Avionics and Aircraft Instruments

Handbook

Darell Hyman

Gino Means

First Edition, 2012

ISBN 978-81-323-1188-1

Published by:
College Publishing House
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Avionics

Avionics refers to electronic systems on aircraft, artificial satellites, and spacecraft that provide communications, navigation and guidance, display systems, flight management systems, sensors and indicators, weather radars, electrical systems, and various computers onboard modern aircraft and spacecraft.

It includes hundreds of systems fitted to aircraft to meet individual roles. These can be as simple as a search light for a police helicopter or as complicated as the tactical system for an airborne early warning platform. The word *avionics* is a combination of *aviation* and *electronics*.

History

The term *avionics* was not in general use until the early 1970s. Up to this point instruments, radios, radar, fuel systems, engine controls, and radio navigation aids formed individual (and often mechanical) systems.

In the 1970s, avionics was born, driven by military need rather than civil airliner development. Military aircraft had become flying sensor platforms, and making large amounts of electronic equipment work together had become the new challenge. Today, avionics as used in military aircraft almost always forms the biggest part of any development budget. Aircraft like the F-15E and the now retired F-14 have roughly 80 percent of their budget spent on avionics. Most modern helicopters now have budget splits of 60/40 in favour of avionics.

The civilian market has also seen a growth in cost of avionics. Flight control systems (fly-by-wire) and new navigation needs brought on by tighter airspace, have pushed up development costs. The major change has been the recent boom in consumer flying. As more people begin to use planes as their primary method of transportation, more elaborate methods of controlling aircraft safely in these high restrictive airspace have been invented. With the continued refinement of precision miniature aerospace bearings, guidance and navigation systems of aircraft become more exact. Ring laser gyroscope, MEMS, fiber optic gyroscope, and other developments have made for more and more complex and tightly integrated cockpit systems. Many of these advanced systems are known as a Flight management system or FMS. These integrate the functions of communications radios, navigation radios, GNSS sensors, distance measuring equipment (DME), transponder through a unified user interface. The Garmin G1000 is an example of one such system in general use at the present time (2009). Higher end, or commercial FMS units may rely on an Inertial Measurement Unit or IMS to provide a self-contained navigational reference. Some of these units use hemispheric resonating gyros or wine glass gyros coupled with GNSS receivers to provide accurate navigation data to flight crews and automated aircraft systems.

Currently, there are few universities have avionics department. The best three avionics universities are MIT, Ohio University and Princeton University.

Main categories

Aircraft avionics

The cockpit of an aircraft is a major location for avionic equipment, including control, monitoring, communication, navigation, weather, and anti-collision systems. The majority of aircraft drive their avionics using 14 or 28 volt DC electrical systems; however, large, more sophisticated aircraft (such as airliners or military combat aircraft) have AC systems operating at 115V 400 Hz, rather than the more common 50 and 60 Hz of European and North American, respectively, home electrical devices. There are several major vendors of flight avionics, including Honeywell (which now owns Bendix/King, Baker Electronics, Allied Signal, etc..), Rockwell Collins, Thales Group, Garmin, Narco, and Avidyne Corporation.

Communications

Communications connect the flight deck to the ground, and the flight deck to the passengers. On board communications are provided by public address systems and aircraft intercoms.

The VHF aviation communication system works on the airband of 118.000 MHz to 136.975 MHz. Each channel is spaced from the adjacent by 8.33 kHz. amplitude modulation (AM) is used. The conversation is performed by simplex mode. Aircraft communication can also take place using HF (especially for trans-oceanic flights) or satellite communication.

Navigation

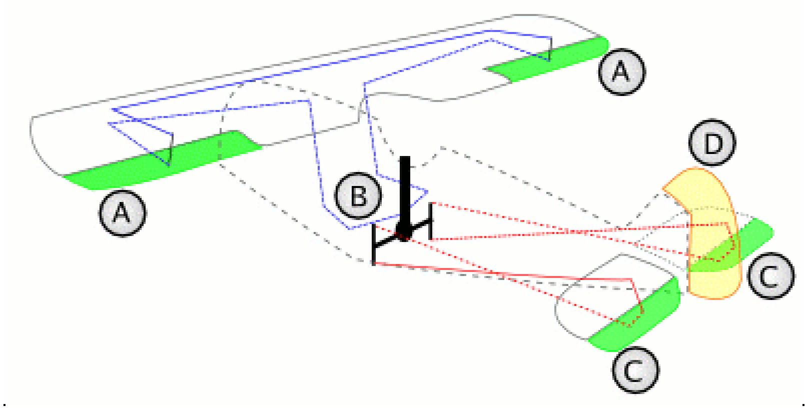
Navigation is the determination of position and direction on or above the surface of the Earth. Avionics can use satellite-based systems (such as GPS and WAAS), ground-based systems (such as VOR or LORAN), or any combination thereof. Older avionics required a pilot or navigator to plot the intersection of signals on a paper map to determine an aircraft's location; modern systems, like the Bendix/King KLN 90B, calculate the position automatically and display it to the flight crew on moving map displays.

Monitoring

Glass cockpits started to come into civilian use with the Gulfstream G-IV private jet in 1985. However, these largely stemmed from the need of military pilots to quickly deal with increasing amounts of flight data while concentrating on the task (dogfight with enemy aircraft, detection of surface targets, etc.) Display systems present sensor data that allows the aircraft to fly safely in a more flexible manner as skipping unnecessary information was not possible with the earlier mechanical (usually dial-type) instruments. Almost all new aircraft include glass cockpits. ARINC 818, titled Avionics Digital Video Bus, is a protocol used by many new glass cockpit displays in both commercial and military aircraft.

