementary and their information can be combined to the mutual benefit of both systems. For example:

- Calibration and correction of INS errors – the GPS enables very accurate calibration and correction of the INS errors in flight by means of a Kalman filter.
- The INS can smooth out the step change in the GPS position output which can occur when switching to another satellite because of the change in inherent errors.
- Jamming resistance – like any radio system, GPS can be jammed, albeit over a local area, although it can be given a high degree of resistance to jamming. The INS, having had its errors previously corrected by the Kalman filter, is able to provide accurate navigation information when the aircraft is flying over areas subjected to severe jamming.

- Antenna obscuration – GPS is a line of sight system and it is possible for the GPS antenna to be obstructed by the terrain or aircraft structure during manoeuvres.
- Antenna location corrections – the GPS derived position is valid at the antenna and needs to be corrected for reference to the INS location. The INS provides attitude information which together with the lever arm constants enables this correction to be made.

- Any errors, however, in deriving the acceleration components from the accelerometer outputs will be integrated with time, producing velocity errors which in turn are integrated with time generating position errors.
- **An accelerometer bias error of 10^–3 g will produce a distance error of 0.45 km after 5 minutes and 1.8 km after 10 minutes, for example.**

- Figure shows a schematic of an inertial navigation processor. This integrates he IMU outputs to produce a position, velocity, and attitude solution.
- **The navigation equations comprise four steps:**
- Attitude update,
- Transformation of the specific force resolving axes,
- Velocity update, and
- Position update.
- In addition, a gravity or gravitation model is needed to transform specific force into acceleration. In basic terms, the attitude is updated by integrating the angular rate measurements; the velocity is updated by integrating acceleration; and the position is updated by integrating velocity.



5.1 Schematic of an inertial navigation processor.

Satellite Navigation Systems.

Global Navigation Satellite System(GNSS)

- GNSS is the generic name used to describe any global system of satellites that transmit signals for navigation purposes on Earth.

As of 2017, the main GNSSs are:

- **GPS.** America's Global Position System.
- **GLONASS.** Russia's Global Orbiting Navigation Satellite System.
- **BeiDou.** A system being developed by China—not currently fully operational.
- **Galileo.** A system being developed by the EU

GPS

GPS SYSTEM

- The space segment
- The control segment
- The user segment

The space segment

- The space segment is concerned with the satellites in orbit. In 2015 the GPS constellation consists **of 32 non-geostationary satellites** in medium Earth orbit.
- The satellites are in **semi synchronous orbits**
- Each satellite orbits once every 11 hours, 58 minutes and 2 seconds at an average altitude of 20,200 km (that’s an orbital radius of 26,571 km).
- The GPS **satellite constellation is arranged into six equally-spaced orbital planes**, with no fewer than four satellites in each plane. This arrangement ensures a minimum of four satellites can be seen 15° above the horizon at almost any time, from any point on the planet—although in reality there are generally more.

- Each satellite contains four highly accurate clocks with a fundamental frequency of 10.23 MHz, and they constantly transmit two carrier waves in the L-Band that travel back to earth at the speed of light. These carrier waves are referred to as L1 and L2.

- **The L1 carrier has a frequency of 1575.42 MHz**
- **The L2 carrier has a frequency of 1227.60 MHz**

Control segment

- The control segment refers to a number of ground stations situated around the globe that are used **to track, control and send information to each of the GPS satellites.**
- The orbit information that is sent up to each satellite is vital too, because we need it in order to work out where the satellite was when the information was sent.
- All of this information is sent up to the satellites, then carried to GPS receiver within the L1 carrier wave navigation message.

Ground Control Segment

- **Continuously**
  - Track GPS satellites (svs)
  - Estimate clock and orbit
  - Keep GPS time
  - Upload data that describes clock & orbit for each sv
- **Infrequently**
  - Command small maneuvers to maintain orbit
  - Command small clock corrections
  - Command major relocations to compensate for any sv failures

GPS Operational Control Segment
(from Anthony Russo CGSIC, 2008)
Alaska
England
Vandenberg AFB
California
USNO
Cape
Canaveral
Hawaii
Kwajalein
Tahiti
Ascension
Diego Garcia
Argentina
South Africa
Australia
New
Zealand
Master Control Station
Backup MCS
Ground Antenna
OCS Monitor Station
NGA Monitor Station
Future Monitor Station

User segment

- This segment includes anyone or anything with a GPS receiver;

sat navs, mobile phones, UAVs, etc.

- Locating a body using GPS is based on a process called **trilateration.**
- **At no point does the GNSS receiver inside your sat nav or phone send any information up to the satellites. The receivers we use today are completely passive—they only receive information.**
- When Europe's Galileo system is operational, its receivers will be slightly different because there will be an emergency function, which will send information when activated, but this won't apply in normal operation.


Stay on 40 WEST
6.3
284
Wade Avenue
TRINITY CIRCLE
TRINITY WOODS
COOPER
11 min

GPS SIGNAL

- The unique signal transmitted from each satellite contains two codes and a message:
  - **C/A code (coarse/acquisition code)**
  - **P code (precision code) [called Y code in its encrypted form]**
  - **Navigation message**
- The P code is encrypted for military use. It’s encrypted to stop spoofing and to control who has access to the system. Incidentally, once the P code has been encrypted, it’s referred to as Y code.

GPS Signal Structure

GPS Carrier Waves Frequencies

L-band - The group of radio frequencies extending from 390 MHz to 1550 MHz.

L1 frequency - 1575.42 MHz

- Civilian frequency

L2 frequency - 1227.60 MHz

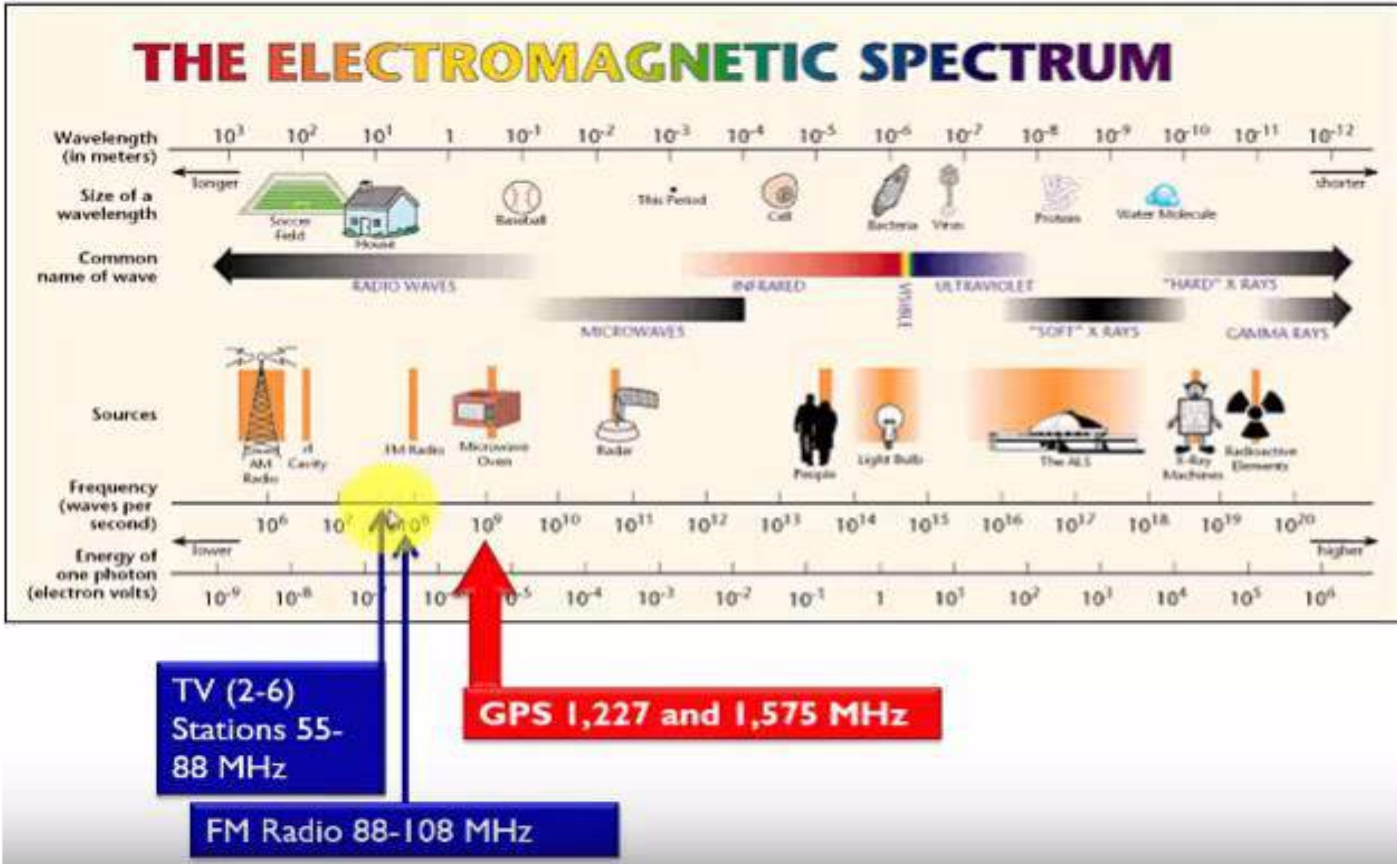
- Military and dual-frequency receivers – increases accuracy by measuring ionosphere delay
- New L2C are new civilian codes – Block IIR-M & II-F

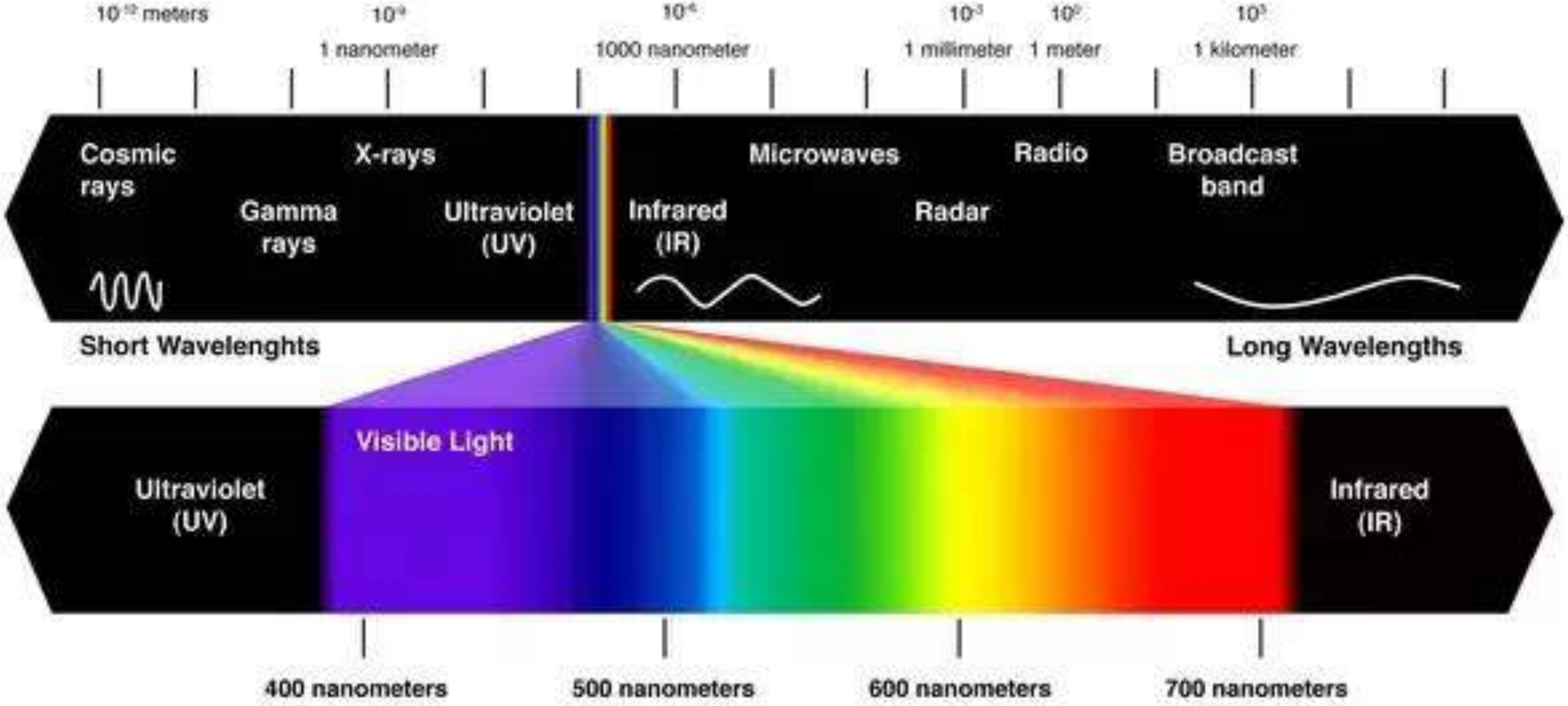
L5 frequency - 1176.45 MHz

- new safety of life signal – latest satellites Block II-F are first to have it

L3 frequency - 1381.05 MHz

- Nuclear detonation detection?


THE ELECTROMAGNETIC SPECTRUM
Wavelength (in meters)
longer
shorter
Size of a wavelength
Soccer Field
House
Baseball
This Period
Cell
Bacteria
Virus
Protein
Water Molecule
Common name of wave
RADIO WAVES
INFRARED
VISIBLE
ULTRAVIOLET
"HARD" X RAYS
MICROWAVES
"SOFT" X RAYS
GAMMA RAYS
Sources
AM Radio
Cavity
FM Radio
Microwave Oven
Radar
People
Light Bulb
The ALS
X-Ray Machines
Radioactive Elements
Frequency (waves per second)
Energy of one photon (electron volts)
lower
higher
TV (2-6) Stations 55-88 MHz
FM Radio 88-108 MHz
GPS 1,227 and 1,575 MHz


10^{-12} meters
10^{-9}
1 nanometer
10^{-6}
1000 nanometer
10^{-3}
1 millimeter
10^{0}
1 meter
10^{3}
1 kilometer
Cosmic rays
Gamma rays
X-rays
Ultraviolet (UV)
Infrared (IR)
Microwaves
Radar
Radio
Broadcast band
Short Wavelenghts
Long Wavelengths
Visible Light
Ultraviolet (UV)
Infrared (IR)
400 nanometers
500 nanometers
600 nanometers
700 nanometers

GPS Signal Structure

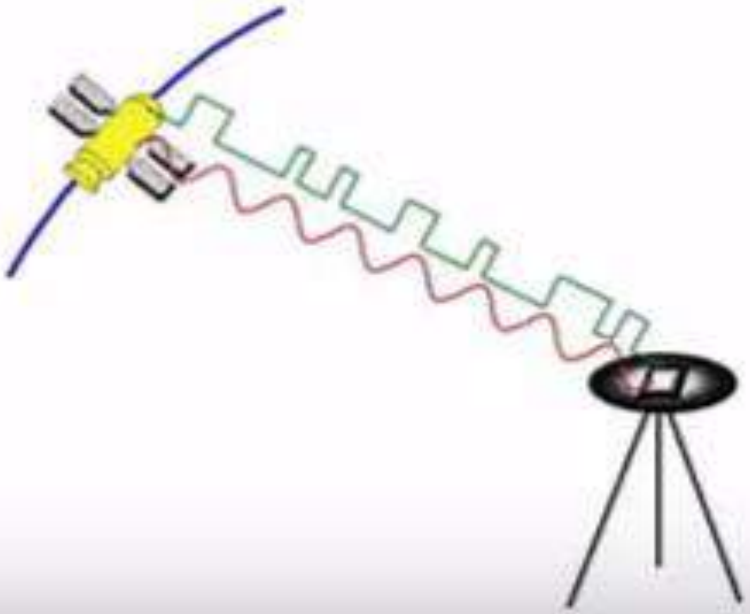
A sine wave (carrier wave) on its own does not transmit information

Carrier wave is **modulated** to carry a code

Modulate the wave in some way to encode information on it.

