

# Navigational Equipments



Ariel Gabbard

First Edition, 2012

ISBN 978-81-323-4193-2



© All rights reserved.

*Published by:*

**White Word Publications**

4735/22 Prakashdeep Bldg,

Ansari Road, Darya Ganj,

Delhi - 110002

Email: [info@wtbooks.com](mailto:info@wtbooks.com)

# Table of Contents

Chapter 1 - GPS Navigation Device

Chapter 2 - Sonar

Chapter 3 - Anemometer

Chapter 4 - Astrocompass and Flight Instruments

Chapter 5 - Decca Navigator System

Chapter 6 - Dynamic Positioning

Chapter 7 - Global Positioning System

Chapter 8 - Inertial Navigation System

Chapter 9 - Octant (Instrument)

Chapter 10 - Reflecting Instrument

## Chapter 1

# GPS Navigation Device



Bicycle navigation with Gosmore, an open source routing software, on a personal navigation assistant with free map data from OpenStreetMap.

A **GPS navigation device** is any device that receives Global Positioning System (GPS) signals for the purpose of determining the device's current location on Earth. GPS devices provide latitude and longitude information, and some may also calculate altitude, although this is not considered sufficiently accurate or continuously available enough (due to the possibility of signal blockage and other factors) to rely on exclusively to pilot

aircraft. GPS devices are used in military, aviation, marine and consumer product applications.

GPS devices may also have additional capabilities such as:

- containing maps, which may be displayed in human readable format via text or in a graphical format
- providing suggested directions to a human in charge of a vehicle or vessel via text or speech
- providing directions directly to an autonomous vehicle such as a robotic probe
- providing information on traffic conditions (either via historical or real time data) and suggesting alternative directions
- providing information on nearby amenities such as restaurants, fueling stations, etc.

In other words, all GPS devices can answer the question "Where am I?", and may also be able to answer:

- which roads or paths are available to me now?
- which roads or paths should I take in order to get my desired destination?
- if some roads are usually busy at this time or are busy right now, what would be a better route to take?
- where can I get something to eat nearby or where can I get fuel for my vehicle?

### ***Consumer applications***

Consumer GPS navigation devices include:

- Dedicated GPS navigation devices
- GPS modules that need to be connected to a computer to be used
- GPS loggers that record trip information for download. Such GPS tracking is useful for trailblazing, mapping by hikers and cyclists, and the production of geocoded photographs.
- Converged devices, including GPS Phones and GPS cameras, in which GPS is a feature rather than the main purpose of the device. Those devices may be assisted GPS or standalone (not network dependent) or both.

## Dedicated GPS navigation devices



A variety of hand-held receivers



A taxi equipped with GPS navigation system

Dedicated devices have various degrees of mobility. *Hand-held, outdoor, or sport* receivers have replaceable batteries that can run them for several hours, making them suitable for hiking, bicycle touring and other activities far from an electric power source. Their screens are small, and some do not show color, in part to save power. Cases are rugged and some are water resistant.

Other receivers, often called *mobile* are intended primarily for use in a car, but have a small rechargeable internal battery that can power them for an hour or two away from the car. Special purpose devices for use in a car may be permanently installed and depend entirely on the automotive electrical system.

The pre-installed embedded software of early receivers did not display maps; 21st century ones commonly show interactive street maps (of certain regions) that may also show points of interest, route information and step-by-step routing directions, often in spoken form with a feature called "text to speech".

Manufacturers include:

- Navman products
- TomTom products
- Garmin products
- Mio products

- Navigon products
- Magellan Navigation consumer products
- TeleType products

## **Mobile phones with GPS capability**

Due in part to regulations encouraging mobile phone tracking, including E911, the majority of GPS receivers are built into mobile telephones, with varying degrees of coverage and user accessibility. Commercial navigation software is available for most 21st century smartphones as well as some Java-enabled phones that allows them to use an internal or external GPS receiver (in the latter case, connecting via serial or Bluetooth). Some phones with GPS capability work by assisted GPS (A-GPS) only, and do not function when out of range of their carrier's cell towers. Others can navigate worldwide with satellite GPS signals as a dedicated portable GPS receiver does, upgrading their operation to A-GPS mode when in range. Still others have a hybrid positioning system that can use other signals when GPS signals are inadequate.

More bespoke solutions also exist for smartphones with inbuilt GPS capabilities. Some such phones can use tethering to double as a wireless modem for a laptop, while allowing GPS-navigation/localisation as well. One such example is marketed by Verizon Wireless in the United States, and is called VZ Navigator. The system uses gpsOne technology to determine the location, and then uses the mobile phone's data connection to download maps and calculate navigational routes. Other products including iPhone are used to provide similar services. Nokia gives Ovi Maps free on its smartphones and maps can be preloaded. According to market research from the independent analyst firm Berg Insight, the sales of GPS-enabled GSM/WCDMA handsets was 150 million units in 2009. while only 40 million separate GPS receivers were sold.

GPS navigation applications for mobile phones include Waze.

## **Laptop PC GPS**

Various software companies have made available GPS road navigation software programs for in-vehicle use on laptop computers. Benefits of GPS on a laptop include larger map overview, ability to use the keyboard to control GPS functions, and some GPS software for laptops offers advanced trip-planning features not available on other platforms. Laptop computers allow for other uses beside GPS.

## GPS modules



A modern SiRFstarIII chip based 20-channel GPS receiver with WAAS/EGNOS support.

Other GPS devices need to be connected to a computer in order to work. This computer can be a home computer, laptop or even a PDAs, or smartphones. Depending on the type of computer and available connectors, connections can be made through a serial or USB cable, as well as Bluetooth, CompactFlash, SD, PCMCIA and the newer ExpressCard. Some PCMCIA/ExpressCard GPS units also include a wireless modem. Devices usually do not come with preinstalled GPS navigation software, thus once purchased the user must install or write their own navigation software. As the user can choose which navigation software to use, it can be better matched to their personal taste. It is very common for a PC-based GPS receiver to come bundled with a navigation software suite. Also, GPS modules are significantly cheaper than complete stand-alone systems (around 50-100 €). The software may include maps only for a particular region, or the entire world (if software such as Google Maps, Networks in Motion's AtlasBook mobile navigation platform, etc. is used).

Some hobbyists have also made some GPS devices and open-sourced the plans. An example is the Elektor GPS units. These are based around a SirFStar 3 chip and are comparable to their commercial counterparts.

## ***Commercial aviation***

Commercial aviation applications include GPS devices that calculate location and feed that information to large multi-input navigational computers for autopilot, course information and correction displays to the pilots, and course tracking and recording devices.

## ***Military***

Military applications include devices similar to consumer sport products for foot soldiers (commanders and regular soldiers), small vehicles and ships, and devices similar to commercial aviation applications for aircraft and missiles. Examples are the US military's **Commander's Digital Assistant** and the **Soldier Digital Assistant**.

WWT

## Chapter 2

# Sonar



French F70 type frigates (here, *La Motte-Picquet*) are fitted with VDS (Variable Depth Sonar) type DUBV43 or DUBV43C towed sonars

**Sonar** (originally an acronym for **SO**und **N**avigation **A**nd **R**anging) is a technique that uses sound propagation (usually underwater, as in Submarine navigation) to navigate, communicate with or detect other vessels. Two types of technology share the name "sonar": *passive* sonar is essentially listening for the sound made by vessels; *active* sonar is emitting pulses of sounds and listening for echoes. Sonar may be used as a means of

acoustic location and of measurement of the echo characteristics of "targets" in the water. Acoustic location in air was used before the introduction of radar. Sonar may also be used in air for robot navigation, and SODAR (an upward looking in-air sonar) is used for atmospheric investigations. The term *sonar* is also used for the equipment used to generate and receive the sound. The acoustic frequencies used in sonar systems vary from very low (infrasonic) to extremely high (ultrasonic). The study of underwater sound is known as underwater acoustics or hydroacoustics.

## **History**

Although some animals (dolphins and bats) have used sound for communication and object detection for millions of years, use by humans in the water is initially recorded by Leonardo Da Vinci in 1490: a tube inserted into the water was said to be used to detect vessels by placing an ear to the tube.

In the 19th century an underwater bell was used as an ancillary to lighthouses to provide warning of hazards.

The use of sound to 'echo locate' underwater in the same way as bats use sound for aerial navigation seems to have been prompted by the *Titanic* disaster of 1912. The world's first patent for an underwater echo ranging device was filed at the British Patent Office by English meteorologist Lewis Richardson a month after the sinking of the *Titanic*, and a German physicist Alexander Behm obtained a patent for an echo sounder in 1913.

The Canadian engineer Reginald Fessenden, while working for the Submarine Signal Company in Boston, built an experimental system beginning in 1912, a system later tested in Boston Harbor, and finally in 1914 from the U.S. Revenue (now Coast Guard) Cutter *Miami* on the Grand Banks off Newfoundland Canada. In that test, Fessenden demonstrated depth sounding, underwater communications (Morse Code) and echo ranging (detecting an iceberg at two miles (3 km) range). The so-called Fessenden oscillator, at ca. 500 Hz frequency, was unable to determine the bearing of the berg due to the 3 metre wavelength and the small dimension of the transducer's radiating face (less than 1 metre in diameter). The ten Montreal-built British H class submarines launched in 1915 were equipped with a Fessenden oscillator.

During World War I the need to detect submarines prompted more research into the use of sound. The British made early use of underwater hydrophones, while the French physicist Paul Langevin, working with a Russian immigrant electrical engineer, Constantin Chilowski, worked on the development of active sound devices for detecting submarines in 1915 using quartz. Although piezoelectric and magnetostrictive transducers later superseded the electrostatic transducers they used, this work influenced future designs. Lightweight sound-sensitive plastic film and fibre optics have been used for hydrophones (acousto-electric transducers for in-water use), while Terfenol-D and PMN (lead magnesium niobate) have been developed for projectors.

## ASDIC

In 1916, under the British Board of Invention and Research, Canadian physicist Robert William Boyle took on the active sound detection project with A B Wood, producing a prototype for testing in mid 1917. This work, for the Anti-Submarine Division of the British Naval Staff, was undertaken in utmost secrecy, and used quartz piezoelectric crystals to produce the world's first practical underwater active sound detection apparatus. To maintain secrecy no mention of sound experimentation or quartz was made - the word used to describe the early work ('supersonics') was changed to 'ASD'ics, and the quartz material to 'ASD'ivite: hence the British acronym *ASDIC*. In 1939, in response to a question from the Oxford English Dictionary, the Admiralty made up the story that it stood for 'Allied Submarine Detection Investigation Committee', and this is still widely believed, though no committee bearing this name has been found in the Admiralty archives.

By 1918, both France and Britain had built prototype active systems. The British tested their ASDIC on HMS *Antrim* in 1920, and started production in 1922. The 6th Destroyer Flotilla had ASDIC-equipped vessels in 1923. An anti-submarine school, HMS *Osprey*, and a training flotilla of four vessels were established on Portland in 1924. The US Sonar QB set arrived in 1931.

By the outbreak of World War II, the Royal Navy had five sets for different surface ship classes, and others for submarines, incorporated into a complete anti-submarine attack system. The effectiveness of early ASDIC was hamstrung by the use of the depth charge as an anti-submarine weapon. This required an attacking vessel to pass over a submerged contact before dropping charges over the stern, resulting in a loss of ASDIC contact in the moments leading up to attack. The hunter was effectively firing blind, during which time a submarine commander could take evasive action. This situation was remedied by using several ships cooperating and by the adoption of "ahead throwing weapons", such as Hedgehog and later Squid, which projected warheads at a target ahead of the attacker and thus still in ASDIC contact. Developments during the war resulted in British ASDIC sets which used several different shapes of beam, continuously covering blind spots. Later, acoustic torpedoes were used.

At the start of World War II, British ASDIC technology was transferred for free to the United States. Research on ASDIC and underwater sound was expanded in the UK and in the US. Many new types of military sound detection were developed. These included sonobuoys, first developed by the British in 1944 under the codename *High Tea*, dipping/dunking sonar and mine detection sonar. This work formed the basis for post war developments related to countering the nuclear submarine. Work on sonar had also been carried out in the Axis countries, notably in Germany, which included countermeasures. At the end of World War II this German work was assimilated by Britain and the US. Sonars have continued to be developed by many countries, including Russia, for both military and civil uses. In recent years the major military development has been the increasing interest in low frequency active systems.

## SONAR

During the 1930s American engineers developed their own underwater sound detection technology and important discoveries were made, such as thermoclines, that would help future development. After technical information was exchanged between the two countries during the Second World War, Americans began to use the term *SONAR* for their systems, coined as the equivalent of RADAR.

### ***Performance factors***

The detection, classification and localisation performance of a sonar depends on the environment and the receiving equipment, as well as the transmitting equipment in an active sonar or the target radiated noise in a passive sonar.

### **Sound propagation**

Sonar operation is affected by variations in sound speed, particularly in the vertical plane. Sound travels more slowly in fresh water than in sea water, though the difference is small. The speed is determined by the water's bulk modulus and mass density. The bulk modulus is affected by temperature, dissolved impurities (usually salinity), and pressure. The density effect is small. The speed of sound (in feet per second) is approximately:

$$4388 + (11.25 \times \text{temperature (in } ^\circ\text{F)}) + (0.0182 \times \text{depth (in feet)}) + \text{salinity (in parts-per-thousand )}.$$

This empirically derived approximation equation is reasonably accurate for normal temperatures, concentrations of salinity and the range of most ocean depths. Ocean temperature varies with depth, but at between 30 and 100 meters there is often a marked change, called the thermocline, dividing the warmer surface water from the cold, still waters that make up the rest of the ocean. This can frustrate sonar, because a sound originating on one side of the thermocline tends to be bent, or refracted, through the thermocline. The thermocline may be present in shallower coastal waters. However, wave action will often mix the water column and eliminate the thermocline. Water pressure also affects sound propagation: higher pressure increases the sound speed, which causes the sound waves to refract away from the area of higher sound speed. The mathematical model of refraction is called Snell's law.

If the sound source is deep and the conditions are right, propagation may occur in the 'deep sound channel'. This provides extremely low propagation loss to a receiver in the channel. This is because of sound trapping in the channel with no losses at the boundaries. Similar propagation can occur in the 'surface duct' under suitable conditions. However in this case there are reflection losses at the surface.

In shallow water propagation is generally by repeated reflection at the surface and bottom, where considerable losses can occur.

Sound propagation is affected by absorption in the water itself as well as at the surface and bottom. This absorption depends upon frequency, with several different mechanisms in sea water. Long-range sonar uses low frequencies to minimise absorption effects.

The sea contains many sources of noise that interfere with the desired target echo or signature. The main noise sources are waves and shipping. The motion of the receiver through the water can also cause speed-dependent low frequency noise.

## Scattering

When active sonar is used, scattering occurs from small objects in the sea as well as from the bottom and surface. This can be a major source of interference. This acoustic scattering is analogous to the scattering of the light from a car's headlights in fog: a high-intensity pencil beam will penetrate the fog to some extent, but broader-beam headlights emit much light in unwanted directions, much of which is scattered back to the observer, overwhelming that reflected from the target ("white-out"). For analogous reasons active sonar needs to transmit in a narrow beam to minimise scattering.

## Target characteristics

The sound *reflection* characteristics of the target of an active sonar, such as a submarine, are known as its target strength. A complication is that echoes are also obtained from other objects in the sea such as whales, wakes, schools of fish and rocks.

Passive sonar detects the target's *radiated* noise characteristics. The radiated spectrum comprises a continuous spectrum of noise with peaks at certain frequencies which can be used for classification.

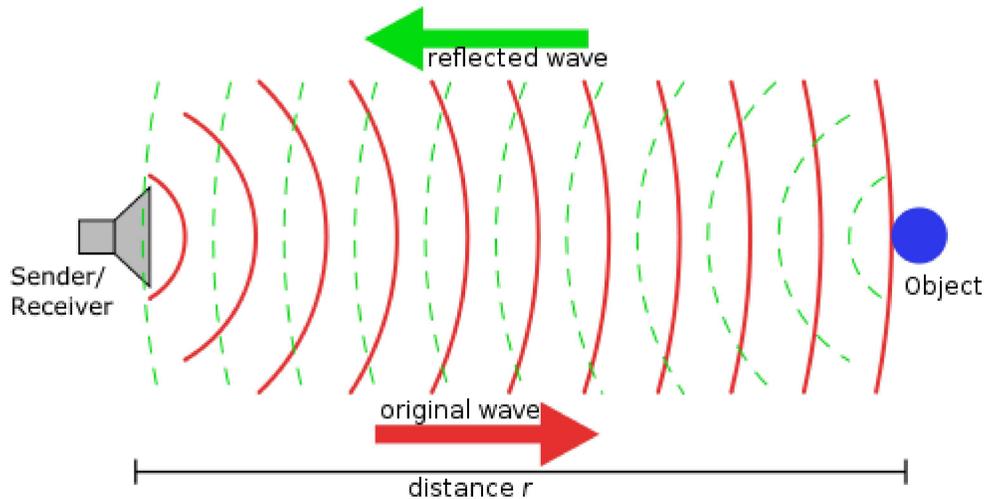
## Countermeasures

*Active* (powered) countermeasures may be launched by a submarine under attack to raise the noise level, provide a large false target, and obscure the signature of the submarine itself.

*Passive* (i.e., non-powered) countermeasures include:

- Mounting noise-generating devices on isolating devices.
- Sound-absorbent coatings on the hulls of submarines, for example anechoic tiles.

## Active sonar



### Principle of an active sonar

Active sonar uses a sound transmitter and a receiver. When the two are in the same place it is monostatic operation. When the transmitter and receiver are separated it is bistatic operation. When more transmitters (or more receivers) are used, again spatially separated, it is multistatic operation. Most sonars are used monostatically with the same array often being used for transmission and reception. Active sonobuoy fields may be operated multistatically.

Active sonar creates a pulse of sound, often called a "ping", and then listens for reflections (echo) of the pulse. This pulse of sound is generally created electronically using a sonar Projector consisting of a signal generator, power amplifier and electro-acoustic transducer/array. A beamformer is usually employed to concentrate the acoustic power into a beam, which may be swept to cover the required search angles. Generally, the electro-acoustic transducers are of the Tonpiliz type and their design may be optimised to achieve maximum efficiency over the widest bandwidth, in order to optimise performance of the overall system. Occasionally, the acoustic pulse may be created by other means, e.g. (1) chemically using explosives, or (2) airguns or (3) plasma sound sources.

To measure the distance to an object, the time from transmission of a pulse to reception is measured and converted into a range by knowing the speed of sound. To measure the bearing, several hydrophones are used, and the set measures the relative arrival time to each, or with an array of hydrophones, by measuring the relative amplitude in beams formed through a process called beamforming. Use of an array reduces the spatial response so that to provide wide cover multibeam systems are used. The target signal (if present) together with noise is then passed through various forms of signal processing, which for simple sonars may be just energy measurement. It is then presented to some form of decision device that calls the output either the required signal or noise. This decision device may be an operator with headphones or a display, or in more

sophisticated sonars this function may be carried out by software. Further processes may be carried out to classify the target and localise it, as well as measuring its velocity.

The pulse may be at constant frequency or a chirp of changing frequency (to allow pulse compression on reception). Simple sonars generally use the former with a filter wide enough to cover possible Doppler changes due to target movement, while more complex ones generally include the latter technique. Since digital processing became available pulse compression has usually been implemented using digital correlation techniques. Military sonars often have multiple beams to provide all-round cover while simple ones only cover a narrow arc, although the beam may be rotated, relatively slowly, by mechanical scanning.

Particularly when single frequency transmissions are used, the Doppler effect can be used to measure the radial speed of a target. The difference in frequency between the transmitted and received signal is measured and converted into a velocity. Since Doppler shifts can be introduced by either receiver or target motion, allowance has to be made for the radial speed of the searching platform.

One useful small sonar is similar in appearance to a waterproof flashlight. The head is pointed into the water, a button is pressed, and the device displays the distance to the target. Another variant is a "fishfinder" that shows a small display with shoals of fish. Some civilian sonars (which are not designed for stealth) approach active military sonars in capability, with quite exotic three-dimensional displays of the area near the boat.

When active sonar is used to measure the distance from the transducer to the bottom, it is known as echo sounding. Similar methods may be used looking upward for wave measurement.

Active sonar is also used to measure distance through water between two sonar transducers or a combination of a hydrophone (underwater acoustic microphone) and projector (underwater acoustic speaker). A transducer is a device that can transmit and receive acoustic signals ("pings"). When a hydrophone/transducer receives a specific interrogation signal it responds by transmitting a specific reply signal. To measure distance, one transducer/projector transmits an interrogation signal and measures the time between this transmission and the receipt of the other transducer/hydrophone reply. The time difference, scaled by the speed of sound through water and divided by two, is the distance between the two platforms. This technique, when used with multiple transducers/hydrophones/projectors, can calculate the relative positions of static and moving objects in water.

In combat situations, an active pulse can be detected by an opponent and will reveal a submarine's position.

A very directional, but low-efficiency, type of sonar (used by fisheries, military, and for port security) makes use of a complex nonlinear feature of water known as non-linear sonar, the virtual transducer being known as a *parametric array*.

## Project ARTEMIS

Project ARTEMIS was a one-of-a-kind low-frequency sonar for surveillance that was deployed off Bermuda for several years in the early 1960s. The active portion was deployed from a World War II tanker, and the receiving array was a built into a fixed position on an offshore bank.

## Transponder

This is an active sonar device that receives a stimulus and immediately (or with a delay) retransmits the received signal or a predetermined one.

## Performance prediction

A sonar target is small relative to the sphere, centred around the emitter, on which it is located. Therefore, the power of the reflected signal is very low, several orders of magnitude less than the original signal. Even if the reflected signal was of the same power, the following example (using hypothetical values) shows the problem: Suppose a sonar system is capable of emitting a  $10,000 \text{ W/m}^2$  signal at 1 m, and detecting a  $0.001 \text{ W/m}^2$  signal. At 100 m the signal will be  $1 \text{ W/m}^2$  (due to the inverse-square law). If the entire signal is reflected from a  $10 \text{ m}^2$  target, it will be at  $0.001 \text{ W/m}^2$  when it reaches the emitter, i.e. just detectable. However, the original signal will remain above  $0.001 \text{ W/m}^2$  until 300 m. Any  $10 \text{ m}^2$  target between 100 and 300 m using a similar or better system would be able to detect the pulse but would not be detected by the emitter. The detectors must be very sensitive to pick up the echoes. Since the original signal is much more powerful, it can be detected many times further than twice the range of the sonar (as in the example).

In active sonar there are two performance limitations, due to noise and reverberation. In general one or other of these will dominate so that the two effects can be initially considered separately.

In noise limited conditions at initial detection:

$$SL - 2TL + TS - (NL - DI) = DT$$

where SL is the source level, TL is the transmission loss (or propagation loss), TS is the target strength, NL is the noise level, DI is the directivity index of the array (an approximation to the array gain) and DT is the detection threshold.

In reverberation limited conditions at initial detection (neglecting array gain):

$$SL - 2TL + TS = RL + DT$$

where RL is the reverberation level and the other factors are as before.

## Marine mammals



A Humpback whale

Active sonar may harm marine animals, although the precise mechanisms for this are not well understood. Some marine animals, such as whales and dolphins, use echolocation systems, sometimes called *biosonar* to locate predators and prey. It is conjectured that active sonar transmitters could confuse these animals and interfere with basic biological functions such as feeding and mating.

## Hand-held sonar for use by a diver



Scuba diver using INSS hand-held sonar

- The LIMIS (= Limpet Mine Imaging Sonar) is a hand-held or ROV-mounted imaging sonar for use by a diver. Its name is because it was designed for patrol divers (combat frogmen or Clearance Divers) to look for limpet mines in low visibility water.

- Abstract of article by the International Society for Optical Engineering
  - Used to find debris from the Space Shuttle Columbia crash
  - Used in fish passage research at hydropower facilities
- The LUIS (= Lensing Underwater Imaging System) is another imaging sonar for use by a diver. Links:
  - Used for counting salmon in a river
- There is or was a small flashlight-shaped handheld sonar for divers, that merely displays range.
- For the INSS = Integrated Navigation Sonar System see:
  - an image.
  - short description
  - description

## ***Passive sonar***

Passive sonar listens without transmitting. It is often employed in military settings, although it is also used in science applications, *e.g.*, detecting fish for presence/absence studies in various aquatic environments. In the very broadest usage, this term can encompass virtually any analytical technique involving remotely generated sound, though it is usually restricted to techniques applied in an aquatic environment.

### **Identifying sound sources**

Passive sonar has a wide variety of techniques for identifying the source of a detected sound. For example, U.S. vessels usually operate 60 Hz alternating current power systems. If transformers or generators are mounted without proper vibration insulation from the hull or become flooded, the 60 Hz sound from the windings can be emitted from the submarine or ship. This can help to identify its nationality, as most European submarines have 50 Hz power systems. Intermittent sound sources (such as a wrench being dropped) may also be detectable to passive sonar. Until fairly recently, an experienced trained operator identified signals, but now computers may do this.

Passive sonar systems may have large sonic databases, but the sonar operator usually finally classifies the signals manually. A computer system frequently uses these databases to identify classes of ships, actions (*i.e.* the speed of a ship, or the type of weapon released), and even particular ships. Publications for classification of sounds are provided by and continually updated by the US Office of Naval Intelligence.

### **Noise limitations**

Passive sonar on vehicles is usually severely limited because of noise generated by the vehicle. For this reason, many submarines operate nuclear reactors that can be cooled without pumps, using silent convection, or fuel cells or batteries, which can also run silently. Vehicles' propellers are also designed and precisely machined to emit minimal noise. High-speed propellers often create tiny bubbles in the water, and this cavitation has a distinct sound.

The sonar hydrophones may be towed behind the ship or submarine in order to reduce the effect of noise generated by the watercraft itself. Towed units also combat the thermocline, as the unit may be towed above or below the thermocline.

The display of most passive sonars used to be a two-dimensional waterfall display. The horizontal direction of the display is bearing. The vertical is frequency, or sometimes time. Another display technique is to color-code frequency-time information for bearing. More recent displays are generated by the computers, and mimic radar-type plan position indicator displays.

## **Performance prediction**

Unlike active sonar, only one way propagation is involved. Because of the different signal processing used, the minimum detectable signal to noise ratio will be different. The equation for determining the performance of a passive sonar is:

$$SL - TL = NL - DI + DT$$

where SL is the source level, TL is the transmission loss, NL is the noise level, DI is the directivity index of the array (an approximation to the array gain) and DT is the detection threshold. The figure of merit of a passive sonar is:

$$FOM = SL + DI - (NL + DT).$$

## **Warfare**

Modern naval warfare makes extensive use of both passive and active sonar from waterborne vessels, aircraft and fixed installations. The relative usefulness of active versus passive sonar depends on the radiated noise characteristics of the target, generally a submarine. Although in World War II active sonar was used by surface craft—submarines avoided emitting pings which revealed their presence and position—with the advent of modern signal-processing passive sonar became preferred for initial detection. Submarines were then designed for quieter operation, and active sonar is now more used. In 1987 a division of Japanese company Toshiba reportedly sold machinery to the Soviet Union that allowed it to mill submarine propeller blades so that they became radically quieter, creating a huge security issue with their newer generation of submarines.

Active sonar gives the exact bearing to a target, and sometimes the range. Active sonar works the same way as radar: a signal is emitted. The sound wave then travels in many directions from the emitting object. When it hits an object, the sound wave is then reflected in many other directions. Some of the energy will travel back to the emitting source. The echo will enable the sonar system or technician to calculate, with many factors such as the frequency, the energy of the received signal, the depth, the water temperature, the position of the reflecting object, etc. Active sonar is used when the platform commander determines that it is more important to determine the position of a possible threat submarine than it is to conceal his own position. With surface ships it

might be assumed that the threat is already tracking the ship with satellite data. Any vessel around the emitting sonar will detect the emission. Having heard the signal, it is easy to identify the sonar equipment used (usually with its frequency) and its position (with the sound wave's energy). Active sonar is similar to radar in that, while it allows detection of targets at a certain range, it also enables the emitter to be detected at a far greater range, which is undesirable.

Since active sonar reveals the presence and position of the operator, and does not allow exact classification of targets, it is used by fast (planes, helicopters) and by noisy platforms (most surface ships) but rarely by submarines. When active sonar is used by surface ships or submarines, it is typically activated very briefly at intermittent periods to minimise the risk of detection. Consequently active sonar is normally considered a backup to passive sonar. In aircraft, active sonar is used in the form of disposable sonobuoys that are dropped in the aircraft's patrol area or in the vicinity of possible enemy sonar contacts.

Passive sonar has several advantages. Most importantly, it is silent. If the target radiated noise level is high enough, it can have a greater range than active sonar, and allows the target to be identified. Since any motorized object makes some noise, it may in principle be detected, depending on the level of noise emitted and the ambient noise level in the area, as well as the technology used. To simplify, passive sonar "sees" around the ship using it. On a submarine, nose-mounted passive sonar detects in directions of about 270°, centered on the ship's alignment, the hull-mounted array of about 160° on each side, and the towed array of a full 360°. The invisible areas are due to the ship's own interference. Once a signal is detected in a certain direction (which means that something makes sound in that direction, this is called broadband detection) it is possible to zoom in and analyze the signal received (narrowband analysis). This is generally done using a Fourier transform to show the different frequencies making up the sound. Since every engine makes a specific sound, it is straightforward to identify the object. Databases of unique engine sounds are part of what is known as *acoustic intelligence* or ACINT.

Another use of passive sonar is to determine the target's trajectory. This process is called Target Motion Analysis (TMA), and the resultant "solution" is the target's range, course, and speed. TMA is done by marking from which direction the sound comes at different times, and comparing the motion with that of the operator's own ship. Changes in relative motion are analyzed using standard geometrical techniques along with some assumptions about limiting cases.

Passive sonar is stealthy and very useful. However, it requires high-tech electronic components and is costly. It is generally deployed on expensive ships in the form of arrays to enhance detection. Surface ships use it to good effect; it is even better used by submarines, and it is also used by airplanes and helicopters, mostly to a "surprise effect", since submarines can hide under thermal layers. If a submarine's commander believes he is alone, he may bring his boat closer to the surface and be easier to detect, or go deeper and faster, and thus make more sound.

Examples of sonar applications in military use are given below. Many of the civil uses given in the following section may also be applicable to naval use.

### **Anti-submarine warfare**



Variable Depth Sonar and its winch

Until recently, ship sonars were usually with hull mounted arrays, either amidships or at the bow. It was soon found after their initial use that a means of reducing flow noise was required. The first were made of canvas on a framework, then steel ones were used. Now domes are usually made of reinforced plastic or pressurised rubber. Such sonars are primarily active in operation. An example of a conventional hull mounted sonar is the SQS-56.

Because of the problems of ship noise, towed sonars are also used. These also have the advantage of being able to be placed deeper in the water. However, there are limitations on their use in shallow water. These are called towed arrays (linear) or variable depth sonars (VDS) with 2/3D arrays. A problem is that the winches required to deploy/recover these are large and expensive. VDS sets are primarily active in operation while towed arrays are passive.

An example of a modern active/passive ship towed sonar is Sonar 2087 made by Thales Underwater Systems.