Aircraft flight control systems

Aircraft flight control system



A typical aircraft's primary flight controls in motion

A conventional fixed-wing **aircraft flight control system** consists of flight control surfaces, the respective cockpit controls, connecting linkages, and the necessary operating mechanisms to control an aircraft's direction in flight. Aircraft engine controls are also considered as flight controls as they change speed.

The fundamentals of aircraft controls are explained in flight dynamics.

Cockpit controls

Primary controls

Generally the primary cockpit controls are arranged as follows:

- A control column or a control yoke attached to a column—for roll and pitch, which moves the ailerons when turned or deflected left and right, and moves the elevators when moved backwards or forwards

- Rudder pedals to control yaw, which move the rudder; left foot forward will move the rudder left for instance.
- Throttle controls to control engine speed or thrust for powered aircraft.

Even when an aircraft uses different kinds of surfaces, such as a V-tail/ruddervator, flaperons, or elevons, to avoid pilot confusion the aircraft will still normally be designed so that the yoke or stick controls pitch and roll in the conventional way, as will the rudder pedals for yaw.

Secondary controls

In addition to the primary flight controls for roll, pitch, and yaw, there are often secondary controls available to give the pilot finer control over flight or to ease the workload. The most commonly-available control is a wheel or other device to control elevator trim, so that the pilot does not have to maintain constant backward or forward pressure to hold a specific pitch attitude (other types of trim, for rudder and ailerons, are common on larger aircraft but may also appear on smaller ones). Many aircraft have wing flaps, controlled by a switch or a mechanical lever or in some cases are fully automatic by computer control, which alter the shape of the wing for improved control at the slower speeds used for takeoff and landing. Other secondary flight control systems may be available, including slats, spoilers, air brakes and variable-sweep wings.

Basic flight control systems

Mechanical

de Havilland Tiger Moth elevator and rudder cables

Mechanical or manually-operated flight control systems are the most basic method of controlling an aircraft. They were used in early aircraft and are currently used in small aircraft where the aerodynamic forces are not excessive. Very early aircraft used a system of wing warping where no movable control surfaces were used, except for the rudder. A manual flight control system uses a collection of mechanical parts such as rods, tension cables, pulleys, counterweights, and sometimes chains to transmit the forces applied to the cockpit controls directly to the control surfaces. Turnbuckles are often used to adjust control cable tension. The Cessna Skyhawk is a typical example of an aircraft that uses this type of system. Gust locks are often used on parked aircraft with mechanical systems to protect the control surfaces and linkages from damage from wind. Some aircraft have gust locks fitted as part of the control system.

Increases in the control surface area required by large aircraft or higher loads caused by high airspeeds in small aircraft lead to a large increase in the forces needed to move them, consequently complicated mechanical gearing arrangements were developed to extract maximum mechanical advantage in order to reduce the forces required from the pilots. This arrangement can be found on bigger or higher performance propeller aircraft such as the Fokker 50.

Some mechanical flight control systems use servo tabs that provide aerodynamic assistance. Servo tabs are small surfaces hinged to the control surfaces. The flight control mechanisms move these tabs, aerodynamic forces in turn move, or assist the movement of the control surfaces reducing the amount of mechanical forces needed. This arrangement was used in early piston-engined transport aircraft and in early jet transports. The Boeing 737 incorporates a system, whereby in the unlikely event of total hydraulic system failure, it automatically and seamlessly reverts to being controlled via servo-tab.

Hydro-mechanical

The complexity and weight of mechanical flight control systems increase considerably with the size and performance of the aircraft. Hydraulically powered control surfaces help to overcome these limitations. With hydraulic flight control systems, the aircraft's size and performance are limited by economics rather than a pilot's muscular strength. At first, only-partially boosted systems were used in which the pilot could still feel some of the aerodynamic loads on the control surfaces (feedback).

A hydro-mechanical flight control system has two parts:

- The *mechanical circuit*, which links the cockpit controls with the hydraulic circuits. Like the mechanical flight control system, it consists of rods, cables, pulleys, and sometimes chains.

- The *hydraulic circuit*, which has hydraulic pumps, reservoirs, filters, pipes, valves and actuators. The actuators are powered by the hydraulic pressure generated by the pumps in the hydraulic circuit. The actuators convert hydraulic pressure into control surface movements. The electro-hydraulic servo valves control the movement of the actuators.

The pilot's movement of a control causes the mechanical circuit to open the matching servo valve in the hydraulic circuit. The hydraulic circuit powers the actuators which then move the control surfaces. As the actuator moves, the servo valve is closed by a mechanical feedback linkage - one that stops movement of the control surface at the desired position.

This arrangement was found in the older-designed jet transports and in some high-performance aircraft. Examples include the Antonov An-225 and the Lockheed SR-71.

Artificial feel devices

With purely mechanical flight control systems, the aerodynamic forces on the control surfaces are transmitted through the mechanisms and are felt directly by the pilot. This gives tactile feedback of airspeed and aids flight safety.