- Examples: Pulse, Frequency, and Amplitude Modulation

Once you modulate a sine wave with information, you can transmit the information!

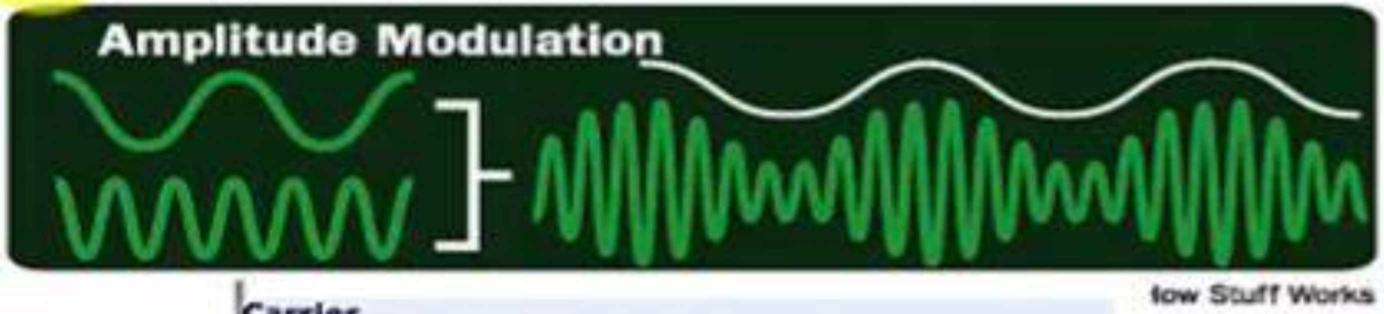


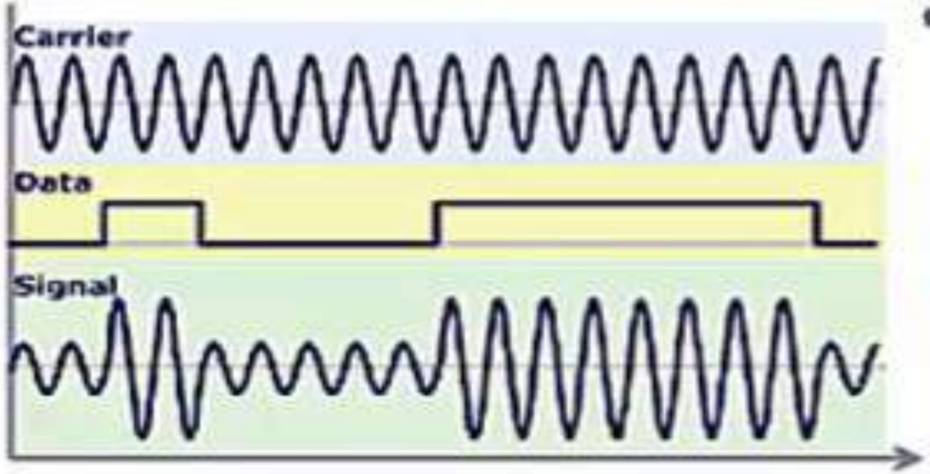
GPS signal modulation is called "bi-phase" modulation

Transmitting information

Amplitude Modulation –

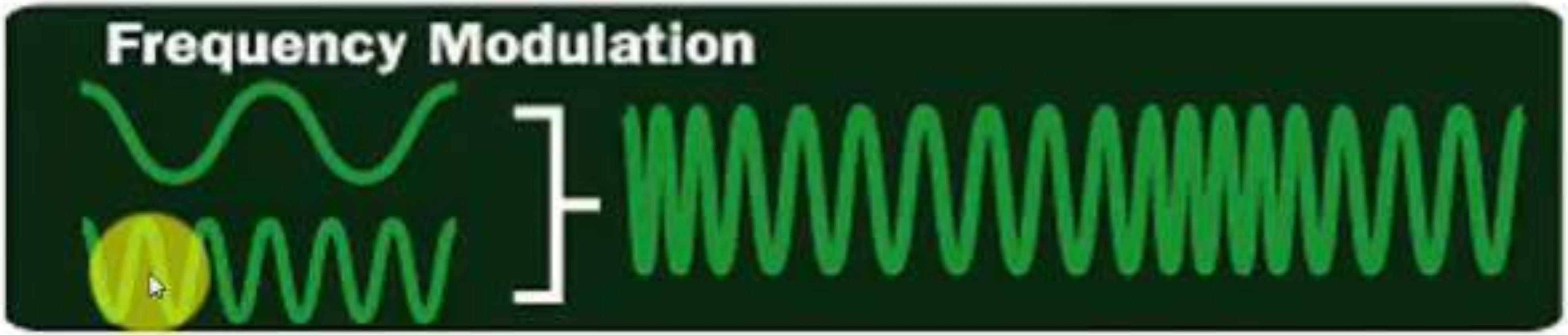
- AM radio stations and the picture part of a TV signal
- the amplitude of the sine wave (its peak-to-peak voltage) changes

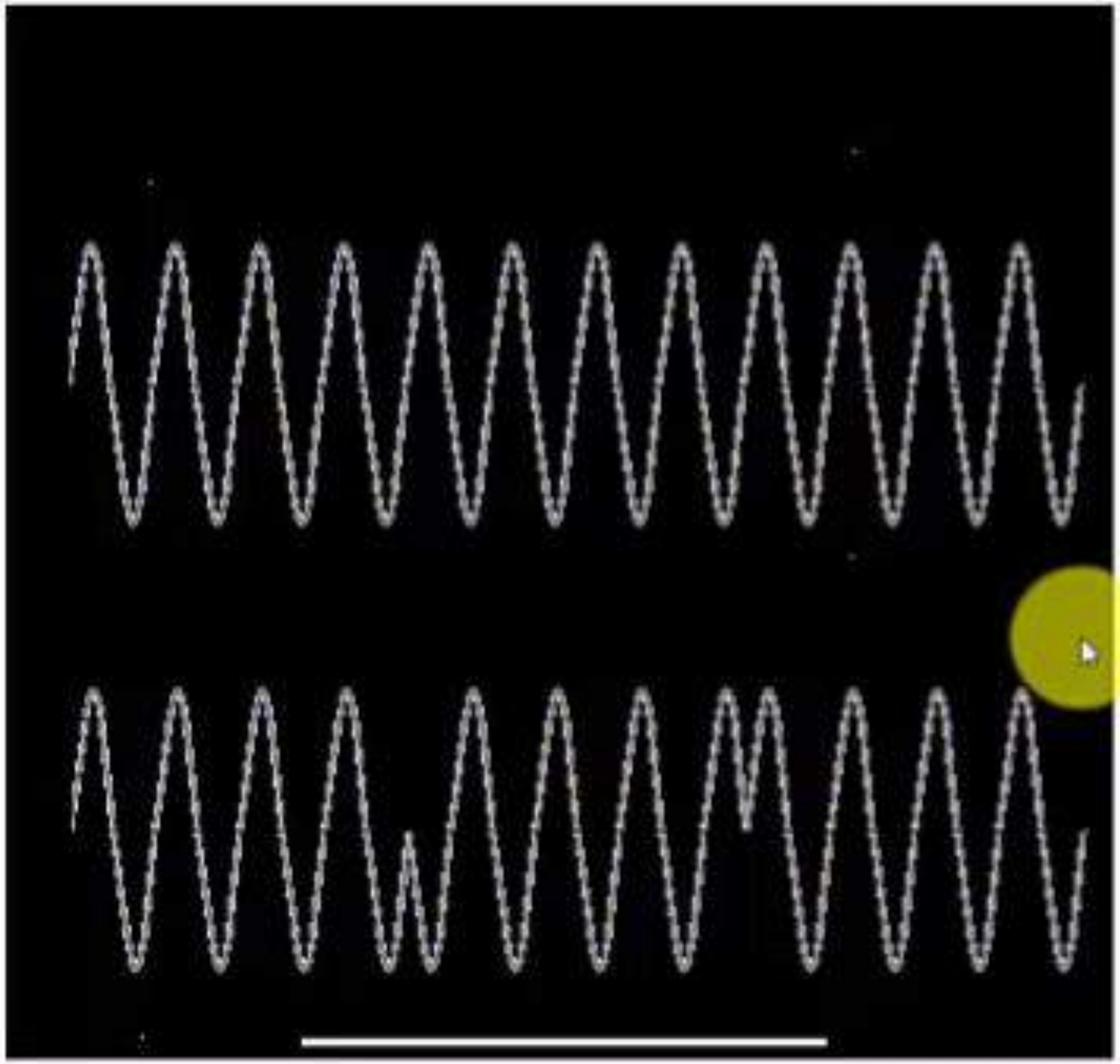




Frequency Modulation

- FM radio stations and hundreds of other wireless technologies (including the sound portion of a TV signal, cordless phones, cell phones, etc.)
- Advantage - largely immune to static
- the transmitter's sine wave frequency changes very slightly based on the information signal.





GPS uses **Biphase modulation**

shifts the carrier phase wave to indicate change from 1 to 0 or vice versa

A Short Repeating PRN Code Sample

GPS Signal Structure

GPS Codes carried by the carrier wave:

C/A – coarse acquisition code
- Standard Positioning Service (SPS)
- civilian

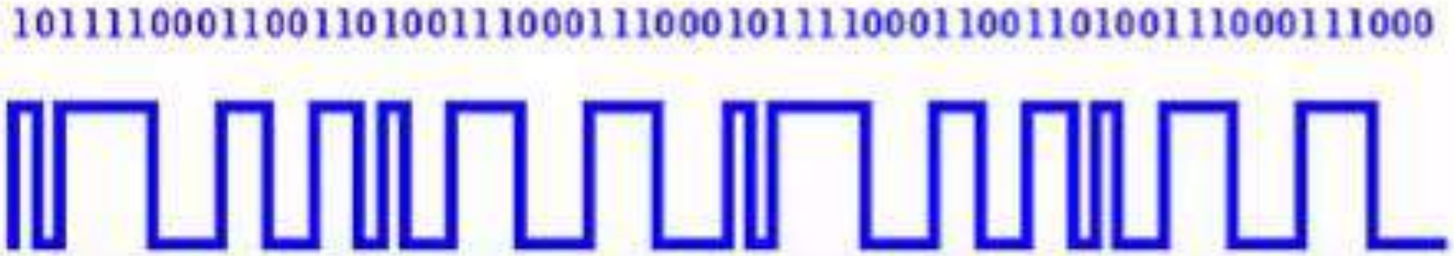
P(Y) – precise military code
- Precise Positioning Service (PPS)
- Military – more impervious to jamming and signal spoofing

Other information

GPS Signals

Pseudo-random noise (PRN) codes

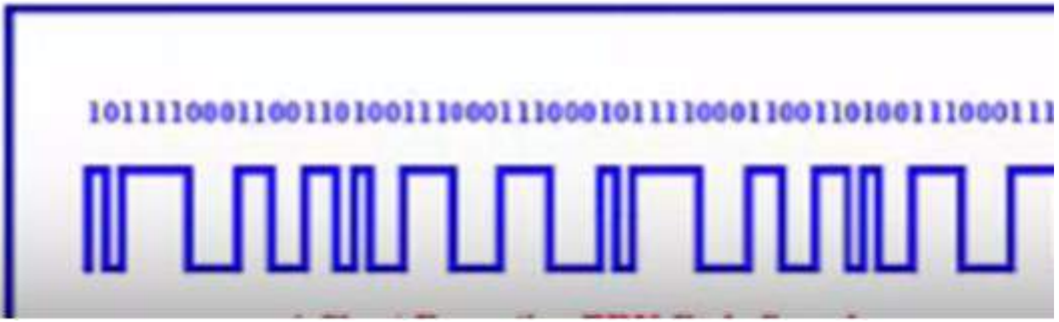
- A signal with random noise-like properties
- A very complicated but repeated pattern of 1's and 0's.
- There are two types of PRN codes used by GPS: C/A code and P(Y) code
- a different PRN code for each SV
  - GPS satellites are identified by their PRN number
  - The number refers to the week of the P(Y) code that the satellite broadcasts



A Short Repeating PRN Code Sample

C/A Code (Coarse Acquisition)

- Stream of 1,023 binary digits (1s and 0s)
- Repeats every millisecond
  - each second 1,023,000 chips (or digits) are generated.
- Taking into account the speed of light the length of one chip can be calculated to be 300 m.
- 37 C/A codes are defined
  - Each sv has a unique C/A code
  - GPS satellites are identified by their PRN number, which is a unique identifier for the PRN code that satellite uses
- Designed for **civilian** use



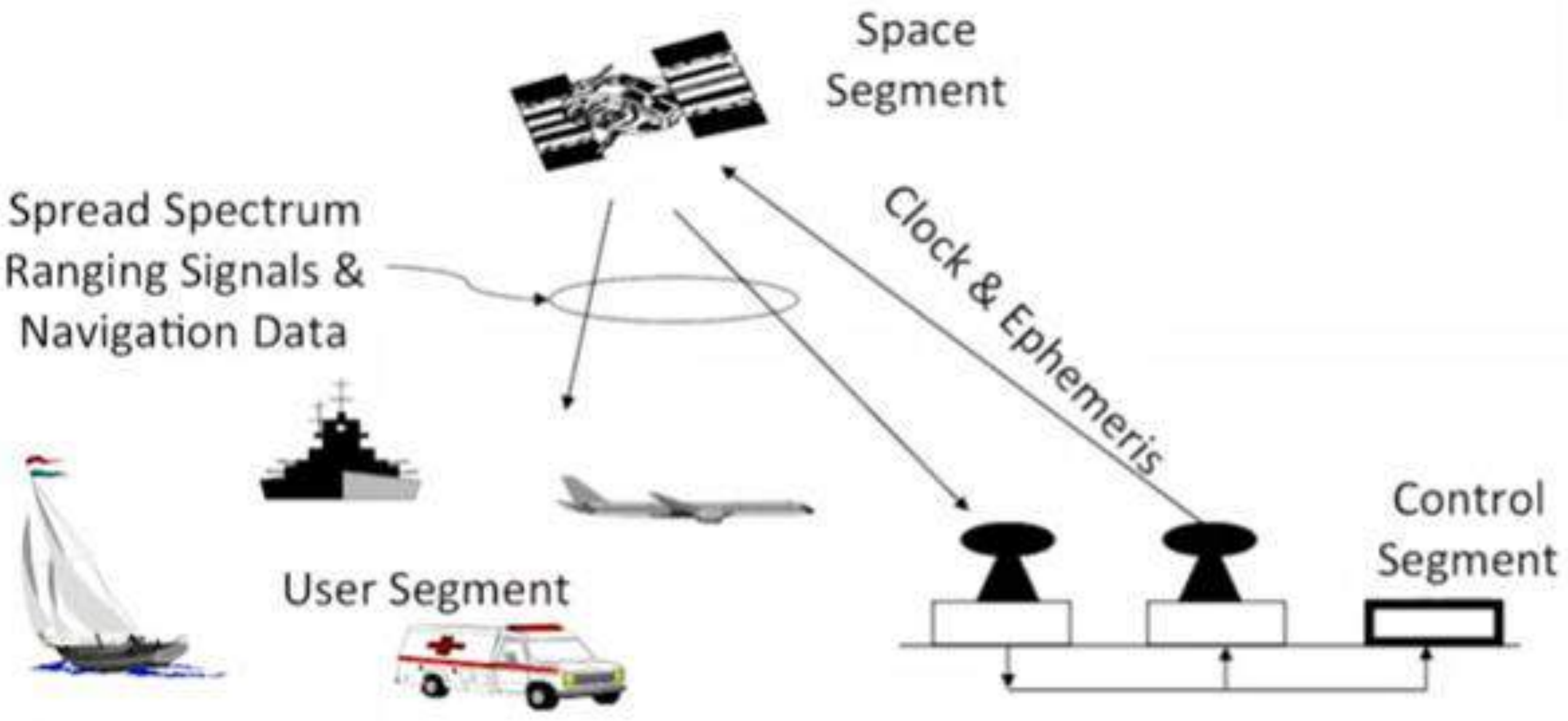
P(y) code (Precise)