## **Torpedoes**

Modern torpedoes are generally fitted with an active/passive sonar. This may be used to home directly on the target, but wake following torpedoes are also used. An early example of an acoustic homer was the Mark 37 torpedo.

Torpedo countermeasures can be towed or free. An early example was the German Sieglinde device while the Pillenwerfer was a chemical device. A widely used US device was the towed Nixie while MOSS submarine simulator was a free device. A modern alternative to the Nixie system is the UK Royal Navy S2170 Surface Ship Torpedo Defence system.

## **Mines**

Mines may be fitted with a sonar to detect, localize and recognize the required target.

## **Mine countermeasures**

Mine Countermeasure (MCM) Sonar, sometimes called "Mine and Obstacle Avoidance Sonar (MOAS)", is a specialised type of sonar used for detecting small objects. Most MCM sonars are hull mounted but a few types are VDS design. An example of a hull mounted MCM sonar is the Type 2193 while the SQQ-32 Mine-hunting sonar and Type 2093 systems are VDS designs.

## **Submarine navigation**

Submarines rely on sonar to a greater extent than surface ships as they cannot use radar at depth. The sonar arrays may be hull mounted or towed. Information fitted on typical fits is given in Oyashio class submarine and Swiftsure class submarine.

## **Aircraft**

Helicopters can be used for antisubmarine warfare by deploying fields of active/passive sonobuoys or can operate dipping sonar, such as the AQS-13. Fixed wing aircraft can also deploy sonobuoys and have greater endurance and capacity to deploy them. Processing from the sonobuoys or dipping sonar can be on the aircraft or on ship.

Helicopters have also been used for mine countermeasure missions using towed sonars such as the AQS-20A



AN/AQS-13 Dipping sonar deployed from an H-3 Sea King.

## **Underwater communications**

Dedicated sonars can be fitted to ships and submarines for underwater communication.

## **Ocean surveillance**

For many years, the United States operated a large set of passive sonar arrays at various points in the world's oceans, collectively called Sound Surveillance System (SOSUS) and later Integrated Undersea Surveillance System (IUSS). A similar system is believed to have been operated by the Soviet Union. As permanently mounted arrays in the deep ocean were utilised, they were in very quiet conditions so long ranges could be achieved. Signal processing was carried out using powerful computers ashore. With the ending of the Cold War a SOSUS array has been turned over to scientific use.

In the United States Navy, a special badge known as the Integrated Undersea Surveillance System Badge is awarded to those who have been trained and qualified in its operation.

## **Underwater security**

Sonar can be used to detect frogmen and other scuba divers. This can be applicable around ships or at entrances to ports. Active sonar can also be used as a deterrent and/or disablement mechanism. One such device is the Cerberus system.

## **Hand-held sonar**

Limpet Mine Imaging Sonar (LIMIS) is a hand-held or ROV-mounted imaging sonar designed for patrol divers (combat frogmen or clearance divers) to look for limpet mines in low visibility water.

The LUIS is another imaging sonar for use by a diver.

Integrated Navigation Sonar System (INSS) is a small flashlight-shaped handheld sonar for divers that displays range.

## **Intercept sonar**

This is a sonar designed to detect and locate the transmissions from hostile active sonars. An example of this is the Type 2082 fitted on the British Vanguard class submarines.

## ***Civilian applications***

### **Fisheries**

Fishing is an important industry that is seeing growing demand, but world catch tonnage is falling as a result of serious resource problems. The industry faces a future of continuing worldwide consolidation until a point of sustainability can be reached. However, the consolidation of the fishing fleets are driving increased demands for sophisticated fish finding electronics such as sensors, sounders and sonars. Historically, fishermen have used many different techniques to find and harvest fish. However, acoustic technology has been one of the most important driving forces behind the development of the modern commercial fisheries.

Sound waves travel differently through fish than through water because a fish's air-filled swim bladder has a different density than seawater. This density difference allows the detection of schools of fish by using reflected sound. Acoustic technology is especially well suited for underwater applications since sound travels farther and faster underwater than in air. Today, commercial fishing vessels rely almost completely on acoustic sonar and sounders to detect fish. Fishermen also use active sonar and echo sounder technology to determine water depth, bottom contour, and bottom composition.



Cabin display of a fish finder sonar

Companies such as Raymarine UK, Marport Canada, Wesmar, Furuno, Krupp, and Simrad make a variety of sonar and acoustic instruments for the deep sea commercial fishing industry. For example, net sensors take various underwater measurements and transmit the information back to a receiver onboard a vessel. Each sensor is equipped with one or more acoustic transducers depending on its specific function. Data is transmitted from the sensors using wireless acoustic telemetry and is received by a hull mounted hydrophone. The analog signals are decoded and converted by a digital acoustic receiver into data which is transmitted to a bridge computer for graphical display on a high resolution monitor.

### **Echo sounding**

An echo-sounder sends an acoustic pulse directly downwards to the seabed and records the returned echo. The sound pulse is generated by a transducer that emits an acoustic pulse and then “listens” for the return signal. The time for the signal to return is recorded and converted to a depth measurement by calculating the speed of sound in water. As the speed of sound in water is around 1,500 metres per second, the time interval, measured in milliseconds, between the pulse being transmitted and the echo being received, allows bottom depth and targets to be measured.

The value of underwater acoustics to the fishing industry has led to the development of other acoustic instruments that operate in a similar fashion to echo-sounders but, because their function is slightly different from the initial model of the echo-sounder, have been given different terms.

## **Net location**

The net sounder is an echo sounder with a transducer mounted on the headline of the net rather than on the bottom of the vessel. Nevertheless, to accommodate the distance from the transducer to the display unit, which is much greater than in a normal echo-sounder, several refinements have to be made. Two main types are available. The first is the cable type in which the signals are sent along a cable. In this case there has to be the provision of a cable drum on which to haul, shoot and stow the cable during the different phases of the operation. The second type is the cable less net-sounder – such as Marport's Trawl Explorer - in which the signals are sent acoustically between the net and hull mounted receiver/hydrophone on the vessel. In this case no cable drum is required but sophisticated electronics are needed at the transducer and receiver.

The display on a net sounder shows the distance of the net from the bottom (or the surface), rather than the depth of water as with the echo-sounder's hull-mounted transducer. Fixed to the headline of the net, the footrope can usually be seen which gives an indication of the net performance. Any fish passing into the net can also be seen, allowing fine adjustments to be made to catch the most fish possible. In other fisheries, where the amount of fish in the net is important, catch sensor transducers are mounted at various positions on the cod-end of the net. As the cod-end fills up these catch sensor transducers are triggered one by one and this information is transmitted acoustically to display monitors on the bridge of the vessel. The skipper can then decide when to haul the net.

Modern versions of the net sounder, using multiple element transducers, function more like a sonar than an echo sounder and show slices of the area in front of the net and not merely the vertical view that the initial net sounders used.

The sonar is an echo-sounder with a directional capability that can show fish or other objects around the vessel.

## **Ship velocity measurement**

Sonars have been developed for measuring a ship's velocity either relative to the water or to the bottom.

## **ROV and UUV**

Small sonars have been fitted to Remotely Operated Vehicles (ROV) and Unmanned Underwater Vehicles (UUV) to allow their operation in murky conditions. These sonars are used for looking ahead of the vehicle. The Long-Term Mine Reconnaissance System is an UUV for MCM purposes.

## **Vehicle location**

Sonars which act as beacons are fitted to aircraft to allow their location in the event of a crash in the sea. Short and Long Baseline sonars may be used for carrying out the location, such as LBL.

## ***Scientific applications***

### **Biomass estimation**

Detection of fish, and other marine and aquatic life, and estimation their individual sizes or total biomass using active sonar techniques. As the sound pulse travels through water it encounters objects that are of different density or acoustic characteristics than the surrounding medium, such as fish, that reflect sound back toward the sound source. These echoes provide information on fish size, location, abundance and behavior. Data is usually processed and analysed using a variety of software such as Echoview.

### **Wave measurement**

An upward looking echo sounder mounted on the bottom or on a platform may be used to make measurements of wave height and period. From this statistics of the surface conditions at a location can be derived.

### **Water velocity measurement**

Special short range sonars have been developed to allow measurements of water velocity.

### **Bottom type assessment**

Sonars have been developed that can be used to characterise the sea bottom into, for example, mud, sand, and gravel. Relatively simple sonars such as echo sounders can be promoted to seafloor classification systems via add-on modules, converting echo parameters into sediment type. Different algorithms exist, but they are all based on changes in the energy or shape of the reflected sounder pings. Advanced substrate classification analysis can be achieved using calibrated (scientific) echosounders and parametric or fuzzy-logic analysis of the acoustic data (See: Acoustic Seabed Classification)

### **Bottom topography measurement**

Side-scan sonars can be used to derive maps of the topography of an area by moving the sonar across it just above the bottom. Low frequency sonars such as GLORIA have been used for continental shelf wide surveys while high frequency sonars are used for more detailed surveys of smaller areas.

## **Sub-bottom profiling**

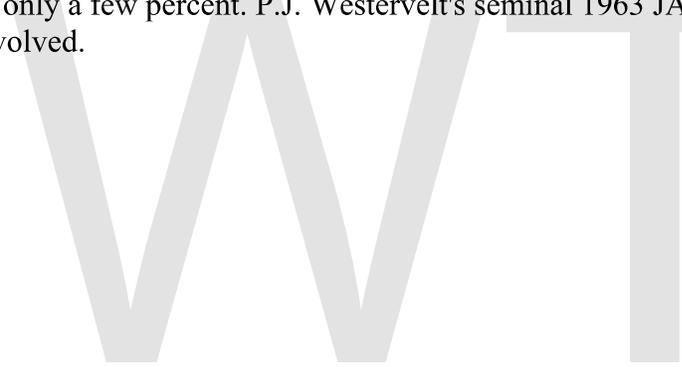
Powerful low frequency echo-sounders have been developed for providing profiles of the upper layers of the ocean bottom.

## **Synthetic aperture sonar**

Various synthetic aperture sonars have been built in the laboratory and some have entered use in mine-hunting and search systems. An explanation of their operation is given in synthetic aperture sonar.

## **Parametric sonar**

Parametric sources use the non-linearity of water to generate the difference frequency between two high frequencies. A virtual end-fire array is formed. Such a projector has advantages of broad bandwidth, narrow beamwidth, and when fully developed and carefully measured it has no obvious sidelobes. Its major disadvantage is very low efficiency of only a few percent. P.J. Westervelt's seminal 1963 JASA paper summarizes the trends involved.



## Chapter 3

# Anemometer



A hemispherical cup anemometer of the type invented in 1846 by John Thomas Romney Robinson



Cup-type anemometer with vertical axis, a sensor on a remote meteorological station deployed on Skagit Bay, Washington July–August, 2009.

An **anemometer** is a device for measuring wind speed, and is a common weather station instrument. The term is derived from the Greek word *anemos*, meaning wind. The first known description of an anemometer was given by Leon Battista Alberti around 1450. They are also very easy to make as a project.

Anemometers can be divided into two classes: those that measure the wind's speed, and those that measure the wind's pressure; but as there is a close connection between the pressure and the speed, an anemometer designed for one will give information about both.

### ***Velocity anemometers***

#### **Cup anemometers**

A simple type of anemometer, invented (1846) by Dr. John Thomas Romney Robinson, of Armagh Observatory. It consisted of four hemispherical cups each mounted on one end of four horizontal arms, which in turn were mounted at equal angles to each other on a vertical shaft. The air flow past the cups in any horizontal direction turned the cups in a manner that was proportional to the wind speed. Therefore, counting the turns of the cups over a set time period produced the average wind speed for a wide range of speeds. On an anemometer with four cups it is easy to see that since the cups are arranged

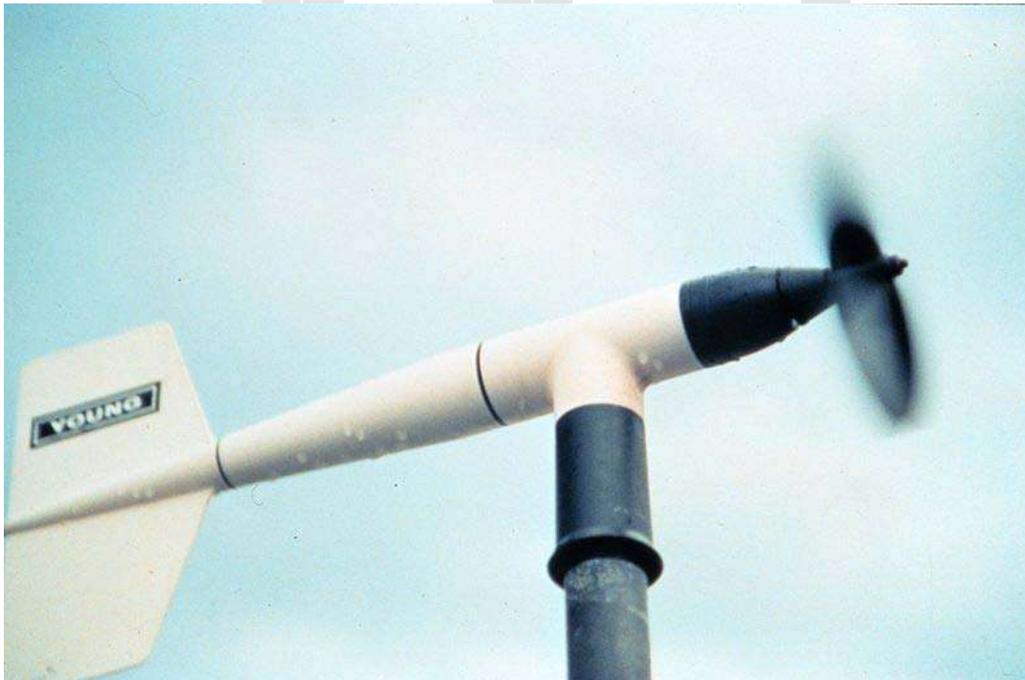
symmetrically on the end of the arms, the wind always has the hollow of one cup presented to it and is blowing on the back of the cup on the opposite end of the cross.

When Robinson first designed his anemometer, he asserted that the cups moved one-third of the speed of the wind, unaffected by the cup size or arm length. This was apparently confirmed by some early independent experiments, but it was incorrect. Instead, the ratio of the speed of the wind and that of the cups, the anemometer factor, depends on the dimensions of the cups and arms, and may have a value between two and a little over three. Every experiment involving an anemometer had to be repeated.

The three cup anemometer developed by the Canadian John Patterson in 1926 and subsequent cup improvements by Brevoort & Joiner of the USA in 1935 led to a cupwheel design which was linear and had an error of less than 3% up to 60 mph (97 km/h). Patterson found that each cup produced maximum torque when it was at 45 degrees to the wind flow. The three cup anemometer also had a more constant torque and responded more quickly to gusts than the four cup anemometer.

The three cup anemometer was further modified by the Australian Derek Weston in 1991 to measure both wind direction and wind speed. Weston added a tag to one cup, which causes the cupwheel speed to increase and decrease as the tag moves alternately with and against the wind. Wind direction is calculated from these cyclical changes in cupwheel speed, while wind speed is as usual determined from the average cupwheel speed.

Three cup anemometers are currently used as the industry standard for wind resource assessment studies.

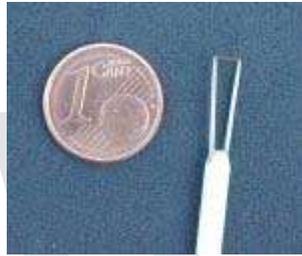


A windmill style of anemometer

## Windmill anemometers

The other forms of mechanical velocity anemometer may be described as belonging to the windmill type or propeller anemometer. In the Robinson anemometer the axis of rotation is vertical, but with this subdivision the axis of rotation must be parallel to the direction of the wind and therefore horizontal. Furthermore, since the wind varies in direction and the axis has to follow its changes, a wind vane or some other contrivance to fulfil the same purpose must be employed. An *aerovane* combines a propeller and a tail on the same axis to obtain accurate and precise wind speed and direction measurements from the same instrument. In cases where the direction of the air motion is always the same, as in the ventilating shafts of mines and buildings for instance, wind vanes, known as air meters are employed, and give most satisfactory results.

## Hot-wire anemometers



Hot-wire sensor

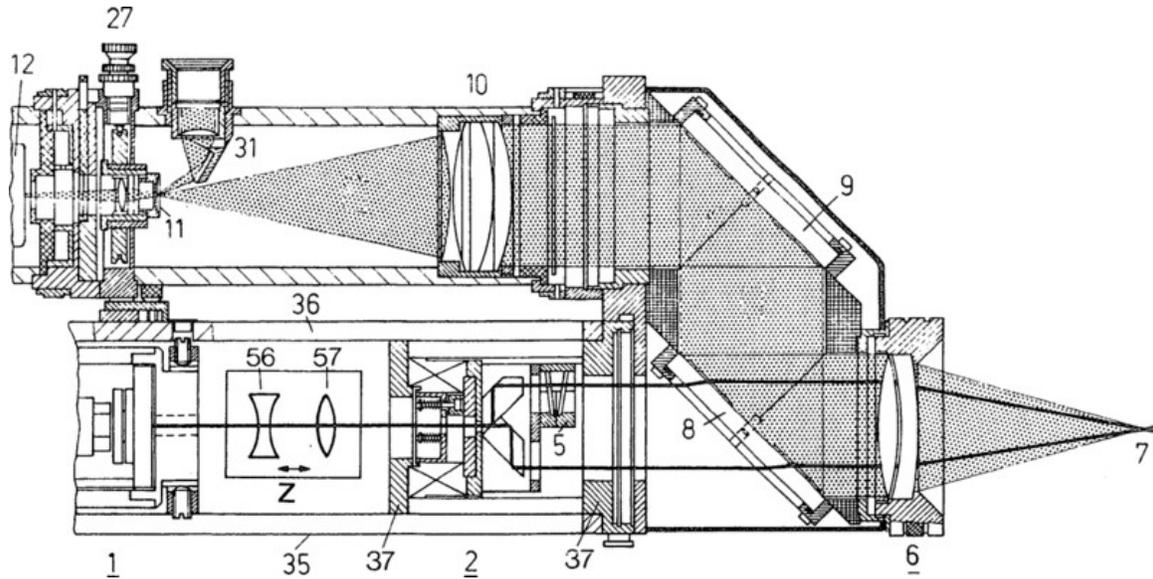
Hot wire anemometers use a very fine wire (on the order of several micrometres) electrically heated up to some temperature above the ambient. Air flowing past the wire has a cooling effect on the wire. As the electrical resistance of most metals is dependent upon the temperature of the metal (tungsten is a popular choice for hot-wires), a relationship can be obtained between the resistance of the wire and the flow speed.

Several ways of implementing this exist, and hot-wire devices can be further classified as CCA (Constant-Current Anemometer), CVA (Constant-Voltage Anemometer) and CTA (Constant-Temperature Anemometer). The voltage output from these anemometers is thus the result of some sort of circuit within the device trying to maintain the specific variable (current, voltage or temperature) constant.

Additionally, PWM (pulse-width modulation) anemometers are also used, wherein the velocity is inferred by the time length of a repeating pulse of current that brings the wire up to a specified resistance and then stops until a threshold "floor" is reached, at which time the pulse is sent again.

Hot-wire anemometers, while extremely delicate, have extremely high frequency-response and fine spatial resolution compared to other measurement methods, and as such are almost universally employed for the detailed study of turbulent flows, or any flow in which rapid velocity fluctuations are of interest.

## Laser Doppler anemometers



Drawing of a laser anemometer. The laser is emitted (1) through the front lens (6) of the anemometer and is backscattered off the air molecules (7). The backscattered radiation (dots) re-enter the device and are reflected and directed into a detector (12).

Laser Doppler anemometers use a beam of light from a laser that is divided into two beams, with one propagated out of the anemometer. Particulates (or deliberately introduced seed material) flowing along with air molecules near where the beam exits reflect, or backscatter, the light back into a detector, where it is measured relative to the original laser beam. When the particles are in great motion, they produce a Doppler shift for measuring wind speed in the laser light, which is used to calculate the speed of the particles, and therefore the air around the anemometer.

## Sonic anemometers



3D ultrasonic anemometer

Sonic anemometers, first developed in the 1970s, use ultrasonic sound waves to measure wind velocity. They measure wind speed based on the time of flight of sonic pulses between pairs of transducers. Measurements from pairs of transducers can be combined to yield a measurement of velocity in 1-, 2-, or 3-dimensional flow. The spatial resolution is given by the path length between transducers, which is typically 10 to 20 cm. Sonic anemometers can take measurements with very fine temporal resolution, 20 Hz or better, which makes them well suited for turbulence measurements. The lack of moving parts makes them appropriate for long term use in exposed automated weather stations and weather buoys where the accuracy and reliability of traditional cup-and-vane anemometers is adversely affected by salty air or large amounts of dust. Their main disadvantage is the distortion of the flow itself by the structure supporting the transducers, which requires a correction based upon wind tunnel measurements to minimize the effect. An international standard for this process, ISO 16622 *Meteorology—Sonic anemometers/thermometers—Acceptance test methods for mean wind measurements* is in general circulation. Another disadvantage is lower accuracy due to precipitation, where rain drops may vary the speed of sound.

Since the speed of sound varies with temperature, and is virtually stable with pressure change, sonic anemometers are also used as thermometers.

Two-dimensional (wind speed and wind direction) sonic anemometers are used in applications such as weather stations, ship navigation, wind turbines, aviation and weather buoys.

### **Ping-pong ball anemometers**

A common anemometer for basic use is constructed from a ping-pong ball attached to a string. When the wind blows horizontally, it presses on and moves the ball; because ping-pong balls are very lightweight, they move easily in light winds. Measuring the angle between the string-ball apparatus and the line normal to the ground gives an estimate of the wind speed.

This type of anemometer is mostly used for middle-school level instruction which most students make themselves, but a similar device was also flown on Phoenix Mars Lander .

### **Pressure anemometers**

The first designs of anemometers which measure the pressure were divided into plate and tube classes.

#### **Plate anemometers**

These are the earliest anemometers and are simply a flat plate suspended from the top so that the wind deflects the plate. In 1450, the Italian art architect Leon Battista Alberti invented the first mechanical anemometer; in 1664 it was re-invented by Robert Hooke (who is often mistakenly considered the inventor of the first anemometer). Later versions

of this form consisted of a flat plate, either square or circular, which is kept normal to the wind by a wind vane. The pressure of the wind on its face is balanced by a spring. The compression of the spring determines the actual force which the wind is exerting on the plate, and this is either read off on a suitable gauge, or on a recorder. Instruments of this kind do not respond to light winds, are inaccurate for high wind readings, and are slow at responding to variable winds. Plate anemometers have been used to trigger high wind alarms on bridges.

### **Tube anemometers**



Helicoid propeller anemometer incorporating a wind vane for orientation.

James Lind's anemometer of 1775 consisted simply of a glass U tube containing liquid, a manometer, with one end bent in a horizontal direction to face the wind and the other

vertical end remains parallel to the wind flow. Though the Lind was not the first it was the most practical and best known anemometer of this type. If the wind blows into the mouth of a tube it causes an increase of pressure on one side of the manometer. The wind over the open end of a vertical tube causes little change in pressure on the other side of the manometer. The resulting liquid change in the U tube is an indication of the wind speed. Small departures from the true direction of the wind causes large variations in the magnitude.

The highly successful metal pressure tube anemometer of William Henry Dines in 1892 utilized the same pressure difference between the open mouth of a straight tube facing the wind and a ring of small holes in a vertical tube which is closed at the upper end. Both are mounted at the same height. The pressure differences on which the action depends are very small, and special means are required to register them. The recorder consists of a float in a sealed chamber partially filled with water. The pipe from the straight tube is connected to the top of the sealed chamber and the pipe from the small tubes is directed into the bottom inside the float. Since the pressure difference determines the vertical position of the float this is a measure of the wind speed.

The great advantage of the tube anemometer lies in the fact that the exposed part can be mounted on a high pole, and requires no oiling or attention for years; and the registering part can be placed in any convenient position. Two connecting tubes are required. It might appear at first sight as though one connection would serve, but the differences in pressure on which these instruments depend are so minute, that the pressure of the air in the room where the recording part is placed has to be considered. Thus if the instrument depends on the pressure or suction effect alone, and this pressure or suction is measured against the air pressure in an ordinary room, in which the doors and windows are carefully closed and a newspaper is then burnt up the chimney, an effect may be produced equal to a wind of 10 mi/h (16 km/h); and the opening of a window in rough weather, or the opening of a door, may entirely alter the registration.

While the Dines anemometer had an error of only 1% at 10 mph (16 km/h) it did not respond very well to low winds due to the poor response of the flat plate vane required to turn the head into the wind. In 1918 an aerodynamic vane with eight times the torque of the flat plate overcame this problem.

### **Effect of density on measurements**

In the tube anemometer the pressure is measured, although the scale is usually graduated as a velocity scale. In cases where the density of the air is significantly different from the calibration value (as on a high mountain, or with an exceptionally low barometer) an allowance must be made. Approximately 1½% should be added to the velocity recorded by a tube anemometer for each 1000 ft (5% for each kilometer) above sea-level.

## Chapter 4

# Astrocompass and Flight Instruments

## Astrocompass



Astrocompass mk. II

An **astrocompass** is a navigational tool for determining the direction of north through the positions of various astronomical bodies.

There are certain circumstances when magnetic compasses and gyrocompasses are unreliable. The most obvious is in polar regions, where the force exerted on the needle of a magnetic compass is nearly vertical and gyrocompasses become unstable due to the rotation of the Earth. Magnetic compasses are also particularly susceptible to magnetic fields, such as those produced by the hulls of some metal vehicles or craft. Before the advent of electronic navigational aids such as GPS the most reliable way to ascertain north in such circumstances was through the use of an astrocompass.

### ***Principle of use***

The Earth's axis of rotation remains, for all intents and purposes, stationary throughout the year. Thus, with knowledge of the current time and geographical position in the form of latitude and longitude, which are set on the instrument using dials, an astrocompass can be sighted on to any astronomical object with a known position to give an extremely accurate reading.

In its most basic form, the astrocompass consists of a base plate marked with the points of the compass, with a mechanism known as an equatorial drum mounted on it. On this drum is a set of adjustable sights and a scale of declination. More advanced versions may have built-in chronometers or default settings for bodies such as the Sun.

To use the compass, the base plate is first levelled with the horizon then pointed roughly to what the user believes to be north. The equatorial drum is then tilted in relation to this base according to the local latitude. The sights are then set using the local hour angle and the declination of whatever astronomical body is being used. Once all these settings have been made, the astrocompass is simply turned until the astronomical body is visible in the sights: it will then be precisely aligned to the points of the compass. Because of this procedure, an astrocompass requires its user to be in possession of a nautical almanac or similar astronomical tables, one of its chief disadvantages.

### ***Historical uses***

Astrocompasses only became useful following the invention of the marine chronometer, without which it is almost useless for navigation. Even then, they saw only limited use, with first magnetic compasses and then gyrocompasses being preferred in almost all cases. Polar exploration was one of the fields in which the astrocompass saw the most use, for the reasons described above. They have also been used throughout history in other climes to check the accuracy of other forms of compasses: they saw use, for example, in the North African Campaign of World War 2.

GPS and other similar forms of electronic navigation aids mean that the astrocompass is now functionally obsolete anywhere except for areas very close to the poles where GPS coverage is not available and there are no current electronic navigation aids.

## Operation

The operation is as follows:

1. Getting the local time (by means of a clock)
2. Setting the latitude
3. Setting (in the alidade) the star's LHA for that day obtained by means of tables or a computer program
4. Pointing the compass pinnula to the star (sun or moon)
5. Reading the true course on the limb

## Flight instruments



The cockpit of a Slingsby T-67 Firefly two-seat light airplane. The flight instruments are visible on the left of the instrument panel

**Flight instruments** are the instruments in the cockpit of an aircraft that provide the pilot with information about the flight situation of that aircraft, such as height, speed and altitude. The flight instruments are of particular use in conditions of poor visibility, such as in clouds, when such information is not available from visual reference outside the aircraft.

The term is sometimes used loosely as a synonym for cockpit instruments as a whole, in which context it can include engine instrument, navigational and communication equipment.

### ***Flight instruments***

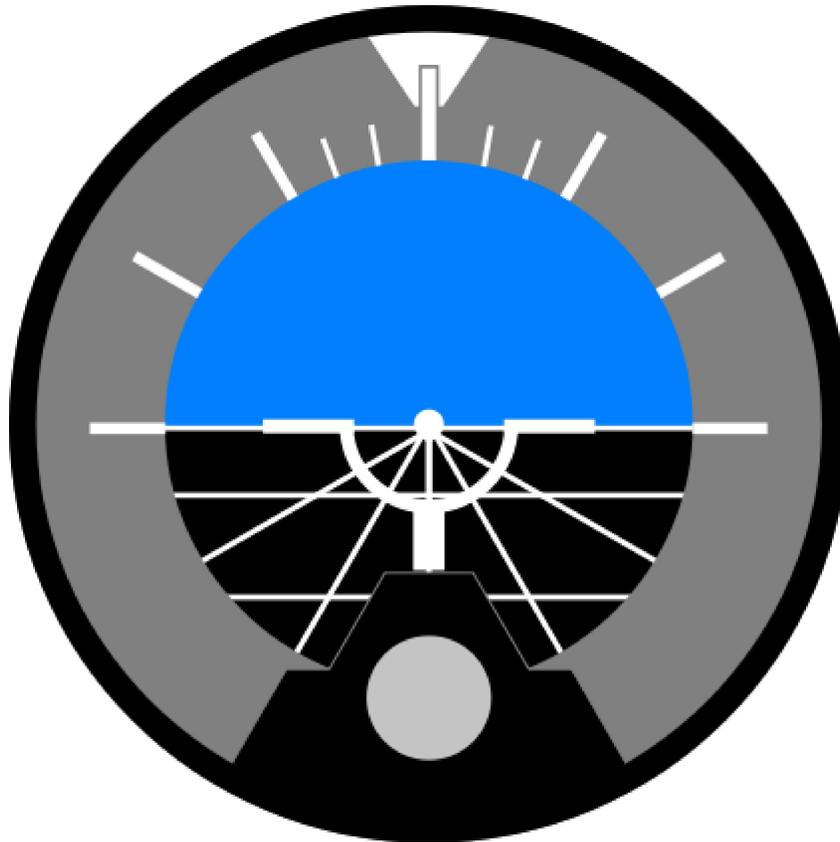
Most aircraft have these flight instruments:

#### **Altimeter**



The altimeter shows the aircraft's altitude above sea-level by measuring the difference between the pressure in a stack of aneroid capsules inside the altimeter and the atmospheric pressure obtained through the static system. It is adjustable for local barometric pressure which must be set correctly to obtain accurate altitude readings. As the aircraft ascends, the capsules expand as the static pressure drops therefore causing the altimeter to indicate a higher altitude. The opposite occurs when descending.