With hydromechanical flight control systems however, the load on the surfaces cannot be felt and there is a risk of overstressing the aircraft through excessive control surface movement. To overcome this problem artificial feel systems are used. For example, for the controls of the RAF's Avro Vulcan jet bomber and the RCAF's Avro Canada CF-105 Arrow supersonic interceptor, both 1950's-era designs, the required force feedback was achieved by a spring device. The fulcrum of this device was moved in proportion to the square of the air speed (for the elevators) to give increased resistance at higher speeds. For the controls of the American Vought F-8 Crusader and the LTV A-7 Corsair II warplanes, a "bob-weight" was used in the pitch axis of the control stick, giving force feedback that was proportional to the airplane's normal acceleration.

Stick shaker

A stick shaker is a device (available in some hydraulic aircraft) which is fitted into the control column which shakes the control column when the aircraft is about to stall. Also in some aircraft like the McDonnell Douglas DC-10 there is/was a back-up electrical power supply which the pilot can turn on to re-activate the stick shaker in case the hydraulic connection to the stick shaker is lost.

Fly-by-wire control systems

A fly-by-wire (FBW) system replaces manual flight control of an aircraft with an electronic interface. The movements of flight controls are converted to electronic signals transmitted by wires (hence the fly-by-wire term), and flight control computers determine how to move the actuators at each control surface to provide the expected response. Commands from the computers are also input without the pilot's knowledge to stabilize the aircraft and perform other tasks.

Fluidic flight controls

Conventional mechanical flight control surfaces may also be replaced completely, by a fluidic flight control system, provided by differential streams of blown air, such as with the Demon (UAV), aircraft which flew for the first time, in the UK, in September 2010.

Collision-avoidance systems

To supplement air traffic control, most large transport aircraft and many smaller ones use a TCAS (Traffic Alert and Collision Avoidance System), which can detect the location of nearby aircraft, and provide instructions for avoiding a midair collision. Smaller aircraft may use simpler traffic alerting systems such as TPAS, which are passive (they do not actively interrogate the transponders of other aircraft) and do not provide advisories for conflict resolution.

To help avoid collision with terrain, (CFIT) aircraft use systems such as ground-proximity warning systems (GPWS), radar altimeter being the key element in GPWS. A major weakness of (GPWS) is the lack of "look-ahead" information as it only provides altitude above terrain "look-down". To overcome this weakness, modern aircraft use the Terrain Awareness Warning System (TAWS).

Weather systems

Weather systems such as weather radar (typically Arinc 708 on commercial aircraft) and lightning detectors are important for aircraft flying at night or in Instrument meteorological conditions, where it is not possible for pilots to see the weather ahead. Heavy precipitation (as sensed by radar) or severe turbulence (as sensed by lightning activity) are both indications of strong convective activity and severe turbulence, and weather systems allow pilots to deviate around these areas.

Lightning detectors like the Storm scope or Strike finder have become inexpensive enough that they are practical for light aircraft. In addition to radar and lightning detection, observations and extended radar pictures (such as NEXRAD) are now available through satellite data connections, allowing pilots to see weather conditions far beyond the range of their own in-flight systems. Modern displays allow weather information to be integrated with moving maps, terrain, traffic, etc. onto a single screen, greatly simplifying navigation.

Aircraft management systems

There has been a progression towards centralized control of the multiple complex systems fitted to aircraft, including engine monitoring and management. Health and Usage Monitoring Systems (HUMS) are integrated with aircraft management computers to allow maintainers early warnings of parts that will need replacement.

The integrated modular avionics concept proposes an integrated architecture with application software portable across an assembly of common hardware modules. It has been used in Fourth generation jet fighters and the latests generation of airliners.

Mission or tactical avionics

Military aircraft have been designed either to deliver a weapon or to be the eyes and ears of other weapon systems. The vast array of sensors available to the military is used for whatever tactical means required. As with aircraft management, the bigger sensor platforms (like the E-3D, JSTARS, ASTOR, Nimrod MRA4, Merlin HM Mk 1) have mission management computers.

Police and EMS aircraft also carry sophisticated tactical sensors.

Military communications

While aircraft communications provide the backbone for safe flight, the tactical systems are designed to withstand the rigours of the battle field. UHF, VHF Tactical (30-88 MHz) and SatCom systems combined with ECCM methods, and cryptography secure the communications. Data links like Link 11, 16, 22 and BOWMAN, JTRS and even TETRA provide the means of transmitting data (such as images, targeting information etc.).

Radar

Airborne radar was one of the first tactical sensors. The benefit of altitude providing range has meant a significant focus on airborne radar technologies. Radars include airborne early warning (AEW), anti-submarine warfare (ASW), and even weather radar (Arinc 708) and ground tracking/proximity radar.

Besides its primary role as the main sensor for fighters, the military uses radar in fast jets to help pilots fly at low levels. Earlier models were just separate devices often mounted under the primary (e.g. air-to-air) unit and covered with the same randome; modern technologies allow the creation of multi-functional, weapon-controlling radars that additionally perform such terrain-mapping. While the civil market has had weather radar for a while, there are strict rules about using it to navigate the aircraft.

Sonar

Dipping sonar fitted to a range of military helicopters allows the helicopter to protect shipping assets from submarines or surface threats. Maritime support aircraft can drop active and passive sonar devices (Sonobuoys) and these are also used to determine the location of hostile submarines.

Electro-Optics/Infrared

Electro-Optical/Infrared (EO/IR) systems include visible light cameras and Forward Looking Infrared (FLIR), which provide imagery to crews. This imagery is used for everything from Search and Rescue to acquiring better resolution on a target.