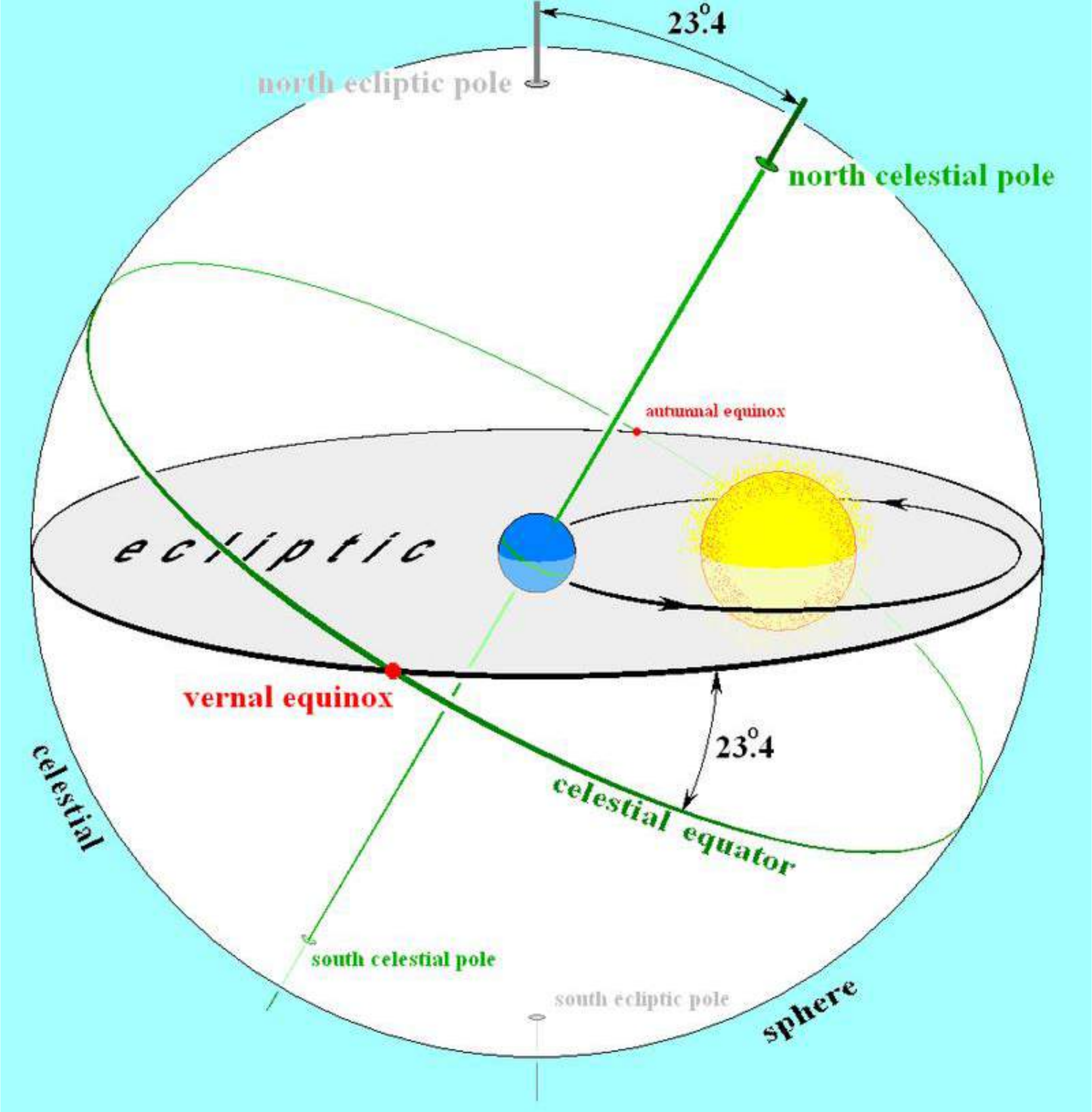
- Very long sequence of binary digits
- Also a PRN code
- Repeats every 266 days
- 2.35×10^{14} chips!
- 38 segments, each 1-week long
- Each satellite transmits a unique one-week segment of the P-code, initialized Sat at midnight (6.1871×10^{12} bits ≈ 720.213 Gb)
- C/A and P-code combinations are inseparable
- PRN number refers to which week of the P-code a satellite transmits
- In 1994 became encrypted by the Y code for antispoofing (encryption by the military)

Navigation message

- Takes 12.5 minutes to transmit
- Includes
  - Satellite time and synchronization signals
  - Precise orbital data (ephemeris)
  - Time correction information to determine the exact satellite time
  - Approximate orbital data for all satellites (almanac)
  - Correction signals to calculate signal transit time
  - Data on the ionosphere
  - Information on the operating status (health) of the satellite
- using the navigation message, the receiver is able to determine the transmission time of each satellite signal and the exact position of the satellite at the time of transmission

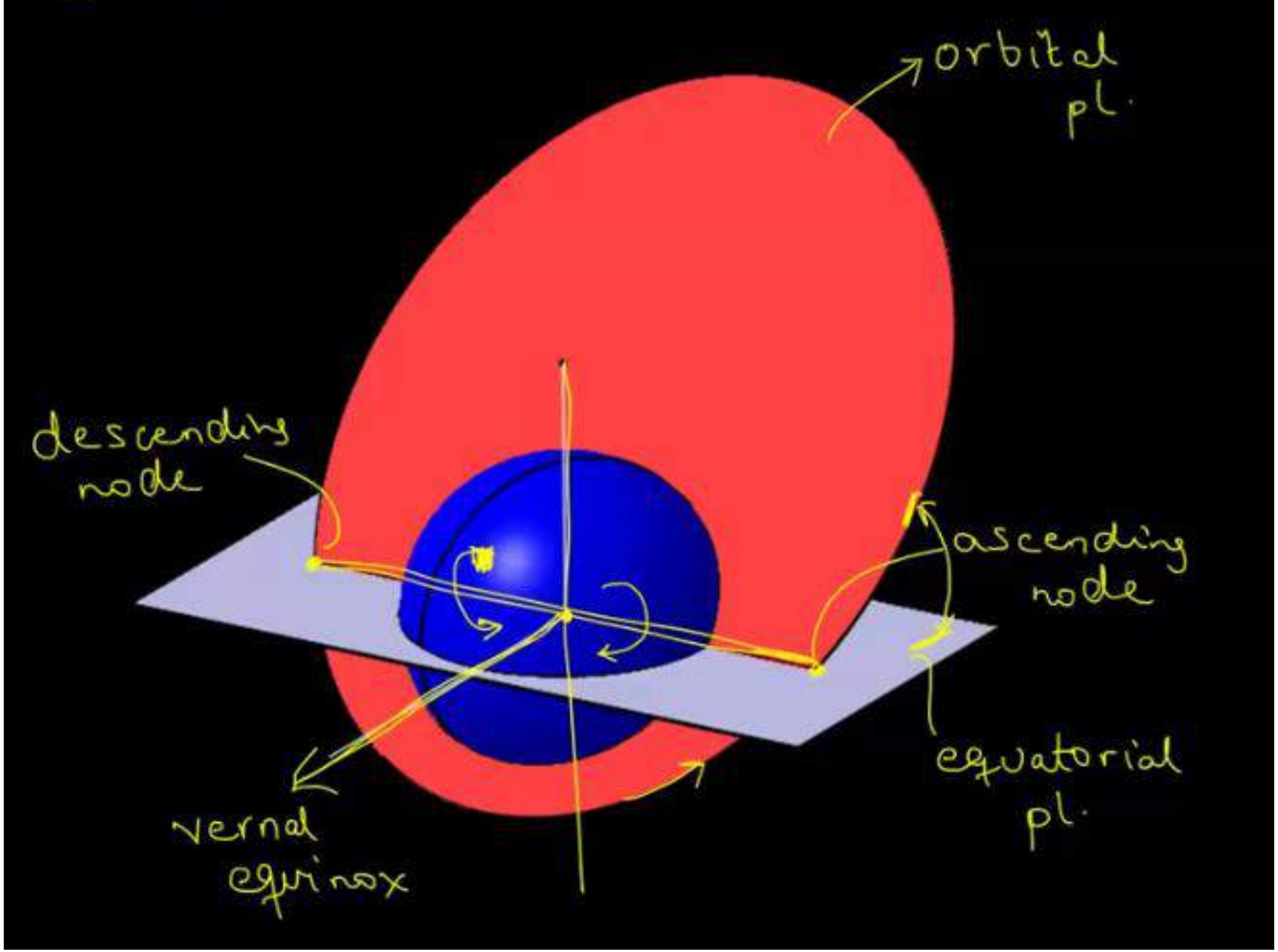
Development & Delivery of the Navigation Message

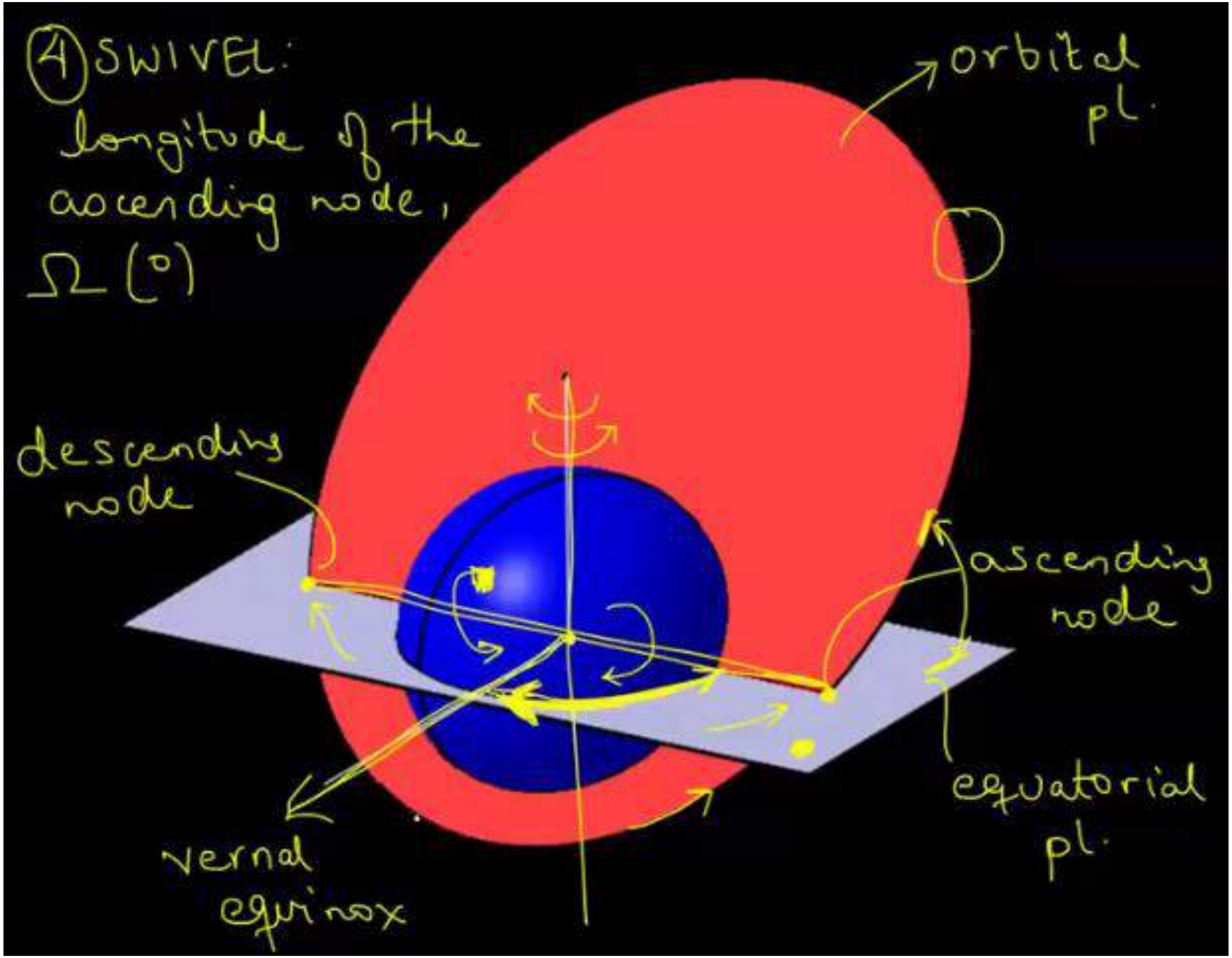




23°.4
north ecliptic pole
north celestial pole
autumnal equinox
ecliptic
vernal equinox
23°.4
celestial equator
celestial
sphere
south celestial pole
south ecliptic pole

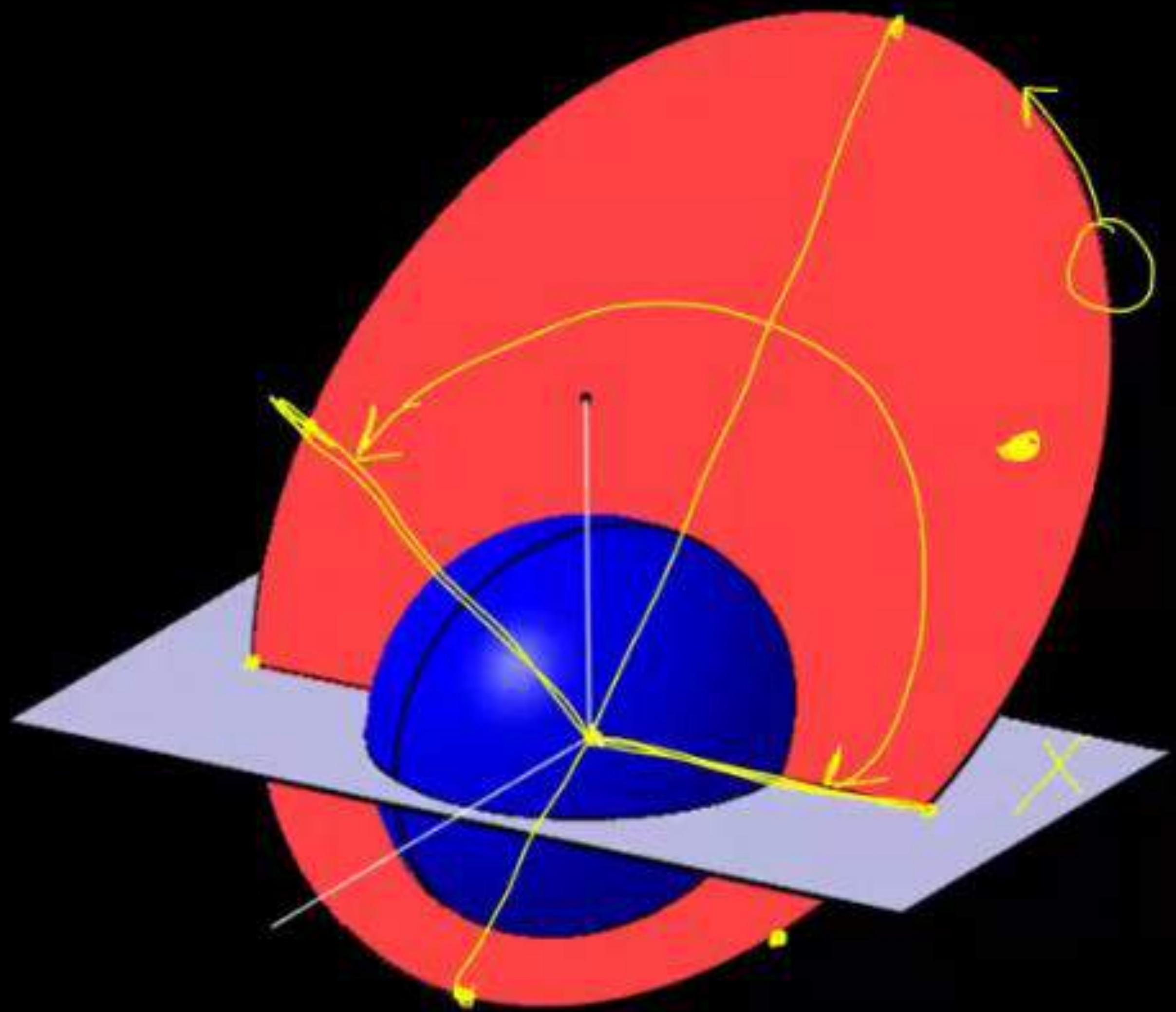
(3) TILT: inclination, i.