### **Attitude indicator**

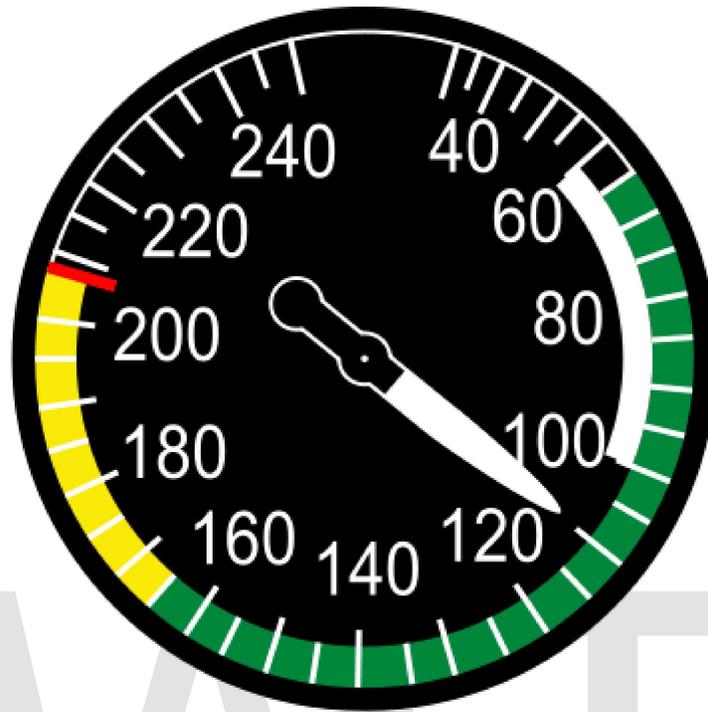


The attitude indicator (also known as an *artificial horizon*) shows the aircraft's attitude relative to the horizon. From this the pilot can tell whether the wings are level and if the aircraft nose is pointing above or below the horizon. This is a primary instrument for instrument flight and is also useful in conditions of poor visibility. Pilots are trained to use other instruments in combination should this instrument or its power fail.



Schempp-Hirth Janus-C glider Instrument panel equipped for "cloud flying". The turn and bank indicator is top center. The heading indicator is replaced by a GPS-driven computer with wind and glide data, driving two electronic variometer displays to the right.

## Airspeed indicator



The airspeed indicator shows the aircraft's speed (usually in knots ) relative to the surrounding air. It works by measuring the ram-air pressure in the aircraft's pitot tube. The indicated airspeed must be corrected for air density (which varies with altitude, temperature and humidity) in order to obtain the true airspeed, and for wind conditions in order to obtain the speed over the ground.

## Magnetic compass



The compass shows the aircraft's heading relative to magnetic north. While reliable in steady level flight it can give confusing indications when turning, climbing, descending, or accelerating due to the inclination of the Earth's magnetic field. For this reason, the heading indicator is also used for aircraft operation. For purposes of navigation it may be necessary to correct the direction indicated (which points to a magnetic pole) in order to obtain direction of true north or south (which points to the Earth's axis of rotation).

## Heading indicator



The heading indicator (also known as the directional gyro, or DG; sometimes also called the gyrocompass, though usually not in aviation applications) displays the aircraft's heading with respect to geographical north. Principle of operation is a spinning gyroscope, and is therefore subject to drift errors (called precession) which must be periodically corrected by calibrating the instrument to the magnetic compass. In many advanced aircraft (including almost all jet aircraft), the heading indicator is replaced by a Horizontal Situation Indicator (HSI) which provides the same heading information, but also assists with navigation

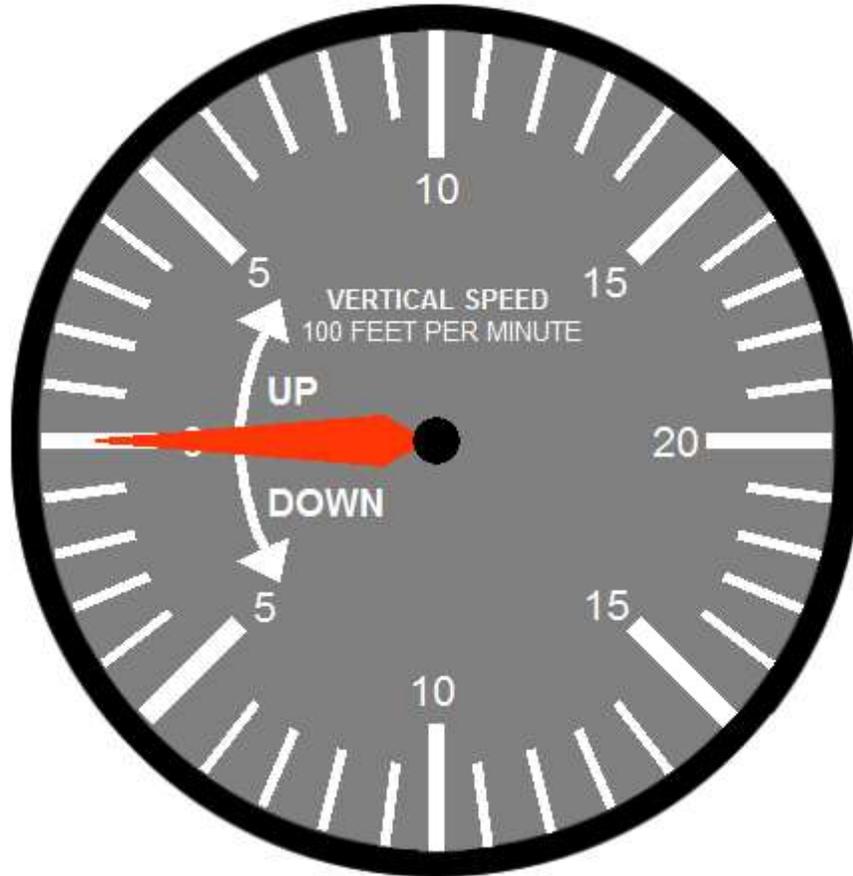
## Turn indicator



The turn indicator displays direction of turn and rate of turn. Internally mounted inclinometer displays 'quality' of turn, i.e. whether the turn is correctly coordinated, as

opposed to an uncoordinated turn, wherein the aircraft would be in either a slip or a skid. The original turn and bank indicator was replaced in the late 1960s and early '70s by the newer turn coordinator, which is responsive to roll as well as rate of turn, the turn and bank is typically only seen in aircraft manufactured prior to that time, or in gliders manufactured in Europe.

### **Vertical speed indicator**



The VSI (also sometimes called a variometer). Senses changing air pressure, and displays that information to the pilot as a rate of climb or descent in feet per minute, meters per second or knots.

Additional panel instruments that may not be found in smaller aircraft include:

## Course deviation indicator



The CDI is an avionics instrument used in aircraft navigation to determine an aircraft's lateral position in relation to a track, which can be provided by a VOR or an Instrument Landing System.

This instrument can also be integrated with the heading indicator in a horizontal situation indicator.

## Radio Magnetic Indicator



An RMI is generally coupled to an *automatic direction finder* (ADF), which provides bearing for a tuned Non-directional beacon (NDB). While simple ADF displays may have only one needle, a typical RMI has two, coupled to different ADF receivers, allowing for position fixing using one instrument.

## Layout



Six basic instruments in a light twin-engine airplane arranged in a "basic-T". From top left: airspeed indicator, attitude indicator, altimeter, turn coordinator, heading indicator, and vertical speed indicator

Most aircraft are equipped with a standard set of **flight instruments** which give the pilot information about the aircraft's attitude, airspeed, and altitude.

## T arrangement

Most aircraft built since about 1953 have four of the **flight instruments** located in a standardized pattern called the T arrangement. The attitude indicator is in the top center, airspeed to the left, altimeter to the right and heading indicator under the attitude indicator. The other two, turn-coordinator and vertical-speed, are usually found under the airspeed and altimeter, but are given more latitude in placement. The magnetic compass will be above the instrument panel, often on the windscreen centerpost. In newer aircraft with glass cockpit instruments the layout of the displays conform to the basic T arrangement.

## Basic Six

In 1937 the Royal Air Force (RAF) chose a set of six essential flight instruments which would remain the standard panel used for flying in Instrument Meteorological Conditions (IMC) for the next 20 years. They were:

- airspeed indicator (knots)

- attitude indicator
- vertical speed indicator (rate of climb)
- altimeter
- directional gyro (compass)
- turn and bank indicator (aircraft attitude)

This panel arrangement was incorporated into every RAF aircraft, from the light Tiger Moth, to the heavy Avro Lancaster, and minimized the type-conversion difficulties associated with Blind Flying, since a pilot trained on one aircraft could quickly become accustomed to any other if the instruments were identical.

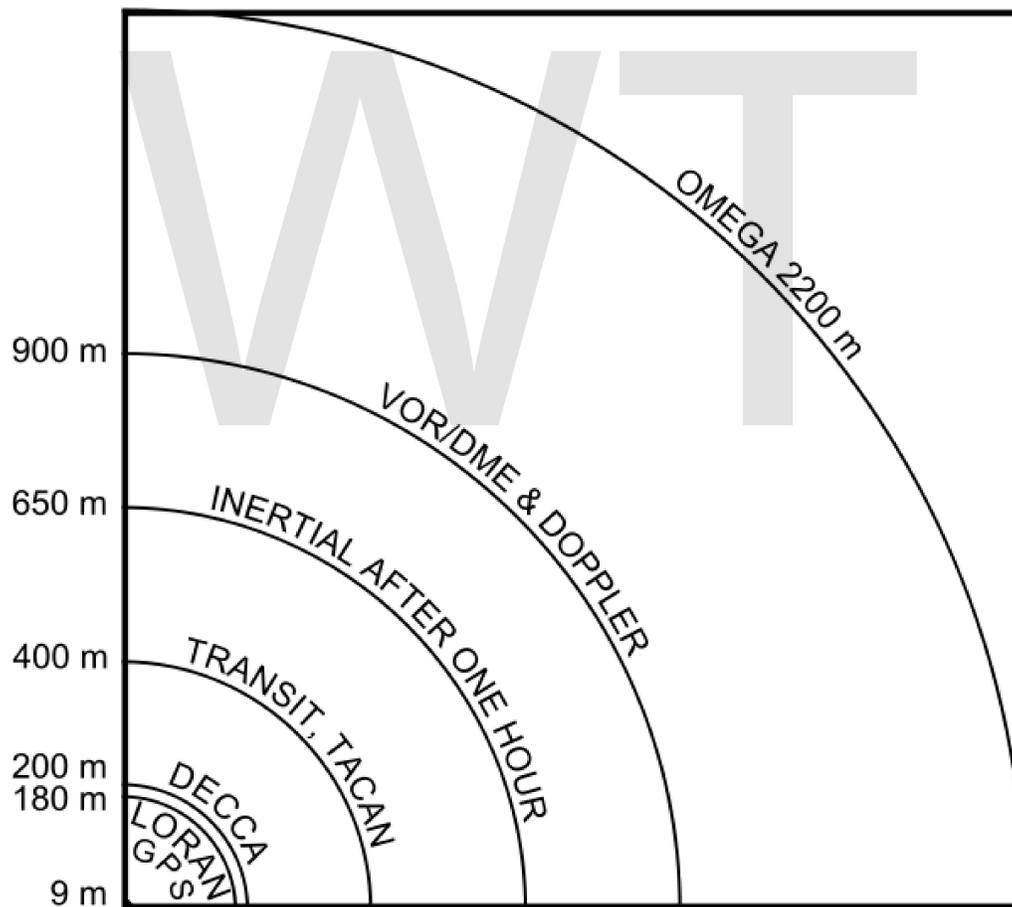
This Basic Six set, also known as a six pack, was also adopted by commercial aviation. After the Second World War the arrangement was changed to: (top row) airspeed, artificial horizon, altimeter, (bottom row) radio compass, direction indicator, vertical speed.

The image shows the letters 'WWT' in a large, bold, sans-serif font. The 'W' is composed of three vertical strokes, and the 'T' is a single vertical stroke with a horizontal top bar. The letters are light gray and centered horizontally.

## Chapter 5

# Decca Navigator System

ACCURACY OF NAVIGATION SYSTEMS  
(2-dimensional)





Decca Navigator Mk 12

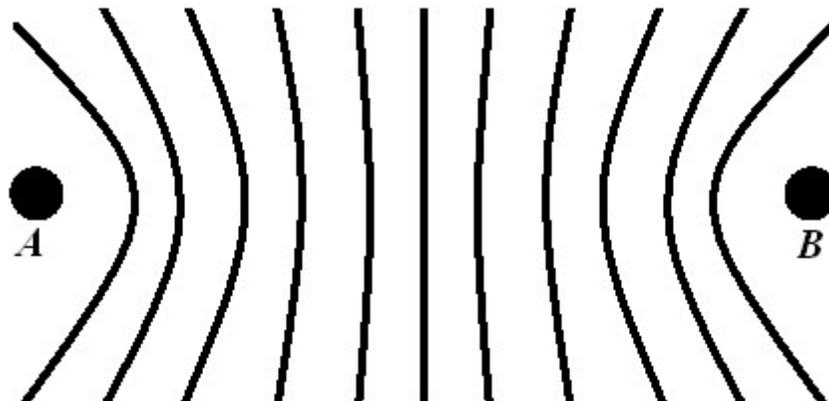
The **Decca Navigator System** was a hyperbolic low frequency radio navigation system (also known as multilateration) that was first deployed during World War II when the Allied forces needed a system which could be used to achieve accurate landings. As was the case with Loran C, its primary use was for ship navigation in coastal waters.

Fishing vessels were major post-war users, but it was also used on aircraft, including a very early (1949) application of moving-map displays. The system was deployed extensively in the North Sea and was used by helicopters operating to oil platforms. After being shut down in the spring of 2000, it has been superseded by systems such as the American GPS and the planned European GALILEO positioning system.

It was deployed in the United Kingdom after World War II and later used in many areas around the world. Decca employees used to joke that DECCA was an acronym for *Dedicated Englishmen Causing Chaos Abroad*.

## ***Principles of Operation***

### **Overview**



*The Decca Navigator principle.*

The phase difference between the signals received from stations A (Master) and B (Slave) is constant along each hyperbolic curve. The foci of the hyperbola are at the transmitting stations, A and B.

The Decca Navigator System consisted of a number of land-based stations organised into *chains*. Each chain consisted of a Master station and three (occasionally two) Slave stations, termed Red, Green and Purple. Ideally, the Slaves would be positioned at the vertices of an equilateral triangle with the Master at the centre. The baseline length, i.e. the Master-Slave distance, was typically 60~120 nautical miles. Each station transmitted a continuous wave signal that, by comparing the phase difference of the signals from the Master and one of the Slaves, resulted in a set of hyperbolic *lines of position* called a *pattern*. As there were three Slaves there were three patterns, termed Red, Green and Purple. The patterns were drawn on nautical charts as a set of hyperbolic lines in the appropriate colour. Receivers identified which hyperbola they were on and a position could be plotted at the intersection of the hyperbola from different patterns, usually by using the pair with the angle of cut closest to orthogonal as possible.

### **Detailed Principles of Operation**

When two stations transmit at the same phase-locked frequency, the difference in phase between the two signals is constant along a hyperbolic path. Of course, if two stations transmit on the same frequency, it is practically impossible for the receiver to separate them; so instead of all stations transmitting at the same frequency, each chain was allocated a nominal frequency,  $1f$ , and each station in the chain transmitted at a harmonic of this base frequency, as follows:

Station	Harmonic	Frequency (kHz)
Master	6f	85.000
Purple Slave	5f	70.833
Red Slave	8f	113.333
Green Slave	9f	127.500

The frequencies given are those for Chain 5B, known as the English Chain, but all chains used similar frequencies between 70 kHz and 129 kHz.

Decca receivers multiplied the signals received from the Master and each Slave by different values to arrive at a common frequency (least common multiple, LCM) for each Master/Slave pair, as follows:

Pattern	Slave Harmonic	Slave Multiplier	Master Harmonic	Master Multiplier	Common Frequency
Purple	5f	×6	6f	×5	30f
Red	8f	×3	6f	×4	24f
Green	9f	×2	6f	×3	18f

It was phase comparison at this common frequency that resulted in the hyperbolic lines of position. The interval between two adjacent hyperbolas on which the signals are in phase was called a *lane*. Since the wavelength of the common frequency was small compared with the distance between the Master and Slave stations there were many possible lines of position for a given phase difference, and so a unique position could not be arrived at by this method.

Other receivers, typically for aeronautical applications, divided the transmitted frequencies down to the basic frequency (1f) for phase comparison, rather than multiplying them up to the LCM frequency.

## Lanes and Zones

Early Decca receivers were fitted with three rotating *Decometers* that indicated the phase difference for each pattern. Each Decometer drove a second indicator that counted the number of lanes traversed – each 360 degrees of phase difference was one lane traversed. In this way, assuming the point of departure was known, a more or less distinct location could be identified.

The lanes were grouped into *zones*, with 18 green, 24 red, or 30 purple lanes in each zone. This meant that on the baseline (the straight line between the Master and its Slave) the zone width was the same for all three patterns of a given chain. Typical lane and zone widths on the baseline are shown in the table below (for chain 5B):

<b>Lane or Zone</b>	<b>Width on Baseline</b>
Purple lane	352.1 m
Red lane	440.1 m
Green lane	586.8 m
Zones (all patterns)	10563 m

The lanes were numbered 0 to 23 for red, 30 to 47 for green and 50 to 79 for purple. The zones were labelled A to J, repeating after J. A Decca position coordinate could thus be written: Red I 16.30; Green D 35.80. Later receivers incorporated a microprocessor and displayed a position in latitude and longitude.

## **Multipulse**

*Multipulse* provided an automatic method of lane and zone identification by using the same phase comparison techniques described above on lower frequency signals.

The nominally continuous wave transmissions were in fact divided into a 20 second cycle, with each station in turn simultaneously transmitting all four Decca frequencies (5f, 6f, 8f and 9f) in a phase-coherent relationship for a brief period of 0.45 seconds each cycle. This transmission, known as Multipulse, allowed the receiver to extract the 1f frequency and so to identify which lane the receiver was in (to a resolution of a zone).

As well as transmitting the Decca frequencies of 5f, 6f, 8f and 9f, an 8.2f signal, known as Orange, was also transmitted. The beat frequency between the 8.0f (Red) and 8.2f (Orange) signals allowed a 0.2f signal to be derived and so resulted in a hyperbolic pattern in which one cycle ( $360^\circ$ ) of phase difference equates to 5 zones.

Assuming that one's position was known to this accuracy, this gave an effectively unique position.

## **Range and Accuracy**

During daylight ranges of around 400 nautical miles (740 km) could be obtained, reducing at night to 200 to 250 nautical miles (460 km), depending on propagation conditions.

The accuracy depended on:

- Width of the lanes
- Angle of cut of the hyperbolic lines of position
- Instrumental errors
- Propagation errors (e.g. Skywave)

By day these errors could range from a few meters on the baseline up to a nautical mile at the edge of coverage. At night, skywave errors were greater and on receivers without multipulse capabilities it was not unusual for the position to jump a lane, sometimes without the navigator knowing.

Although in the days of differential GPS this range and accuracy may appear poor, in its day the Decca system was one of the few, if not the only, position fixing system available to many mariners. Since the need for an accurate position is less when the vessel is further from land, the reduced accuracy at long ranges was not a great problem.

## ***History***

### **Origins**

In 1936 William J. O'Brien, an American engineer, contracted tuberculosis which put his career on hold for a period of two years. During this period he had the idea of position fixing by means of phase comparison of continuous wave transmissions. The initial market envisaged was for aircraft and some experiments were carried out in California in 1938. However both the American Army and Navy considered the idea too complicated.

O'Brien had a friend, Harvey F. Schwarz, who was chief engineer of the Decca Record company in England, and in 1939 sent him details of the system so it could be put forward to the British military. Initially Robert Watson-Watt reviewed the system but he did not follow it up. However, in October 1941 the British Admiralty Signal Establishment (ASE) became interested in the system, which was then classified as *Admiralty Outfit QM*. O'Brien came over to the UK and conducted the first marine trials between Anglesey and the Isle of Man, at frequencies of 305/610 kHz, on 16 September 1942. These were successful and further trials were conducted in the northern Irish Sea in April 1943 at 70/130 kHz. A three-station trial was held in conjunction with a large-scale assault and landing exercise in the Moray Firth in February/March 1944.

The success of the trials and the relative ease of use and accuracy of the system resulted in Decca receiving an order for 27 *Admiralty Outfit QM* receivers. The receiver consisted of an electronics unit with two dials and was known to its operators as the "Blue Gasmeter Job". A Decca chain was set up, consisting of a master station at Chichester and slaves at Swanage and Beachy Head. A fourth, decoy, transmitter was located in the Thames Estuary as part of the deception that the invasion would be focussed on the Calais area.

Twenty-one minesweepers and other vessels were fitted with *Admiralty Outfit QM* and on 5 June 1944 they used it to accurately navigate across the English Channel and to sweep the minefields in the planned areas. The swept areas were marked with buoys in preparation for the Normandy Landings.

After the initial ship tests, Decca conducted tests in cars, driving in the Kingston By-Pass area to verify receiver accuracy. In the car installation, it was found possible to navigate

within an individual traffic lane. The company entertained high hopes that the system could be used in aircraft, to permit much more precise navigation in the critical airspace around airports and urban centers where traffic density was highest.

## Deployment

After the end of World War II the Decca Navigator Co. Ltd. was formed (1945) and the system expanded rapidly, particularly in areas of British influence; at its peak it was deployed in many of the world's major shipping areas. More than 15,000 receiving sets were in use aboard ships in 1970. There were 4 chains around England, 1 in Ireland and 2 in Scotland, 12 in Scandinavia (5 each in Norway and Sweden and 1 each in Denmark and Finland), a further 4 elsewhere in northern Europe and 2 in Spain. In the late 1950s an experimental Decca chain was set up in the United States, in the New York area, to be used for navigating the Vertol 107 helicopters of New York Airways. These helicopters were operating from the principal local airports—John F. Kennedy Airport on Long Island, Newark Airport in New Jersey, LaGuardia Airport in the Borough of Queens, nearer to Manhattan, and a site on the top of the (then) PanAm Building on Park Avenue. Use of Decca was essential because its signals could be received down to sea level, were not subject to the line-of-sight limitations of VOR/DME and did not suffer the slant-range errors that create problems with VOR/DME close to the transmitters. The Decca installations in the New York Airways helicopters included the unique Decca 'roller map' displays that enabled the pilot to see his or her position at a glance, a concept infeasible with VOR/DME. This chain installation was considered highly controversial at the time, for political reasons. This led to the U.S. Coast Guard, under instructions from the Treasury Department to which it reported, banning the use of Decca receivers in ships entering New York harbor for fear that the system might create a de facto standard (as it had become in other areas of the world). It also served to protect the marketing interests of the Hoffman Electronics division of ITT, a principal supplier of VOR/DME systems, that Decca might have been poised to usurp.

This situation was exacerbated by the workload problems of the Air Traffic Controllers Association (ATCA), under its executive director Francis McDermott, whose members were forced to use radar data on aircraft positions, relaying those positions by radio to the aircraft from their control locations. An example of the problem, cited by experts, was the collision of a Boeing 707 and a Lockheed Constellation over Staten Island, New York, that—according to some experts—could have been avoided if the aircraft had been Decca equipped and could not only have determined their positions more precisely but would not have suffered from the rho-theta position errors inherent in VOR/DME.

Other chains were established in Japan (6 chains); Namibia and South Africa (5 chains); India and Bangladesh (4 chains); Canada (4 chains around Newfoundland and Nova Scotia); North-West Australia (2 chains); the Persian Gulf (1 chain with stations in Qatar and the United Arab Emirates and a second chain in the north of the Gulf with stations in Iran) and the Bahamas (1 chain). Four chains were planned for Nigeria but only 2 chains were built and these did not enter into public service. Two chains in Vietnam were used during the Vietnam War for helicopter navigation, with limited success. During the Cold

War period, following WWII, the R.A.F. established a confidential chain in Germany. The Master station was in Bad Iburg near Osnabrück and there were two Slaves. The purpose of this chain was to provide accurate air navigation for the corridor between Western Germany and Berlin in the event that a mass evacuation of allied personnel may be required. In order to maintain secrecy, frequencies were changed at irregular intervals.



Decca Navigator Mk 21

### **Decca, Racal, and the closedown**

Decca Navigator was headquartered at New Malden, Surrey, just off the Kingston bypass. There was a Decca School, at Brixham, Devon, where employees were sent on courses from time to time.

Racal, the UK weapons and communications company, acquired Decca in 1980. Claiming the acquisition was to acquire Decca's radar company, rather than the avionics side of the business, it sold off parts including Decca Navigator.

The monopoly on leased, not purchased, receivers by Decca generated great wealth for the company. This monopoly was later broken in the early 1980s when receivers could be purchased by users, thereby reducing the cost following the lapse of the patent on the basic system technology.



Decca receiver from the 1980s by Philips, which could be purchased instead of leased. A Danish company started manufacturing receivers for fishing boats which employed Decca's navigation charts, but users didn't pay rental for using the system.

In the ensuing court battle Decca lost the monopoly, and that signalled the beginning of the end. Income dwindled and eventually, the UK Ministry of Transport stepped in, having the lighthouse authorities take responsibility for operating the system in the early 1990s.

A ruling from the European Union forced the UK government to withdraw funding - for fishermen users - and started the process which eventually resulted in the system being closed down and the installations scrapped.

The Decca Navigator System provided by the General Lighthouse Authorities ceased to operate at midnight on 31 March 2000. The Irish chain provided by Bórd Iascaigh Mhara continued transmitting until 19 May 2000.

Japan holds the distinction of being the last bastion of Decca having closed down the Hokkaidō chain (9C) in March 2001. This was one year later than the UK closure. Hokkaidō was also the first Decca chain to open in Japan in the year 1967.

## ***Other Applications***

A more accurate system named Hi-Fix was developed using signalling in the 1.6 MHz range. It was used for specialised applications such as precision measurements involved with oil-drilling and by the Royal Navy for detailed mapping and surveying of coasts and harbours. The Hi-Fix equipment was leased for a period with temporary chains established to provide coverage of the area required, Hi-Fix was commercialised by Racal Survey in the early 1980s. An experimental chain was installed with coverage of central London and receivers placed in London buses and other vehicles to demonstrate an early vehicle location and tracking system. Each vehicle would report its location automatically via a conventional VHF two-way radio link, the data added to a voice channel. A forerunner to a system employed by London black cabs in the 1990s using GPS and now the widespread use of GSM, 3G and Wi-Fi geolocation applications.

An interesting characteristic of the Decca VLF signal discovered on BOAC, later British Airways, test flights to Moscow, was that the carrier switching could not be detected even though the carrier could be received with sufficient strength to provide navigation. Such testing, involving civilian aircraft, is quite common and may well not be in the knowledge of a pilot.

The 'low frequency' signalling of the Decca system also permitted its use on submarines. One 'enhancement' of the Decca system was to offer the potential of keying the signal, using Morse code, to signal the onset of nuclear war. This was never optioned by the UK government. Messages were clandestinely sent, however, between Decca stations thereby bypassing international telephone calls, especially in non-UK chains.

A long range trans North Atlantic system was in operation from the mid nineteen fifties. It was called DECTRA. It utilised two stations in Newfoundland and two in Scotland. The transmissions used normal "pattern" transmitters of a much higher power than on standard DECCA frequencies. It was intended as an air navigational aid.

### ***Special DECCA towers***

- Puckeridge DECCA tower
- Zeven DECCA-transmitter

## Chapter 6

# Dynamic Positioning



Offshore Support Vessel *Toisa Perseus* with, in the background, the fifth-generation deepwater drillship *Discoverer Enterprise*, over the Thunder Horse Oil Field. Both are equipped with DP systems.

**Dynamic positioning (DP)** is a computer controlled system to automatically maintain a vessel's position and heading by using its own propellers and thrusters. Position reference sensors, combined with wind sensors, motion sensors and gyro compasses, provide information to the computer pertaining to the vessel's position and the magnitude and direction of environmental forces affecting its position. Examples of vessel types that employ DP include, but are not limited to, ships and semi-submersible Mobile Offshore Drilling Units (MODU) and Oceanographic Research Vessels.

The computer program contains a mathematical model of the vessel that includes information pertaining to the wind and current drag of the vessel and the location of the thrusters. This knowledge, combined with the sensor information, allows the computer to calculate the required steering angle and thruster output for each thruster. This allows operations at sea where mooring or anchoring is not feasible due to deep water, congestion on the sea bottom (pipelines, templates) or other problems.

Dynamic positioning may either be absolute in that the position is locked to a fixed point over the bottom, or relative to a moving object like another ship or an underwater vehicle. One may also position the ship at a favourable angle towards wind, waves and current, called weathervaning.

Dynamic positioning is utilized by much of the offshore oil industry, for example in the North Sea, Persian Gulf, Gulf of Mexico, West Africa, and off the coast of Brazil. There are currently more than 1000 DP ships.

## **History**

Dynamic positioning started in the 1960s for offshore drilling. With drilling moving into ever deeper waters, Jack-up barges could not be used any more and anchoring became less economical.

In 1961 the drillship *Cuss I* was fitted with four steerable propellers, in an attempt to drill the first Moho well. It was possible to keep the ship in position above the well off La Jolla, California, at a depth of 948 meters.

After this, off the coast of Guadalupe, Mexico, five holes were drilled, the deepest at 183 m (601 ft) below the sea floor in 3,500 m (11,700 ft) of water, while maintaining a position within a radius of 180 meters. The ship's position was determined by radar ranging to buoys and sonar ranging from subsea beacons.

Whereas the *Cuss I* was kept in position manually, later in the same year Shell launched the drilling ship *Eureka* that had an analogue control system interfaced with a taut wire, making it the first true DP ship.

While the first DP ships had analogue controllers and lacked redundancy, since then vast improvements have been made. Besides that, DP nowadays is not only used in the oil industry, but also on various other types of ships. In addition, DP is not limited to

maintaining a fixed position any more. One of the possibilities is sailing an exact track, useful for cablelay, pipelay, survey and other tasks.

### **Comparison between position-keeping options**

Other methods of position-keeping are the use of an anchor spread and the use of a jack-up barge. All have their own advantages and disadvantages.

<b>Comparison position-keeping options</b>		
<b>Jack-up Barge</b>	<b>Anchoring</b>	<b>Dynamic Positioning</b>
<p><b>Advantages:</b></p> <ul style="list-style-type: none"> <li>• No complex systems with thrusters, extra generators and controllers.</li> <li>• No chance of running off position by system failures or blackouts.</li> <li>• No underwater hazards from thrusters.</li> </ul>	<p><b>Advantages:</b></p> <ul style="list-style-type: none"> <li>• No complex systems with thrusters, extra generators and controllers.</li> <li>• No chance of running off position by system failures or blackouts.</li> <li>• No underwater hazards from thrusters.</li> </ul>	<p><b>Advantages:</b></p> <ul style="list-style-type: none"> <li>• Manoeuvring is excellent; it is easy to change position.</li> <li>• No anchor handling tugs are required.</li> <li>• Not dependent on waterdepth.</li> <li>• Quick set-up.</li> <li>• Not limited by obstructed seabed.</li> </ul>
<p><b>Disadvantages:</b></p> <ul style="list-style-type: none"> <li>• No manoeuvrability once positioned.</li> <li>• Limited to water depths of ~150 meters.</li> </ul>	<p><b>Disadvantages:</b></p> <ul style="list-style-type: none"> <li>• Limited manoeuvrability once anchored.</li> <li>• Anchor handling tugs are required.</li> <li>• Less suitable in deep water.</li> <li>• Time to anchor out varies between several hours to several days.</li> <li>• Limited by obstructed seabed (pipelines, seabed).</li> </ul>	<p><b>Disadvantages:</b></p> <ul style="list-style-type: none"> <li>• Complex systems with thrusters, extra generators and controllers.</li> <li>• High initial costs of installation.</li> <li>• High fuel costs.</li> <li>• Chance of running off position by system failures or blackouts.</li> <li>• Underwater hazards from thrusters for divers and ROVs.</li> <li>• Higher maintenance of the mechanical systems.</li> </ul>

Although all methods have their own advantages, dynamic positioning has made many operations possible that were not feasible before.