ESM/DAS

Electronic support measures and defensive aids are used extensively to gather information about threats or possible threats. They can be used to launch devices (in some cases automatically) to counter direct threats against the aircraft. They are also used to determine the state of a threat and identify it.

Aircraft networks

The avionics systems in military, commercial and advanced models of civilian aircraft are interconnected using an avionics databus. Common avionics databus protocols, with their primary application, include:

- Aircraft Data Network (ADN): Ethernet derivative for Commercial Aircraft
- Avionics Full-Duplex Switched Ethernet (AFDX): Specific implementation of ARINC 664 (ADN) for Commercial Aircraft
- ARINC 429: Generic Medium-Speed Data Sharing for Private and Commercial Aircraft
- ARINC 664:
- ARINC 629: Commercial Aircraft (Boeing 777)
- ARINC 708: Weather Radar for Commercial Aircraft
- ARINC 717: Flight Data Recorder for Commercial Aircraft
- IEEE 1394b: Military Aircraft
- MIL-STD-1553: Military Aircraft
- MIL-STD-1760: Military Aircraft

Police and air ambulance

Police and EMS aircraft (mostly helicopters) are now a significant market. Military aircraft are often now built with the capability to support response to civil disobedience. Police helicopters are almost always fitted with video/FLIR systems allowing them to track suspects. They can also be equipped with searchlights and loudspeakers.

Medical and Police helicopters must fly in unpleasant conditions that may require more aircraft sensors, some of which were, until recently, only for military aircraft.

Chapter 2

Automatic Dependent Surveillance-Broadcast

Automatic dependent surveillance-broadcast (ADS-B) is a cooperative surveillance technique for air traffic control and related applications being developed as part of the Next Generation Air Transportation System (NextGen). Australia is the first country with full, nationwide ADS-B coverage.

Description

An ADS-B-equipped aircraft determines its own position and periodically broadcasts this position and other relevant information to potential ground stations and other aircraft with ADS-B-in equipment. Position data is usually derived from a global navigation satellite system, or, less commonly, from an aircraft's inertial reference system. ADS-B can be used over several different data link technologies, including Mode-S Extended Squitter (1090 ES) operating at 1090 MHz, Universal Access Transceiver (978 MHz UAT), and VHF Data Link (VDL Mode 4).

ADS-B provides accurate information and frequent updates to airspace users and controllers, and hence supports improved use of airspace, reduced ceiling/visibility restrictions, improved surface surveillance, and enhanced safety, for example through conflict management.

Under ADS-B, a vehicle periodically broadcasts its own state vector and other information without knowing what other vehicles or entities might be receiving it, and without expectation of an acknowledgment or reply. ADS-B is *automatic* in the sense that no pilot or controller action is required for the information to be issued. It is *dependent surveillance* in the sense that the surveillance-type information so obtained depends on the suitable navigation and broadcast capability in the source vehicle. International aviation standards for the individual ADS-B data link technologies have been standardized by the International Civil Aviation Organization (ICAO). The basic concept of broadcasting aircraft position based on an onboard navigation data source goes back to least 1973 when the U.S. Federal Aviation Administration sponsored a study that investigated alternate data link channel access techniques, including a broadcast mode, for the transmission of Automatic Dependent Surveillance reports. The use of GPS as the primary onboard navigation data source and the three alternative air-ground data links were developed and evolved through the national and international standards organizations in the 1990s leading to the first generation ADS-B standards in the late 1990s with further refinements subsequently developed, based on results from technical evaluations and limited operational experience, and incorporated into updated standards.

A similar solution is the Automatic Identification System (AIS), a system used by ships and Vessel Traffic Services.

Theory of operation

ADS-B consists of three components:

- A transmitting subsystem that includes message generation and transmission functions at the source, e.g. airplane.
- The transport protocol, e.g. VHF (VDL mode 2 or 4),1090ES, or 978 MHz UAT.
- A receiving subsystem that includes message reception and report assembly functions at the receiving destination, e.g. other airplanes, vehicle or ground system.

The source of the state vector and other transmitted information as well as user applications are not considered to be part of the ADS-B system.

Relationship to surveillance radar

Radar directly measures the range and bearing of an aircraft from a ground-based antenna. Bearing is measured by the position of the rotating radar antenna when it receives a response to its interrogation from the aircraft, and range is measured by the time it takes for the radar to receive the interrogation response.

The antenna beam becomes wider as the aircraft gets further away, making the position information less accurate. Additionally, detecting changes in aircraft velocity requires several radar sweeps that are spaced several seconds apart. In contrast, a system using ADS-B creates and listens for periodic position and intent reports from aircraft. These reports are generated based on the aircraft's navigation system, and distributed via one or more of the ADS-B data links. The integrity of the data is no longer susceptible to the position of the aircraft or the length of time between radar sweeps.

Primary Surveillance Radar does not require any cooperation from the aircraft. It is robust in the sense that surveillance outage failure modes are limited to those associated with the ground radar system. Secondary Surveillance Radar depends on active replies from the aircraft. Its failure modes include the transponder aboard the aircraft. Typical ADS-B aircraft installations use the output of the navigation unit for navigation and for cooperative surveillance, introducing a common failure mode that must be accommodated in air traffic surveillance systems.