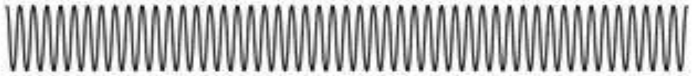


4 SWIVEL:
longitude of the ascending node,
Ω (°)
orbital pl.
descending node
ascending node
equatorial pl.
vernal equinox

⑤ LOCATION: argument of the periapsis
ω (δ)



The GPS signal

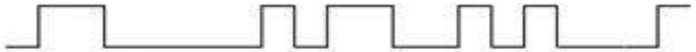


Carrier wave

Two carrier waves are used:
L1 = 1575.42 MHz
L2 = 1227.60 MHz

C/A code

Each satellite modulates its own unique codes onto the carrier waves. The C/A code that civilians can access is sent on L1. It is made up of 1,023 bits. The entire code takes one millisecond to transmit and repeats endlessly.



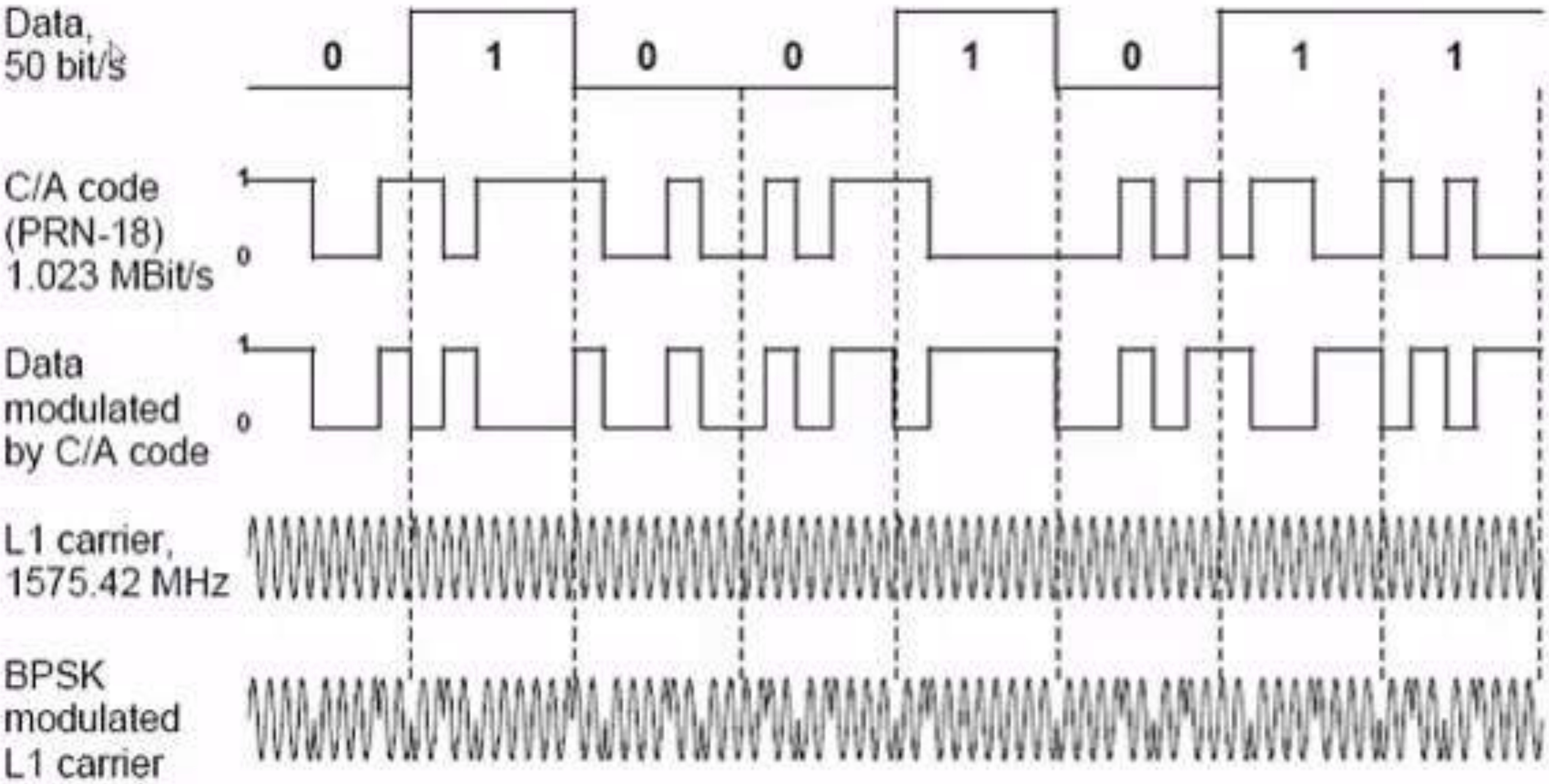
Navigation message

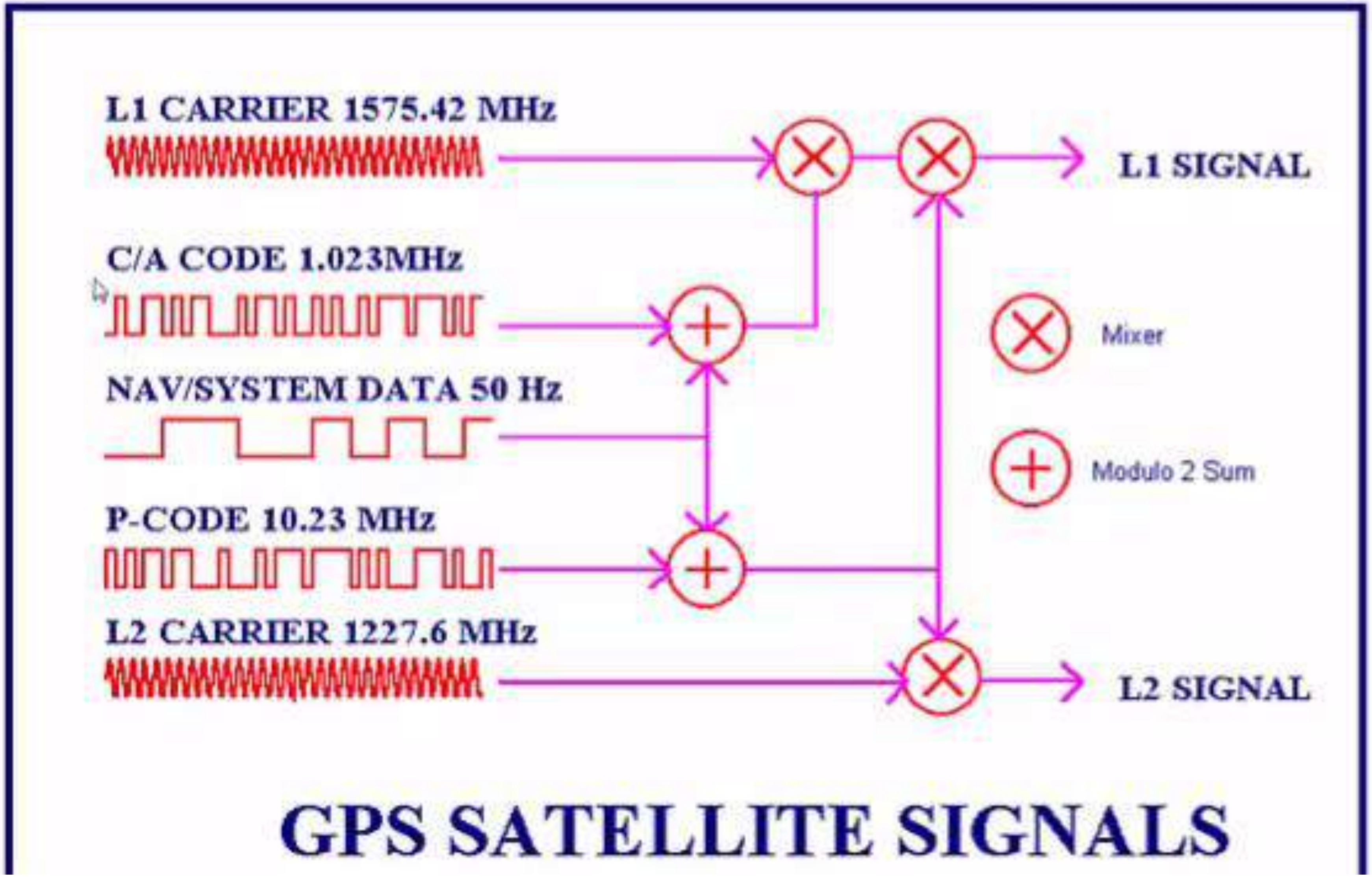
As well as the C/A code, a navigation message is also modulated onto the L1 carrier. This message contains lots of vital information and is quite long. However, because of the relatively slow rate that it is sent, it takes 12.5 minutes to send one complete message.

P Code

A second unique code is modulated onto both the L1 and L2 carriers. This code is encrypted for military use and cannot be used by civilians. It contains many more bits and is sent at a higher speed, which allows authorised users to calculate position accuracy with much greater accuracy.

GPS Bi-Phase Modulation




L1 CARRIER 1575.42 MHz
L1 SIGNAL
C/A CODE 1.023MHz
Mixer
NAV/SYSTEM DATA 50 Hz
Modulo 2 Sum
P-CODE 10.23 MHz
L2 CARRIER 1227.6 MHz
L2 SIGNAL
GPS SATELLITE SIGNALS

- The C/A code each satellite transmits is unique to that satellite, and that while it appears to be a random string of bits, it is not. The code actually follows a precise deterministic pattern, and so it is often called pseudo-random noise (PRN).
- The navigation message is different to C/A code because it contains data. This includes information about clock corrections, satellite health, ephemeris (precise orbit) data, ionosphere model parameters and almanac (general information about all satellites in the constellation) data. Because of the amount of information contained in the navigation message, and the relatively slow rate that it’s transmitted, it takes 12.5 minutes to send the whole message.

- The final part of the GPS signal we need to look at is the carrier itself. As we’ve already said, each satellite transmits two frequencies—L1 at 1575.42 MHz and L2 at 1227.60 MHz. All three elements (the C/A code, Y code and navigation message) are modulated onto the L1 carrier, while only the Y code is modulated onto the L2 carrier. However, while civilian users can’t demodulate the Y code from the L2 carrier, we can make use of the L2 carrier wave itself in differential corrections.
- The C/A code is modulated onto the carrier wave at 1.023 Mb/s. The Navigation message is modulated onto the carrier at 50 b/s.

How GPS Works?

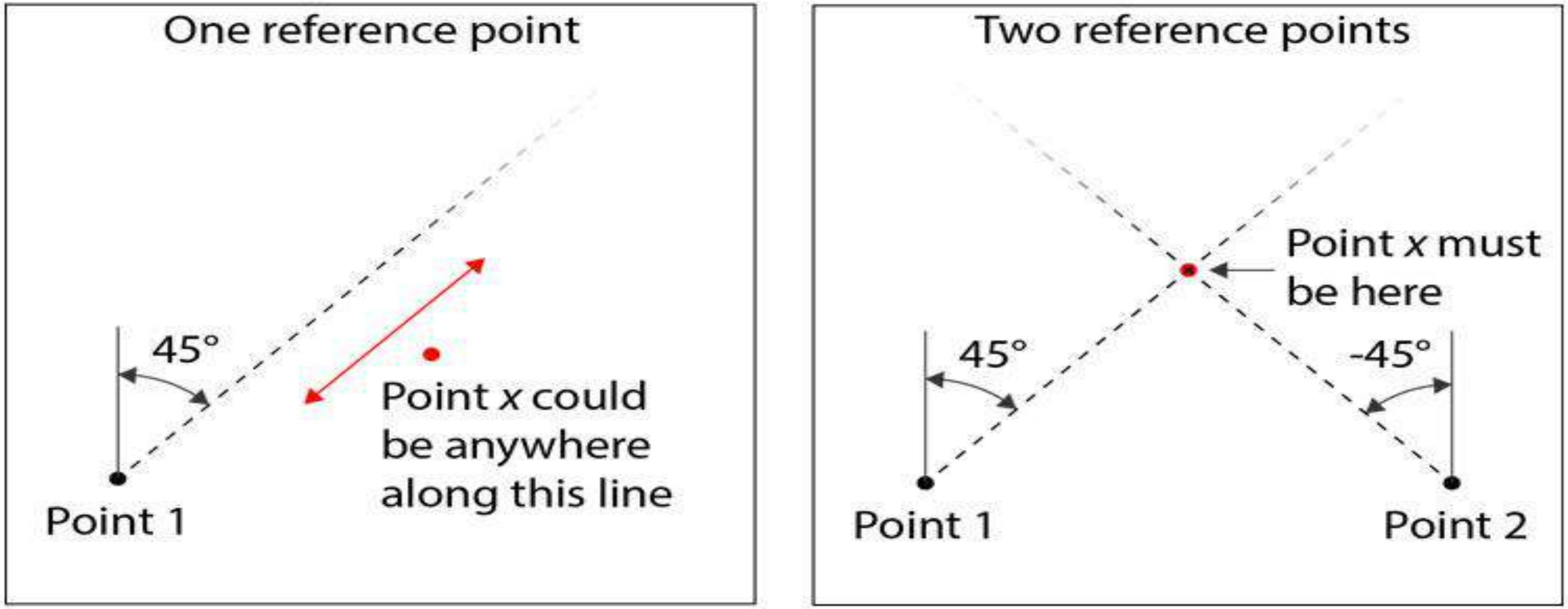
- The GNSS receiver on the vehicle is receiving signals from the satellites and working out where it is.
- The signals that are used depend on the type of receiver. A GPS receiver can only make use of signals from the GPS satellites, while a GLONASS receiver can only use signals from GLONASS satellites.
- There is another kind of receiver that can actually take signals from both types of satellites though (GPS and GLONASS), to augment its measurements.