The costs are falling due to newer and cheaper technologies and the advantages are becoming more compelling as offshore work enters ever deeper water and the environment (coral) is given more respect. With container operations, crowded ports can be made more efficient by quicker and more accurate berthing techniques. Cruise ship

operations benefit from faster berthing and non-anchored "moorings" off beaches or inaccessible ports.

## ***Applications***



SBX underway

Important applications include:

- Servicing Aids to Navigation (ATON)
- Cable-laying
- Crane vessels
- Cruise ships

- Diving support vessels
- Dredging
- Drillships
- FPSOs
- Flotels
- Landing Platform Docks
- Maritime research
- Mine sweepers
- Pipe-laying ship
- Platform supply vessels
- Rockdumping
- Sea Launch
- Sea-based X-band Radar
- Shuttle tankers
- Survey ships

## **Scope**

A ship can be considered to have six degrees of freedom in its motion, i.e., it can move in any of six axes.

Three of these involve translation:

- surge (forward/astern)
- sway (starboard/port)
- heave (up/down)

and the other three rotation:

- roll (rotation about surge axis)
- pitch (rotation about sway axis)
- yaw (rotation about heave axis)

Dynamic positioning is concerned primarily with control of the ship in the horizontal plane, i.e., the three axis surge, sway and yaw.

## **Requirements**

A ship that is to be used for DP requires:

- to maintain position and heading, first of all the position and heading need to be known.
- a control computer to calculate the required control actions to maintain position and correct for position errors.
- thrust elements to apply forces to the ship as demanded by the control system.

For most applications, the position reference systems and thrust elements must be carefully considered when designing a DP ship. In particular, for good control of position in adverse weather, the thrust capability of the ship in three axes must be adequate.

## **Reference systems**

### **Position reference systems**

There are several means to determine a ship's position at sea. Most traditional methods used for ships navigation are not accurate enough. For that reason, several positioning systems have been developed during the past decades. Producers of DP systems are: Kongsberg, Navis Engineering Oy, Converteam, EMI, Deep Down Marine Technologies, L3, MT-div.Chouest, Rolls Royce, Nautronix, and others. The applications and availability depends on the type of work and water depth. The most common Position reference/Measuring systems /Equipment (PRS/PME) are:



GPS satellite in orbit.

- **DGPS, Differential GPS.** The position obtained by GPS is not accurate enough for use by DP. The position is improved by use of a fixed ground based reference

station (differential station) that compares the GPS position to the known position of the station. The correction is sent to the DGPS receiver by long wave radio frequency. For use in DP an even higher accuracy and reliability is needed. Companies such as Fugro or C&C Technologies supply differential signals via satellite, enabling the combination of several differential stations. The advantage of DGPS is that it is almost always available. Disadvantages are degrading of the signal because of sunspots or atmospheric disturbances, blockage of satellites by cranes or structures and deterioration of the signal at high altitudes. There are also systems installed on vessels that use various different Augmentation systems, as well as combining GPS position with GLONASS.

- **Acoustics.** This system consists of one or more transponders placed on the seabed and a transducer placed in the ship's hull. The transducer sends an acoustic signal (by means of piezoelectric elements) to the transponder, which is triggered to reply. As the velocity of sound through water is known (preferably a soundprofile is taken regularly), the distance is known. Because there are many elements on the transducer, the direction of the signal from the transponder can be determined. Now the position of the ship relative to the transponder can be calculated. Disadvantages are the vulnerability to noise by thrusters or other acoustic systems. Furthermore, the use is limited in shallow waters because of ray bending that occurs when sound travels through water horizontally. Three types of HPR systems are commonly used:
  - **Ultra- or Super- Short Base Line, USBL or SSBL.** This works as described above. Because the angle to the transponder is measured, a correction needs to be made for the ship's roll and pitch. These are determined by Motion Reference Units. Because of the nature of angle measurement, the accuracy deteriorates with increasing water depth.
  - **Long Base Line, LBL.** This consists of an array of at least three transponders. The initial position of the transponders is determined by USBL and/ or by measuring the baselines between the transponders. Once that is done, only the ranges to the transponders need to be measured to determine a relative position. The position should theoretically be located at the intersection of imaginary spheres, one around each transponder, with a radius equal to the time between transmission and reception multiplied by the speed of sound through water. Because angle measurement is not necessary, the accuracy in large water depths is better than USBL.
  - **Short Baseline, SBL.** This works with an array of transducers in the ship's hull. These determine their position to a transponder, so a solution is found in the same way as with LBL. As the array is located on the ship, it needs to be corrected for roll and pitch.

- **Riser Angle Monitoring.** On drillships, riser angle monitoring can be fed into the DP system. It may be an electrical inclinometer or based on USBL, where a riser angle monitoring transponder is fitted to the riser and a remote inclinometer unit is installed on the Blow Out Preventer (BOP) and interrogated through the ship's HPR.



Light Taut Wire on the *HOS Achiever*

- **Light Taut Wire, LTW.** The oldest position reference system used for DP is still very accurate in relatively shallow water. A clumpweight is lowered to the seabed. By measuring the amount of wire paid out and the angle of the wire by a gimbal head, the relative position can be calculated. Care should be taken not to let the wire angle become too large to avoid dragging. For deeper water the

system is less favourable, as current will curve the wire. There are however systems that counteract this with a gimbal head on the clumpweight. Horizontal LTW's are also used when operating close to a structure. Objects falling on the wire are a risk here.

- **Fanbeam** and **CyScan**. These are laser based position reference systems. They are very straightforward system, as only a small prism needs to be installed on a nearby structure or ship. Risks are the system locking on other reflecting objects and blocking of the signal. Range depends on the weather, but is typically more than 500 meters.
- **Artemis**. A radar based system. A unit is placed on a nearby structure and aimed at the unit on board the ship. The range is several kilometres. Advantage is the reliable, all-weather performance. Disadvantage is that the unit is rather heavy.
- **DARPS, Differential, Absolute and Relative Positioning System**. Commonly used on shuttle tankers while loading from a FPSO. Both will have a GPS receiver. As the errors are the same for the both of them, the signal does not need to be corrected. The position from the FPSO is transmitted to the shuttle tanker, so a range and bearing can be calculated and fed into the DP system.
- **RADius** and **RadaScan**. These are radar based system, but have no moving parts as Artemis. Another advantage is that the transponders are much smaller than the Artemis unit. The range is typically 500 – 1000 meters.
- **Inertial navigation** is used in combination with GPS (Seapath) and Hydroacoustics (HAIN).

## Heading reference systems

- **Gyrocompasses** are normally used to determine heading.

More advanced methods are:

- **Ring-Laser gyroscopes**
- **Fibre optic gyroscopes**
- **Seapath**, a combination of GPS and inertial sensors.

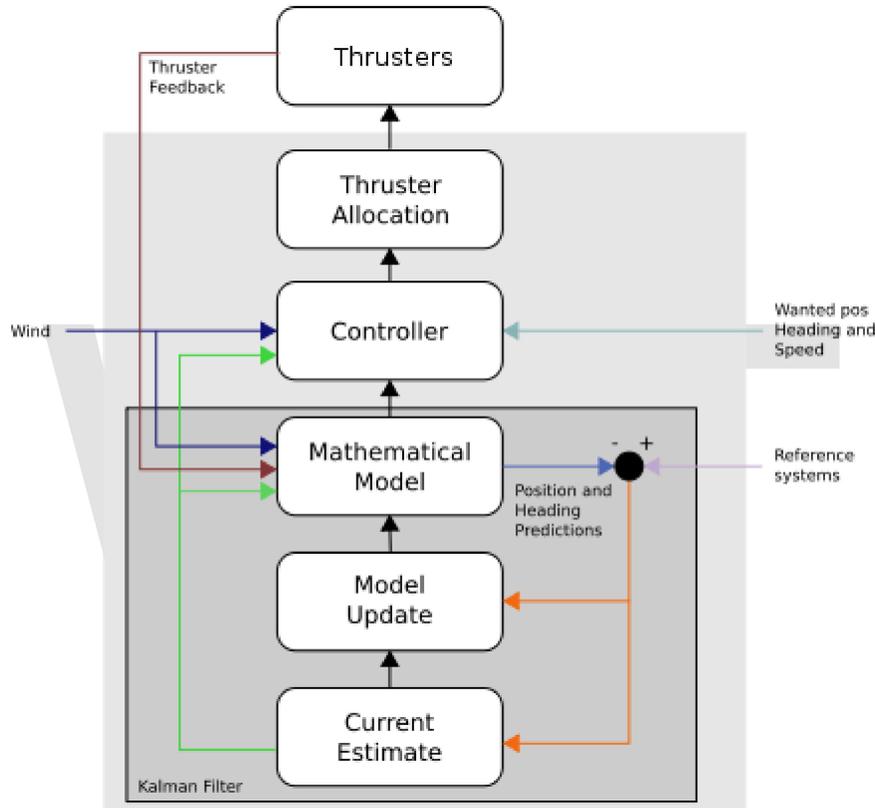
## Reference systems

Besides position and heading, other variables are fed into the DP system through sensors:

- **Motion Reference Units, Vertical Reference Units or Vertical Reference Sensors, VRU's or MRU's or VRS's**, determine the ship's roll, pitch and heave.
- **Wind sensors** are fed into the DP system feed-forward, so the system can anticipate wind gusts before the ship is blown off position.

- **Draught sensors**, since a change of draught influences the effect of wind and current on the hull.
- Other sensors depend on the kind of ship. A pipelay ship may measure the force needed to pull on the pipe, large crane vessels will have sensors to determine the cranes position, as this changes the wind model, enabling the calculation of a more accurate model.

## Control systems



Block diagram of control system

In the beginning PID controllers were used and today are still used in the simpler DP systems. But modern controllers use a mathematical model of the ship that is based on a hydrodynamic and aerodynamic description concerning some of the ship's characteristics such as mass and drag. Of course, this model is not entirely correct. The ship's position and heading are fed into the system and compared with the prediction made by the model. This difference is used to update the model by using Kalman filtering technique. For this reason, the model also has input from the windsensors and feedback from the thrusters. This method even allows not having input from any PRS for some time, depending on the quality of the model and the weather.

The accuracy and precision of the different PRS's is not the same. While a DGPS has a high accuracy and precision, a USBL can have a much lower precision. For this reason, the PRS's are weighted. Based on variance a PRS receives a weight between 0 and 1.

## ***Power and propulsion systems***

To maintain position azimuth thrusters (L-drive or Z-drive), azipods, bow thrusters, stern thrusters, water jets, rudders and propellers are used. DP ships are usually at least partially diesel-electric, as this allows a more flexible set-up and is better able to handle the large changes in power demand, typical for DP operations.

The set-up depends on the DP class of the ship. A Class 1 can be relatively simple, whereas the system of a Class 3 ship is quite complex.

On Class 2 and 3 ships, all computers and reference systems should be powered through a UPS.

## ***Class Requirements***

Based on IMO (International Maritime Organization) publication 645 the Classification Societies have issued rules for Dynamic Positioned Ships described as Class 1, Class 2 and Class 3.

- Equipment Class 1 has no redundancy.  
Loss of position may occur in the event of a single fault.
- Equipment Class 2 has redundancy so that no single fault in an active system will cause the system to fail.  
Loss of position should not occur from a single fault of an active component or system such as generators, thruster, switchboards, remote controlled valves etc., but may occur after failure of a static component such as cables, pipes, manual valves etc.
- Equipment Class 3 which also has to withstand fire or flood in any one compartment without the system failing.  
Loss of position should not occur from any single failure including a completely burnt fire sub division or flooded watertight compartment.

Classification Societies have their own Class notations:

Description	IMO Equipment Class	LR Equipment Class	DnV Equipment Class	GL Equipment Class	ABS Equipment Class
Manual position control and automatic heading control under specified maximum environmental conditions	-	DP(CM)	DYNPOS-AUTS	-	-
Automatic and manual position and heading control under specified maximum environmental conditions	Class 1	DP(AM)	DYNPOS-AUT	DP 1	DPS-0, DPS-1
Automatic and manual position and heading control under specified maximum environmental conditions, during and following any single fault excluding loss of a compartment. (Two independent computer systems).	Class 2	DP(AA)	DYNPOS-AUTR	DP 2	DPS-2
Automatic and manual position and heading control under specified maximum environmental conditions, during and following any single fault including loss of a compartment due to fire or flood. (At least two independent computer systems with a separate backup system separated by A60 class division).	Class 3	DP(AAA)	DYNPOS-AUTRO	DP 3	DPS-3

## **NMD**

Where IMO leaves the decision of which Class applies to what kind of operation to the operator of the DP ship and its client, the Norwegian Maritime Directorate (NMD) has specified what Class should be used in regard to the risk of an operation. In the NMD Guidelines and Notes No. 28, enclosure A four classes are defined:

- Class 0 Operations where loss of position keeping capability is not considered to endanger human lives, or cause damage.
- Class 1 Operations where loss of position keeping capability may cause damage or pollution of small consequence.
- Class 2 Operations where loss of position keeping capability may cause personnel injury, pollution, or damage with large economic consequences.
- Class 3 Operations where loss of position keeping capability may cause fatal accidents, or severe pollution or damage with major economic consequences.

Based on this the type of ship is specified for each operation:

- Class 1 DP units with equipment class 1 should be used during operations where loss of position is not considered to endanger human lives, cause significant damage or cause more than minimal pollution.

- Class 2 DP units with equipment class 2 should be used during operations where loss of position could cause personnel injury, pollution or damage with great economic consequences.
- Class 3 DP units with equipment class 3 should be used during operations where loss of position could cause fatal accidents, severe pollution or damage with major economic consequences.

## ***Redundancy***

Redundancy is the ability to cope with a single failure without loss of position. A single failure can be, amongst others:

- Thruster failure
- Generator failure
- Powerbus failure (when generators are combined on one powerbus)
- Control computer failure
- Position reference system failure
- Reference system failure

For certain operations redundancy is not required. For instance, if a survey ship loses its DP capability, there is normally no risk of damage or injuries. These operations will normally be done in Class 1.

For other operations, such as diving and heavy lifting, there is a risk of damage or injuries. Depending on the risk, the operation is done in Class 2 or 3. This means at least three Position reference systems should be selected. This allows the principle of voting logic, so the failing PRS can be found. For this reason, there are also three DP control computers, three gyrocompasses, three MRU's and three wind sensors on Class 3 ships. If a single fault occurs that jeopardizes the redundancy, i.e., failing of a thruster, generator or a PRS, and this cannot be resolved immediately, the operation should be abandoned as quickly as possible.

To have sufficient redundancy, enough generators and thrusters should be on-line so the failure of one does not result in a loss of position. This is left to the judgement of the DP operator. For Class 2 and Class 3 a Consequence Analyses should be incorporated in the system to assist the DPO in this process.

Disadvantage is that a generator can never operate at full load, resulting in less economy and fouling of the engines.

The redundancy of a DP ship should be judged by an failure mode and effects analysis (FMEA) study and proved by FMEA trials. Besides that, annual trials are done and normally DP function tests are completed prior to each project.

## ***DP Operator***

The DP operator (DPO) judges whether there is enough redundancy available at any given moment of the operation. IMO issued MSC/Circ.738 (Guidelines for dynamic positioning system (DP) operator training) on 24-06-1996. This refers to IMCA (International Marine Contractors Association) M 117 as acceptable standard.

To qualify as a DP operator the following path should be followed:

1. a DP Induction course
2. a minimum of 30 days seagoing DP familiarisation
3. a DP Advanced course
4. a minimum of 6 months watchkeeping on a DP ship
5. a statement of suitability by the master of a DP ship

When the watchkeeping is done on a Class 1 DP ship, a limited certificate will be issued; otherwise a full certificate will be issued.

The DP Training and Certification scheme is operated by The Nautical Institute (NI). The NI issue logbooks to trainees, they accredit training centres and control the issuance of certification.

With ever more DP ships and with increasing manpower demands, the position of DPO is gaining increasing prominence. This shifting landscape led to the creation of The International Dynamic Positioning Operators Association (IDPOA) in 2009.

IDPOA membership is made up of certified DPO's who qualify for fellowship (fDPO), while Members (mDPO) are those with DP experience or who may already be working within the DP certification scheme.

## ***IMCA***

The International Marine Contractors Association was formed in April 1995 from the amalgamation of AODC (originally the International Association of Offshore Diving Contractors), founded in 1972, and DPVOA (the Dynamic Positioning Vessel Owners Association), founded in 1990. It represents offshore, marine and underwater engineering contractors. Acergy, Allseas, Heerema Marine Contractors, Helix Energy Solutions Group, J. Ray McDermott, Saipem, Subsea 7 and Technip have representation on IMCA's Council and provide the president. Previous presidents are:

- 1995-6 - Derek Leach, Coflexip Stena Offshore
- 1997-8 - Hein Mulder, Heerema Marine Contractors
- 1999/2000 - Donald Carmichael, Coflexip Stena Offshore
- 2001-2 - John Smith, Halliburton Subsea/Subsea 7
- 2003-4 - Steve Preston, - Heerema Marine Contractors
- 2005 - Frits Janmaat, Allseas Group

(2005 Vice-President - Knut Boe, Technip)

While it started with the collection and analysis of DP Incidents, since then it has produced publications on different subjects to improve standards for DP systems. It also works with IMO and other regulatory bodies.

WWT

## Chapter 7

# Global Positioning System



Artist's conception of GPS Block II-F satellite in orbit



Civilian GPS receiver ("GPS navigation device") in a marine application.



Automotive navigation system in a taxicab.



GPS receivers are now integrated in many mobile phones.

The **Global Positioning System (GPS)** is a space-based global navigation satellite system (GNSS) that provides reliable location and time information in all weather and at all times and anywhere on or near the Earth when and where there is an unobstructed line of sight to four or more GPS satellites. It is maintained by the United States government and is freely accessible by anyone with a GPS receiver.

The GPS project was developed in 1973 to overcome the limitations of previous navigation systems, integrating ideas from several predecessors, including a number of classified engineering design studies from the 1960s. GPS was created and realized by the U.S. Department of Defense (USDOD) and was originally run with 24 satellites. It became fully operational in 1994.

In addition to GPS, other systems are in use or under development. The Russian GLObal NAVigation Satellite System (GLONASS) was in use by the Russian military only until it was made fully available to civilians in 2007. There are also the planned Chinese Compass navigation system and the European Union's Galileo positioning system.

## History



The design of GPS is based partly on similar ground-based radio navigation systems, such as LORAN and the Decca Navigator developed in the early 1940s, and used during World War II. In 1956 Friedwardt Winterberg proposed a test of general relativity using accurate atomic clocks placed in orbit in artificial satellites. To achieve accuracy requirements, GPS uses principles of general relativity to correct the satellites' atomic clocks. Additional inspiration for GPS came when the Soviet Union launched the first man-made satellite, Sputnik in 1957. A team of U.S. scientists led by Dr. Richard B. Kershner were monitoring Sputnik's radio transmissions. They discovered that, because of the Doppler effect, the frequency of the signal being transmitted by Sputnik was higher as the satellite approached, and lower as it continued away from them. They realized that because they knew their exact location on the globe, they could pinpoint where the satellite was along its orbit by measuring the Doppler distortion.

The first satellite navigation system, Transit, used by the United States Navy, was first successfully tested in 1960. It used a constellation of five satellites and could provide a navigational fix approximately once per hour. In 1967, the U.S. Navy developed the Timation satellite that proved the ability to place accurate clocks in space, a technology required by GPS. In the 1970s, the ground-based Omega Navigation System, based on phase comparison of signal transmission from pairs of stations, became the first worldwide radio navigation system. Limitations of these systems drove the need for a more universal navigation solution with greater accuracy.

While there were wide needs for accurate navigation in military and civilian sectors, almost none of those were seen as justification for the billions of dollars it would cost in research, development, deployment, and operation for a constellation of navigation satellites. During the Cold War arms race, the nuclear threat to the existence of the United States was the one need that did justify this cost in the view of the United States Congress. This deterrent effect is why GPS was funded. The nuclear triad consisted of the United States Navy's submarine-launched ballistic missiles (SLBMs) along with United States Air Force (USAF) strategic bombers and intercontinental ballistic missiles (ICBMs). Considered vital to the nuclear deterrence posture, accurate determination of the SLBM launch position was a force multiplier.

Precise navigation would enable United States submarines to get an accurate fix of their positions prior to launching their SLBMs. The USAF with two-thirds of the nuclear triad also had requirements for a more accurate and reliable navigation system. The Navy and Air Force were developing their own technologies in parallel to solve what was essentially the same problem. To increase the survivability of ICBMs, there was a

proposal to use mobile launch platforms so the need to fix the launch position had similarity to the SLBM situation.

In 1960, the Air Force proposed a radio-navigation system called MOSAIC (Mobile System for Accurate ICBM Control) that was essentially a 3-D LORAN. A follow-on study called Project 57 was worked in 1963 and it was "in this study that the GPS concept was born." That same year the concept was pursued as Project 621B, which had "many of the attributes that you now see in GPS" and promised increased accuracy for Air Force bombers as well as ICBMs. Updates from the Navy Transit system were too slow for the high speeds of Air Force operation. The Navy Research Laboratory continued advancements with their Timation (Time Navigation) satellites, first launched in 1967, and with the third one in 1974 carrying the first atomic clock into orbit.

With these parallel developments in the 1960s, it was realized that a superior system could be developed by synthesizing the best technologies from 621B, Transit, Timation, and SECOR in a multi-service program.

During Labor Day weekend in 1973, a meeting of about 12 military officers at the Pentagon discussed the creation of a *Defense Navigation Satellite System (DNSS)*. It was at this meeting that "the real synthesis that became GPS was created." Later that year, the DNSS program was named *Navstar*. With the individual satellites being associated with the name Navstar (as with the predecessors Transit and Timation), a more fully encompassing name was used to identify the constellation of Navstar satellites, *Navstar-GPS*, which was later shortened simply to GPS.

After Korean Air Lines Flight 007, carrying 269 people, was shot down in 1983 after straying into the USSR's prohibited airspace, in the vicinity of Sakhalin and Moneron Islands, President Ronald Reagan issued a directive making GPS freely available for civilian use, once it was sufficiently developed, as a common good. The first satellite was launched in 1989, and the 24th satellite was launched in 1994.

Initially, the highest quality signal was reserved for military use, and the signal available for civilian use was intentionally degraded ("Selective Availability", SA). This changed with United States President Bill Clinton ordering Selective Availability turned off at midnight May 1, 2000, improving the precision of civilian GPS from 100 meters (about 300 feet) to 20 meters (about 65 feet). The United States military by then had the ability to deny GPS service to potential adversaries on a regional basis.

GPS is owned and operated by the United States Government as a national resource. Department of Defense (USDOD) is the steward of GPS. *Interagency GPS Executive Board (IGEB)* oversaw GPS policy matters from 1996 to 2004. After that the *National Space-Based Positioning, Navigation and Timing Executive Committee* was established by presidential directive in 2004 to advise and coordinate federal departments and agencies on matters concerning the GPS and related systems. The executive committee is chaired jointly by the deputy secretaries of defense and transportation. Its membership includes equivalent-level officials from the departments of state, commerce, and

homeland security, the joint chiefs of staff, and NASA. Components of the executive office of the president participate as observers to the executive committee, and the FCC chairman participates as a liaison.

USDOD is required by law to "maintain a Standard Positioning Service (as defined in the federal radio navigation plan and the standard positioning service signal specification) that will be available on a continuous, worldwide basis," and "develop measures to prevent hostile use of GPS and its augmentations without unduly disrupting or degrading civilian uses."

## Timeline and modernization

### Summary of satellites

Block	Launch Period	Satellite launches				Currently in orbit and healthy
		Suc-cess	Fail-ure	In prep-ara-tion	Plan-ned	
<b>I</b>	1978–1985	10	1	0	0	0
<b>II</b>	1989–1990	9	0	0	0	0
<b>IIA</b>	1990–1997	19	0	0	0	10
<b>IIR</b>	1997–2004	12	1	0	0	12
<b>IIR-M</b>	2005–2009	8	0	0	0	7
<b>IIF</b>	2010–2011	1	0	11	0	1
<b>IIIA</b>	2014–?	0	0	0	12	0
<b>IIIB</b>	Theoretical	0	0	0	8	0
<b>IIIC</b>	Theoretical	0	0	0	16	0
<b>Total</b>		59	2	11	36	30

(Last update: 24 May 2010)

PRN 01 from Block IIR-M is unhealthy  
 PRN 25 from Block IIA is unhealthy  
 PRN 32 from Block IIA is unhealthy

- In 1972, the USAF Central Inertial Guidance Test Facility (Holloman AFB), conducted developmental flight tests of two prototype GPS receivers over White Sands Missile Range, using ground-based pseudo-satellites.
- In 1978, the first experimental Block-I GPS satellite was launched.
- In 1983, after Soviet interceptor aircraft shot down the civilian airliner KAL 007 that strayed into prohibited airspace because of navigational errors, killing all 269 people on board, U.S. President Ronald Reagan announced that GPS would be made available for civilian uses once it was completed.
- By 1985, ten more experimental Block-I satellites had been launched to validate the concept.
- On February 14, 1989, the first modern Block-II satellite was launched.
- The Gulf War from 1990 to 1991, was the first conflict where GPS was widely used.
- In 1992, the 2nd Space Wing, which originally managed the system, was deactivated and replaced by the 50th Space Wing.
- By December 1993, GPS achieved initial operational capability (IOC), indicating a full constellation (24 satellites) was available and providing the Standard Positioning Service (SPS).
- Full Operational Capability (FOC) was declared by Air Force Space Command (AFSPC) in April 1995, signifying full availability of the military's secure Precise Positioning Service (PPS).
- In 1996, recognizing the importance of GPS to civilian users as well as military users, U.S. President Bill Clinton issued a policy directive declaring GPS to be a dual-use system and establishing an Interagency GPS Executive Board to manage it as a national asset.
- In 1998, United States Vice President Al Gore announced plans to upgrade GPS with two new civilian signals for enhanced user accuracy and reliability, particularly with respect to aviation safety and in 2000 the United States Congress authorized the effort, referring to it as *GPS III*.
- In 1998, GPS technology was inducted into the Space Foundation Space Technology Hall of Fame.
- On May 2, 2000 "Selective Availability" was discontinued as a result of the 1996 executive order, allowing users to receive a non-degraded signal globally.
- In 2004, the United States Government signed an agreement with the European Community establishing cooperation related to GPS and Europe's planned Galileo system.
- In 2004, United States President George W. Bush updated the national policy and replaced the executive board with the National Executive Committee for Space-Based Positioning, Navigation, and Timing.
- November 2004, QUALCOMM announced successful tests of assisted GPS for mobile phones.
- In 2005, the first modernized GPS satellite was launched and began transmitting a second civilian signal (L2C) for enhanced user performance.
- On September 14, 2007, the aging mainframe-based Ground Segment Control System was transferred to the new Architecture Evolution Plan.

- On May 19, 2009, the United States Government Accountability Office issued a report warning that some GPS satellites could fail as soon as 2010.
- On May 21, 2009, the Air Force Space Command allayed fears of GPS failure saying "There's only a small risk we will not continue to exceed our performance standard."
- On January 11, 2010, an update of ground control systems caused a software incompatibility with 8000 to 10000 military receivers manufactured by a division of Trimble Navigation Limited of Sunnyvale, Calif.
- The most recent launch was on May 28, 2010. The oldest GPS satellite still in operation was launched on November 26, 1990, and became operational on December 10, 1990.

## **Awards**

On February 10, 1993, the National Aeronautic Association selected the GPS Team as winners of the 1992 Robert J. Collier Trophy, the nation's most prestigious aviation award. This team combines researchers from the Naval Research Laboratory, the USAF, the Aerospace Corporation, Rockwell International Corporation, and IBM Federal Systems Company. The citation honors them "for the most significant development for safe and efficient navigation and surveillance of air and spacecraft since the introduction of radio navigation 50 years ago."

Two GPS developers received the National Academy of Engineering Charles Stark Draper Prize for 2003:

- Ivan Getting, emeritus president of The Aerospace Corporation and an engineer at the Massachusetts Institute of Technology, established the basis for GPS, improving on the World War II land-based radio system called LORAN (*Long-range Radio Aid to Navigation*).
- Bradford Parkinson, professor of aeronautics and astronautics at Stanford University, conceived the present satellite-based system in the early 1960s and developed it in conjunction with the U.S. Air Force. Parkinson served twenty-one years in the Air Force, from 1957 to 1978, and retired with the rank of colonel.

GPS developer Roger L. Easton received the National Medal of Technology on February 13, 2006.

Francis X. Kane (Col. USAF, ret.) was inducted into the U.S. Air Force Space and Missile Pioneers Hall of Fame at Lackland A.F.B., San Antonio, Texas, March 2, 2010 for his role in space technology development and the engineering design concept of GPS conducted as part of Project 621B.

## ***Basic concept of GPS***

A GPS receiver calculates its position by precisely timing the signals sent by GPS satellites high above the Earth. Each satellite continually transmits messages that include

- the time the message was transmitted
- precise orbital information (the ephemeris)
- the general system health and rough orbits of all GPS satellites (the almanac).

The receiver uses the messages it receives to determine the transit time of each message and computes the distance to each satellite. These distances along with the satellites' locations are used with the possible aid of trilateration, depending on which algorithm is used, to compute the position of the receiver. This position is then displayed, perhaps with a moving map display or latitude and longitude; elevation information may be included. Many GPS units show derived information such as direction and speed, calculated from position changes.

Three satellites might seem enough to solve for position since space has three dimensions and a position near the Earth's surface can be assumed. However, even a very small clock error multiplied by the very large speed of light — the speed at which satellite signals propagate — results in a large positional error. Therefore receivers use four or more satellites to solve for the receiver's location and time. The very accurately computed time is effectively hidden by most GPS applications, which use only the location. A few specialized GPS applications do however use the time; these include time transfer, traffic signal timing, and synchronization of cell phone base stations.