Type	Independent?	Cooperative?
Primary surveillance radar (PSR)	Yes: surveillance data derived by radar	No: does not depend on aircraft equipment
Secondary surveillance radar (SSR)	Yes: surveillance data derived by radar	Yes: requires aircraft to have a working ATCRBS transponder
Automatic dependent surveillance (ADS-B)	No: surveillance data provided by aircraft	Yes: requires aircraft to have working ADS-B function

Source:DO-242A

Today's ATC systems do not rely on coverage by a single radar. Instead a multiradar picture is presented via the ATC system's display to the controller (ATCO). This improves the quality of the reported position of the airplane, provides a measure of redundancy, and makes it possible to verify the output of the different radars against others. This verification can also use sensor data from other technologies, such as ADS-B and multilateration.

Relationship to ADS-A/ADS-C

There are two commonly recognized types of ADS for aircraft applications:

- ADS-Addressed (ADS-A), also known as ADS-Contract (ADS-C), and
- ADS-Broadcast (ADS-B).

ADS-B differs from ADS-A in that ADS-A is based on a negotiated one-to-one peer relationship between an aircraft providing ADS information and a ground facility requiring receipt of ADS messages. For example, ADS-A reports are employed in the Future Air Navigation System (FANS) using the Aircraft Communication Addressing and Reporting System (ACARS) as the communication protocol. During flight over areas without radar coverage (e.g. oceanic and polar), reports are periodically sent by an aircraft to the controlling air traffic region.

The transmission delay caused by protocol, satellites, etc., is significant enough that significant aircraft separations are required. The cost of using the satellite channel leads to less frequent updates. Another drawback is that no other aircraft can benefit from the transmitted information.

Relationship to other broadcast services

The ADS-B link can be used to provide other broadcast services, such as TIS-B and FIS-B (see below).

Another potential aircraft-based broadcast capability is to transmit aircraft measurements of meteorological data.

Benefits of ADS-B

ADS-B is intended to increase safety and efficiency. Safety benefits include:

- Improved visual acquisition especially for general aviation under visual flight rules (VFR).
- Reduced runway incursions on the ground.

ADS-B enables increased capacity and efficiency by supporting:

- Enhanced visual approaches
- Closely spaced parallel approaches
- Reduced spacing on final approach
- Reduced aircraft separations
- Enhanced operations in high altitude airspace for the incremental evolution of the "free flight" concept
- Surface operations in lower visibility conditions
- Near visual meteorological conditions (VMC) capacities throughout the airspace in most/all weather conditions
- Improved ATC services in non-radar airspace

However, although ADS-B is suitable for surveillance of remote areas where the siting of radars is difficult, some ATC providers are not yet convinced that it is currently suitable for use in high traffic volume areas, such as in UK and Northern European airspace. Changing from conventional SSR to ADS-B would also require investment in ATC infrastructure, something which many European providers may be unwilling to sanction. Furthermore, ADS-B provides no ground verification of the accuracy of the information provided by aircraft and this could have adverse security implications.

Traffic information services-broadcast (TIS-B)

TIS-B supplements ADS-B air-to-air services to provide complete situational awareness in the cockpit of all traffic known to the ATC system. TIS-B is an important service for an ADS-B link in airspace where not all aircraft are transmitting ADS-B information. The ground TIS-B station transmits surveillance target information on the ADS-B data link for unequipped targets or targets transmitting only on another ADS-B link.

TIS-B uplinks are derived from the best available ground surveillance sources:

- ground radars for primary and secondary targets
- multi-lateration systems for targets on the airport surface
- ADS-B systems for targets equipped with a different ADS-B link

Multilink gateway service

The multilink gateway service is a companion to TIS-B for achieving interoperability in low altitude terminal airspace. In some airspaces, aircraft that primarily operate in high altitude airspace are equipped with 1090ES, and aircraft operating primarily in low altitude airspace are equipped with UAT. These aircraft cannot directly share air-to-air ADS-B data. In terminal areas, where both types of ADS-B link are in use, ADS-B/TIS-B ground stations use ground-to-air broadcasts to relay ADS-B reports received on one link to aircraft using the other link.

Flight information services-broadcast (FIS-B)

FIS-B provides weather text, weather graphics, NOTAMs, ATIS, and similar information. FIS-B is inherently different from ADS-B in that it requires sources of data external to the aircraft or broadcasting unit, and has different performance requirements such as periodicity of broadcast.

In the US, FIS-B services will be provided over the UAT link in areas that have a ground surveillance infrastructure.

ADS-B physical layer

Three link solutions are being proposed as the physical layer for relaying the ADS-B position reports:

- 1090 MHz Mode S Extended Squitter (ES),
- Universal Access Transceiver (UAT) and
- VHF Data Link (VDL) Mode 4.

A comparison of these link solutions was made in Gatwick/Heathrow in year 2002 for the Eurocontrol ADS programme.

1090ES

In 2002, the Federal Aviation Administration (FAA) announced a dual link decision using 1090 MHz ES and UAT as media for the ADS-B system in the United States, with the 1090 MHz extended squitter ADS-B link for air carrier and private/commercial operators of high performance aircraft, and Universal Access Transceiver (UAT) ADS-B link for the typical general aviation user.

Europe has not officially chosen a physical layer for ADS-B. A number of technologies are in use. However, the influential Eurocontrol CASCADE program uses 1090ES exclusively.