How does a GPS receiver work out the location?

- It calculates how long it takes the signal from each satellite it can see to arrive.
- It multiplies that time by the speed of light to calculate the distance to each satellite.
- It then calculates its position relative to no fewer than three satellites using trilateration.
- Because the receiver knows the precise position of each satellite when the signal was sent, it can translate its own relative position into an Earth-based co-ordinate system.

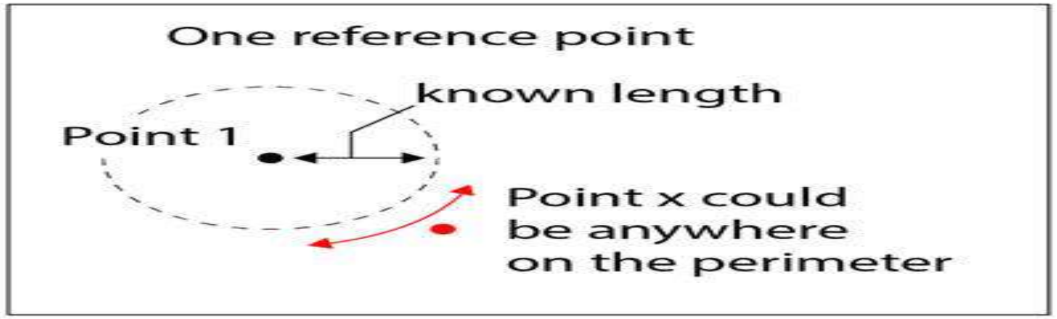
Triangulation

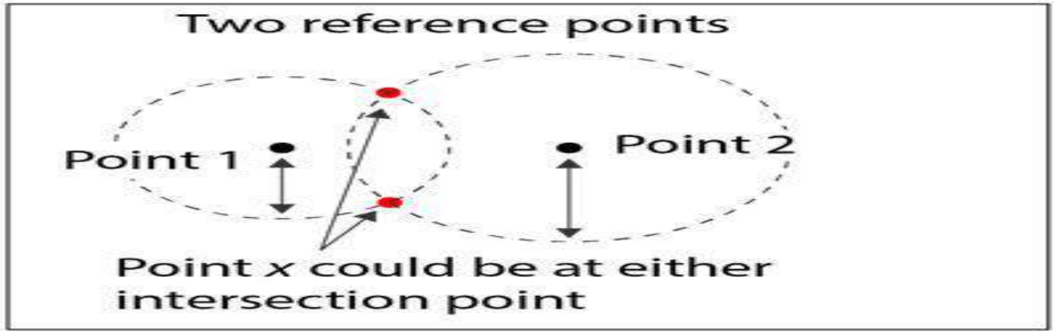
Triangulation projects lines of unknown length along known angles to find a point. As long as there is more than one reference point we can identify the location of a new point. So, if we know point *x* is located at angle of 45° from point 1, and at an -45° from known point 2, the point at which those projected lines intersect must be the location of point *x*.

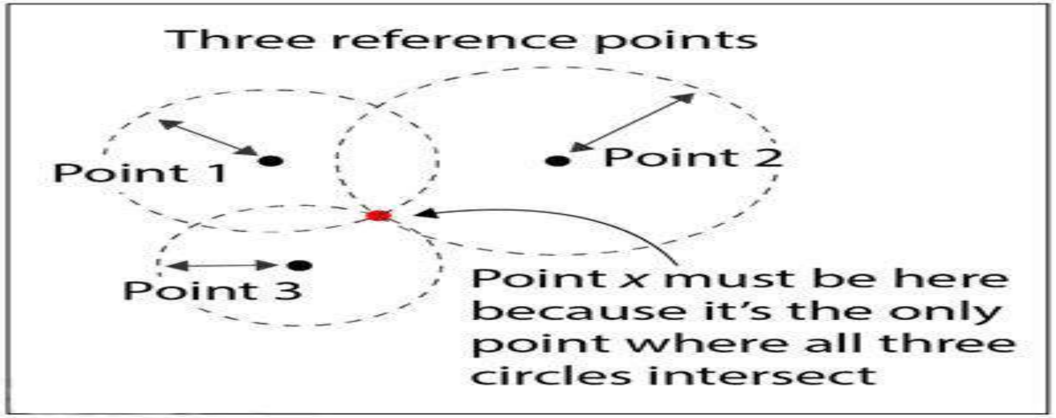


Trilateration

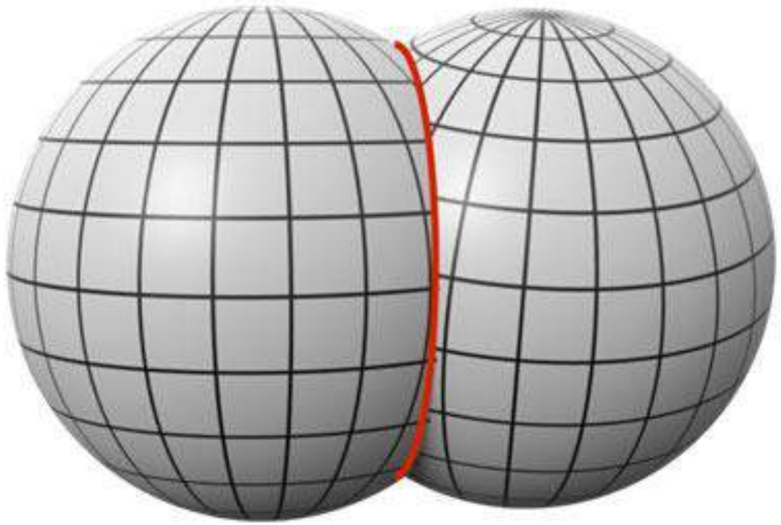
Trilateration uses lines of known length but unknown angle (circles in other words) to find a point. As long as there are more than two reference points we can identify the location of a new point.
So, if we know point *x* is 1 metre from point 1, 1.5 meters from point 2 and 0.75 meters from point 3, the point at which those circles intersect must be the location of point *x*.



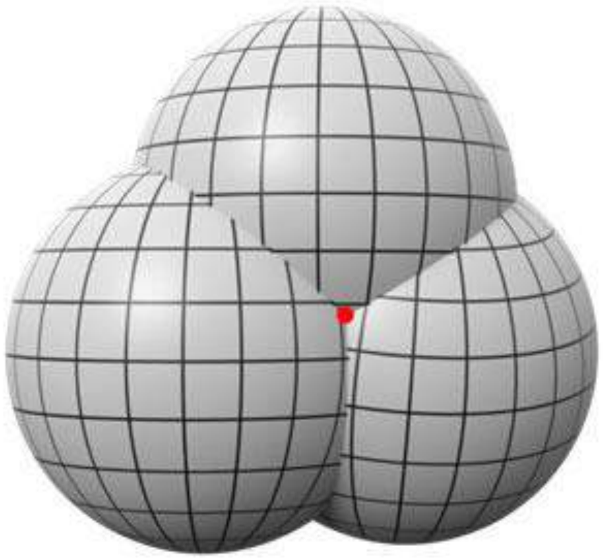




- GPS uses the same technique, but has to approach things slightly differently. The circles we’ve been looking at are two-dimensional. In real life things are far more three-dimensional, and that means our intersecting circles become intersecting spheres.
- When two spheres intersect, you don’t end up with two points of intersection, you end up with a ring of intersection (imagine two bubbles joined together). If a third bubble joins in, it creates two points where all three intersection rings meet. This is shown below.



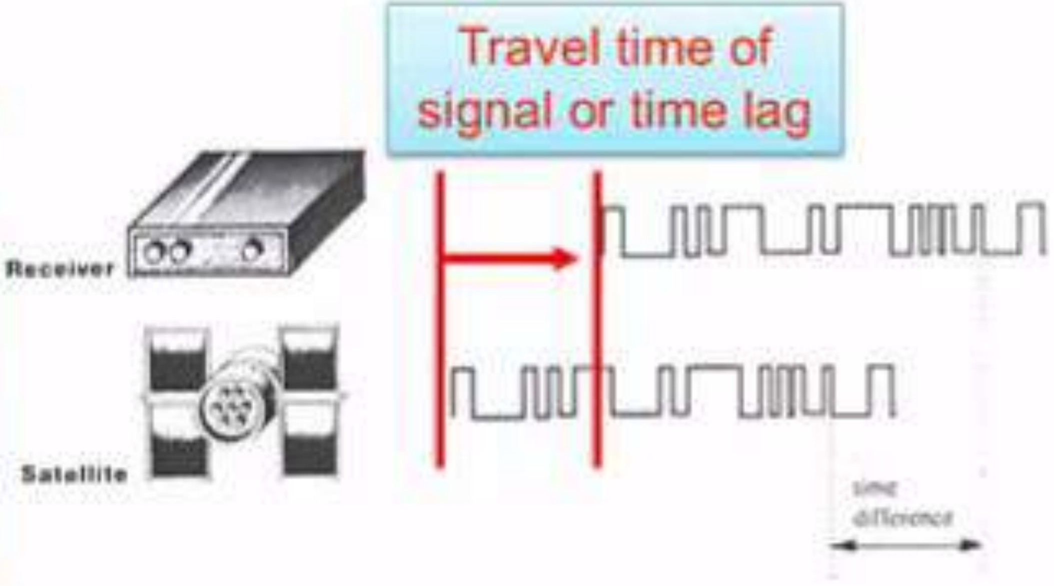
When two spheres intersect, the interestion creates a circle (red).



When three spheres intersect, there are only two points common to all three spheres (one each side).

- At this point you might recall hearing that GPS only needs three satellites in order to generate a position measurement. This statement is both right and wrong. It’s clear from the image above that in order to use trilateration to arrive a single point, a fourth sphere is required (to show us which point is the correct one). However, as GPS uses an Earth centred, Earth fixed coordinate system, one of the points can be ruled out immediately as it would be well outside Earth’s atmosphere. The other point must therefore be the correct one.
- So GPS can generate a position measurement using only three satellites, but in order to do that, its internal clock must be accurate otherwise it can’t calculate the distance properly. And correcting the internal clock requires four satellites! That’s why the statement was both right and wrong. As long as the system has used four satellites to correct its internal clock, it can drop back to only three satellites and still make a position estimate—it just can’t do it forever.

Satellite and receiver generate code at same time, but the travel time delays the signal's arrival at the receiver.



The time lag in the signal is the travel time of the signal

If signal traveled in a straight line at the speed of light:

- Distance to satellite = Speed of light * time lag

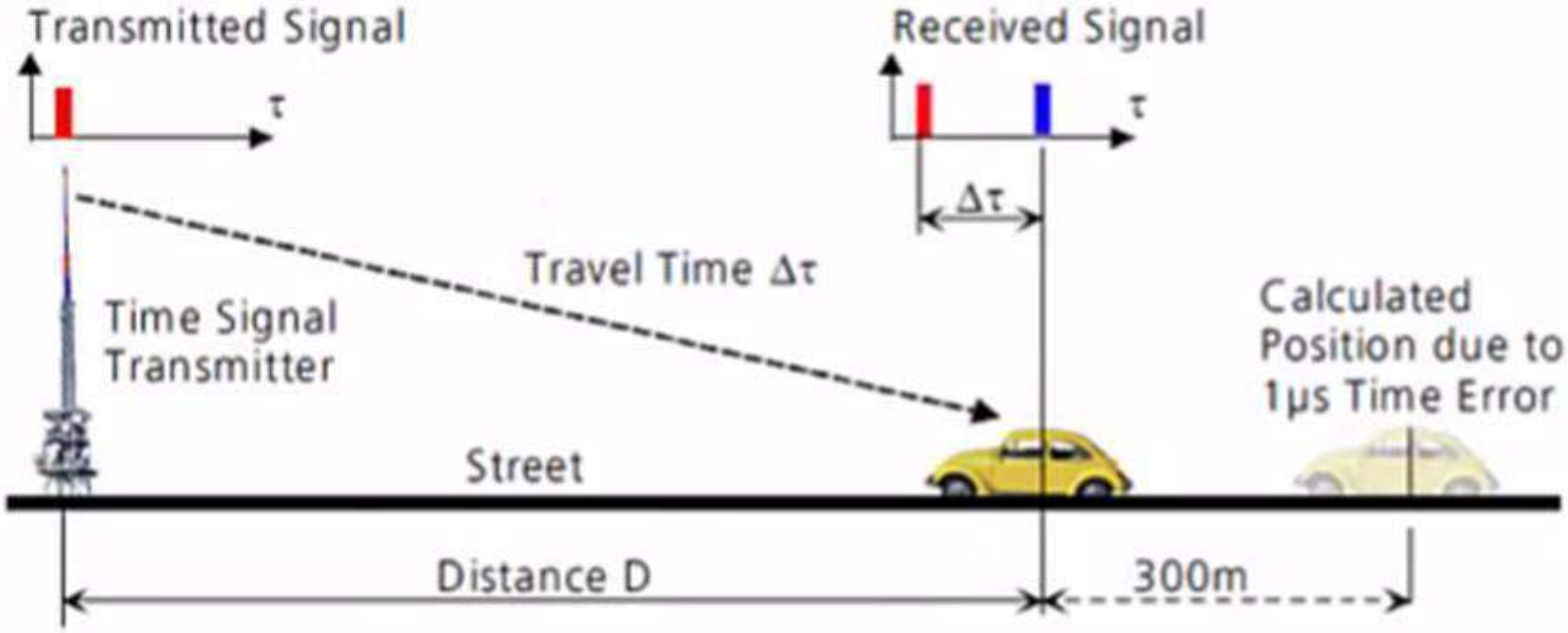


Figure 3: In the simplest case distance is determined by measuring the travel time

The distance D is calculated by multiplying the travel time $\Delta\tau$ by the velocity of light c.

$$D = \Delta\tau \bullet c$$

Need **extreme accuracy** in measuring time!

On order of nanoseconds (10^{-9})

- **Atomic clocks**
- Cost $50-100,000!!!
- Too expensive for consumers of GPS receivers
- Lose a max of 1 sec every 30,000 to 1,000,000 years
- Regularly adjusted or synchronized from controls points on Earth

A GPS receiver actually keeps extremely accurate time by adjusting its quartz clock to match the atomic clocks in the satellites!

The atom in an atomic clock oscillates at a rate of about 9.2 billion times per second.

From *Integrating GIS and the Global Positioning System*

This atomic clock was built for the first GPS satellites in the late 1970s. (National Air and Space Museum, Smithsonian Institution.)