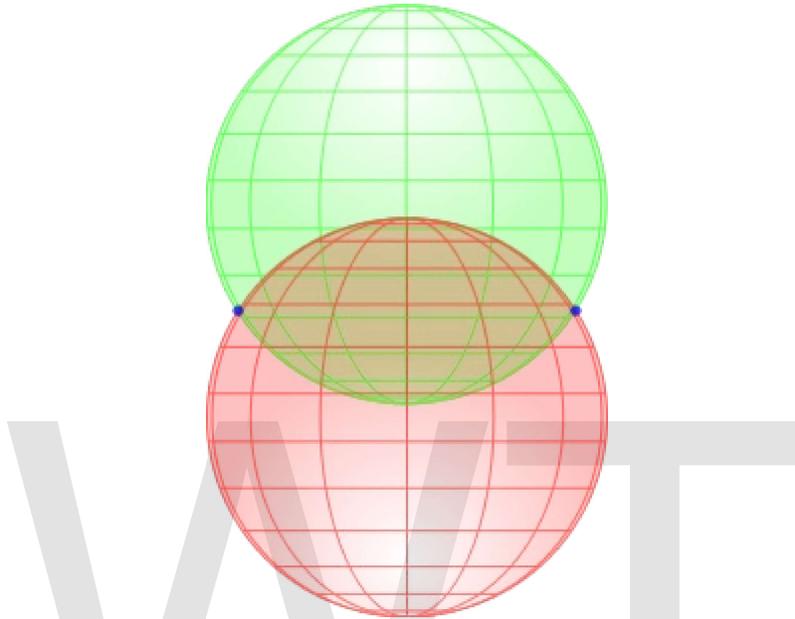
Although four satellites are required for normal operation, fewer apply in special cases. If one variable is already known, a receiver can determine its position using only three satellites. For example, a ship or aircraft may have known elevation. Some GPS receivers may use additional clues or assumptions (such as reusing the last known altitude, dead reckoning, inertial navigation, or including information from the vehicle computer) to give a less accurate (degraded) position when fewer than four satellites are visible.

## Position calculation introduction

To provide an introductory description of how a GPS receiver works, error effects are deferred to a later section. Using messages received from a minimum of four visible satellites, a GPS receiver is able to determine the times sent and then the satellite positions corresponding to these times sent. The x, y, and z components of position, and the time sent, are designated as  $[x_i, y_i, z_i, t_i]$  where the subscript  $i$  is the satellite number and has the value 1, 2, 3, or 4. Knowing the indicated, or uncorrected, time the message was received  $t_{r, \text{uncorr}}$ , the GPS receiver can compute the uncorrected transit time of the message as  $(t_{r, \text{uncorr}} - t_i)$ . Assuming the message traveled at the speed of light,  $c$ , the uncorrected distance traveled or pseudorange,  $P_i$  can be computed as  $(t_{r, \text{uncorr}} - t_i)c$ .

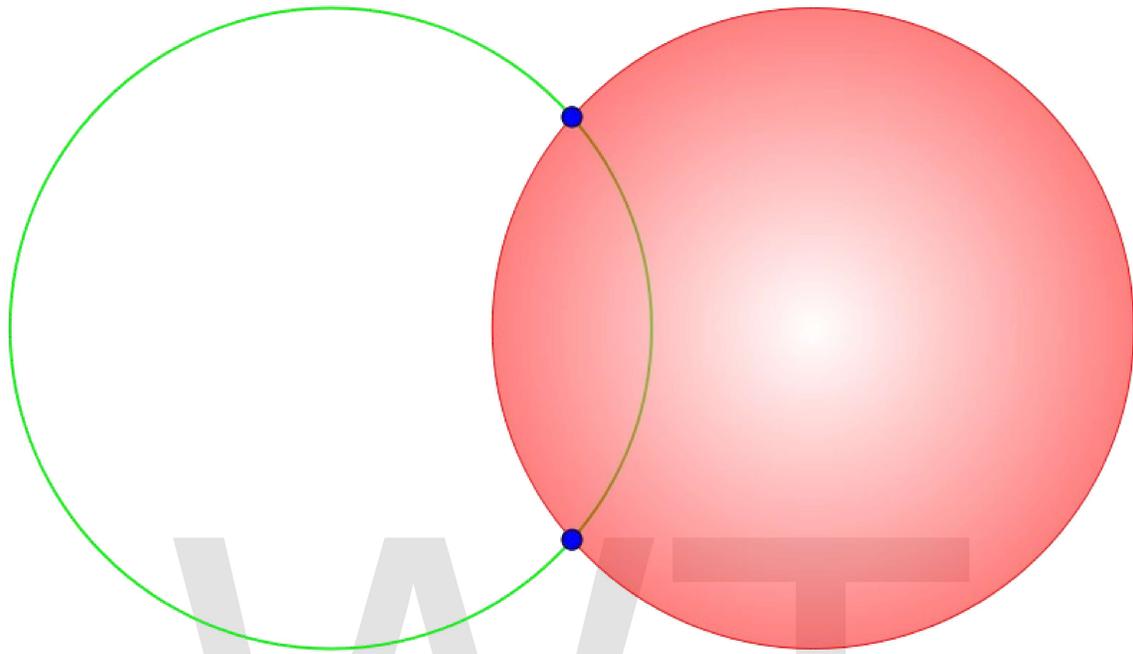
A satellite's position and pseudorange define a sphere, centered on the satellite, with radius equal to the pseudorange. The position of the receiver is somewhere on the surface of this sphere. Thus with four satellites, the indicated position of the GPS receiver is at or near the intersection of the surfaces of four spheres. In the ideal case of no errors, the GPS receiver would be at a precise intersection of the four surfaces.

If the surfaces of two spheres intersect at more than one point, they intersect in a circle. The trilateration shows this mathematically. A figure, *Two Sphere Surfaces Intersecting in a Circle*, is shown below. Two points where the surfaces of the spheres intersect are clearly shown in the figure. The distance between these two points is the diameter of the circle of intersection.



Two sphere surfaces intersecting in a circle

The intersection of a third spherical surface with the first two will be its intersection with that circle; in most cases of practical interest, this means they intersect at two points. Another figure, *Surface of Sphere Intersecting a Circle (not a solid disk) at Two Points*, illustrates the intersection. The two intersections are marked with dots. Again the trilateration clearly shows this mathematically.



Surface of sphere Intersecting a circle (not a solid disk) at two points

For automobiles and other near-earth vehicles, the correct position of the GPS receiver is the intersection closest to the Earth's surface. For space vehicles, the intersection farthest from Earth may be the correct one.

The correct position for the GPS receiver is also the intersection closest to the surface of the sphere corresponding to the fourth satellite.

### **Correcting a GPS receiver's clock**

One of the most significant error sources is the GPS receiver's clock. Because of the very large value of the speed of light,  $c$ , the estimated distances from the GPS receiver to the satellites, the pseudoranges, are very sensitive to errors in the GPS receiver clock; for example an error of one microsecond (0.000 001 second) corresponds to an error of 300 metres (980 ft). This suggests that an extremely accurate and expensive clock is required for the GPS receiver to work. Because manufacturers prefer to build inexpensive GPS receivers for mass markets, the solution for this dilemma is based on the way sphere surfaces intersect in the GPS problem.

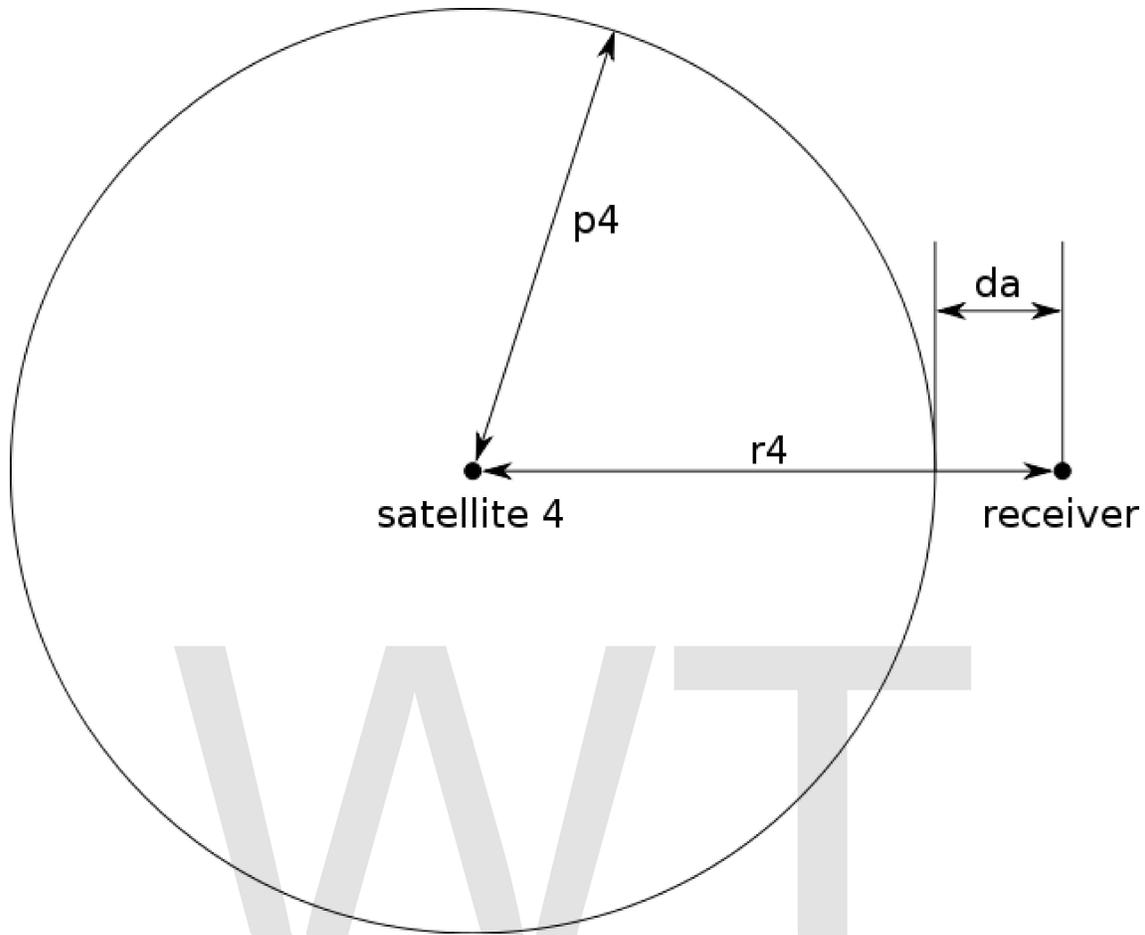


Diagram depicting satellite 4, sphere,  $p_4$ ,  $r_4$ , and  $da$

It is likely that the surfaces of the three spheres intersect, because the circle of intersection of the first two spheres is normally quite large, and thus the third sphere surface is likely to intersect this large circle. It is very unlikely that the surface of the sphere corresponding to the fourth satellite will intersect either of the two points of intersection of the first three, because any clock error could cause it to miss intersecting a point. However, the distance from the valid estimate of GPS receiver position to the surface of the sphere corresponding to the fourth satellite can be used to compute a clock correction. Let  $r_4$  denote the distance from the valid estimate of GPS receiver position to the fourth satellite and let  $p_4$  denote the pseudorange of the fourth satellite. Let  $da = r_4 - p_4$ .  $da$  is the distance from the computed GPS receiver position to the surface of the sphere corresponding to the fourth satellite. Thus the quotient,  $b_r = -da/c$ , provides an estimate of

(time indicated by the receiver's on-board clock) - (correct time),

and the GPS receiver clock can be advanced if  $b_r$  is positive or delayed if  $b_r$  is negative. However, it should be kept in mind that a less simple function of  $da$  may be needed to

estimate the time error in an iterative algorithm as discussed in the Navigation equations section.

## **Structure**

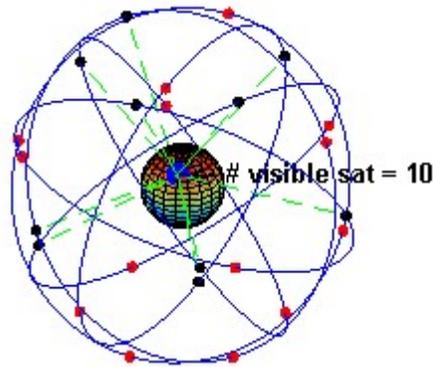
The current GPS consists of three major segments. These are the space segment (SS), a control segment (CS), and a user segment (U.S.). The U.S. Air Force develops, maintains, and operates the space and control segments. GPS satellites broadcast signals from space, and each GPS receiver uses these signals to calculate its three-dimensional location (latitude, longitude, and altitude) and the current time.

The space segment is composed of 24 to 32 satellites in medium Earth orbit and also includes the payload adapters to the boosters required to launch them into orbit. The control segment is composed of a master control station, an alternate master control station, and a host of dedicated and shared ground antennas and monitor stations. The user segment is composed of hundreds of thousands of U.S. and allied military users of the secure GPS Precise Positioning Service, and tens of millions of civil, commercial, and scientific users of the Standard Positioning Service.

## **Space segment**



Unlaunched GPS satellite on display at the San Diego Air & Space Museum



A visual example of the GPS constellation in motion with the Earth rotating. Notice how the number of *satellites in view* from a given point on the Earth's surface, in this example at 45°N, changes with time.

The space segment (SS) is composed of the orbiting GPS satellites, or Space Vehicles (SV) in GPS parlance. The GPS design originally called for 24 SVs, eight each in three circular orbital planes, but this was modified to six planes with four satellites each. The orbital planes are centered on the Earth, not rotating with respect to the distant stars. The six planes have approximately 55° inclination (tilt relative to Earth's equator) and are separated by 60° right ascension of the ascending node (angle along the equator from a reference point to the orbit's intersection). The orbits are arranged so that at least six satellites are always within line of sight from almost everywhere on Earth's surface. The result of this objective is that the four satellites are not evenly spaced (90 degrees) apart within each orbit. In general terms, the angular difference between satellites in each orbit is 30, 105, 120, and 105 degrees apart which, of course, sum to 360 degrees.

Orbiting at an altitude of approximately 20,200 kilometers (about 12,550 miles or 10,900 nautical miles; orbital radius of approximately 26,600 km (about 16,500 mi or 14,400 NM)), each SV makes two complete orbits each sidereal day, repeating the same ground track each day. This was very helpful during development because even with only four satellites, correct alignment means all four are visible from one spot for a few hours each day. For military operations, the ground track repeat can be used to ensure good coverage in combat zones.

As of March 2008, there are 31 actively broadcasting satellites in the GPS constellation, and two older, retired from active service satellites kept in the constellation as orbital spares. The additional satellites improve the precision of GPS receiver calculations by providing redundant measurements. With the increased number of satellites, the constellation was changed to a nonuniform arrangement. Such an arrangement was shown to improve reliability and availability of the system, relative to a uniform system, when multiple satellites fail. About eight satellites are visible from any point on the ground at any one time.

## Control segment



Ground monitor station used from 1984 to 2007, on display at the Air Force Space & Missile Museum

The control segment is composed of

1. a master control station (MCS),
2. an alternate master control station,
3. four dedicated ground antennas and
4. six dedicated monitor stations

The MCS can also access U.S. Air Force Satellite Control Network (AFSCN) ground antennas (for additional command and control capability) and NGA (National Geospatial-Intelligence Agency) monitor stations. The flight paths of the satellites are tracked by dedicated U.S. Air Force monitoring stations in Hawaii, Kwajalein, Ascension Island, Diego Garcia, Colorado Springs, Colorado and Cape Canaveral, along with shared NGA monitor stations operated in England, Argentina, Ecuador, Bahrain, Australia and Washington DC. The tracking information is sent to the Air Force Space Command's MCS at Schriever Air Force Base 25 km (16 miles) ESE of Colorado Springs, which is operated by the 2nd Space Operations Squadron (2 SOPS) of the U.S. Air Force. Then 2 SOPS contacts each GPS satellite regularly with a navigational update using dedicated or shared (AFSCN) ground antennas (GPS dedicated ground antennas are located at Kwajalein, Ascension Island, Diego Garcia, and Cape Canaveral). These updates synchronize the atomic clocks on board the satellites to within a few nanoseconds of each

other, and adjust the ephemeris of each satellite's internal orbital model. The updates are created by a Kalman filter that uses inputs from the ground monitoring stations, space weather information, and various other inputs.

Satellite maneuvers are not precise by GPS standards. So to change the orbit of a satellite, the satellite must be marked *unhealthy*, so receivers will not use it in their calculation. Then the maneuver can be carried out, and the resulting orbit tracked from the ground. Then the new ephemeris is uploaded and the satellite marked healthy again.

## User segment



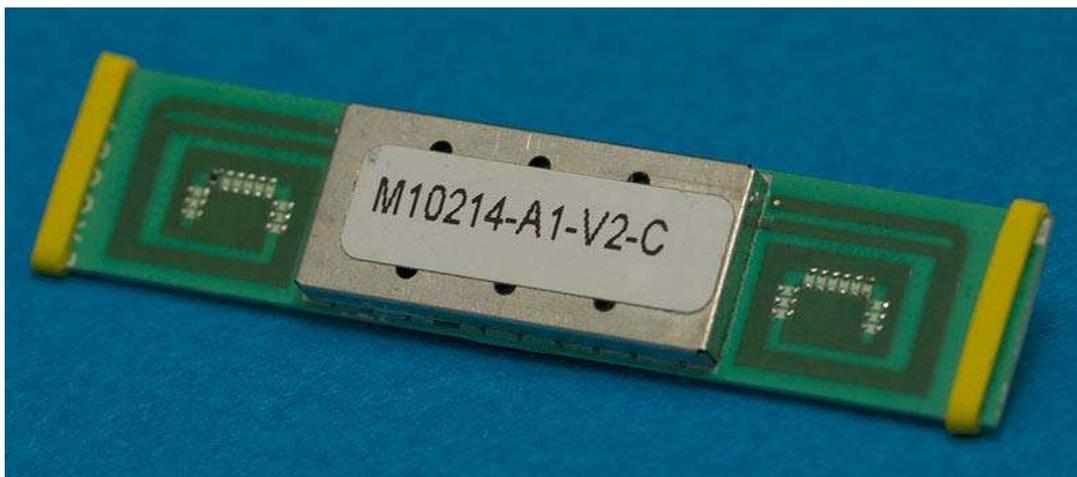
GPS receivers come in a variety of formats, from devices integrated into cars, phones, and watches, to dedicated devices such as those shown here from manufacturers Trimble, Garmin and Leica (left to right).

The user segment is composed of hundreds of thousands of U.S. and allied military users of the secure GPS Precise Positioning Service, and tens of millions of civil, commercial and scientific users of the Standard Positioning Service. In general, GPS receivers are composed of an antenna, tuned to the frequencies transmitted by the satellites, receiver-processors, and a highly stable clock (often a crystal oscillator). They may also include a display for providing location and speed information to the user. A receiver is often described by its number of channels: this signifies how many satellites it can monitor simultaneously. Originally limited to four or five, this has progressively increased over the years so that, as of 2007, receivers typically have between 12 and 20 channels.



A typical OEM GPS receiver module measuring 15×17 mm.

GPS receivers may include an input for differential corrections, using the RTCM SC-104 format. This is typically in the form of an RS-232 port at 4,800 bit/s speed. Data is actually sent at a much lower rate, which limits the accuracy of the signal sent using RTCM. Receivers with internal DGPS receivers can outperform those using external RTCM data. As of 2006, even low-cost units commonly include Wide Area Augmentation System (WAAS) receivers.



A typical GPS receiver with integrated antenna.

Many GPS receivers can relay position data to a PC or other device using the NMEA 0183 protocol. Although this protocol is officially defined by the National Marine Electronics Association (NMEA), references to this protocol have been compiled from public records, allowing open source tools like `gpsd` to read the protocol without violating intellectual property laws. Other proprietary protocols exist as well, such as the SiRF and MTK protocols. Receivers can interface with other devices using methods including a serial connection, USB, or Bluetooth.

## ***Applications***

While originally a military project, GPS is considered a *dual-use* technology, meaning it has significant military and civilian applications.

GPS has become a widely deployed and useful tool for commerce, scientific uses, tracking, and surveillance. GPS's accurate time facilitates everyday activities such as banking, mobile phone operations, and even the control of power grids by allowing well synchronized hand-off switching.



## Civilian



This antenna is mounted on the roof of a hut containing a scientific experiment needing precise timing.

Many civilian applications use one or more of GPS's three basic components: absolute location, relative movement, and time transfer.

- Cellular telephony: Clock synchronization enables time transfer, which is critical for synchronizing its spreading codes with other base stations to facilitate inter-cell handoff and support hybrid GPS/cellular position detection for mobile emergency calls and other applications. The first handsets with integrated GPS

launched in the late 1990s. The U.S. Federal Communications Commission (FCC) mandated the feature in either the handset or in the towers (for use in triangulation) in 2002 so emergency services could locate 911 callers. Third-party software developers later gained access to GPS APIs from Nextel upon launch, followed by Sprint in 2006, and Verizon soon thereafter.

- Disaster relief/emergency services: Depend upon GPS for location and timing capabilities.
- Geofencing: Vehicle tracking systems, person tracking systems, and pet tracking systems use GPS to locate a vehicle, person, or pet. These devices are attached to the vehicle, person, or the pet collar. The application provides continuous tracking and mobile or Internet updates should the target leave a designated area.
- Geotagging: Applying location coordinates to digital objects such as photographs and other documents for purposes such as creating map overlays.
- GPS Aircraft Tracking
- GPS tours: Location determines what content to display; for instance, information about an approaching point of interest.
- Map-making: Both civilian and military cartographers use GPS extensively.
- Navigation: Navigators value digitally precise velocity and orientation measurements.
- Phasor measurement units: GPS enables highly accurate timestamping of power system measurements, making it possible to compute phasors.
- Recreation: For example, geocaching, geodashing, GPS drawing and waymarking.
- Surveying: Surveyors use absolute locations to make maps and determine property boundaries.
- Tectonics: GPS enables direct fault motion measurement in earthquakes.

## **Restrictions on civilian use**

The U.S. Government controls the export of some civilian receivers. All GPS receivers capable of functioning above 18 kilometers (11 mi) altitude and 515 metres per second (1,001 kn) are classified as munitions (weapons) for which U.S. State Department export licenses are required. These limits attempt to prevent use of a receiver in a ballistic missile. They would not prevent use in a cruise missile because their altitudes and speeds are similar to those of ordinary aircraft.

This rule applies even to otherwise purely civilian units that only receive the L1 frequency and the C/A (Clear/Acquisition) code and cannot correct for Selective Availability (SA), etc.

Disabling operation above these limits exempts the receiver from classification as a munition. Vendor interpretations differ. The rule targets operation given the combination of altitude and speed, while some receivers stop operating even when stationary. This has

caused problems with some amateur radio balloon launches that regularly reach 30 kilometers (19 mi).

## Military



Attaching a GPS guidance kit to a 'dumb' bomb, March 2003.

As of 2009, military applications of GPS include:

- Navigation: GPS allows soldiers to find objectives, even in the dark or in unfamiliar territory, and to coordinate troop and supply movement. In the United States armed forces, commanders use the *Commanders Digital Assistant* and lower ranks use the *Soldier Digital Assistant*.
- Target tracking: Various military weapons systems use GPS to track potential ground and air targets before flagging them as hostile. These weapon systems pass target coordinates to precision-guided munitions to allow them to engage targets accurately. Military aircraft, particularly in air-to-ground roles, use GPS to find targets (for example, gun camera video from AH-1 Cobras in Iraq show GPS co-ordinates that can be viewed with special software.)
- Missile and projectile guidance: GPS allows accurate targeting of various military weapons including ICBMs, cruise missiles and precision-guided munitions. Artillery projectiles. Embedded GPS receivers able to withstand accelerations of 12,000 g or about 118 km/s<sup>2</sup> have been developed for use in 155 millimeters (6.1 in) howitzers.
- Search and Rescue: Downed pilots can be located faster if their position is known.
- Reconnaissance: Patrol movement can be managed more closely.

- GPS satellites carry a set of nuclear detonation detectors consisting of an optical sensor (Y-sensor), an X-ray sensor, a dosimeter, and an electromagnetic pulse (EMP) sensor (W-sensor), that form a major portion of the United States Nuclear Detonation Detection System.

## Communication

The navigational signals transmitted by GPS satellites encode a variety of information including satellite positions, the state of the internal clocks, and the health of the network. These signals are transmitted on two separate carrier frequencies that are common to all satellites in the network. Two different encodings are used, a public encoding that enables lower resolution navigation, and an encrypted encoding used by the U.S. military.

### Message format

GPS message format	
Subframes	Description
1	Satellite clock, GPS time relationship
2-3	Ephemeris (precise satellite orbit)
4-5	Almanac component (satellite network synopsis, error correction)

Each GPS satellite continuously broadcasts a *navigation message* at a rate of 50 bits per second. Each complete message is composed of 30-second frames, distinct groupings of 1,500 bits of information. Each frame is further subdivided into 5 subframes of length 6 seconds and with 300 bits each. Each subframe contains 10 words of 30 bits with length 0.6 seconds each. Each 30 second frame begins precisely on the minute or half minute as indicated by the atomic clock on each satellite.

The first part of the message encodes the week number and the time within the week, as well as the data about the health of the satellite. The second part of the message, the *ephemeris*, provides the precise orbit for the satellite. The last part of the message, the *almanac*, contains coarse orbit and status information for all satellites in the network as well as data related to error correction.

All satellites broadcast at the same frequencies. Signals are encoded using code division multiple access (CDMA) allowing messages from individual satellites to be distinguished from each other based on unique encodings for each satellite (that the receiver must be aware of). Two distinct types of CDMA encodings are used: the coarse/acquisition (C/A)

code, which is accessible by the general public, and the precise (P) code, that is encrypted so that only the U.S. military can access it.

The ephemeris is updated every 2 hours and is generally valid for 4 hours, with provisions for updates every 6 hours or longer in non-nominal conditions. The almanac is updated typically every 24 hours. Additionally data for a few weeks following is uploaded in case of transmission updates that delay data upload.

## Satellite frequencies

GPS frequency overview		
Band	Frequency	Description
L1	1575.42 MHz	Course-acquisition (C/A) and encrypted precision P(Y) codes, plus the L1 civilian (L1C) and military (M) codes on future Block III satellites.
L2	1227.60 MHz	P(Y) code, plus the L2C and military codes on the Block IIR-M and newer satellites.
L3	1381.05 MHz	Used for nuclear detonation (NUDET) detection.
L4	1379.913 MHz	Being studied for additional ionospheric correction.
L5	1176.45 MHz	Proposed for use as a civilian safety-of-life (SoL) signal.

All satellites broadcast at the same two frequencies, 1.57542 GHz (L1 signal) and 1.2276 GHz (L2 signal). The satellite network uses a CDMA spread-spectrum technique where the low-bitrate message data is encoded with a high-rate pseudo-random (PRN) sequence that is different for each satellite. The receiver must be aware of the PRN codes for each satellite to reconstruct the actual message data. The C/A code, for civilian use, transmits data at 1.023 million chips per second, whereas the P code, for U.S. military use, transmits at 10.23 million chips per second. The L1 carrier is modulated by both the C/A and P codes, while the L2 carrier is only modulated by the P code. The P code can be encrypted as a so-called P(Y) code that is only available to military equipment with a proper decryption key. Both the C/A and P(Y) codes impart the precise time-of-day to the user.

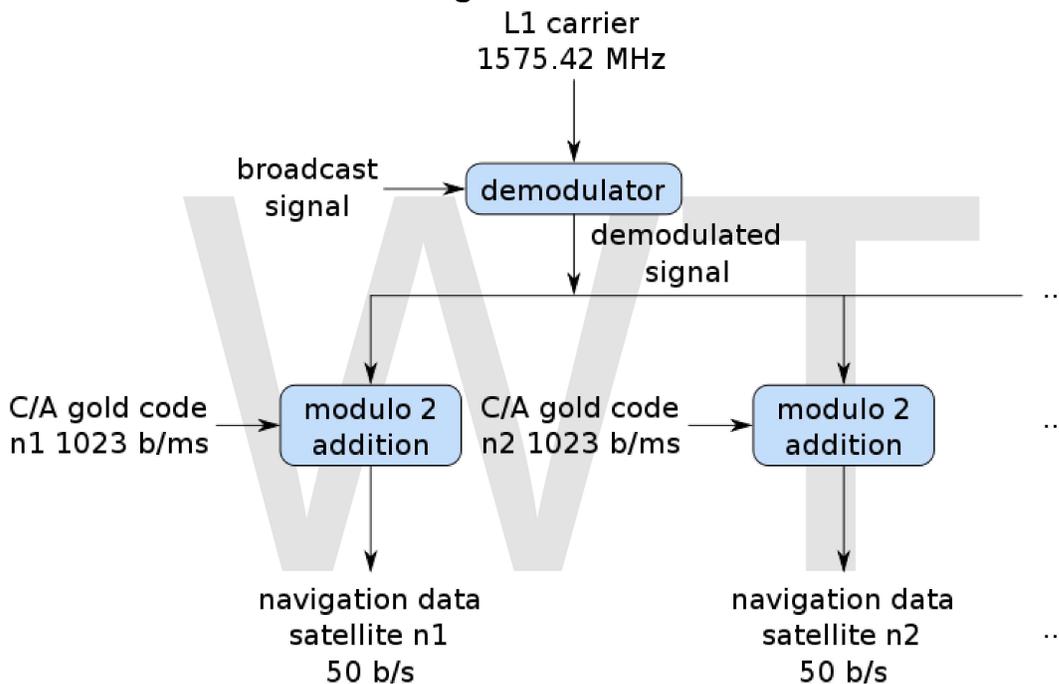
The L3 signal at a frequency of 1.38105 GHz is used by the United States Nuclear Detonation (NUDET) Detection System (USNDS) to detect, locate, and report nuclear detonations (NUDETs) in the Earth's atmosphere and near space. One usage is the enforcement of nuclear test ban treaties.

The L4 band at 1.379913 GHz is being studied for additional ionospheric correction.

The L5 frequency band at 1.17645 GHz was added in the process of GPS modernization. This frequency falls into an internationally protected range for aeronautical navigation, promising little or no interference under all circumstances. The first Block IIF satellite that would provide this signal is set to be launched in 2009. The L5 consists of two carrier components that are in phase quadrature with each other. Each carrier component is bi-phase shift key (BPSK) modulated by a separate bit train.

A waiver has recently been granted to LightSquared to operate a terrestrial broadband service in the L1 band. There is some concern that this will seriously degrade the GPS signal for many consumer uses.

## Demodulation and decoding



Demodulating and Decoding GPS Satellite Signals using the Coarse/Acquisition Gold code.

Because all of the satellite signals are modulated onto the same L1 carrier frequency, the signals must be separated after demodulation. This is done by assigning each satellite a unique binary sequence known as a Gold code. The signals are decoded after demodulation using addition of the Gold codes corresponding to the satellites monitored by the receiver.

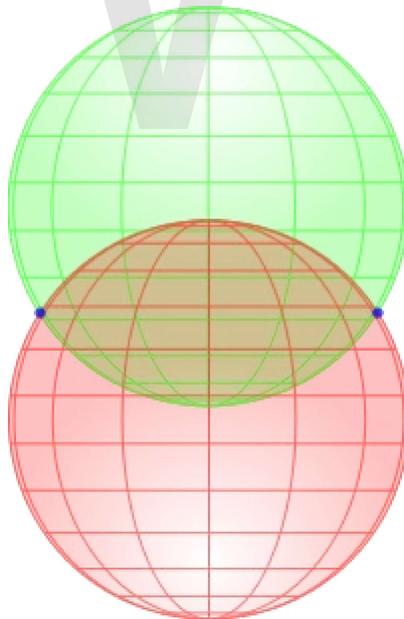
If the almanac information has previously been acquired, the receiver picks the satellites to listen for by their PRNs, unique numbers in the range 1 through 32. If the almanac information is not in memory, the receiver enters a search mode until a lock is obtained on one of the satellites. To obtain a lock, it is necessary that there be an unobstructed line of sight from the receiver to the satellite. The receiver can then acquire the almanac and determine the satellites it should listen for. As it detects each satellite's signal, it identifies

it by its distinct C/A code pattern. There can be a delay of up to 30 seconds before the first estimate of position because of the need to read the ephemeris data.