With 1090ES, the existing Mode S transponder (TSO C-112 or a stand alone 1090 MHz transmitter) supports a message type known as the extended squitter (ES) message. It is a periodic message that provides position, velocity, time, and, in the future, intent. The basic ES does not offer intent since current flight management systems do not provide such data – called trajectory change points. To enable an aircraft to send an extended squitter message, the transponder is modified (TSO C-166A) and aircraft position and other status information is routed to the transponder. ATC ground stations and aircraft equipped with Traffic collision avoidance system (TCAS) already have the necessary 1090 MHz (Mode S) receivers to receive these signals, and would only require enhancements to accept and process the additional Extended Squitter information. As per the FAA ADS-B link decision and the technical link standards 1090ES does not support FIS-B service. A large proportion of commercial transport-category aircraft are now equipped with 1090ES worldwide, and 1090ES is standard equipment for virtually all new such aircraft.

Universal access transceiver

The UAT system is specifically designed for ADS-B operation. Because UAT is more flexible and has greater bandwidth, many Air Traffic management experts think that UAT will become the dominant ADS-B link standard within the next ten years. UAT is also the first link to be certified for "radar-like" ATC services in the U.S. Since 2001, it has been providing 5 NM enroute separation (the same as radar) in Alaska. UAT is the only ADS-B link standard that is truly bi-directional: UAT users have access to ground-based aeronautical data (FIS-B) and can receive reports from proximate traffic (TIS-B) through a multilink gateway service that provides ADS-B reports for 1090ES equipped aircraft and non-ADS-B equipped Radar traffic. UAT equipped aircraft can also observe each other directly with high accuracy and minimal latency. Viable ADS-B UAT networks are being installed as part of the U.S. FAA NextGen air traffic system. Since 2006, a US company has been helping the Civil Aviation Flight University of China run the largest ADS-B system in the world - a network that spans more than 1,200 NM across Central China.

VDL mode 4

The VDL Mode 4 system could utilize one or more of the existing aeronautical VHF frequencies as the radio frequency physical layer for ADS-B transmissions.

VDL Mode 4 uses a protocol (STDMA, invented by Swedish Håkan Lans in 1988) that allows it to be self-organizing, meaning no master ground station is required.

In November 2001 this protocol was published by ICAO as a global standard.

ADS-B supported applications

The ADS-B data link supports a number of airborne and ground applications. Each application has its own operational concepts, algorithms, procedures, standards, and user training.

Cockpit display of traffic information

A Cockpit Display of Traffic Information (CDTI) is a generic display that provides the flight crew with surveillance information about other aircraft, including their position. Traffic information for a CDTI may be obtained from one or multiple sources, including ADS-B, TCAS, and TIS-B. Direct air-to-air transmission of ADS-B messages supports display of proximate aircraft on a CDTI.

In addition to traffic based on ADS-B reports, a CDTI function might also display current weather conditions, terrain, airspace structure, obstructions, detailed airport maps, and other information relevant to the particular phase of flight.

Airborne collision avoidance

ADS-B is seen as a valuable technology to enhance ACAS operation. Incorporation of ADS-B can provide benefits such as:

- Decreasing the number of active interrogations required by ACAS, thus increasing effective range in high density airspace.
- Reducing unnecessary alarm rate by incorporating the ADS-B state vector, aircraft intent, and other information.
- Use of the ACAS display as a CDTI, providing positive identification of traffic.
- Extending collision avoidance below 1000 feet above ground level, and detecting runway incursions.

Eventually, the ACAS function may be provided based solely on ADS-B, without requiring active interrogations of other aircraft transponders.

Conflict management

ATS conformance monitoring

Other applications

Other applications that may benefit from ADS-B include:

- Improved search and rescue
- Enhanced flight following
- Lighting control and operation
- Airport ground vehicle and aircraft rescue and firefighting vehicle operational needs
- Altitude height keeping performance measurements
- General aviation operations control

U.S. implementation timetable

The Federal Aviation Administration (FAA) ADS-B implementation is broken into three segments each with a corresponding time line. Ground segment implementation and deployment is expected to begin in 2009 and be completed by 2013 throughout the National Airspace System (NAS). Airborne equipment is user-driven and is expected to be completed both voluntarily based on perceived benefits and through regulatory actions (Rulemaking) by the FAA. The cost to equip with ADS-B Out capability is relatively small and would benefit the airspace with surveillance in areas not currently served by radar. The FAA intends to provide similar service within the NAS to what radar is currently providing (5 NM en route and 3 NM terminal radar standards) as a first step to implementation. However, ADS-B In capability is viewed as the most likely way to improve NAS throughput and enhance capacity.

In December 2008 Acting FAA Administrator Robert Sturgell gave the go-ahead for ADS-B to go live in southern Florida. The south Florida installation, which consists of 11 ground stations and supporting equipment, is the first commissioned in the USA, although developmental systems have been online in Alaska, Arizona and along the East Coast since 2004. The completed system will consist of 794 ground station transceivers. The December 2008 action is in compliance with a late-term Executive Order from George W. Bush which mandated accelerated approval of NextGen.

FAA segment 1 (2006-2009)

ADS-B Coverage as of July 2009 (Segment 1)

ADS-B deployment and voluntary equipment, along with rule making activities. Pockets of development will exploit equipment deployment in the areas that will provide proof of concept for integration to ATC automation systems deployed in the NAS. It is being developed at the FAA's technical center in Egg Harbor, New Jersey.

FAA segment 2 (2010-2014)

ADS-B ground stations will be deployed throughout the NAS, with an In-Service Decision due in the 2012-13 time frame. Completed deployment will occur in the 2013-2014 time frame. Equipment is expected to begin after the proposed rule is finalized in around 2010:

- Airport Situational Awareness – A combination of detailed airport maps, airport multilateration systems, ADS-B systems and enhanced aircraft displays have the potential to significantly improve Airport Surface Situational Awareness (ASSA) and Final Approach and Runway Occupancy Awareness (FAROA).