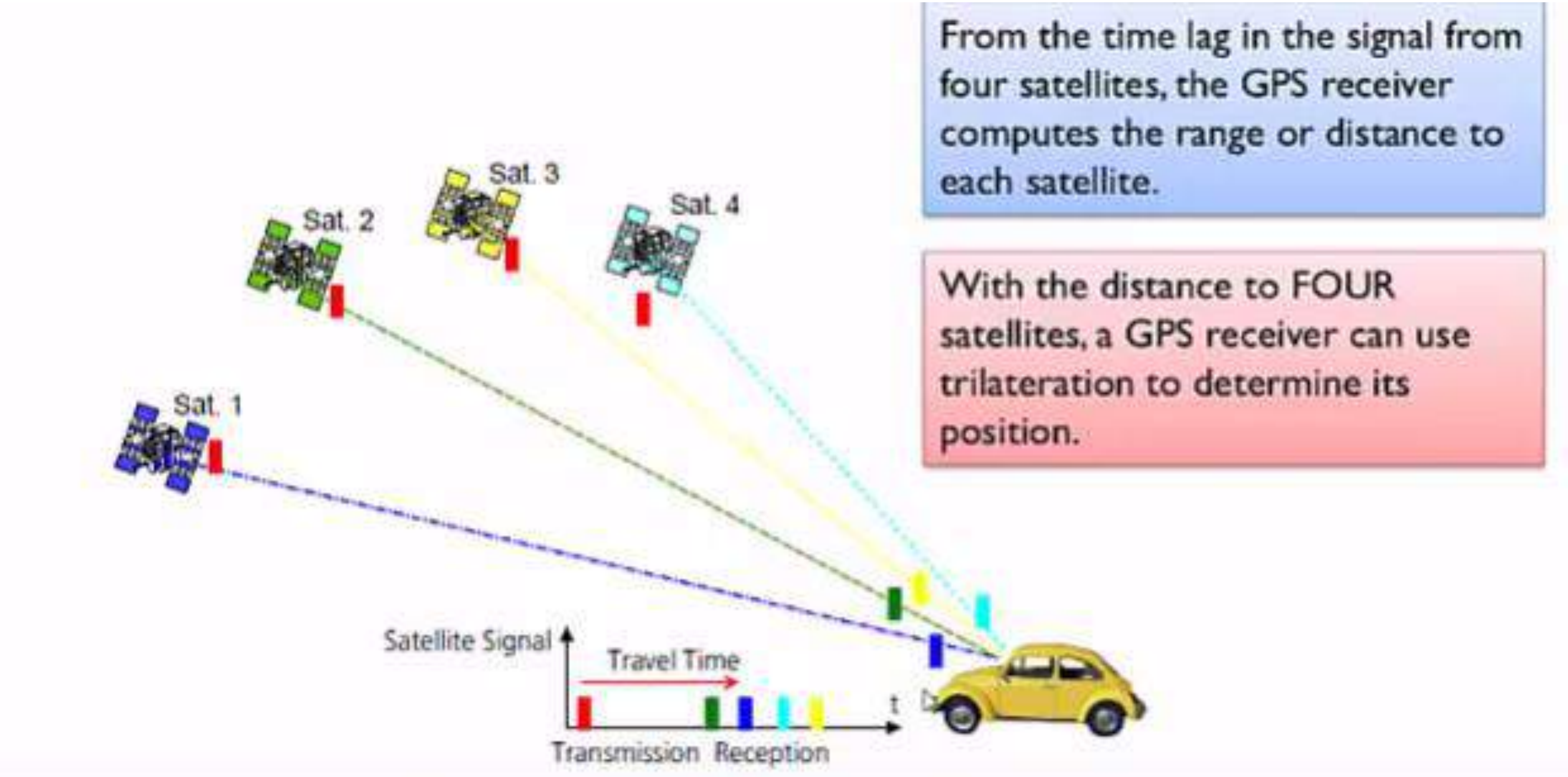


Figure 5: Four satellites are needed to determine longitude, latitude, altitude and time

Another problem…variation in the speed of light…

The speed of the GPS signal when it travels through the Earth's atmosphere changes, which introduces errors into the position calculated by the GPS receiver!

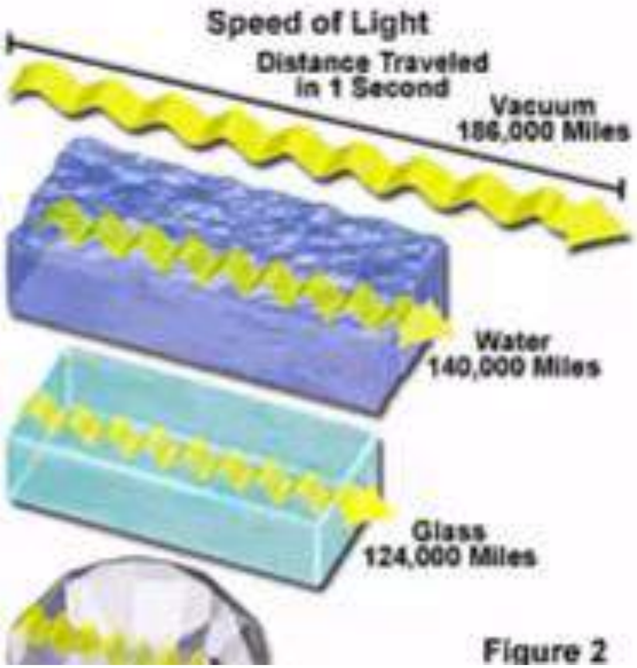


Figure 2

Finding the Satellites

Crucial component of GPS calculations - knowledge of where the satellites are

- Satellite orbits are very high and predictable
  - far enough from the Earth (11,000 miles) that **they are not affected by our atmosphere.**
- But, **satellite orbits do vary**
  - the pull of the moon and the sun

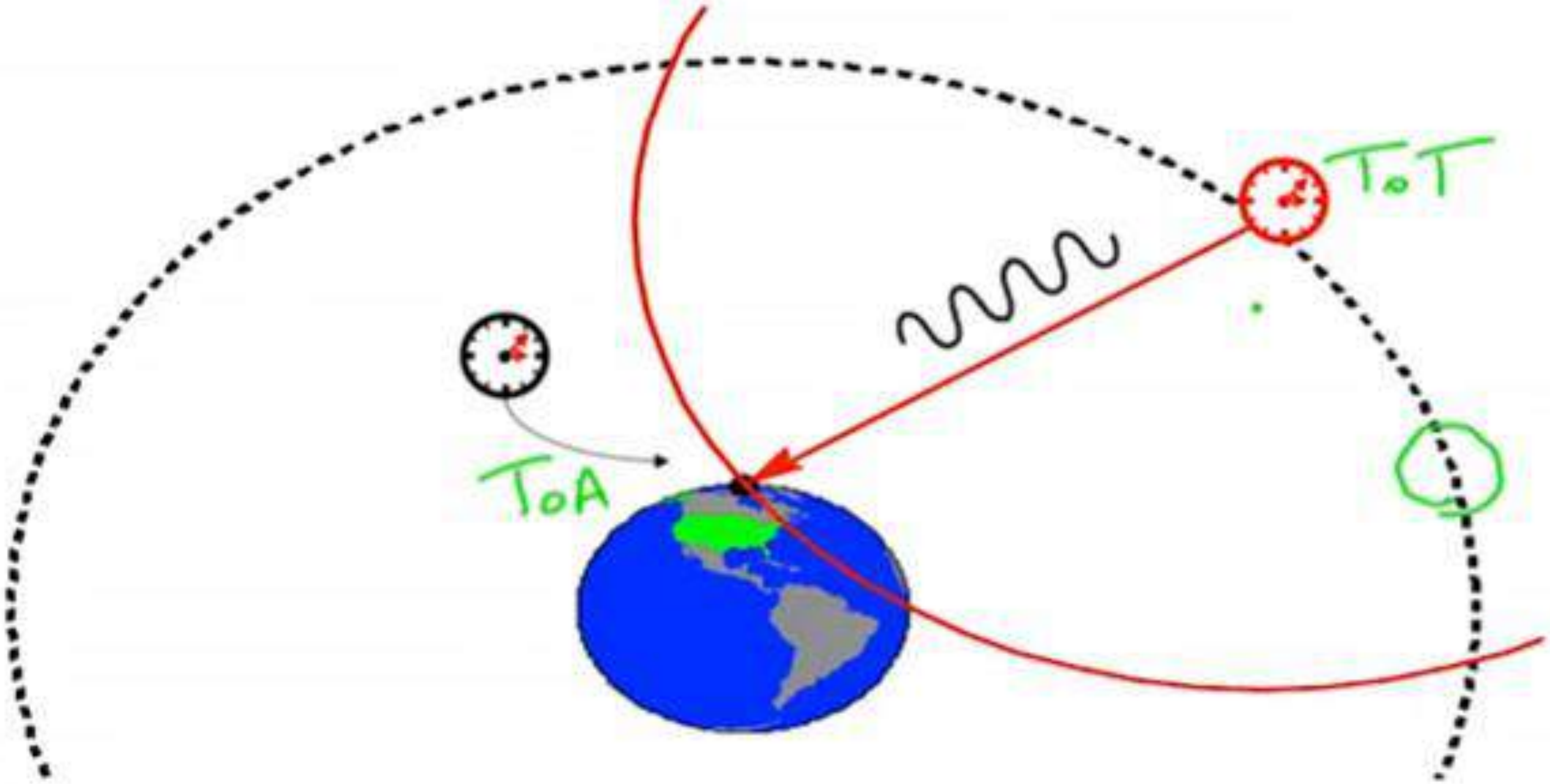
Location of satellites stored in **Navigation Message or Almanac** that is transmitted by the satellites

Finding Position (summary)

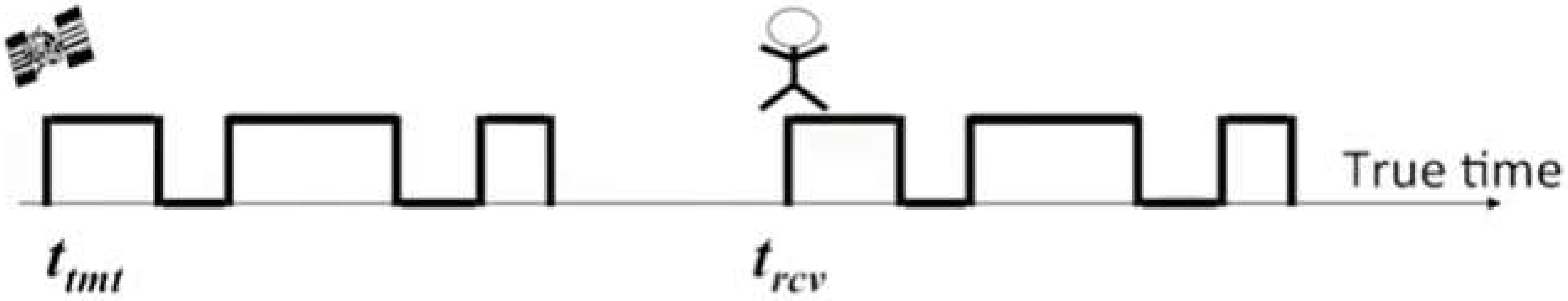
GPS receiver

- Picks up the transmissions of at least four satellites
- Measures the time delay in the signal transmission
- Computes distance to satellite from time delay (assuming the speed of light)
- Uses the distance to four satellites with the information i in an **electronic almanac (includes precise location of each satellite)**
- To mathematically determine the receiver's position on Earth.
  - latitude, longitude and altitude (or some similar measurement)

Each Satellite Stamps the **Transmission Time.**
GPS Receiver Measures the **Arrival Time.**