Processing of the navigation message enables the determination of the time of transmission and the satellite position at this time.

### ***Navigation equations***

The receiver uses messages received from four satellites to determine the satellite positions and time sent. The  $x$ ,  $y$ , and  $z$  components of position and the time sent are designated as  $[x_i, y_i, z_i, t_i]$  where the subscript  $i$  denotes the satellite and has the value 1, 2, 3, or 4. Knowing when the message was received  $t_r$ , the receiver computes the message's transit time as  $(t_r - t_i)$ . Assuming the message traveled at the speed of light ( $c$ ) the distance traveled is  $(t_r - t_i)c$ . Knowing the distance from receiver to satellite and the satellite's position implies that the receiver is on the surface of a sphere centered at the satellite's position. Thus the receiver is at or near the intersection of the surfaces of four spheres. In the ideal case of no errors, the receiver is at the intersection of the surfaces of four spheres. Excluding the unrealistic case (for GPS purposes) of two coincident spheres, the surfaces of two intersecting spheres is either a point (if they merely touch) or a circle as depicted in the illustration below. Two of the points where the surfaces of the spheres intersect are clearly marked on the figure. The distance between these two points is the diameter of the circle of intersection.

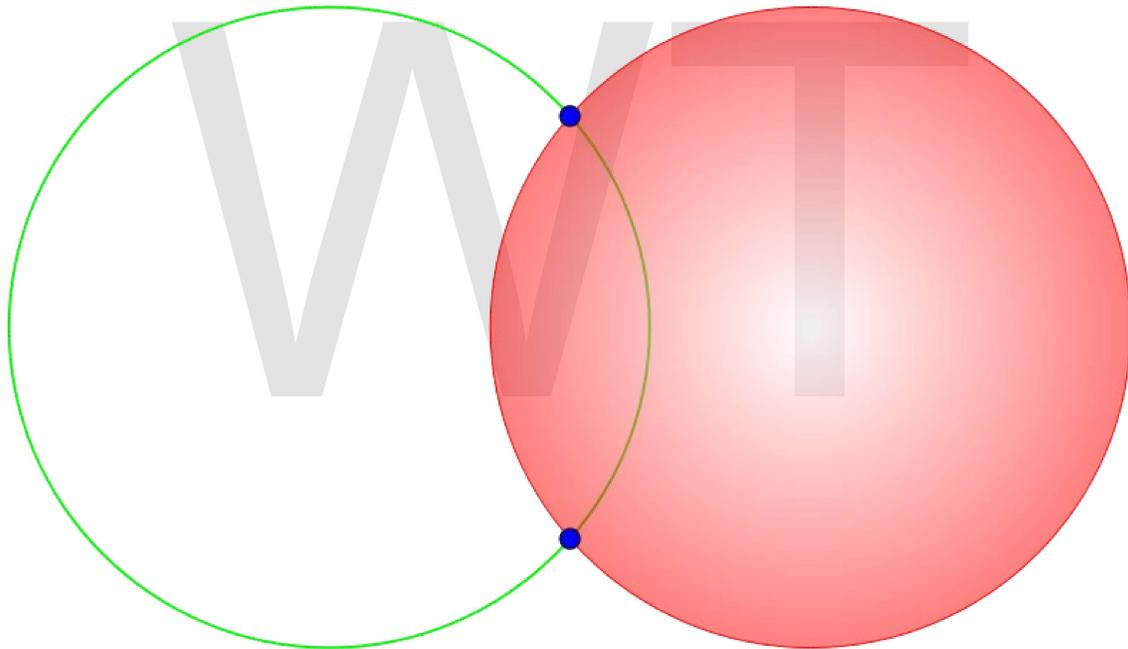


Two sphere surfaces intersecting in a circle

This can be seen more clearly by considering a side view of the intersecting spheres. This view would match the figure because of the symmetry of the spheres. A view from any

horizontal direction would look exactly the same. Therefore the diameter as seen from all directions is the same and thus the surfaces actually do intersect in a circle. The trilateration algebraically confirms this geometric argument that the two sphere surfaces intersect in a circle.

Having found that two sphere surfaces intersect in a circle, we now consider how the intersection of the first two sphere surfaces, the circle, intersect with the third sphere. A circle and sphere surface intersect at zero, one or two points. For the GPS problem we are concerned with the case of two points of intersection. Another figure, Surface of Sphere Intersecting a Circle (not a solid disk) at Two Points, is shown below to aid in visualizing this intersection. Trilateration algebraically confirms this geometric observation. The ambiguity of two points of intersection of three sphere surfaces can be resolved by noting the point that is closest to the fourth sphere surface.



Surface of a sphere intersecting a circle (i.e., the edge of a disk) at two points

Having provided a discussion of how sphere surfaces intersect, we now formulate the equations for the case when errors are present.

Let  $b$  denote the clock error or bias, the amount that the receiver's clock is off. The receiver has four unknowns, the three components of GPS receiver position and the clock

bias  $[x, y, z, b]$ . The equation of the sphere surfaces are given by:

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = ([t_r + b - t_i]c)^2,$$

$$i = 1, 2, 3, 4$$

Another useful form of these equations is in terms of *pseudoranges*, which are the approximate ranges based on the receiver clock's uncorrected time so that  $p_i = (t_r - t_i) c$ . Then the equations becomes:

$$p_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} - bc, \quad i = 1, 2, 3, 4.$$

## Methods of solution of navigation equations

The navigation equations can be solved by an algebraic method, called the Bancroft Method or by numerical methods involving trilateration or multidimensional root finding.

### Bancroft's method

Bancroft's method is perhaps the most important method of solving the navigation equations because it involves an algebraic as opposed to numerical method. The method requires at least four satellites but more can be used.

### Trilateration

The receiver can use trilateration and one dimensional numerical root finding. Trilateration is used to determine the intersection of the surfaces of three spheres. In the usual case of two intersections, the point nearest the surface of the sphere corresponding to the fourth satellite is chosen. The Earth's surface can also sometimes be used instead, especially by civilian GPS receivers, because it is illegal in the United States to track vehicles more than 60,000 feet (18,000 m) in altitude. Let  $d_4$  denote the signed magnitude of the vector from the receiver position to the fourth satellite (i.e.  $d_4 = r_4 - p_4$ ) as defined in the section, Correcting a GPS receiver's clock.  $d_4$  is a function of the correction because the correction changes the satellite transmission times and thus the pseudoranges. The notation,  $d_4(\text{correction})$  denotes this function. The problem is to determine the correction such that

$$d_4(\text{correction}) = 0.$$

This is the familiar problem of finding the zeroes of a one dimensional non-linear function of a scalar variable. Iterative numerical methods, such as those found in the chapter on root finding in *Numerical Recipes* can solve this type of problem. One advantage of this method is that it involves one dimensional as opposed to multidimensional numerical root finding.

## Multidimensional Newton-Raphson calculations

- Alternatively, multidimensional root finding method such as Newton-Raphson method can be used. The approach is to linearize around an approximate solution, say  $[x_r^{(k)}, y_r^{(k)}, z_r^{(k)}, b_r^{(k)}]$  from iteration  $k$ , then solve four linear equations derived from the quadratic equations above to obtain  $[x_r^{(k+1)}, y_r^{(k+1)}, z_r^{(k+1)}, b_r^{(k+1)}]$ . The Newton-Raphson method is more rapidly convergent than other methods of numerical root finding. A disadvantage of this multidimensional root finding method as compared to single dimensional root finding is that, "There are no good general methods for solving systems of more than one nonlinear equations."
- When more than four satellites are available, the calculation can use the four best or more than four, considering number of channels, processing capability, and geometric dilution of precision (GDOP). Using more than four is an over-determined system of equations with no unique solution, which must be solved by least-squares or a similar technique. If all visible satellites are used, the results are as good as or better than using the four best. Errors can be estimated through the residuals. With each combination of four or more satellites, a GDOP factor can be calculated, based on the relative sky directions of the satellites used. As more satellites are picked up, pseudorange from various 4-way combinations can be processed to add more estimates to the location and clock offset. The receiver then takes the weighted average of these positions and clock offsets. After the final location and time are calculated, the location is expressed in a specific coordinate system such as latitude and longitude, using the WGS 84 geodetic datum or a country-specific system.
- Finally, results from other positioning systems such as GLONASS or the upcoming Galileo can be incorporated or used to check the result. (By design, these systems use the same frequency bands, so much of the receiver circuitry can be shared, though the decoding is different.)

## Error sources and analysis

Error analysis for the Global Positioning System is an important aspect for determining what errors and their magnitude are to be expected. GPS errors are affected by geometric dilution of precision and depend on signal arrival time errors, numerical errors, atmospheric effects, ephemeris errors, multipath errors and other effects.

## Accuracy enhancement and surveying

### Augmentation

Integrating external information into the calculation process can materially improve accuracy. Such augmentation systems are generally named or described based on how the information arrives. Some systems transmit additional error information (such as clock

drift, ephemeris, or ionospheric delay), others characterize prior errors, while a third group provides additional navigational or vehicle information.

Examples of augmentation systems include the Wide Area Augmentation System (WAAS), European Geostationary Navigation Overlay Service (EGNOS), Differential GPS, Inertial Navigation Systems (INS) and Assisted GPS.

## Precise monitoring

Accuracy can be improved through precise monitoring and measurement of existing GPS signals in additional or alternate ways.

The largest remaining error is usually the unpredictable delay through the ionosphere. The spacecraft broadcast ionospheric model parameters, but errors remain. This is one reason GPS spacecraft transmit on at least two frequencies, L1 and L2. Ionospheric delay is a well-defined function of frequency and the total electron content (TEC) along the path, so measuring the arrival time difference between the frequencies determines TEC and thus the precise ionospheric delay at each frequency.

Military receivers can decode the P(Y)-code transmitted on both L1 and L2. Without decryption keys, it is still possible to use a *codeless* technique to compare the P(Y) codes on L1 and L2 to gain much of the same error information. However, this technique is slow, so it is currently available only on specialized surveying equipment. In the future, additional civilian codes are expected to be transmitted on the L2 and L5 frequencies. Then all users will be able to perform dual-frequency measurements and directly compute ionospheric delay errors.

A second form of precise monitoring is called *Carrier-Phase Enhancement* (CPGPS). This corrects the error that arises because the pulse transition of the PRN is not instantaneous, and thus the correlation (satellite-receiver sequence matching) operation is imperfect. CPGPS uses the L1 carrier wave, which has a period of

$\frac{1 \text{ sec}}{1575.42 * 10^6} = 0.63475 \text{ nanoseconds} \approx 1 \text{ nanosecond}$ , which is about one-thousandth of the C/A Gold code bit period of

$\frac{1 \text{ sec}}{1023 * 10^3} = 977.5 \text{ nanosecond} \approx 1000 \text{ nanosecond}$ , to act as an additional clock signal and resolve the uncertainty. The phase difference error in the normal GPS amounts to 2–3 metres (6.6–9.8 ft) of ambiguity. CPGPS working to within 1% of perfect transition reduces this error to 3 centimeters (1.2 in) of ambiguity. By eliminating this error source, CPGPS coupled with DGPS normally realizes between 20–30 centimetres (7.9–12 in) of absolute accuracy.

*Relative Kinematic Positioning* (RKP) is a third alternative for a precise GPS-based positioning system. In this approach, determination of range signal can be resolved to a precision of less than 10 centimeters (3.9 in). This is done by resolving the number of

cycles that the signal is transmitted and received by the receiver by using a combination of differential GPS (DGPS) correction data, transmitting GPS signal phase information and ambiguity resolution techniques via statistical tests—possibly with processing in real-time (real-time kinematic positioning, RTK).

## **Timekeeping**

### **Timekeeping and leap seconds**

While most clocks are synchronized to Coordinated Universal Time (UTC), the atomic clocks on the satellites are set to *GPS time*. The difference is that GPS time is not corrected to match the rotation of the Earth, so it does not contain leap seconds or other corrections that are periodically added to UTC. GPS time was set to match Coordinated Universal Time (UTC) in 1980, but has since diverged. The lack of corrections means that GPS time remains at a constant offset with International Atomic Time (TAI) ( $TAI - GPS = 19$  seconds). Periodic corrections are performed on the on-board clocks to correct relativistic effects and keep them synchronized with ground clocks.

The GPS navigation message includes the difference between GPS time and UTC, which as of 2011 is 15 seconds because of the leap second added to UTC December 31, 2008. Receivers subtract this offset from GPS time to calculate UTC and specific timezone values. New GPS units may not show the correct UTC time until after receiving the UTC offset message. The GPS-UTC offset field can accommodate 255 leap seconds (eight bits) that, given the current period of the Earth's rotation (with one leap second introduced approximately every 18 months), should be sufficient to last until approximately the year 2300.

### **Timekeeping accuracy**

GPS time is accurate to about 14ns.

### **Timekeeping format**

As opposed to the year, month, and day format of the Gregorian calendar, the GPS date is expressed as a week number and a seconds-into-week number. The week number is transmitted as a ten-bit field in the C/A and P(Y) navigation messages, and so it becomes zero again every 1,024 weeks (19.6 years). GPS week zero started at 00:00:00 UTC (00:00:19 TAI) on January 6, 1980, and the week number became zero again for the first time at 23:59:47 UTC on August 21, 1999 (00:00:19 TAI on August 22, 1999). To determine the current Gregorian date, a GPS receiver must be provided with the approximate date (to within 3,584 days) to correctly translate the GPS date signal. To address this concern the modernized GPS navigation message uses a 13-bit field that only repeats every 8,192 weeks (157 years), thus lasting until the year 2137 (157 years after GPS week zero).

## Carrier phase tracking (surveying)

Another method that is used in surveying applications is carrier phase tracking. The period of the carrier frequency times the speed of light gives the wavelength, which is about 0.19 meters for the L1 carrier. Accuracy within 1% of wavelength in detecting the leading edge, reduces this component of pseudorange error to as little as 2 millimeters. This compares to 3 meters for the C/A code and 0.3 meters for the P code.

However, 2 millimeter accuracy requires measuring the total phase—the number of waves times the wavelength plus the fractional wavelength, which requires specially equipped receivers. This method has many surveying applications.

Triple differencing followed by numerical root finding, and a mathematical technique called least squares can estimate the position of one receiver given the position of another. First, compute the difference between satellites, then between receivers, and finally between epochs. Other orders of taking differences are equally valid. Detailed discussion of the errors is omitted.

The satellite carrier total phase can be measured with ambiguity as to the number of cycles. Let  $\phi(r_i, s_j, t_k)$  denote the phase of the carrier of satellite  $j$  measured by receiver  $i$  at time  $t_k$ . This notation shows the meaning of the subscripts  $i, j$ , and  $k$ . The receiver ( $r$ ), satellite ( $s$ ), and time ( $t$ ) come in alphabetical order as arguments of  $\phi$  and to balance readability and conciseness, let  $\phi_{i,j,k} = \phi(r_i, s_j, t_k)$  be a concise abbreviation. Also we define three functions,  $\Delta^r, \Delta^s, \Delta^t$ , which return differences between receivers, satellites, and time points, respectively. Each function has variables with three subscripts as its arguments. These three functions are defined below. If  $\alpha_{i,j,k}$  is a function of the three integer arguments,  $i, j$ , and  $k$  then it is a valid argument for the functions,  $\Delta^r, \Delta^s, \Delta^t$ , with the values defined as

$$\begin{aligned}\Delta^r(\alpha_{i,j,k}) &= \alpha_{i+1,j,k} - \alpha_{i,j,k}, \\ \Delta^s(\alpha_{i,j,k}) &= \alpha_{i,j+1,k} - \alpha_{i,j,k}, \text{ and} \\ \Delta^t(\alpha_{i,j,k}) &= \alpha_{i,j,k+1} - \alpha_{i,j,k}.\end{aligned}$$

Also if  $\alpha_{i,j,k}$  and  $\beta_{l,m,n}$  are valid arguments for the three functions and  $a$  and  $b$  are constants then  $(a \alpha_{i,j,k} + b \beta_{l,m,n})$  is a valid argument with values defined as

$$\begin{aligned}\Delta^r(a \alpha_{i,j,k} + b \beta_{l,m,n}) &= a \Delta^r(\alpha_{i,j,k}) + b \Delta^r(\beta_{l,m,n}), \\ \Delta^s(a \alpha_{i,j,k} + b \beta_{l,m,n}) &= a \Delta^s(\alpha_{i,j,k}) + b \Delta^s(\beta_{l,m,n}), \text{ and} \\ \Delta^t(a \alpha_{i,j,k} + b \beta_{l,m,n}) &= a \Delta^t(\alpha_{i,j,k}) + b \Delta^t(\beta_{l,m,n}).\end{aligned}$$

Receiver clock errors can be approximately eliminated by differencing the phases measured from satellite 1 with that from satellite 2 at the same epoch. This difference is designated as  $\Delta^s(\phi_{1,1,1}) = \phi_{1,2,1} - \phi_{1,1,1}$

Double differencing computes the difference of receiver 1's satellite difference from that of receiver 2. This approximately eliminates satellite clock errors. This double difference is:

$$\Delta^r(\Delta^s(\phi_{1,1,1})) = \Delta^r(\phi_{1,2,1} - \phi_{1,1,1}) = \Delta^r(\phi_{1,2,1}) - \Delta^r(\phi_{1,1,1}) = (\phi_{2,2,1} - \phi_{1,2,1}) - (\phi_{2,1,1} - \phi_{1,1,1})$$

Triple differencing subtracts the receiver difference from time 1 from that of time 2. This eliminates the ambiguity associated with the integral number of wave lengths in carrier phase provided this ambiguity does not change with time. Thus the triple difference result eliminates practically all clock bias errors and the integer ambiguity. Atmospheric delay and satellite ephemeris errors have been significantly reduced. This triple difference is:

$$\Delta^t(\Delta^r(\Delta^s(\phi_{1,1,1})))$$

Triple difference results can be used to estimate unknown variables. For example if the position of receiver 1 is known but the position of receiver 2 unknown, it may be possible to estimate the position of receiver 2 using numerical root finding and least squares. Triple difference results for three independent time pairs quite possibly will be sufficient to solve for receiver 2's three position components. This may require the use of a numerical procedure. An approximation of receiver 2's position is required to use such a numerical method. This initial value can probably be provided from the navigation message and the intersection of sphere surfaces. Such a reasonable estimate can be key to successful multidimensional root finding. Iterating from three time pairs and a fairly good initial value produces one observed triple difference result for receiver 2's position. Processing additional time pairs can improve accuracy, overdetermining the answer with multiple solutions. Least squares can estimate an overdetermined system. Least squares determines the position of receiver 2 which best fits the observed triple difference results for receiver 2 positions under the criterion of minimizing the sum of the squares.

## ***Other systems***

Other satellite navigation systems in use or various states of development include:

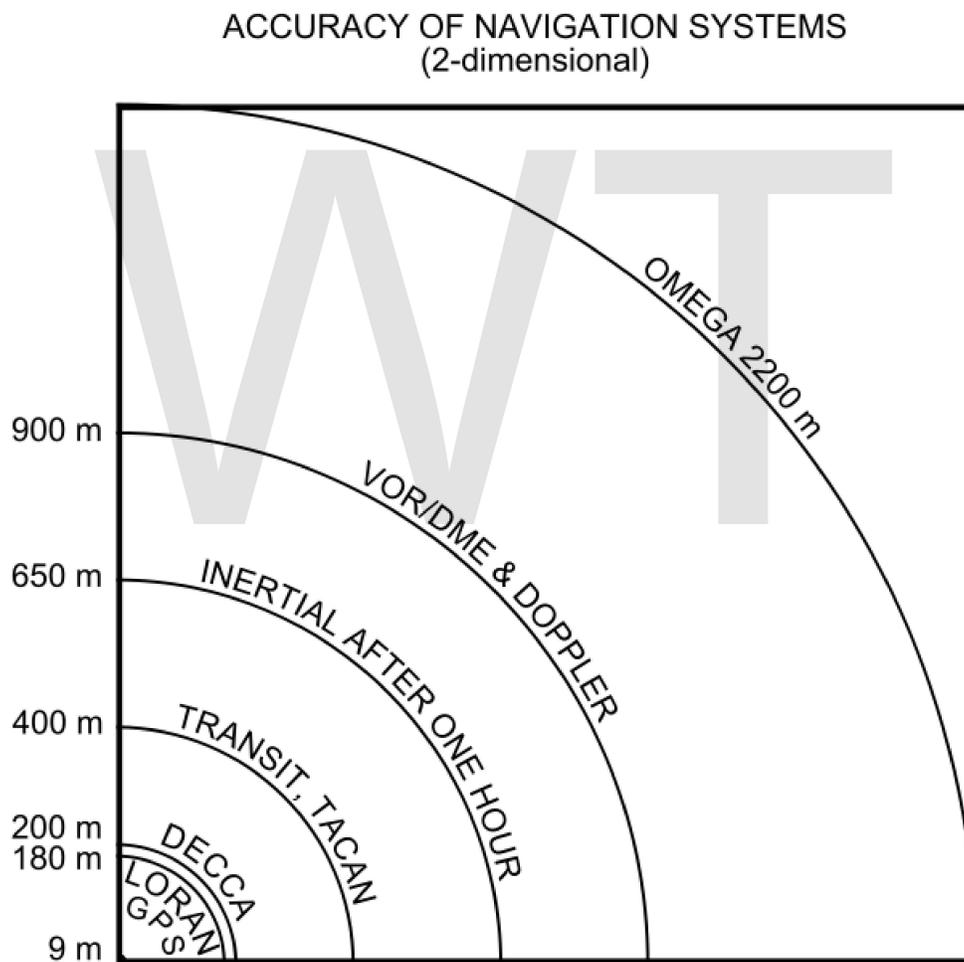
-  Galileo – a global system being developed by the European Union and other partner countries, planned to be operational by 2014
-  Beidou – People's Republic of China's regional system, covering Asia and the West Pacific
-  COMPASS – People's Republic of China's global system, planned to be operational by 2020
-  GLONASS – Russia's global navigation system

-  IRNSS – India's regional navigation system, planned to be operational by 2012, covering India and Northern Indian Ocean
-  QZSS – Japanese regional system covering Asia and Oceania

WWT

## Chapter 8

# Inertial Navigation System



An **inertial navigation system** (INS) is a navigation aid that uses a computer, motion sensors (accelerometers) and rotation sensors (gyroscopes) to continuously calculate via dead reckoning the position, orientation, and velocity (direction and speed of movement) of a moving object without the need for external references. It is used on vehicles such as ships, aircraft, submarines, guided missiles, and spacecraft. Other terms used to refer to

inertial navigation systems or closely related devices include **inertial guidance system**, **inertial reference platform**, **inertial instrument**, and many other variations.

## **Overview**

An inertial navigation system includes at least a computer and a platform or module containing accelerometers, gyroscopes, or other motion-sensing devices. The INS is initially provided with its position and velocity from another source (a human operator, a GPS satellite receiver, etc.), and thereafter computes its own updated position and velocity by integrating information received from the motion sensors. The advantage of an INS is that it requires no external references in order to determine its position, orientation, or velocity once it has been initialized.

An INS can detect a change in its geographic position (a move east or north, for example), a change in its velocity (speed and direction of movement), and a change in its orientation (rotation about an axis). It does this by measuring the linear and angular accelerations applied to the system. Since it requires no external reference (after initialization), it is immune to jamming and deception.

Inertial-navigation systems are used in many different moving objects, including vehicles—such as aircraft, submarines, spacecraft—and guided missiles. However, their cost and complexity place constraints on the environments in which they are practical for use.

Gyroscopes measure the angular velocity of the system in the inertial reference frame. By using the original orientation of the system in the inertial reference frame as the initial condition and integrating the angular velocity, the system's current orientation is known at all times. This can be thought of as the ability of a blindfolded passenger in a car to feel the car turn left and right or tilt up and down as the car ascends or descends hills. Based on this information alone, he knows what direction the car is facing but not how fast or slow it is moving, or whether it is sliding sideways.

Accelerometers measure the linear acceleration of the system in the inertial reference frame, but in directions that can only be measured relative to the moving system (since the accelerometers are fixed to the system and rotate with the system, but are not aware of their own orientation). This can be thought of as the ability of a blindfolded passenger in a car to feel himself pressed back into his seat as the vehicle accelerates forward or pulled forward as it slows down; and feel himself pressed down into his seat as the vehicle accelerates up a hill or rise up out of his seat as the car passes over the crest of a hill and begins to descend. Based on this information alone, he knows how the vehicle is accelerating relative to itself, that is, whether it is accelerating forward, backward, left, right, up (toward the car's ceiling), or down (toward the car's floor) measured relative to the car, but not the direction relative to the Earth, since he did not know what direction the car was facing relative to the Earth when he felt the accelerations.

However, by tracking both the current angular velocity of the system and the current linear acceleration of the system measured relative to the moving system, it is possible to determine the linear acceleration of the system in the inertial reference frame. Performing integration on the inertial accelerations (using the original velocity as the initial conditions) using the correct kinematic equations yields the inertial velocities of the system, and integration again (using the original position as the initial condition) yields the inertial position. In our example, if the blindfolded passenger knew how the car was pointed and what its velocity was before he was blindfolded, and if he is able to keep track of both how the car has turned and how it has accelerated and decelerated since, he can accurately know the current orientation, position, and velocity of the car at any time.

## **Error**

All inertial navigation systems suffer from **integration drift**: small errors in the measurement of acceleration and angular velocity are integrated into progressively larger errors in velocity, which are compounded into still greater errors in position. Since the new position is calculated from the previous calculated position and the measured acceleration and angular velocity, these errors are cumulative and increase at a rate roughly proportional to the time since the initial position was input. Therefore the position must be periodically corrected by input from some other type of navigation system. The inaccuracy of a good-quality navigational system is normally less than 0.6 nautical miles per hour in position and on the order of tenths of a degree per hour in orientation.

Accordingly, inertial navigation is usually used to supplement other navigation systems, providing a higher degree of accuracy than is possible with the use of any single system. For example, if, in terrestrial use, the inertially tracked velocity is intermittently updated to zero by stopping, the position will remain precise for a much longer time, a so-called *zero velocity update*.

Control theory in general and Kalman filtering in particular, provide a theoretical framework for combining information from various sensors. One of the most common alternative sensors is a satellite navigation radio, such as GPS. By properly combining the information from an INS and the GPS system (GPS/INS), the errors in position and velocity are stable. Furthermore, INS can be used as a short-term fallback while GPS signals are unavailable, for example when a vehicle passes through a tunnel.

## **History**

Inertial navigation systems were originally developed for rockets. American rocket pioneer Robert Goddard experimented with rudimentary gyroscopic systems. Dr. Goddard's systems were of great interest to contemporary German pioneers including Wernher von Braun. The systems entered more widespread use with the advent of spacecraft, guided missiles, and commercial airliners.

Early German World War II V2 guidance systems combined two gyroscopes and a lateral accelerometer with a simple analog computer to adjust the azimuth for the rocket in flight. Analog computer signals were used to drive four external rudders on the tail fins for flight control. The GN&C (Guidance, Navigation, and Control) system for V2 provided many innovations as an integrated platform with closed loop guidance. At the end of the war Von Braun engineered the surrender of 500 of his top rocket scientists, along with plans and test vehicles, to the Americans. They arrived at Fort Bliss, Texas in 1945 under the provisions of Operation Paperclip and were subsequently moved to Huntsville, Alabama, in 1950 where they worked for U.S. military rocket research programs.

In the early 1950s, the US government wanted to insulate itself against over dependency on the German team for military applications. Among the areas that were domestically "developed" was missile guidance. In the early 1950s the MIT Instrumentation Laboratory (later to become the Charles Stark Draper Laboratory, Inc.) was chosen by the Air Force Western Development Division to provide a self-contained guidance system backup to Convair in San Diego for the new Atlas intercontinental ballistic missile (Construction and testing were completed by Arma Division of AmBosch Arma). The technical monitor for the MIT task was a young engineer named Jim Fletcher who later served as the NASA Administrator. The Atlas guidance system was to be a combination of an on-board autonomous system, and a ground-based tracking and command system. This was the beginning of a philosophic controversy, which, in some areas, remains unresolved. The self-contained system finally prevailed in ballistic missile applications for obvious reasons. In space exploration, a mixture of the two remains.

In the summer of 1952, Dr. Richard Battin and Dr. J. Halcombe "Hal" Laning, Jr., researched computational based solutions to guidance. Dr. Laning, with the help of Phil Hankins and Charlie Werner, initiated work on MAC, an algebraic programming language for the IBM 650, which was completed by early spring of 1958. MAC became the work-horse of the MIT lab. MAC is an extremely readable language having a three-line format, vector-matrix notations and mnemonic and indexed subscripts. Today's Space Shuttle (STS) language called HAL/S, (developed by Intermetrics, Inc.) is a direct offshoot of MAC. Since the principal architect of HAL was Jim Miller, who co-authored a report on the MAC system with Hal Laning, it is probable the Space Shuttle language is named for Laning and not, as some have suggested, for the electronic star of Stanley Kubrick's *2001: A Space Odyssey*.

Hal Laning and Richard Battin undertook the initial analytical work on the Atlas inertial guidance in 1954. Other key figures at Convair were Charlie Bossart, the Chief Engineer, and Walter Schweidetzky, head of the guidance group. Schweidetzky had worked with Wernher von Braun at Peenemuende during World War II.

The initial Delta guidance system assessed the difference in position from a reference trajectory. A velocity to be gained (VGO) calculation is made to correct the current trajectory with the objective of driving VGO to zero. The mathematics of this approach were fundamentally valid, but dropped because of the challenges in accurate inertial

guidance and analog computing power. The challenges faced by the Delta efforts were overcome by the Q system of guidance. The Q system's revolution was to bind the challenges of missile guidance (and associated equations of motion) in the matrix Q. The Q matrix represents the partial derivatives of the velocity with respect to the position vector. A key feature of this approach allowed for the components of the vector cross product to be used as the basic autopilot rate signals—a technique that became known as *cross-product steering*. The Q-system was presented at the first Technical Symposium on Ballistic Missiles held at the Ramo-Wooldridge Corporation in Los Angeles on June 21 and 22, 1956. The Q system was classified information through the 1960s. Derivations of this guidance are used for today's missiles.

## ***Guidance in Human spaceflight***

In Feb of 1961 NASA Awarded MIT a contract for preliminary design study of a guidance and navigation system for Apollo. MIT and the Delco Electronics Div. of General Motors Corp. were awarded the joint contract for design and production of the Apollo Guidance and Navigation systems for the Command Module and the Lunar Module. Delco produced the Inertial Measurement Units for these systems, Kollsman Instrument Corp. produced the Optical Systems, and the Apollo Guidance Computer was built by Raytheon under subcontract.