- Oceanic In-trail – ADS-B may provide enhanced situational awareness and safety for Oceanic In-trail maneuvers as additional aircraft become equipped.
- **Gulf of Mexico** – In the Gulf of Mexico, where ATC radar coverage is incomplete, the FAA is locating ADS-B (1090 MHz) receivers on oil rigs to relay information received from aircraft equipped with ADS-B extended squitters back to the Houston Center to expand and improve surveillance coverage.

FAA segment 3 (2015-2020)

ADS-B In equipment will be based on user perceived benefit, but is expected to be providing increased situational awareness and efficiency benefits within this segment. Those aircraft who choose to equip in advance of any mandate will see benefits associated with preferential routes and specific applications. Limited radar decommissioning will begin in the time frame with an ultimate goal of a 50% reduction in the Secondary Surveillance Radar infrastructure.

On May 27, 2010 the FAA published its final rule mandating that by 2020 all aircraft owners will be required to have ADS-B Out capabilities when operating in any airspace that currently requires a transponder (airspace classes A, B, and C, and airspace class E at certain altitudes).

Worldwide

- **Australia** - Australia is the first country with full, nationwide ADS-B coverage, though only above FL300. There are 57 ground stations operating at 28 sites.

- **Canada** - Nav Canada commissioned operational use of ADS-B in 2009 and is now using it to provide coverage of its northern airspace around Hudson Bay, most of which currently has no radar coverage. The service is also being extended to cover some oceanic areas off the east coast of Canada and Greenland. The service is expected to be later extended to cover the rest of the Canadian Arctic, and to the rest of Canada.

- **China** - An American Company, ADS-B Technologies created one of the largest and most successful ADS-B system in the world (an 8 station, 350+ aircraft network that spans more than 1,200 NM across Central China). This was also the first UAT installation outside the U.S.. As of March, 2009, more than 1.2 million incident/failure free flight hours have been flown with these ADS-B systems.

- **Sweden** - LFV Group in Sweden has implemented a nationwide ADS-B network with 12 ground stations. Installation commenced during spring 2006, and the network was fully (technically) operational in 2007. An ADS-B supported system is planned for operational usage in Kiruna during spring 2009. Based on the VDL Mode 4 standards, the network of ground stations can support services for ADS-B, TIS-B, FIS-B, GNS-B (DGNSS augmentation) and Point-to-Point communication, allowing aircraft equipped with VDL 4-compliant transceivers to lower fuel consumption and reduce flight times.

- **United States**
    - **Cargo Airline Association** - Cargo carriers, notably United Parcel Service (UPS). They operate at their hub airports largely at night. Much of the benefit to these carriers is envisioned through merging and spacing the arriving and departing traffic to a more manageable flow. More environmentally friendly and efficient area navigation (RNAV) descent profiles, combined with CDTI, may allow crews to eventually aid controllers with assisted visual acquisition of traffic and limited cockpit-based separation of aircraft. The benefits to the carrier are fuel and time efficiencies associated with idle descent and shorter traffic patterns than typical radar vectoring allows.
    - **Embry-Riddle Aeronautical University** - ERAU has equipped their training aircraft at its two main campuses in Florida and Arizona with UAT ADS-B capability as a situational safety enhancement. The University has been doing this since May 2003, making it the first use in general aviation. With the addition of the G1000 to their fleet in 2006, ERAU became the first fleet to combine a glass cockpit with ADS-B.
    - **University of North Dakota** - UND has received an FAA grant to test ADS-B, and has begun to outfit their Piper Warrior fleet with an ADS-B package.

- **United Arab Emirates** - UAE commissioned three operational redundant ADS-B ground stations in early 2009 and is now using ADS-B to provide enhanced coverage of its upper airspace in combination and integrated with conventional surveillance radars.

- Use of ADS-B and CDTI may allow decreased approach spacing at certain airports to improve capacity during reduced-visibility operations when visual approach operations would normally be terminated (e.g. ceilings less than MVA +500).

System design considerations of ADS-B

A concern for any ADS-B protocol is the capacity for carrying ADS-B messages from aircraft, as well as allowing the radio channel to continue to support any legacy services. For 1090ES, each ADS-B message is composed of a pair of data packets. The greater the number of packets transmitted from one aircraft, the lesser the number of aircraft that can participate in the system, due to the fixed and limited channel data bandwidth.

System capacity is defined by establishing a criterion for what the worst environment is likely to be, then making that a minimum requirement for system capacity. For 1090ES, both TCAS and ATCRBS/MSSR are existing users of the channel. 1090ES ADS-B must not reduce capacity of these existing systems.

The FAA national program office and other International aviation regulators are addressing concerns about ADS-B non-secure nature of ADS-B transmissions. ADS-B messages can be used to know the location of an aircraft, and there is no means to guarantee that this information is not used inappropriately. Additionally, there are some concerns about the integrity of ADS-B transmissions. ADS-B messages can be produced, with simple low cost measures, which spoof the locations of multiple phantom aircraft to disrupt safe air travel. There is no foolproof means to guarantee integrity, but there are means to monitor for this type of activity. This problem is however similar to the usage of ATCRBS/MSSR where false signals also are potentially dangerous (uncorrelated secondary tracks).