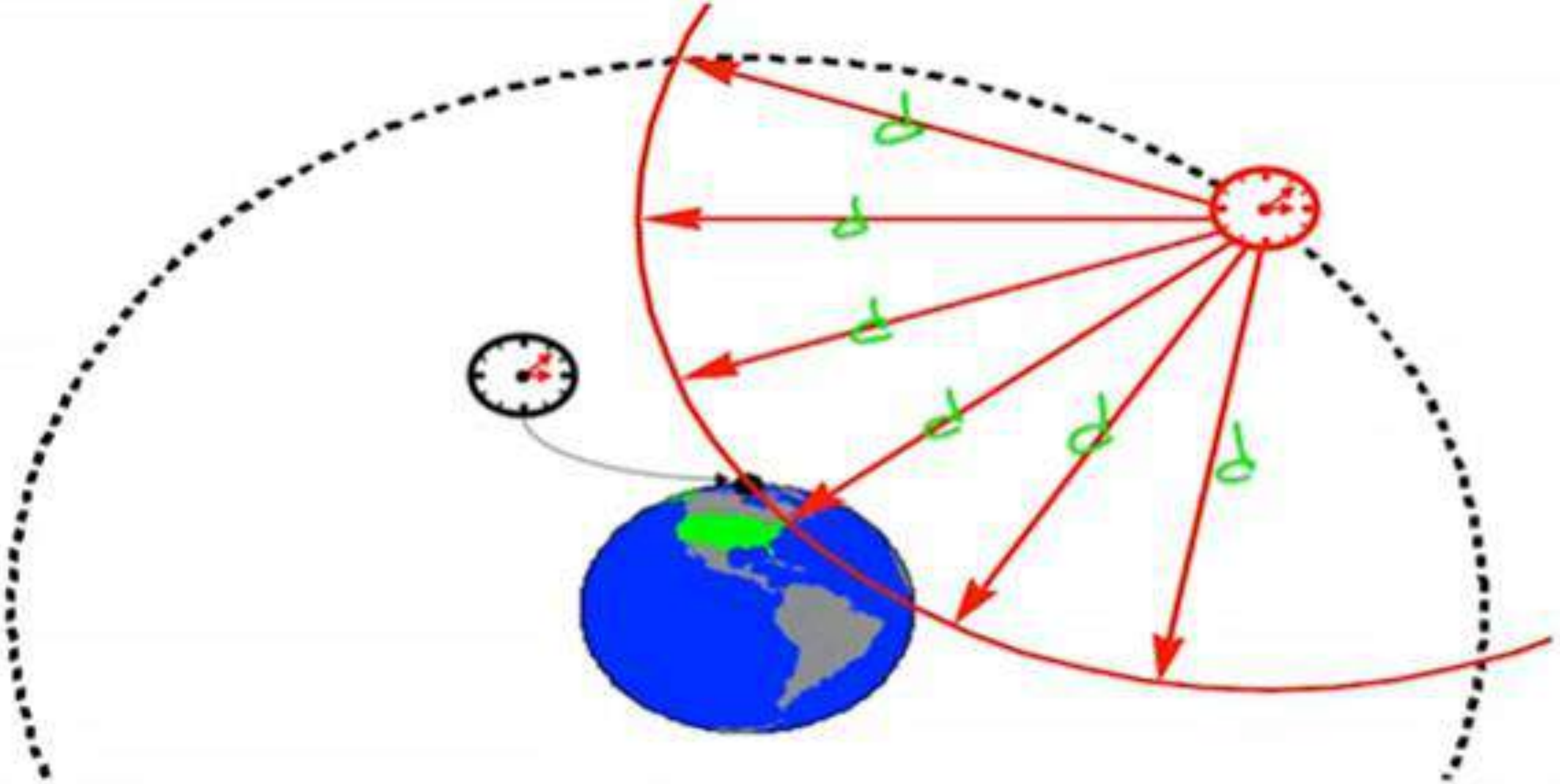
Travel Time in a Vacuum



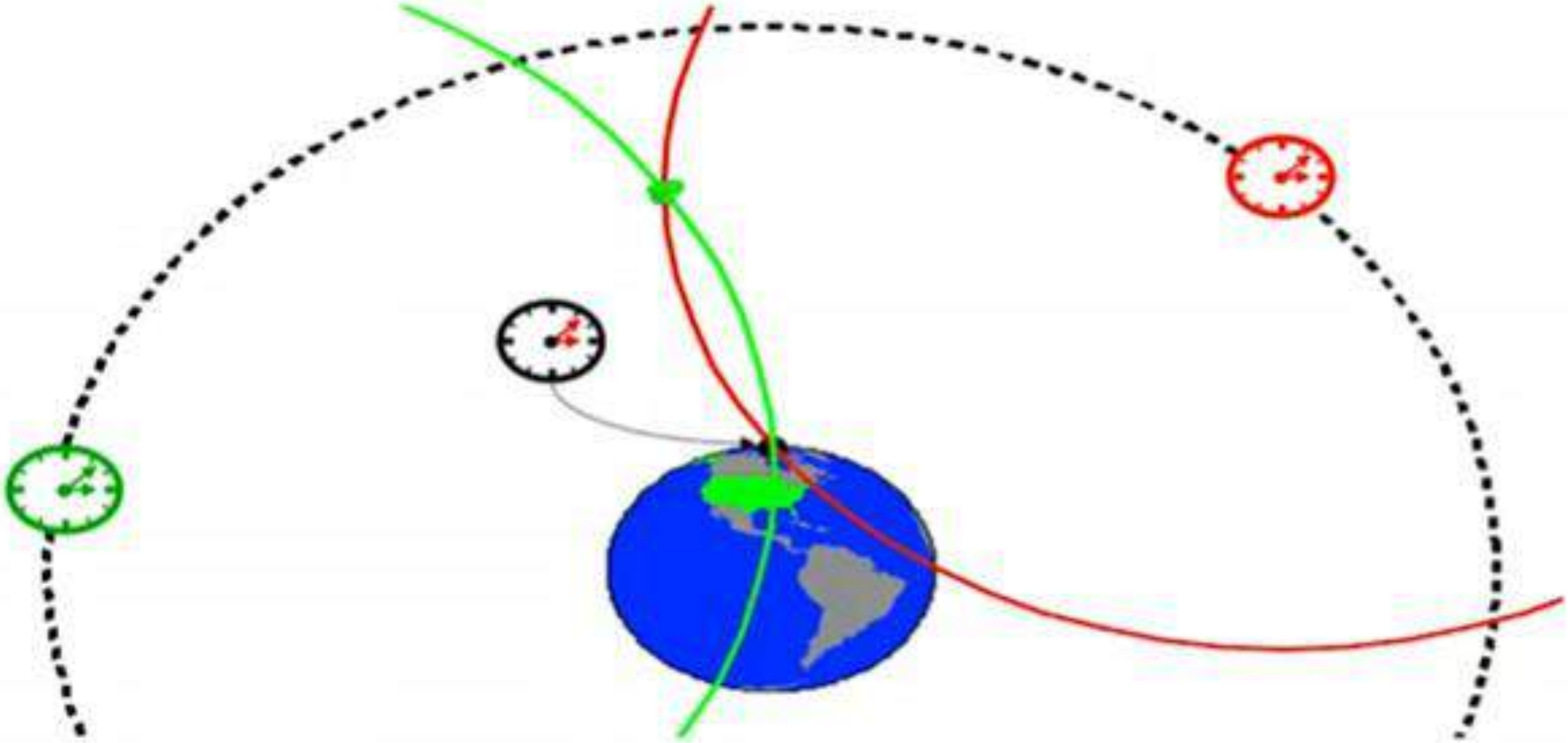
True Travel Time

$$t_{rcv} - t_{tmt} = \frac{d}{c}$$

1 satellite defines a circle (sphere) of possible locations:

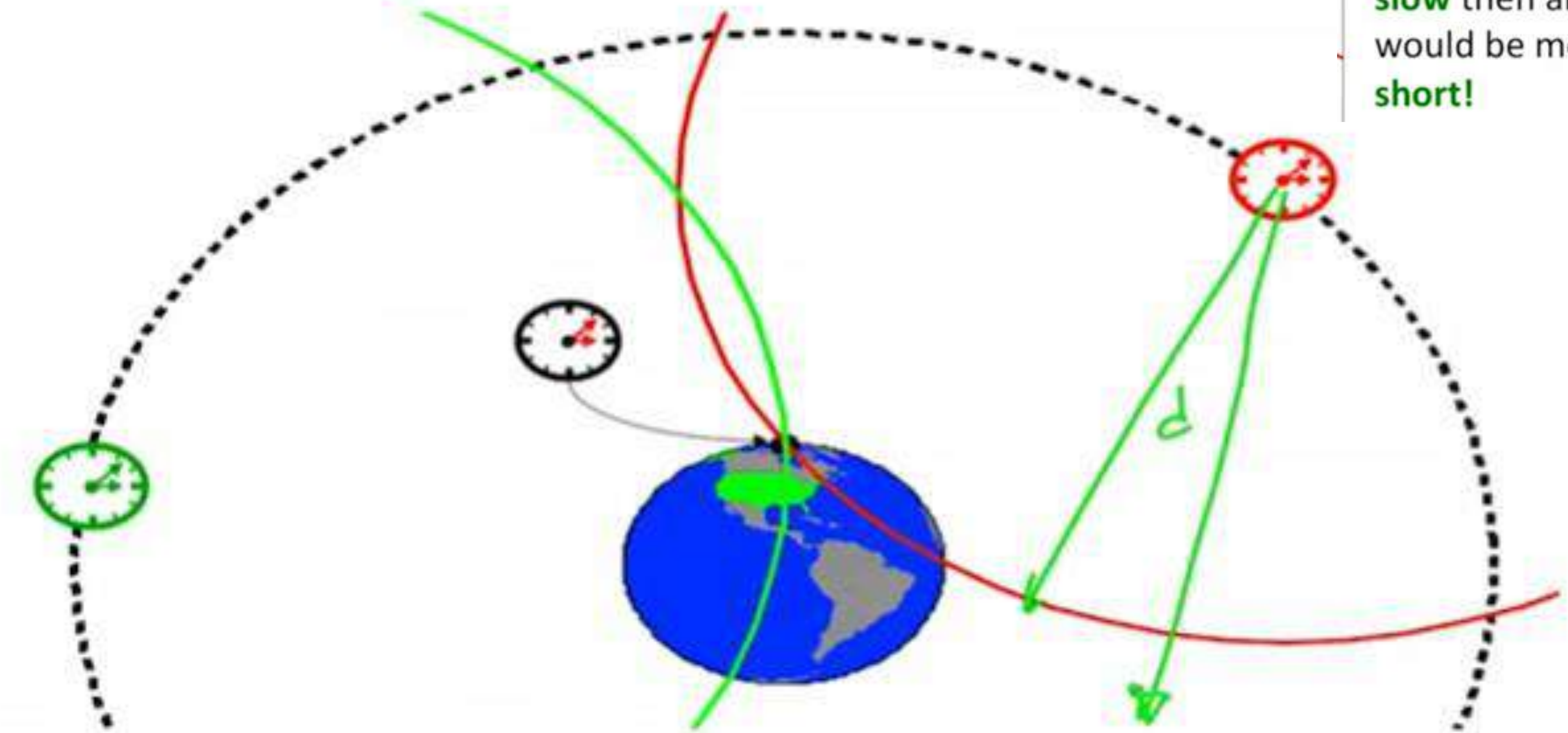


2 satellites seem to be enough.
3 satellites for three dimensions.



If the <u>User Clock</u> is **fast** then all ranges would be measured **long!**

If the <u>User Clock</u> is **slow** then all ranges would be measured **short!**



Some Maths

Measuring the spherical ranges from four satellites as shown in Figure 6.32 enables the user's position to be determined and yields four equations containing the four unknowns, viz the three position co-ordinates of the user and the time bias in the user's clock. The position co-ordinates of the user can thus be determined together with very accurate time information. Figure 6.33 shows the data transmission waveforms and illustrates the user time bias ΔT, and the time delays Δ_{t_1}, Δ_{t_2}, Δ_{t_3} and Δ_{t_4} for the signals transmitted from the satellites to reach the user.

Four pseudo ranges R_{1p}, R_{2p}, R_{3p}, R_{4p} to the four satellites $S1$, $S2$, $S3$, $S4$ can be determined, viz.

$$
\begin{aligned}
R_{1p} &= c\Delta t_1 \\
R_{2p} &= c\Delta t_2 \\
R_{3p} &= c\Delta t_3 \\
R_{4p} &= c\Delta t_4
\end{aligned}
\tag{6.55}
$$

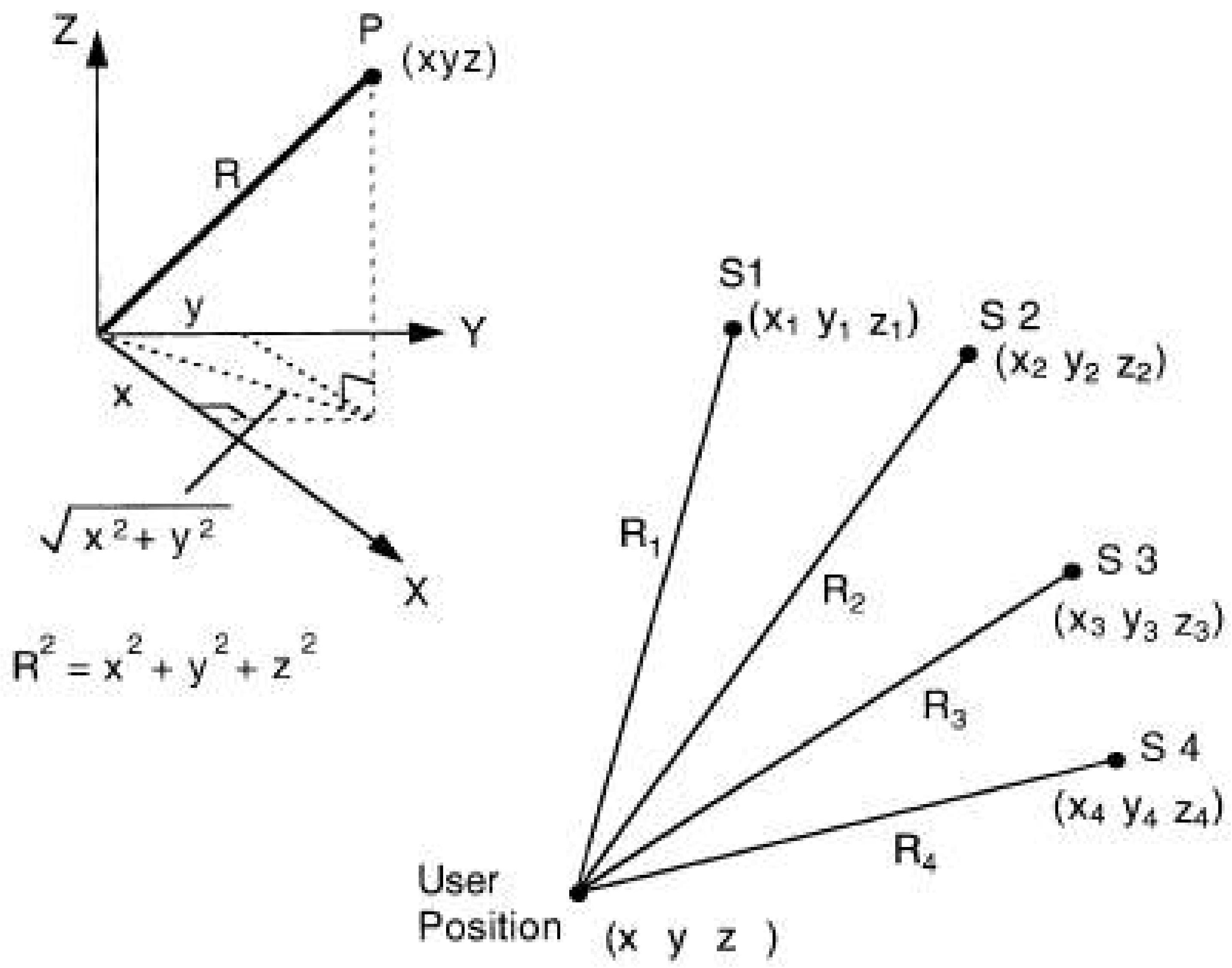


Fig. 6.32 User–satellite geometry.

Let the range equivalent of the user's clock offset be T, i.e.

$$T = c_{\Delta} T$$

Hence, from basic 3D co-ordinate geometry (see Figure 6.31)

$$R_1 = [(X - X_1)^2 + (Y - Y_1)^2 + (Z - Z_1)^2]^{1/2} = R_{1p} - T \tag{6.56}$$

$$R_2 = [(X - X_2)^2 + (Y - Y_2)^2 + (Z - Z_2)^2]^{1/2} = R_{2p} - T \tag{6.57}$$

$$R_3 = [(X - X_3)^2 + (Y - Y_3)^2 + (Z - Z_3)^2]^{1/2} = R_{3p} - T \tag{6.58}$$

$$R_4 = [(X - X_4)^2 + (Y - Y_4)^2 + (Z - Z_4)^2]^{1/2} = R_{4p} - T \tag{6.59}$$

when R_1, R_2, R_3, R_4 are the actual ranges from the user's position to the four satellites S1, S2, S3, S4 and the coordinates of these satellites are $(X_1\ Y_1\ Z_1)$, $(X_2\ Y_2\ Z_2)$, $(X_3\ Y_3\ Z_3)$, $(X_4\ Y_4\ Z_4)$ respectively.

These four equations with four unknowns can thus be solved and yield the user's position coordinates (X, Y, Z) and the user's time offset, ΔT.

The assumption of perfect satellite clocks made initially in the discussion, however, is not a valid one and in fact the clocks are slowly but steadily drifting away from each other. The satellite clocks are therefore, mathematically synchronised to a defined GPS Master time which is maintained at the Master Control Station. This GPS Master time is continuously monitored and related to the Universal Time Co-ordinate (UTC) maintained by the United States Naval Observatory.

GPS ERRORS

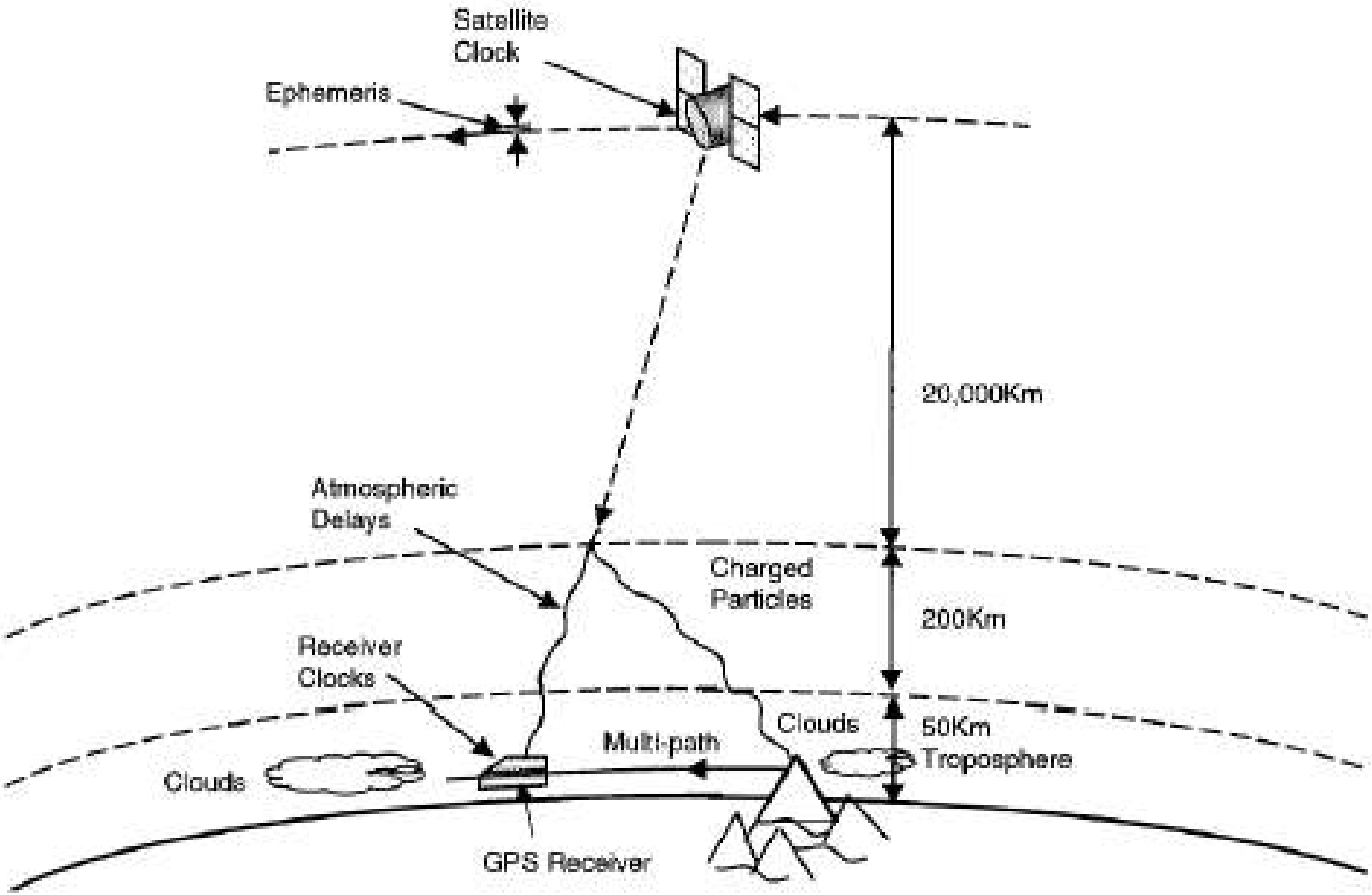


Fig. 6.36 GPS error sources.