For the space shuttle, an open loop (no feedback) guidance is used to guide the shuttle from lift off until Solid Rocket Booster (SRB) separation. After SRB separation the primary space shuttle guidance is named PEG4 (Powered Explicit Guidance). PEG4 takes into account both the Q system and the predictor-corrector attributes of the original "Delta" System (PEG Guidance). Although many updates to the shuttle's navigation system have taken place over the last 30 years (ex. GPS in the OI-22 build), the guidance core of today's Shuttle GN&C system has evolved little. Within a manned system, there is a human interface needed for the guidance system. As Astronauts are the customer for the system, many new teams are formed that touch GN&C as it is a primary interface to "fly" the vehicle.

## ***Aircraft inertial guidance***

One example of a popular INS for commercial aircraft was the Delco Carousel, which provided partial automation of navigation in the days before complete flight management systems became commonplace. The Carousel allowed pilots to enter a series of waypoints, and then guided the aircraft from one waypoint to the next using an INS to determine aircraft position and velocity. Boeing Corporation subcontracted the Delco Electronics Div. of General Motors to design and build the first production Carousel systems for the early models (-100, -200, and -300) models of the 747 aircraft. The 747 utilized three Carousel systems operating in concert for reliability purposes. The Carousel system and derivatives thereof were subsequently adopted for use in many other commercial and military aircraft. The USAF C-141 was the first military aircraft to utilize the Carousel in a dual system configuration, followed by the C-5A which utilized

the triple INS configuration, similar to the 747. The KC-135 fleet was fitted with a dual Carousel system that was aided by a Doppler radar.

***Inertial navigation systems in detail***





Inertial navigation unit of french IRBM S3.

INNs have angular and linear accelerometers (for changes in position); some include a gyroscopic element (for maintaining an absolute angular reference).

Angular accelerometers measure how the vehicle is rotating in space. Generally, there's at least one sensor for each of the three axes: pitch (nose up and down), yaw (nose left and right) and roll (clockwise or counter-clockwise from the cockpit).

Linear accelerometers measure non-gravitational accelerations of the vehicle. Since it can move in three axes (up & down, left & right, forward & back), there is a linear accelerometer for each axis.

A computer continually calculates the vehicle's current position. First, for each of the six degrees of freedom ( $x, y, z$  and  $\theta_x, \theta_y$  and  $\theta_z$ ), it integrates over time the sensed acceleration, together with an estimate of gravity, to calculate the current velocity. Then it integrates the velocity to calculate the current position.

Inertial guidance is difficult without computers. The desire to use inertial guidance in the Minuteman missile and Project Apollo drove early attempts to miniaturize computers.

Inertial guidance systems are now usually combined with satellite navigation systems through a digital filtering system. The inertial system provides short term data, while the satellite system corrects accumulated errors of the inertial system.

An inertial guidance system that will operate near the surface of the earth must incorporate Schuler tuning so that its platform will continue pointing towards the center of the earth as a vehicle moves from place to place.

## ***Basic schemes***

### **Gimballed gyrostabilized platforms**

Some systems place the linear accelerometers on a gimballed gyrostabilized platform. The gimbals are a set of three rings, each with a pair of bearings initially at right angles. They let the platform twist about any rotational axis (or, rather, they let the platform keep the same orientation while the vehicle rotates around it). There are two gyroscopes (usually) on the platform.

Two gyroscopes are used to cancel gyroscopic precession, the tendency of a gyroscope to twist at right angles to an input force. By mounting a pair of gyroscopes (of the same rotational inertia and spinning at the same speed) at right angles the precessions are cancelled, and the platform will resist twisting.

This system allows a vehicle's roll, pitch, and yaw angles to be measured directly at the bearings of the gimbals. Relatively simple electronic circuits can be used to add up the linear accelerations, because the directions of the linear accelerometers do not change.

The big disadvantage of this scheme is that it uses many expensive precision mechanical parts. It also has moving parts that can wear out or jam, and is vulnerable to gimbal lock. The primary guidance system of the Apollo spacecraft used a three-axis gyrostabilized platform, feeding data to the Apollo Guidance Computer. Maneuvers had to be carefully planned to avoid gimbal lock.

### **Fluid-suspended gyrostabilized platforms**

Gimbal lock constrains maneuvering, and it would be beneficial to eliminate the slip rings and bearings of the gimbals. Therefore, some systems use fluid bearings or a flotation chamber to mount a gyrostabilized platform. These systems can have very high precisions (e.g., Advanced Inertial Reference Sphere). Like all gyrostabilized platforms, this system runs well with relatively slow, low-power computers.

The fluid bearings are pads with holes through which pressurized inert gas (such as Helium) or oil press against the spherical shell of the platform. The fluid bearings are very slippery, and the spherical platform can turn freely. There are usually four bearing pads, mounted in a tetrahedral arrangement to support the platform.

In premium systems, the angular sensors are usually specialized transformer coils made in a strip on a flexible printed circuit board. Several coil strips are mounted on great circles around the spherical shell of the gyrostabilized platform. Electronics outside the platform uses similar strip-shaped transformers to read the varying magnetic fields produced by the transformers wrapped around the spherical platform. Whenever a magnetic field changes shape, or moves, it will cut the wires of the coils on the external transformer strips. The cutting generates an electric current in the external strip-shaped coils, and electronics can measure that current to derive angles.

Cheap systems sometimes use bar codes to sense orientations, and use solar cells or a single transformer to power the platform. Some small missiles have powered the platform with light from a window or optic fibers to the motor. A research topic is to suspend the platform with pressure from exhaust gases. Data is returned to the outside world via the transformers, or sometimes LEDs communicating with external photodiodes.

## **Strapdown systems**

Lightweight digital computers permit the system to eliminate the gimbals, creating *strapdown* systems, so called because their sensors are simply strapped to the vehicle. This reduces the cost, eliminates gimbal lock, removes the need for some calibrations, and increases the reliability by eliminating some of the moving parts. Angular rate sensors called *rate gyros* measure how the angular velocity of the vehicle changes.

A strapdown system has a dynamic measurement range several hundred times that required by a gimballed system. That is, it must integrate the vehicle's attitude changes in pitch, roll and yaw, as well as gross movements. Gimballed systems could usually do well with update rates of 50–60 Hz. However, strapdown systems normally update about 2000 Hz. The higher rate is needed to keep the maximum angular measurement within a practical range for real rate gyros: about 4 milliradians. Most rate gyros are now laser interferometers.

The data updating algorithms (*direction cosines* or quaternions) involved are too complex to be accurately performed except by digital electronics. However, digital computers are now so inexpensive and fast that rate gyro systems can now be practically used and mass-produced. The Apollo lunar module used a strapdown system in its backup Abort Guidance System (AGS).

Strapdown systems are nowadays commonly used in commercial and tactical applications (aircraft, missiles, etc.). However they are still not widespread in applications where superb accuracy is required (like submarine navigation or strategic ICBM guidance).

## **Motion-based alignment**

The orientation of a gyroscope system can sometimes also be inferred simply from its position history (e.g., GPS). This is, in particular, the case with planes and cars, where the velocity vector usually implies the orientation of the vehicle body.

For example, Honeywell's *Align in Motion* is an initialization process where the initialization occurs while the aircraft is moving, in the air or on the ground. This is accomplished using GPS and an inertial reasonableness test, thereby allowing commercial data integrity requirements to be met. This process has been FAA certified to recover pure INS performance equivalent to stationary align procedures for civilian flight times up to 18 hours. It avoids the need for gyroscope batteries on aircraft.

## **Vibrating gyros**

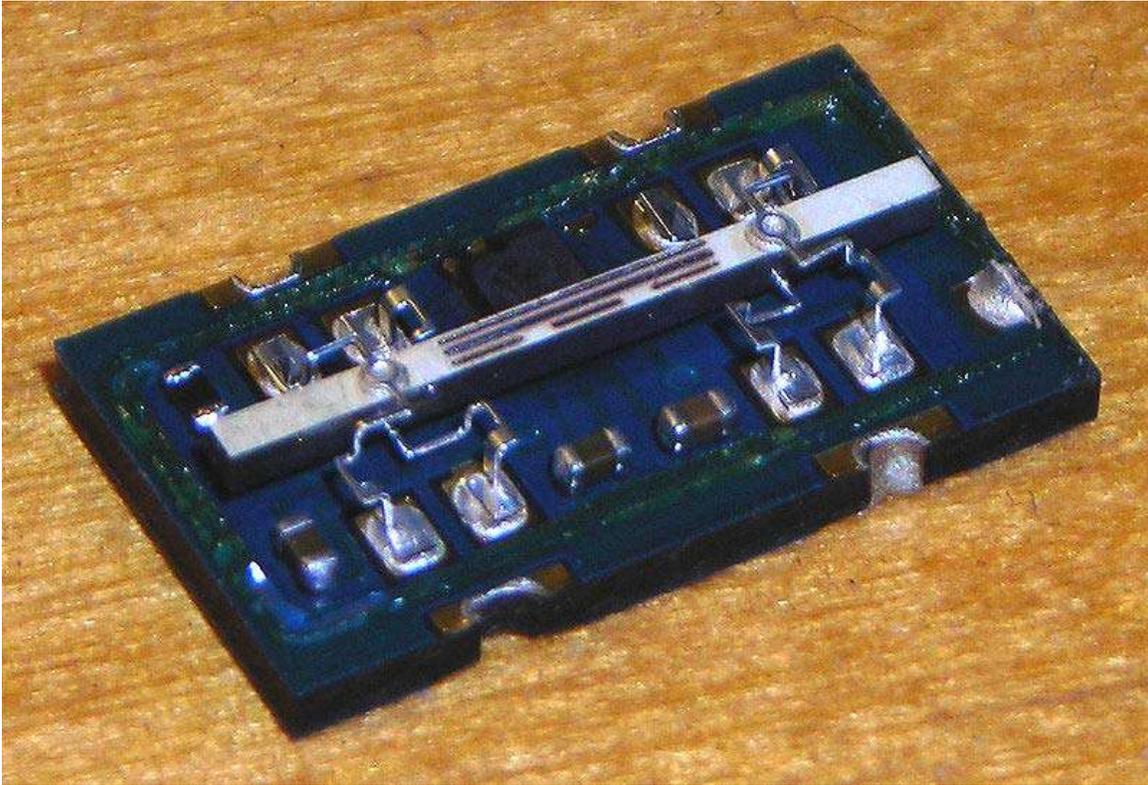
Less-expensive navigation systems, intended for use in automobiles, may use a vibrating structure gyroscope to detect changes in heading, and the odometer pickup to measure distance covered along the vehicle's track. This type of system is much less accurate than a higher-end INS, but it is adequate for the typical automobile application where GPS is the primary navigation system, and dead reckoning is only needed to fill gaps in GPS coverage when buildings or terrain block the satellite signals.

## **Hemispherical Resonator Gyros (wine glass or mushroom gyros)**

If a standing wave is induced in a hemispheric resonant cavity, and then the resonant cavity is rotated, the spherical harmonic standing wave rotates through an angle different than the quartz resonator structure due to the Coriolis force. The movement of the outer case with respect to the standing wave pattern is proportional to the total rotation angle and can be sensed by appropriate electronics. The system resonators are machined from quartz due to its excellent mechanical properties. The electrodes that drive and sense the standing waves are deposited directly onto separate quartz structures that surround the resonator. These gyros can operate in either a whole angle mode (which gives them nearly unlimited rate capability) or a force rebalance mode that holds the standing wave in a fixed orientation with respect to the gyro housing (which gives them much better accuracy).

This system has almost no moving parts, and is very accurate. However it is still relatively expensive due to the cost of the precision ground and polished hollow quartz hemispheres. Northrop Grumman currently manufactures Inertial Measurement Units for spacecraft that use HRGs. These IMUs have demonstrated extremely high reliability since their initial use in 1996.

## Quartz rate sensors



The quartz rate sensor inside an E-Sky model helicopter

This system is usually integrated on a silicon chip. It has two mass-balanced quartz tuning forks, arranged "handle-to-handle" so forces cancel. Aluminum electrodes evaporated onto the forks and the underlying chip both drive and sense the motion. The system is both manufacturable and inexpensive. Since quartz is dimensionally stable, the system can be accurate.

As the forks are twisted about the axis of the handle, the vibration of the tines tends to continue in the same plane of motion. This motion has to be resisted by electrostatic forces from the electrodes under the tines. By measuring the difference in capacitance between the two tines of a fork, the system can determine the rate of angular motion.

These products include 'tuning fork gyros'. Gyro is designed as an electronically-driven tuning fork, often fabricated out of a single piece of quartz or silicon. Such gyros operate in accordance with the dynamic theory that when an angle rate is applied to a translating body, a Coriolis force is generated.

Current state of the art non-military technology (as of 2005) can build small solid state sensors that can measure human body movements. These devices have no moving parts, and weigh about 50 grams.

Solid state devices using the same physical principles are used for image stabilization in small cameras or camcorders. These can be extremely small ( $\approx 5$  mm) and are built with Microelectromechanical systems (MEMS) technologies.

## MHD sensor

Sensors based on magnetohydrodynamic principles can be used to measure angular velocities.

## Laser gyros

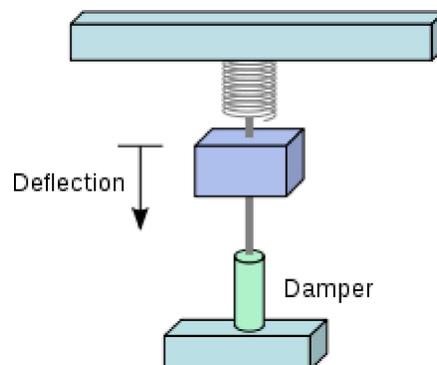
A ring laser gyro splits a beam of laser light into two beams in opposite directions through narrow tunnels in a closed optical circular path around the perimeter of a triangular block of temperature-stable Cervit glass with reflecting mirrors placed in each corner. When the gyro is rotating at some angular rate, the distance traveled by each beam becomes different—the shorter path being opposite to the rotation. The phase-shift between the two beams can be measured by an interferometer, and is proportional to the rate of rotation (Sagnac effect).

In practice, at low rotation rates the output frequency can drop to zero after the result of back scattering causing the beams to synchronise and lock together. This is known as a *lock-in*, or *laser-lock*. The result is that there is no change in the interference pattern, and therefore no measurement change.

To unlock the counter-rotating light beams, laser gyros either have independent light paths for the two directions (usually in fiber optic gyros), or the laser gyro is mounted on a piezo-electric dither motor that rapidly vibrates the laser ring back and forth about its input axis through the lock-in region to decouple the light waves.

The shaker is the most accurate, because both light beams use exactly the same path. Thus laser gyros retain moving parts, but they do not move as far.

## Pendular accelerometers



Principle of open loop accelerometer. Acceleration in the upward direction causes the mass to deflect downward.



## Chapter 9

# Octant (Instrument)



Octant. This instrument, labelled *Crichton - London, Sold by J Berry, Aberdeen*, appears to have an ebony frame with ivory scale, vernier and signature plate. The index arm and mirror supports are brass. Rather than use a sighting telescope, this instrument has a sighting pinnula.

The **octant**, also called **reflecting quadrant**, is a measuring instrument used primarily in navigation. It is a type of reflecting instrument.

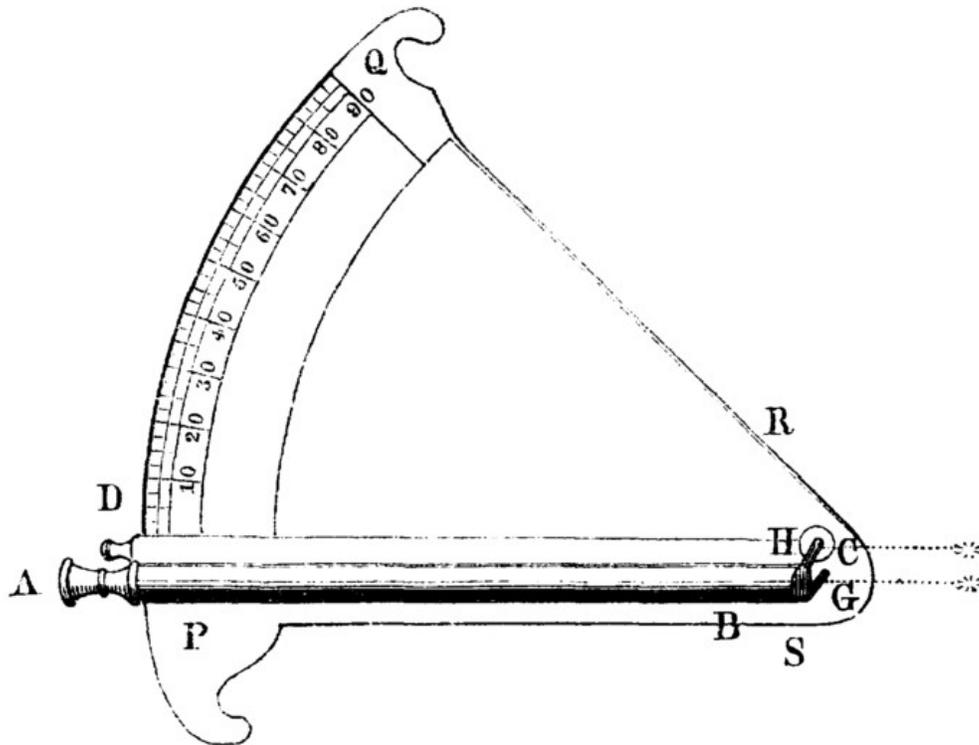
## Etymology

The name *octant* derives from the Latin *octans* meaning *eighth part of a circle*, because the instrument's arc is one eighth of a circle.

*Reflecting quadrant* derives from the instrument using mirrors to reflect the path of light to the observer and, in doing so, doubles the angle measured. This allows the instrument to use a one-eighth of a turn to measure a quarter-turn or quadrant.

## Origin of the octant

### Newton's reflecting quadrant



Drawing of Newton's Reflecting Quadrant. From *Memoirs of the Life, Writings, and Discoveries of Sir Isaac Newton* by David Brewster, 1855 A-B - sighting telescope

C-D - index arm

G - horizon mirror

H - index mirror

P-Q - graduated arc

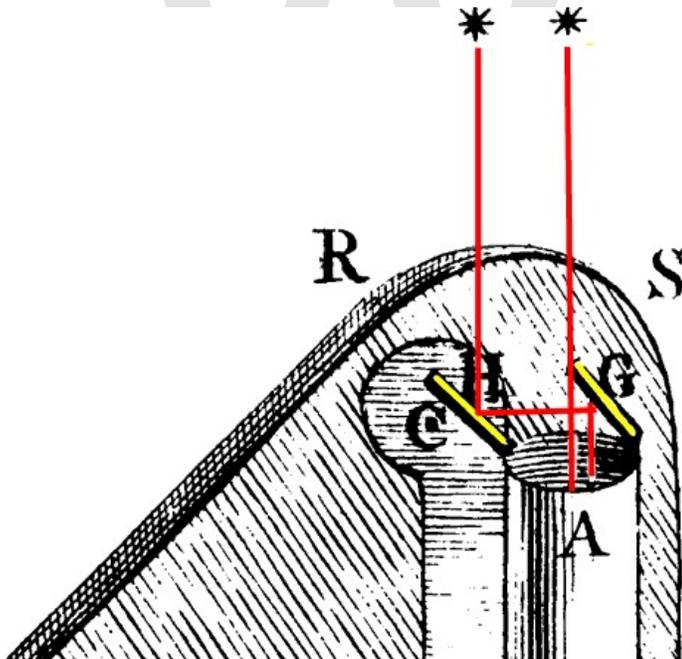
Isaac Newton's reflecting quadrant was invented around 1699. A detailed description of the instrument was given to Edmond Halley, but the description was not published until after Halley's death in 1742. It is not known why Halley did not publish the information during his life, however this prevents Newton from getting the credit for the invention that is generally given to John Hadley and Godfrey.

One copy of this instrument was constructed by Thomas Heath and may have been shown in Heath's shop window prior to its being published by the Royal Society in 1742.

Newton's instrument used two mirrors, but they were used in an arrangement somewhat different than the two mirrors found in modern octants and sextants. The diagram on the right shows the configuration of the instrument.

The  $45^\circ$  arc of the instrument (P-Q), was graduated with 90 divisions of a half-degree each. Each such division was subdivided into 60 parts and each part further divided into sixths. This results in the arc being marked in degrees, minutes and sixths of a minute (10 seconds). Thus the instrument could have readings interpolated to 5 seconds of arc. This fineness of graduation is only possible due to the large size of the instrument - the sighting telescope alone was three to four feet long.

A *sighting telescope* (A-B), three or four feet long, was mounted along one side of the instrument. A *horizon mirror*, was fixed at a  $45^\circ$  angle in front of the telescope's objective lens (G). This mirror was small enough to allow the observer to see the image in the mirror on one side and to see directly ahead on the other. The index arm (C-D) held an index mirror (H), also at  $45^\circ$  to the edge of the index arm. The reflective sides of the two mirrors nominally faced each other, so that the image seen in the first mirror is that reflected from the second.



Details of the mirrors on Newton's reflecting quadrant, showing the light paths (red) through the instrument. This image is turned  $90^\circ$  anticlockwise relative to the one above.

With the two mirrors parallel, the index reads  $0^\circ$ . The view through the telescope sees directly ahead on one side and the view from the mirror G sees the same image reflected from mirror H. When the index arm is moved from zero to a large value, the *index mirror*

reflects an image that is in a direction away from the direct line of sight. As the index arm movement increases, the line of sight for the index mirror moves toward S (to the right in the detail image). This shows a slight deficiency with this mirror arrangement. The horizon mirror will block the view of the index mirror at angles approaching  $90^\circ$ .

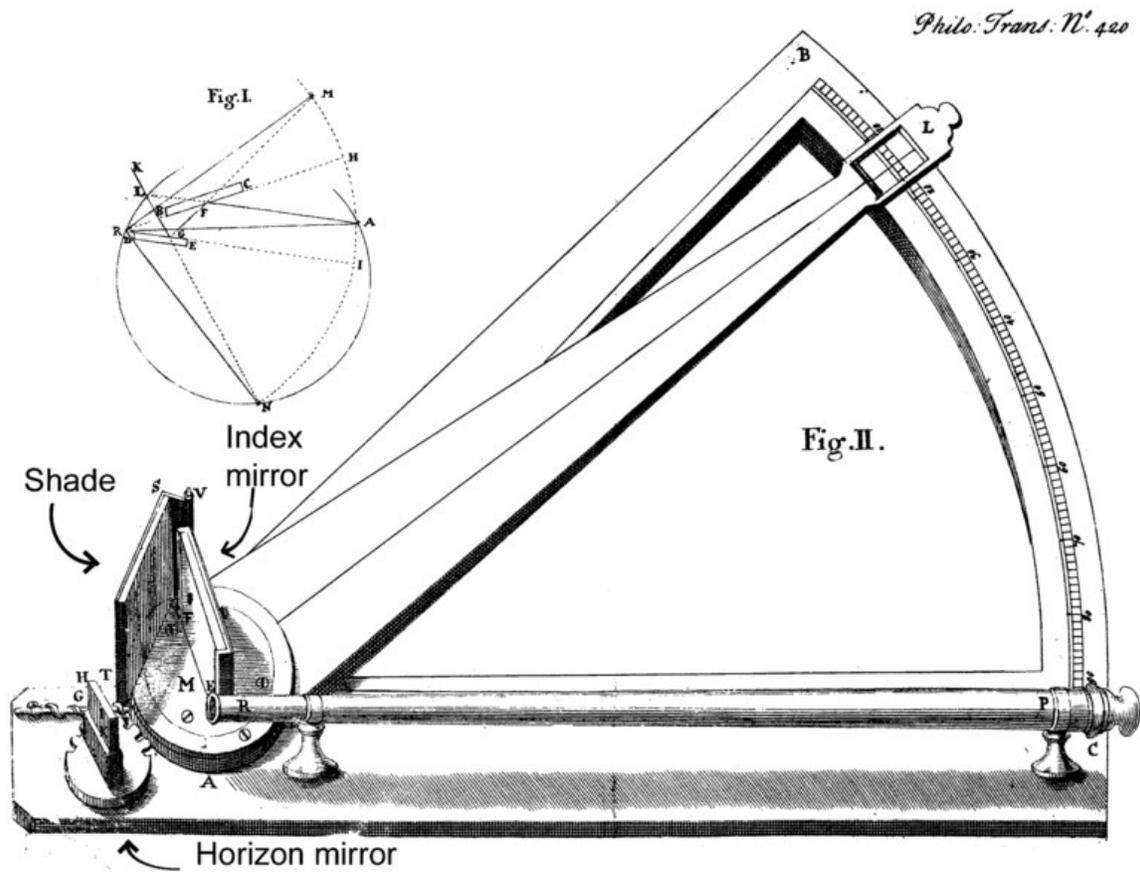
The length of the sighting telescope seems remarkable, given the small size of the telescopes on modern instruments. This was likely Newton's choice of a way to reduce chromatic aberrations. Short-focal length telescopes, prior to the development of achromatic lenses, produced an objectionable degree of aberration, so much so that it could affect the perception of a star's position. Long focal lengths were the solution, and this telescope would likely have had both a long-focal length objective lens and a long-focal length eyepiece. This would decrease aberrations without excessive magnification.

### **The inventors of the octant**

Two men independently developed the octant around 1730: John Hadley (1682–1744), an English mathematician, and Thomas Godfrey (1704–1749), a glazier in Philadelphia. While both have a legitimate and equal claim to the invention, Hadley generally gets the greater share of the credit. This reflects the central role that London and the Royal Society played in the history of scientific instruments in the eighteenth century.

Two others who created octants during this period were Caleb Smith, an English insurance broker with a strong interest in astronomy (in 1734), and Jean-Paul Fouchy, a mathematics professor and astronomer in France (in 1732).

## Hadley's versions



Hadley's reflecting quadrant. This instrument follows the form of Newton's reflecting quadrant from 1699

Hadley produced two versions of the reflecting quadrant. Only the second is well known and is the familiar octant.

### Hadley's reflecting quadrant

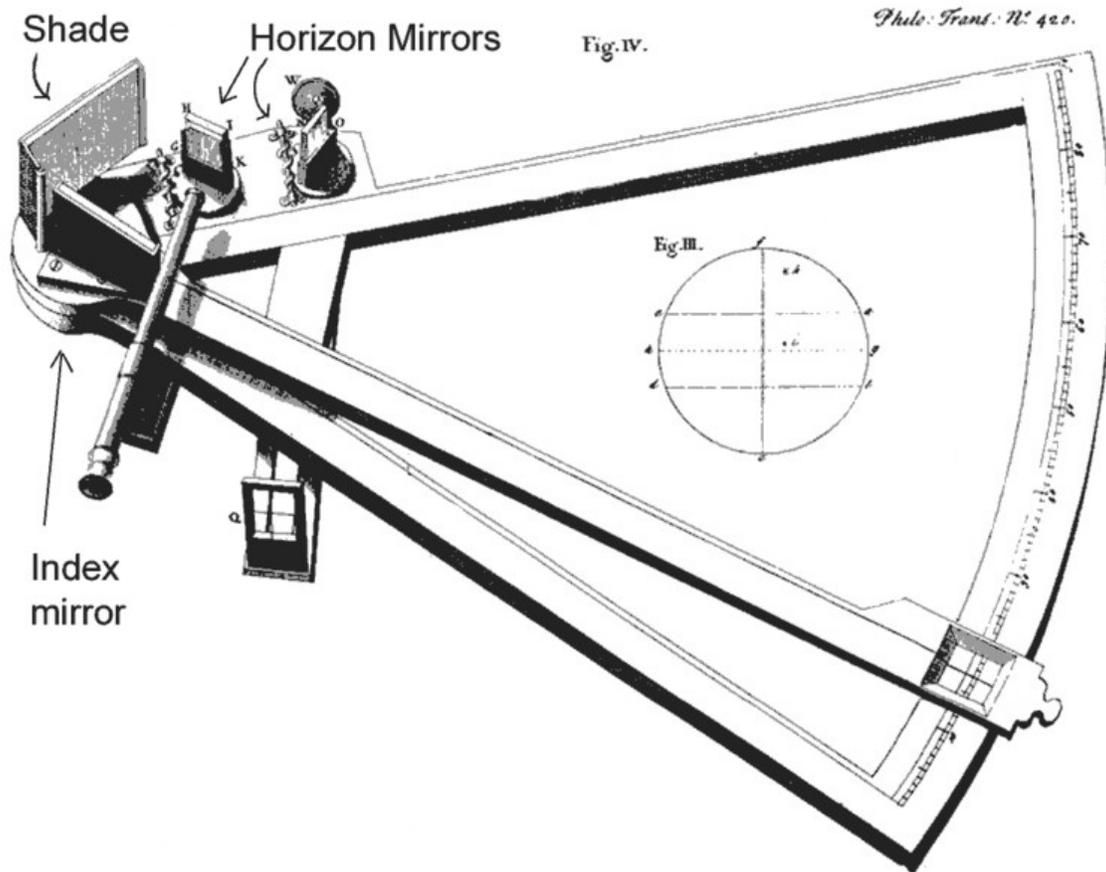
Hadley's first reflecting quadrant was a simple device with a frame spanning a  $45^\circ$  arc. In the image at the right, from Hadley's article in the Philosophical Transactions of the Royal Society, you can see the nature of his design. A small sighting telescope was mounted on the frame along one side. One large *index mirror* was mounted at the point of rotation of the index arm. A second, smaller *horizon mirror* was mounted on the frame in the line of sight of the telescope. The horizon mirror allows the observer to see the image of the index mirror in one half of the view and to see a distant object in the other half. A shade was mounted at the vertex of the instrument to allow one to observe a bright object. The shade pivots to allow it to move out of the way for stellar observations.

Observing through the telescope, the navigator would sight one object directly ahead. The second object would be seen by reflection in the horizon mirror. The light in the horizon

mirror is reflected from the index mirror. By moving the index arm, the index mirror can be made to reveal any object up to 90° from the direct line of sight. When both objects are in the same view, aligning them together allows the navigator to measure the angular distance between them.

Very few of the original reflecting quadrant designs were ever produced. One, constructed by Baradelle, is in the collection of the Musée de la Marine, Paris.

### Hadley's octant



Hadley's Octant. This is in the form familiar to those who have seen a sextant.

Hadley's second design had the form familiar to modern navigators. The image to the right, also taken from his Royal Society publication, shows the details.

He placed an *index mirror* on the index arm. Two *horizon mirrors* were provided. The upper mirror, in the line of the sighting telescope, was small enough to allow the telescope to see directly ahead as well as seeing the reflected view. The reflected view was that of the light from the index mirror. As in the previous instrument, the arrangement of the mirrors allowed the observer to simultaneously see an object straight ahead and to see one reflected in the index mirror to the horizon mirror and then into the

telescope. Moving the index arm allowed the navigator to see any object within 90° of the direct view.