- A predicted correction factor for the Earth's atmosphere path is made in the receiver but this is based on a statistical model and there are inevitable residual errors present.
- ***Multi-path errors*****.** The GPS satellite signal is received by the direct line of sight (LOS) path, but the signal may also be received as the result of reflections off local obstructions. The reflected signals arrive slightly delayed from the direct LOS signal and are termed multi-path signals. The resulting noise is called multi-path error. The multi-path errors experienced in a moving receiver occur in a random fashion which results in a noise like pseudo-range error. The problem is considerably alleviated by the use of an early-late delay-lock loop and suitable filtering techniques.
- **Receiver clock**. Internal noise in the GPS receiver clock introduces a small error. DGPS enables most of the above errors to be counteracted, as they are common to both the Reference Station receiver and the user receiver.

- ***GPS satellite clocks***. GPS satellites are equipped with very accurate atomic clocks and corrections are made via the Ground Stations. Even so, very small timing errors are present and so contribute to the overall position uncertainty. Selective Availability deliberately introduced noise equivalent to around 30 m in the individual satellite clock signals.
- ***Satellite ephemeris errors*****.** The satellite position is the starting point for all the positioning computations, so that errors in the Ephemeris data directly affect the system accuracy. GPS satellites are injected into very high orbits and so are relatively free from the perturbing effects of the Earth's upper atmosphere. Even so, they still drift slightly from their predicted orbits and so contribute to the system error.
- ***Atmospheric errors***. Radio waves slow down slightly from the speed of light in vacuum as they travel through the ionosphere and the Earth's atmosphere. This is due respectively to the charged particles in the ionosphere and the water vapour and neutral gases present in the troposphere. These delays translate directly into a position error.
- The use of different frequencies in the L1 and L2 transmissions enables a significant correction to be made for ionospheric delays.

SPS

- **The term SPS stands for Standard Positioning Service, and describes GPS position measurements that are based only on the C/A code.**
- SPS provides the **lowest accuracy GPS position measurements**, normally in the region of 3–10 meters. To make SPS measurements the GPS receiver locks onto four or more satellites, and then uses the C/A code to estimate the distance to each satellite. These estimates are called pseudo-range measurements.
- Using SPS, the accuracy of the measurement comes from the receiver's ability to correctly align its internally generated C/A code with that received at the antenna. Although even if it were able to do this perfectly, the accuracy is still limited by the design of the system.
- **Greater accuracy is required in many other applications, so other measurement techniques such as DGPS (Differential GPS) and Real Time Kinematics (RTK) are often required.**

Differential GPS

- Concerns about potential enemies using GPS to deliver missiles and other weapons against the US had led to a policy of **accuracy denial, generally known as Selective Availability**.
- The GPS ground stations deliberately introduced satellite timing errors to reduce the positioning accuracy available to civil users to a horizontal positioning accuracy of 100 m to a 95% probability level. This was deemed adequate for general navigation use, but in practice it did not satisfy the accuracy or integrity requirements for land or hydrographic surveying, coastal navigation or airborne navigation.
- It should be noted that even the 16 m accuracy, now available, is insufficiently accurate for many applications. For example, positioning of off-shore oil drilling rigs or automatic landing in the case of airborne applications.
- A supplementary navigation method known as *Differential GPS* (DGPS) has therefore been developed to improve the positioning accuracy for the growing number of civil applications.

DGPS can be defined as:

- The positioning of a mobile station in real-time by corrected GPS pseudo ranges. The corrections are determined at a static ‘reference station’ and transmitted to the mobile station. A monitor station may be part of the system, as a quality check on the reference station transmissions.

- The basic principle underlying DGPS is the fact that the errors experienced by two receivers simultaneously tracking a satellite at two locations fairly close to each other will largely be common to both receivers.
- The basic differential GPS concept is illustrated in Figure . The position of the stationary GPS Reference Station is known to very high accuracy so that the satellite ranges can be very accurately determined.

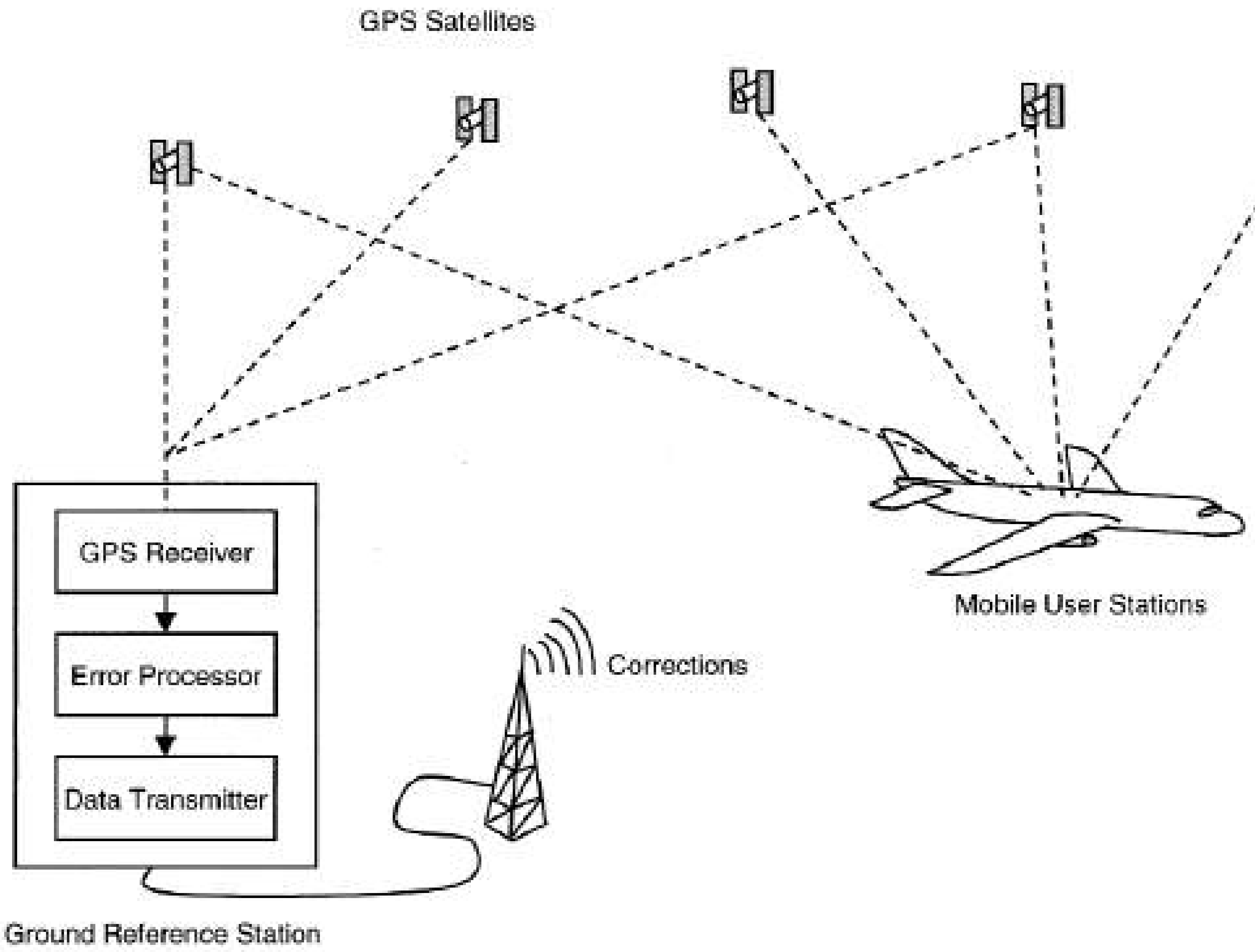


Fig. 6.35 The differential GPS concept.

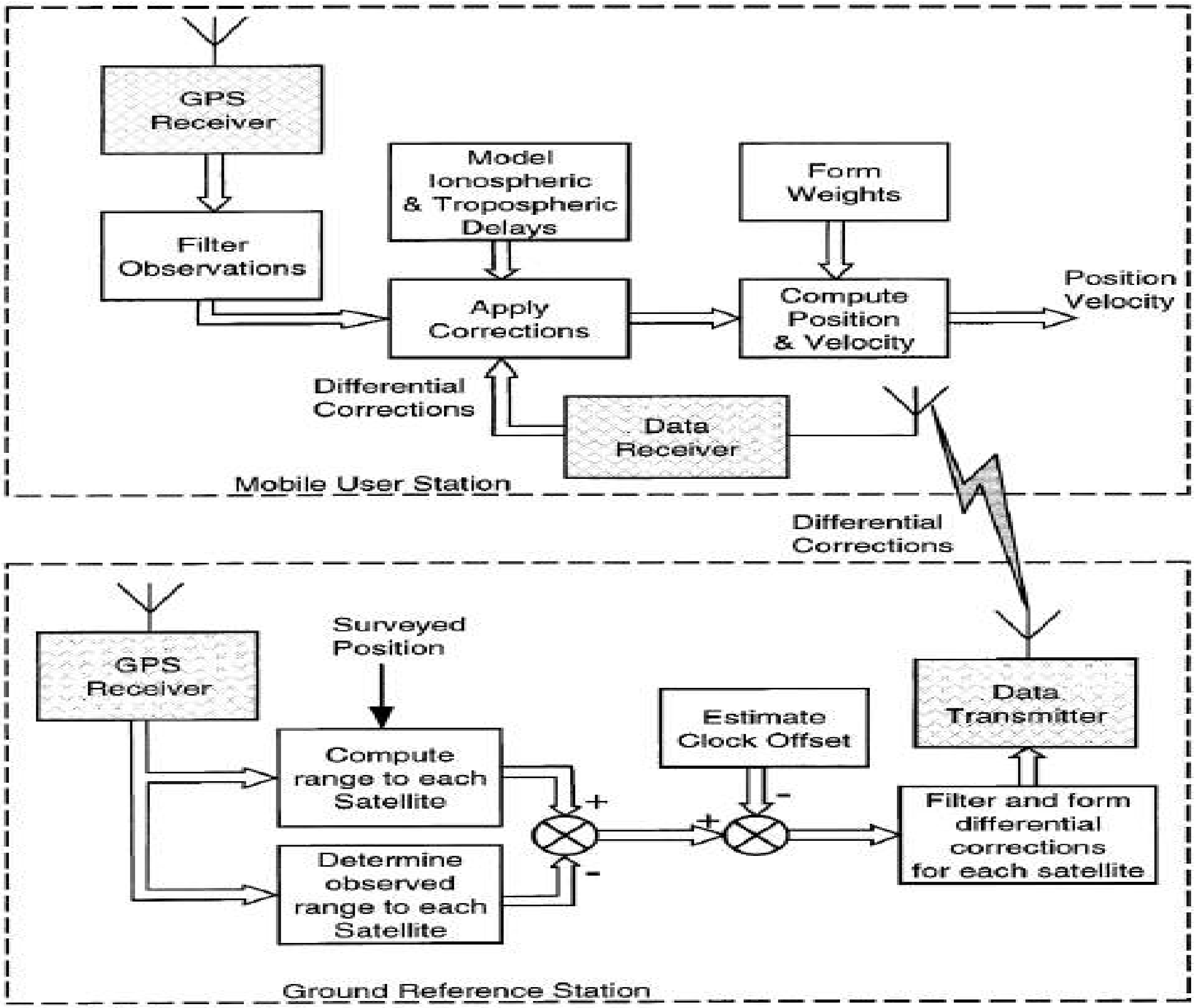


Fig. 6.37 Simplified functional diagram of a generic differential GPS system.

- **There are also other ways to achieve DGPS.** Rather than using your own local base-station it is possible to receive differential corrections via a **web-based service and data connection**.
- In this case your receiver works out where it is in the normal way.
- It then transmits its location to a central server that works out where the nearest reference station is to you, before sending back the differential corrections from that location.

- Another method is to use satellite-based correction services called **SBAS (Satellite Based Augmentation System).**
- Ground stations on earth at known locations work out the differential corrections in different geographic locations.
- These corrections are then uploaded to the satellite where they can be broadcast down to any receiver that is enabled to receive them.
- However, as well as needing a receiver with the correct hardware, you often have to pay a subscription fee in order to unlock the corrections.

- The ultimate aim of RTK is to establish how many carrier waves there are between the antenna and the satellite. The reason for this is simple. We’ve already seen that each satellite broadcasts a unique C/A code made up of 1,023 bits. The code is sent at a rate of 1.023 Mb/s, which means one bit is sent about every microsecond. In one microsecond the radio signal from the satellite covers a distance of about 300 meters.

RTK

- RTK stands for **Real Time Kinematic**, and is another technique that improves the accuracy of GPS position measurements.
- RTK is the next step on from DGPS, and there are two versions of it; **RTK float**, which achieves decimeter level accuracy and **RTK integer,** which achieves centimeter level accuracy.
- It is based on the carrier waves themselves, not on the C/A code or navigation message they carry.

- The carrier wave that the C/A code is modulated onto is at a much higher frequency however—1575.42 MHz. This means a single wave covers about 19 cm**. If we could work out how many full waves there are were between the satellite and the antenna, then it would be possible to calculate the distance much more accurately**. In fact if we know how many full waves there are, and it’s possible to measure partial waves (the phase angle) too, then we can be very precise.

- **RTK Float aims to identify your probable location (improving on the current DGPS accuracy) using statistical methods**. It requires a minimum of four common satellites with the base station, and (in lay-terms), looks for a point in a circle around the current position measurement about which the satellites are revolving.

Velocity measurements

- It's a common misconception that GPS velocity measurements are made by measuring the time taken to move between two consecutive position measurements.
- **To work out velocity, GPS receivers actually use Doppler shift.** So they look at the frequency shift in the carrier wave from each satellite and calculate a velocity vector using that information.

What are the limitations of GNSS?

- Need a clear and uninterrupted view of the sky.
- Also, while a GPS receiver is great at capturing position and velocity measurements it’s not much good if you’re also interested in roll, pitch and yaw.

GPS Modernization

New Block IIR-M satellites

- Two new civil codes on L2 carrier
- Two new military codes on both L1 and L2

Block II-F

- New L5 carrier wave to avoid interference with ground radars that operate near L2 band
- Will carry two new codes

Thank You