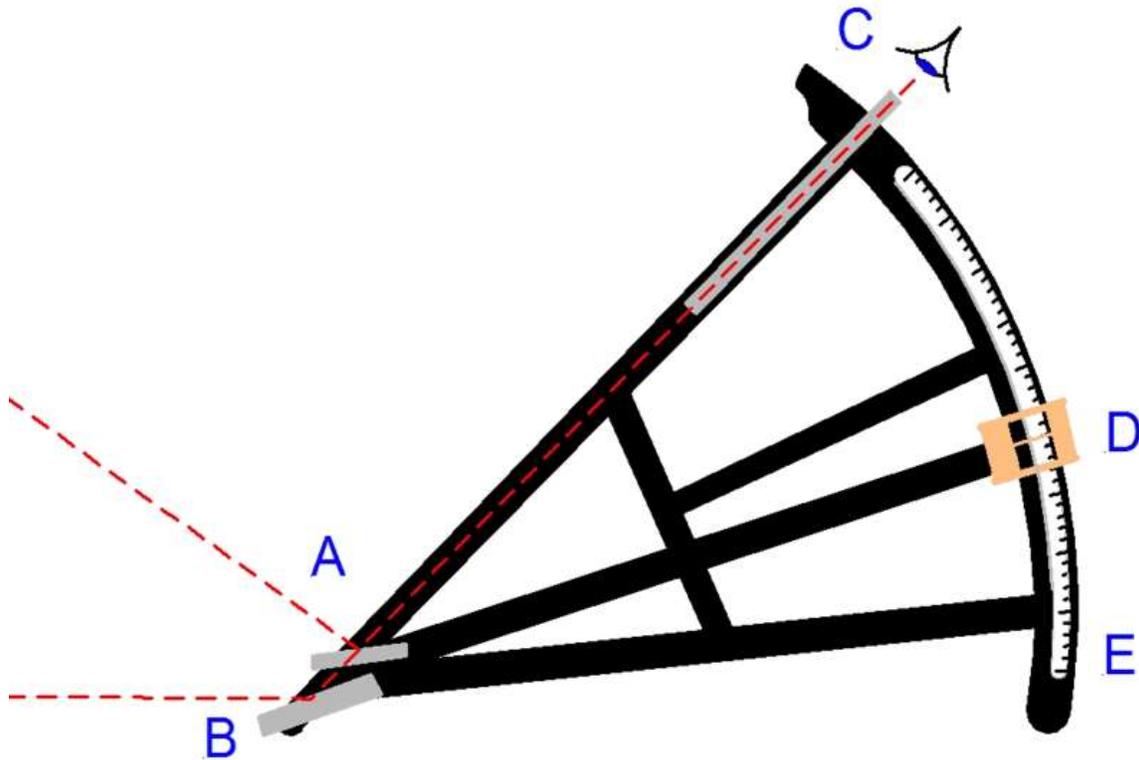
The significant difference with this design was that the mirrors allowed the instrument to be held vertically rather than horizontally and it provided more room for configuring the mirrors without suffering from mutual interference.

The second horizon mirror was an interesting innovation. The telescope was removable. It could be remounted so that the telescope viewed the second horizon mirror from the opposite side of the frame. By mounting the two horizon mirrors at right angles to each other and permitting the movement of the telescope, the navigator could measure angles from 0 to 90° with one horizon mirror and from 90° to 180° with the other. This made the instrument very versatile. For unknown reasons, this feature was not implemented on octants in general use.

Comparing this instrument to the photo of a typical octant at the top, one can see that the only significant differences in the more modern design are:

- The location of the horizon mirror and telescope or sighting pinnula is lower.
- The internal bracing of the frame is more central and robust.
- The position of the shades for the index mirror is in the path between the index and horizon mirrors rather than at the top of the instrument.
- Multiple shades are used to allow for different levels of shading.
- Separate shades are provided on the horizon mirror for sighting a low sun position with a very bright horizon.
- The second horizon mirror and accompanying alidade is not provided.

## Smith's Astroscope



Drawing of Smith's Astroscope or Sea-quadrant

Caleb Smith, an English insurance broker with a strong interest in astronomy, had created an octant in 1734. He called it an *Astroscope* or *Sea-Quadrant*. His used a fixed prism in addition to an index mirror to provide reflective elements. Prisms provide advantages over mirrors in an era when polished speculum metal mirrors were inferior and both the silvering of a mirror and the production of glass with flat, parallel surfaces was difficult.

In the drawing to the right, the horizon element (B) could be a mirror or a prism. On the index arm, the index mirror (A) rotated with the arm. A sighting telescope was mounted on the frame (C). The index did not use a vernier or other device at the scale (D). Smith called the instrument's index arm a *label*, in the manner of Elton for his mariner's quadrant.

Various design elements of Smith's instrument made it inferior to Hadley's octant and it was not used significantly. For example, one problem with the Astroscope was that angle of the observer's line of sight. By looking down, he had greater difficulty in observing than an orientation with his head in a normal orientation.

## ***Advantages of the octant***



Octant reverse side. This side is not seen in photographs very often. On the right, the thumbscrew to adjust the horizon mirror can be seen. At the top, one of the feet on which the octant rests in its case is just below the index arm axis. On the left the notepad is clearly visible. This small keystone-shaped piece of ivory, scarcely larger than a thumbnail, was used by the navigator to record his readings.

The octant provided a number of advantages over previous instruments.

The sight was easy to align because the horizon and the star seem to move together as the ship pitched and rolled. This also created a situation where the error in observation was less dependent on the observer, as he could directly see both objects at once.

With the use of the manufacturing techniques available in the 18<sup>th</sup> century, the instruments were capable of reading very accurately. The size of the instruments was reduced with no loss of accuracy. An octant could be half the size of a Davis quadrant with no increase in error.

Using shades over the light paths, one could observe the sun directly, while moving the shades out of the light path allowed the navigator to observe faint stars. This made the instrument usable both night and day.

By 1780, the octant and sextant had almost completely eliminated all previous instruments.

### ***Production of the octant***

Early octants were constructed primarily in wood, with later versions incorporating ivory and brass components. The earliest mirrors were polished metal, since the technology to produce silvered glass mirrors with flat, parallel surfaces was limited. As glass polishing techniques improved, glass mirrors began to be provided. These used coatings of mercury-containing tin amalgam; coatings of silver or aluminum were not available until the 19<sup>th</sup> century. The poor optical quality of the early polished speculum metal mirrors meant that telescopic sights were not practical. For that reason, most early octants employed a simple naked-eye sighting pinnula instead.



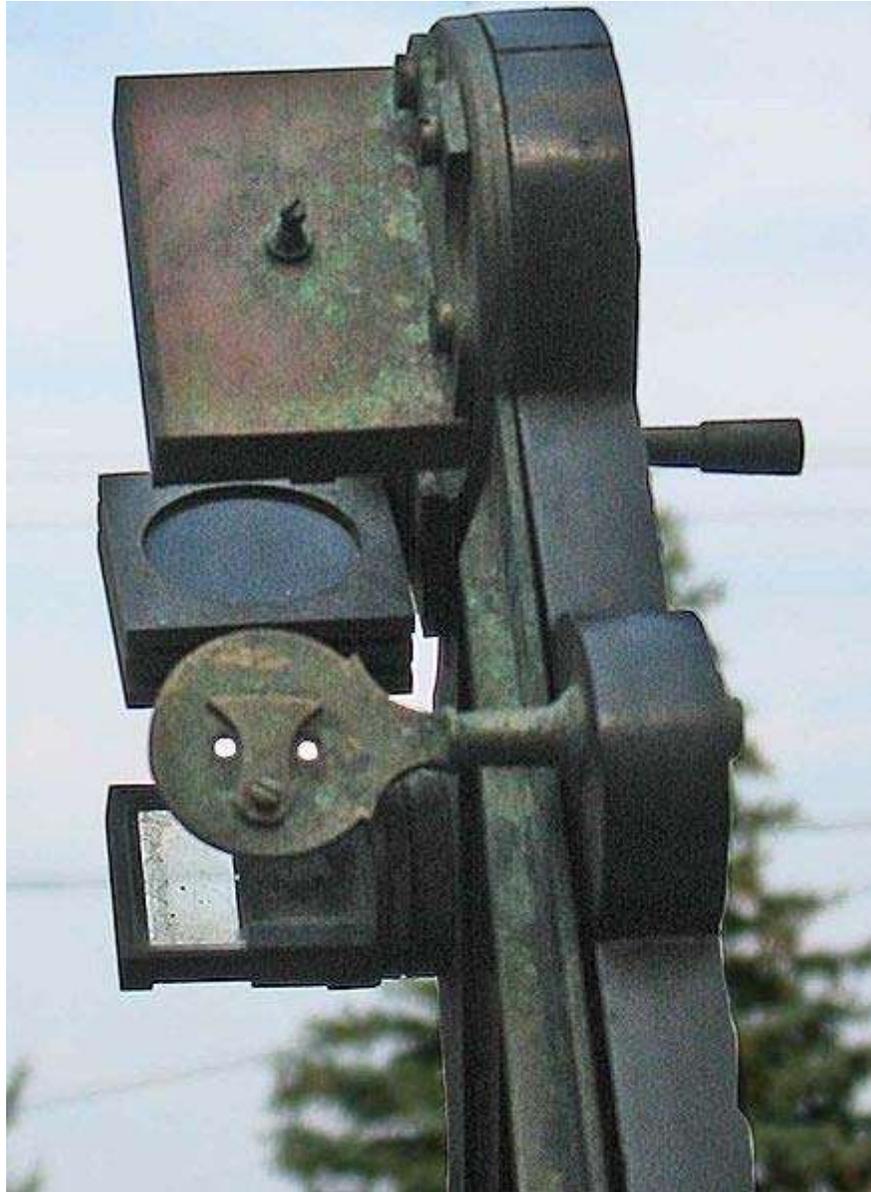
Details on an octant. This photo shows the graduated scale and the end of the index arm with the vernier. The thumbscrew used to lock the index arm position is seen below the index arm while the thumbscrew used for fine adjustment of the arm is on the left. To the

right of the value 50 on the main scale, the SBR logo is engraved. The scale is directly graduated in degrees and thirds of a degree (20'). The vernier can divide the 20' intervals to the nearest minute of arc.

Early octants retained some of the features common to backstaves, such as transversals on the scale. However, as engraved, they showed the instrument to have an apparent accuracy of only two minutes of arc while the backstaff appeared to be accurate to one minute. The use of the vernier scale allowed the scale to be read to one minute, so improved the marketability of the instrument. This and the ease in making verniers compared to transversals, lead to adoption of the vernier on octants produced later in the 18<sup>th</sup> century.

Octants were produced in large numbers. In wood and ivory, their relatively low price compared to an all-brass sextant made them a popular instrument. The design was standardized with many manufacturers using the identical frame style and components. Different shops could make different components, with woodworkers specializing in frames and others in the brass components. For example, Spencer, Browning and Rust, a manufacturer of scientific instruments in England from 1787 to 1840 (operating as *Spencer, Browning and Co.* after 1840) used a Ramsden dividing engine to produce graduated scales in ivory. These were widely used by others and the *SBR* initials could be found on octants from many other manufacturers.

Examples of these very similar octants are in the photos here. The image at the top is essentially the same instrument as the one in the detail photos. However, they are from two different instrument makers - the upper is labelled *Crichton - London, Sold by J Berry Aberdeen* while the detail images are of an instrument from *Spencer, Browning & Co. London*. The only obvious difference is the presence of horizon shades on the Crichton octant that are not on the other.



Octant details showing the double-holed sighting pinulla. Also visible is the small cover that can block one or the other of the holes. The horizon mirror is on the opposite side of the instrument. The left side is transparent while the tin amalgam on the mirrored side has completely corroded and no longer reflects light. The back of the index mirror's holder is at the top and the three circular glass shades in square frames are between the two mirrors.

These octants were available with many options. A basic octant with graduations directly on the wood frame were least expensive. These dispensed with a telescopic sight, using a single- or double-holed sighting pinnula instead. Ivory scales would increase the price, as would the use of a brass index arm or a vernier.

## ***Demise of the octant***

In 1767 the first edition of the Nautical Almanac tabulated lunar distances, enabling navigators to find the current time from the angle between the sun and the moon. This angle is sometimes larger than 90°, and thus not possible to measure with an octant. For that reason, Admiral John Campbell, who conducted shipboard experiments with the lunar distance method, suggested a larger instrument and the sextant was developed.

From that time onward, the sextant was the instrument that experienced significant development and improvements and was the instrument of choice for naval navigators. The octant continued to be produced well into the 19th century, though it was generally a less accurate and less expensive instrument. The lower price of the octant, including versions without telescope, made it a practical instrument for ships in the merchant and fishing fleets.

One common practice among navigators up to the late nineteenth century was to use both a sextant and an octant. The sextant was used with great care and only for lunars, while the octant was used for routine meridional altitude measurements of the sun every day. This protected the very accurate and pricier sextant, while using the more affordable octant where it performs well.

## ***The bubble octant***

From the early 1930s through the end of the 1950s, several types of civilian and military *bubble octant* instruments were produced for use aboard aircraft. All were fitted with an artificial horizon in the form of a bubble, which was centered to align the horizon for a navigator flying thousands of feet above the earth; some had recording features.

## ***Use and adjustment***

Use and adjustment of the octant is essentially identical to the navigator's sextant, which see for information on these topics.

## ***Other reflecting instruments***

Hadley's was not the first reflecting quadrant. Robert Hooke invented a reflecting quadrant in 1684 and had written about the concept as early as 1666. Hooke's was a single-reflecting instrument. Other octants were developed by Jean-Paul Fouchy and Caleb Smith in the early 1730s, however, these did not become significant in the history of navigation instruments.

## Chapter 10

# Reflecting Instrument

**Reflecting instruments** are those that use mirrors to enhance their ability to make measurements. In particular, the use of mirrors permits one to observe two objects simultaneously while measuring the angular distance between the objects. While they are used in many professions, they are primarily associated with celestial navigation, as the need to solve navigation problems, in particular the problem of the longitude, was the primary motivation in their development.

### ***Objectives of the instruments***

The purpose of reflecting instruments was to allow an observer to measure the altitude of a celestial object or measure the angular distance between two objects. The driving force behind the developments discussed here was the solution to the problem of finding one's longitude at sea. The solution to this problem was seen to require an accurate means of measuring angles and the accuracy was seen to rely on the observer's ability to measure this angle by simultaneously observing two objects at once.

The deficiency of prior instruments was well known. By requiring the observer to observe two objects with two divergent lines of sight increased the likelihood of an error. Those that considered the problem realized that the use of specula (mirrors in modern parlance) could permit two objects to be observed in a single view. What followed is a series of inventions and improvements that refined the instrument to the point that its accuracy exceeded that which was required for determining longitude. Any further improvements required a completely new technology.

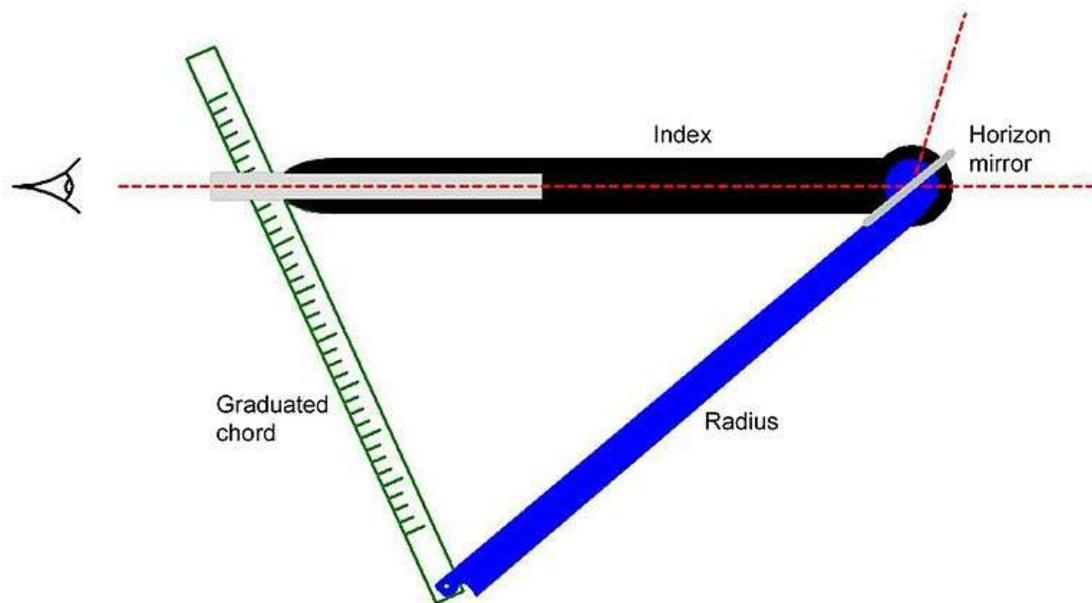
### ***Early reflecting instruments***

Some of the early reflecting instruments were proposed by scientists such as Robert Hooke and Isaac Newton. These were little used or may not have been built or tested extensively. The van Breen instrument was the exception, in that it was used by the Dutch. However, it had little influence outside of the Netherlands.

## Joost van Breen's reflecting cross-staff

Invented in 1660 by the Dutch Joost van Breen, the **spiegelboog** (mirror-bow) was a reflecting cross staff. This instrument appears to have been used for approximately 100 years, mainly in the Zeeland Chamber of the VOC (The Dutch East India Company).

## Robert Hooke's Single-reflecting instrument



Representative drawing of Hooke's reflecting instrument. It does not accurately depict the fine details of the instrument but rather the basic functionality.

The index with telescope mounted is shown in black, the radius arm with the mirror (grey) attached in blue and the chord in green on white. The lines of sight are represented by the red dashed line.

Hooke's instrument was a single-reflecting instrument. It used a single mirror to reflect the image of an astronomical object to the observer's eye. This instrument was first described in 1666 and a working model was presented by Hooke at a meeting of the Royal Society some time later.

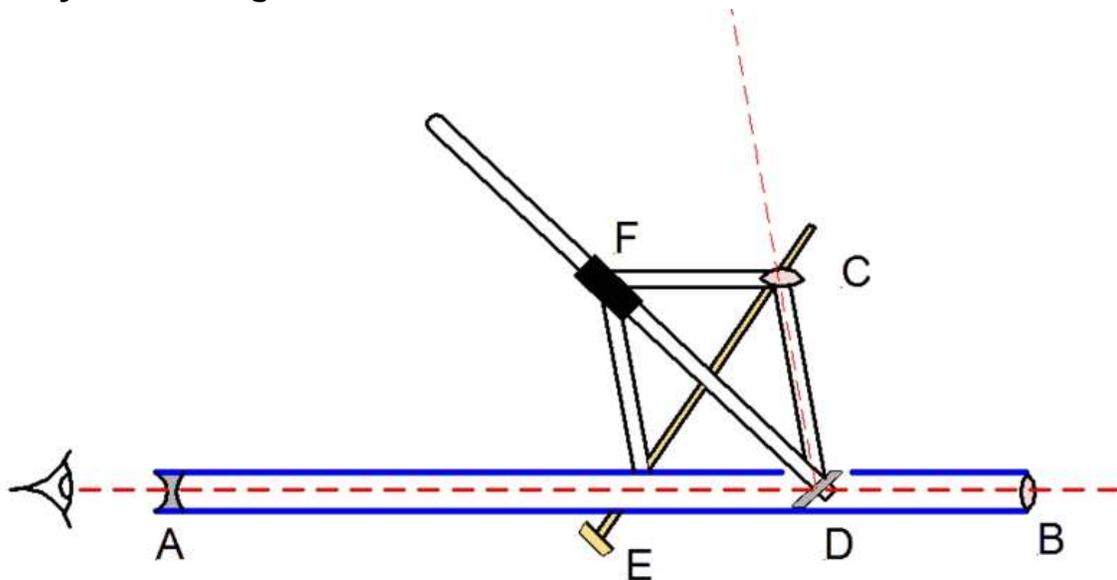
The device consisted of three primary components, an index arm, a radial arm and a graduated chord. The three were arranged in a triangle as in the image on the right. A telescopic sight was mounted on the index arm. At the point of rotation of the radial arm, a single mirror was mounted. This point of rotation allowed the angle between the index arm and the radial arm to be changed. The graduated chord was connected to the opposite end of the radial arm and the chord was permitted to rotate about the end. The chord was held against the distant end of the index arm and slid against it. The graduations on the

chord were uniform and, by using it to measure the distance between the ends of the index arm and the radial arm, the angle between those arms could be determined. A table of chords was used to convert a measurement of distance to a measurement of angle. The use of the mirror resulted in the measured angle being twice the angle included by the index and the radius arm.

The mirror on the radial arm was small enough that the observer could see the reflection of an object in half the telescope's view while seeing straight ahead in the other half. This allowed the observer to see both objects at once. Aligning the two objects together in the telescope's view resulted in the angular distance between them to be represented on the graduated chord.

While Hooke's instrument was novel and attracted some attention at the time, there is no evidence that it was subjected to any tests at sea. The instrument was little used and did not have any significant effect on astronomy or navigation.

### Halley's reflecting instrument



A drawing of Halley's reflecting instrument. The telescope is represented by the blue lines (as if cut open) and the mirrors and lenses are grey. The red dashed lines represent lines of sight.

In 1692, Edmond Halley presented the design of a reflecting instrument to the Royal Society.

This is an interesting instrument, combining the functionality of a radio latino with a double telescope. The telescope (AB in the image to the right), has an eyepiece at one end and a mirror (D) partway along its length with one objective lens at the far end (B). The mirror only obstructs half the field (either left or right) and permits the objective to be seen on the other. Reflected in the mirror is the image from the second objective lens (C).

This permits the observer to see both images, one straight through and one reflected, simultaneously besides each other. It is essential that the focal lengths of the two objective lenses be the same and that the distances from the mirror to either lens be identical. If this condition is not met, the two images cannot be brought to a common focus.

The mirror is mounted on the staff (DF) of the radio latino portion of the instrument and rotates with it. The angle this side of the radio latino's rhombus makes to the telescope can be set by adjusting the rhombus' diagonal length. In order to facilitate this and allow for fine adjustment of the angle, a screw (EC) is mounted so as to allow the observer to change the distance between the two vertexes (E and C).

The observer sights the horizon with the direct lens' view and sights a celestial object in the mirror. Turning the screw to bring the two images directly adjacent sets the instrument. The angle is determined by taking the length of the screw between E and C and converting this to an angle in a table of chords.

Halley specified that the telescope tube be rectangular in cross section. This makes construction easy, but is not a requirement as other cross section shapes can be accommodated. The four sides of the radio latino portion (CD, DE, EF, FC) must be equal in length in order for the angle between the telescope and the objective lens side (AD-DC) to be precisely twice the angle between the telescope and the mirror (AD-DF) (or in other words - to enforce the angle of incidence being equal to the angle of reflection). Otherwise, instrument collimation will be compromised and the resulting measurements would be in error.

The celestial object's elevation angle could have been determined by reading from graduations on the staff at the slider, however, that's not how Halley designed the instrument. This may suggest that the overall design of the instrument was coincidentally like a radio latino and that Halley may not have been familiar with that instrument.

There is no knowledge of whether this instrument was ever tested at sea.

## **Newton's reflecting quadrant**

Newton's reflecting quadrant was similar in many respects to Hadley's first reflecting quadrant that followed it.

Newton had communicated the design to Edmund Halley around 1699. However, Halley did not do anything with the document and it remained in his papers only to be discovered after his death. However, Halley did discuss Newton's design with members of the Royal Society when Hadley presented his reflecting quadrant in 1731. Halley noted that Hadley's design was quite similar to the earlier Newtonian instrument.

As a result of this inadvertent secrecy, Newton's invention played little role in the development of reflecting instruments.

## ***The octant***

What is remarkable about the octant is the number of persons who independently invented the device in a short period of time. John Hadley and Thomas Godfrey both get credit for inventing the octant. They independently developed the same instrument around 1731. They were not the only ones, however.

In Hadley's case, two instruments were designed. The first was an instrument very similar to Newton's reflecting quadrant. The second had essentially the same form as the modern sextant. Few of the first design were constructed, while the second became the standard instrument from which the sextant derived and, along with the sextant, displaced all prior navigation instruments used for celestial navigation.

Caleb Smith, an English insurance broker with a strong interest in astronomy, had created an octant in 1734. He called it an *Astroscope* or *Sea-Quadrant*. His used a fixed prism in addition to an index mirror to provide reflective elements. Prisms provide advantages over mirrors in an era when polished speculum metal mirrors were inferior and both the silvering of a mirror and the production of glass with flat, parallel surfaces was difficult. However, the other design elements of Smith's instrument made it inferior to Hadley's octant and it was not used significantly.

Jean-Paul Fouchy, a mathematics professor and astronomer in France invented an octant in 1732. His was essentially the same as Hadley's. Fouchy did not know of the developments in England at the time, since communications between the two country's instrument makers was limited and the publications of the Royal Society, particularly the *Philosophical Transactions*, were not being distributed in France. Fouchy's octant was overshadowed by Hadley's.

## ***The sextant***



A sextant that has been used for over a half century. This frame shows one standard design - that with three rings. This is one design that has been used to avoid thermal expansion problems while retaining adequate stiffness.

The origin of the sextant is straightforward and not in dispute. Admiral John Campbell, having used Hadley's octant in sea trials of the method of lunar distances, found that it was wanting. The  $90^\circ$  angle subtended by the arc of the instrument was insufficient to measure some of the angular distances required for the method. He suggested that the angle be increased to  $120^\circ$ , yielding the sextant. John Bird made the first such sextant in 1757.

With the development of the sextant, the octant became something of a second class instrument. The octant, while occasionally constructed entirely of brass, remained primarily a wooden-framed instrument. Most of the developments in advanced materials and construction techniques were reserved for the sextant.

There are examples of sextants made with wood, however most are made from brass. In order to ensure the frame was stiff, instrument makers used thicker frames. This had a drawback in making the instrument heavier, which could influence the accuracy due to hand-shaking as the navigator worked against its weight. In order to avoid this problem, the frames were modified. Edward Troughton patented the double-framed sextant in 1788. This used two frames held in parallel with spacers. The two frames were about a centimetre apart. This significantly increased the stiffness of the frame. An earlier version had a second frame that only covered the upper part of the instrument, securing the mirrors and telescope. Later versions used two full frames. Since the spacers looked like little pillars, these were also called *pillar sextants*.

Troughton also experimented with alternative materials. The scales were plated with silver, gold or platinum. Gold and platinum both minimized corrosion problems. The platinum-plated instruments were expensive, due to the scarcity of the metal, though less expensive than gold. Troughton knew William Hyde Wollaston through the Royal Society and this gave him access to the precious metal. Instruments from Troughton's company that used platinum can be easily identified by the word *Platina* engraved on the frame. These instruments remain highly valued as collector's items and are as accurate today as when they were constructed.

As the developments in dividing engines progressed, the sextant was more accurate and could be made smaller. In order to permit easy reading of the vernier, a small magnifying lens was added. In addition, to reduce glare on the frame, some had a diffuser surrounding the magnifier to soften the light. As accuracy increased, the circular arc vernier was replaced with a drum vernier.

Frame designs were modified over time to create a frame that would not be adversely affected by temperature changes. These frame patterns became standardized and one can see the same general shape in many instruments from many different manufacturers.

In order to control costs, modern sextants are now available in precision-made plastic. These are light, affordable and of high quality.

### ***Types of sextants***

While most people think of navigation when they hear the term *sextant*, the instrument has been used in other professions.

#### Navigator's sextant

The common type of instrument most people think of when they hear the term *sextant*.

### Sounding sextants

These are sextants that were constructed for use horizontally rather than vertically and were developed for use in hydrographic surveys.

### Surveyor's sextants

These were constructed for use exclusively on land for horizontal angular measurements. Instead of a handle on the frame, they had a socket to allow the attachment of a surveyor's Jacob's staff.

### Box or pocket sextants

These are small sextants entirely contained within a metal case. First developed by Edward Troughton, they are usually all brass with most of the mechanical components inside the case. The telescope extends from an opening in the side. The index and other parts are completely covered when the case cover is slipped on. Popular with surveyors for their small size (typically only 6.5–8 cm in diameter and 5 cm deep), their accuracy was enabled by improvements in the dividing engines used to graduate the arcs. The arcs are so small that magnifiers are attached to allow them to be read.

In addition to these types, there are terms used for various sextants.

A *pillar sextant* can be either:

1. A double-frame sextant as patented by Edward Troughton in 1788.
2. A surveyor's sextant with a socket for a surveyor's staff (the pillar).

The former is the most common use of the term.

## ***Beyond the sextant***

### **Quintant and others**

Several makers offered instruments with sizes other than one-eighth or one-sixth of a circle. One of the most common was the *quintant* or fifth of a circle ( $72^\circ$  arc reading to  $144^\circ$ ). Other sizes were also available, but the odd sizes never became common. Many instruments are found with scales reading to, for example,  $135^\circ$ , but they are simply referred to as sextants. Similarly, there are  $100^\circ$  octants, but these are not separated as unique types of instruments.

There was interest in much larger instruments for special purposes. In particular a number of full circle instruments were made, categorized as reflecting circles and repeating circles.

## Reflecting circles



Borda's reflecting circle, on display at Toulon naval museum



Mendonça's reflecting circle on display at the Musée national de la Marine.

The reflecting circle was invented by the German geometer and astronomer Tobias Mayer in 1752, with details published in 1767. His development preceded the sextant and was motivated by the need to create a superior surveying instrument.

The reflecting circle is a complete circular instrument graduated to  $720^\circ$  (To measure distances between heavenly bodies, there is no need to read an angle greater than  $180^\circ$ , since the minimum distance will always be less than  $180^\circ$ ). Mayer presented a detailed description of this instrument to the Board of Longitude and John Bird used the information to construct one sixteen inches in diameter for evaluation by the Royal Navy. This instrument was one of those used by Admiral John Campbell during his evaluation of the lunar distance method. It differed in that it was graduated to  $360^\circ$  and was so heavy

that it was fitted with a support that attached to a belt. It was not considered better than the Hadley octant and was less convenient to use. As a result, Campbell recommended the construction of the sextant.

Jean-Charles de Borda further developed the reflecting circle. He modified the position of the telescopic sight in such a way that the mirror could be used to receive an image from either side relative to the telescope. This eliminated the need to ascertain that the mirrors were precisely parallel when reading zero. This simplified the use of the instrument. Further refinements were performed with the help of Etienne Lenoir. The two of them refined the instrument to its definitive form in 1777. This instrument was so distinctive it was given the name *Borda circle*.

Josef de Mendoza y Ríos redesigned Borda's reflecting circle (London, 1801). The goal was to use it together with his Lunar Tables published by the Royal Society (London, 1805). He made a design with two concentric circles and a vernier scale and recommended averaging three sequential readings to reduce the error. Borda's system was not based on a circle of 360° but 400 grads (Borda spent years calculating his tables with a circle divided in 400°). Mendoza's lunar tables have been used through almost the entire nineteenth century.

Edward Troughton also modified the reflecting circle. He created a design with three index arms and verniers. This permitted three simultaneous readings to average out the error.

As a navigation instrument, the reflecting circle was more popular with the French navy than with the British.

One instrument derived from the reflecting circle is the **repeating circle**. Invented by Lenoir in 1784, Borda and Lenoir developed the instrument for geodetic surveying. Since it was not used for the celestial measures, it did not use double reflection and substituted two telescope sights. As such, it was not a reflecting instrument. It was notable as being the equal of the great theodolite created by the renowned instrument maker, Jesse Ramsden.

## **Bris Sextant**

The **Bris sextant** is not a true sextant, but it is a true reflecting instrument based on the principle of double reflection and subject to the same rules and errors as common octants and sextants. Unlike common octants and sextants, the **Bris sextant** is a fixed angle instrument capable of accurately measuring a few specific angles unlike other reflecting instruments which can measure any angle within the range of the instrument. It is particularly suited to determining the altitude of the sun or moon.