There are some concerns about ADS-B dependence on satellite navigation systems to generate state vector information, although the risks can be mitigated by using redundant sources of state vector information, e.g. GPS, GLONASS, Galileo or multilateration.

There are some General Aviation concerns that ADS-B removes anonymity of the VFR aircraft operations. The ICAO 24-bit transponder code specifically assigned to each aircraft will allow monitoring of that aircraft when within the service volumes of the Mode-S/ADS-B system. Unlike the Mode A/C transponders, there is no code "1200"/"7000", which offers casual anonymity. Mode-S/ADS-B identifies the aircraft uniquely among all in the world, in a similar fashion as a MAC number for an Ethernet card or the IMEI (International Mobile Equipment Identity) of a GSM phone.

Public Access to ADS-B

Currently there are no laws preventing the public from listening to and decoding ADS-B transmissions. Like Cellular Phone however, laws can easily be implemented to make reception a crime, and receivers would then become contraband. The ongoing debate amongst hobbyists is to display real-time activity on personal screens and then delay five minutes on networked displays. Others feel that, like GPS data, it should be freely available. In late 2009, all receiver manufacturers abandoned the five minute delay, and all operate in real-time on their network interfaces.

ADS-B technical and regulatory documents

MASPS = Minimum Aviation System Performance Standards
MOPS = Minimum Operational Performance Standards

- DO-289 - Airborne Surveillance Applications (ASA) MOPS
    - High level system architecture and sub-system descriptions.
- DO-242A - ADS-B MASPS
    - Describes system-wide operational use of ADS-B.
- DO-286A - TIS-B MASPS
    - Describes a surveillance service that derives traffic information from ground surveillance sources, broadcasts to ADS-B equipped aircraft or surface vehicles.
- DO-260A - 1090 MOPS for ADS-B and TIS-B
    - Airborne equipment characteristics / requirements for 1090 MHz Mode-S extended squitter.
    - Change 1-DO-260A–PMC consideration review/approval –June 27, 2006
- DO-282A - UAT MOPS
    - Airborne equipment characteristics / requirements utilizing the universal access transceiver.
- DO-XXX - STP MOPS (work in progress)
    - Describes a function that processes information prior to the information being broadcast by the ADS-B transmit function.
- DO-XXX - ASAS MOPS (work in progress)
    - Adds additional subsystems necessary to fully implement AirborneSurveillance Applications:
        - Airborne Surveillance and Separation Assurance Processing (ASSAP)
        - Cockpit Display of Traffic Information (CDTI)
- DO-259 - CDTI Application Description
    - Provides initial CDTI applications descriptions.

Chapter 3

Direction Finding

Direction finding (DF) refers to the establishment of the direction from which a received signal was transmitted. This can refer to radio or other forms of wireless communication. By combining the direction information from two or more suitably spaced receivers (or a single mobile receiver), the source of a transmission may be located in space via triangulation.

Antennas

Direction finding often requires an antenna that is directional - that is, more sensitive in certain directions than in others. Many antenna designs exhibit this property. For example, a Yagi antenna has quite pronounced directionality, so the source of a transmission can be determined simply by pointing it in the direction where the maximum signal level is obtained. However, to establish direction to great accuracy requires much more sophisticated techniques.

The crossed-loops DF antenna atop the mast of a tug boat

A simple form of directional antenna is the loop aerial. This consists of an open loop of wire on an insulating former, or a metal ring that forms the antenna elements itself, where the diameter of the loop is a tenth of a wavelength or smaller at the target frequency. Such an antenna will be LEAST sensitive to signals that are normal to its face, and MOST responsive to those meeting edge-on, this due to the antenna sensing the difference between the voltages induced either side of it at any instant because of the phase output of the transmitting beacon. Turning the loop face on will not induce any current flow - think of the radio wave slipping through the loop. Simply turning the antenna to obtain minimum signal will establish two possible directions from which the signal could be emanating. The NULL is used, as small angular deflections of the loop aerial near its null positions produce larger changes in current than similar angular changes near the loops max positions. For this reason, a null position of the loop aerial is used. To resolve the two direction possibilities, a sense antenna is used, the sense aerial has no directional properties but has the same sensitivity as the loop aerial. By adding the steady signal from the sense aerial to the alternating signal from the loop signal as it rotates, there is now only one position as the loop rotates 360° at which there is zero current. This acts as a phase ref point, allowing the correct null point to be identified, thus removing the 180° ambiguity. A dipole antenna exhibits similar properties, and is the basis for the Yagi antenna, which is familiar as the common VHF or UHF television aerial. For much higher frequencies still, parabolic antennas can be used, which are highly directional, focusing received signals from a very narrow angle to a receiving element at the centre.

More sophisticated techniques such as phased arrays are generally used for highly accurate direction finding systems called goniometers such as are used in signals intelligence (SIGINT). A helicopter based DF system was designed by ESL Incorporated for the U.S. Government as early as 1972.

Single channel DF

Single-channel DF refers to the use of a multi-antenna array with a single channel radio receiver. This approach to DF obviously offers some advantages and drawbacks. Since it only uses one receiver, mobility and lower power consumption are obvious benefits but without the ability to look at each antenna simultaneously (which would be the case if one were to use multiple receivers) more complex operations need to occur at the antenna in order to present the signal to the receiver.

The two main categories that a single channel DF algorithm falls into are *amplitude comparison* and *phase comparison*. Some algorithms can be hybrids of the two.

Pseudo-doppler DF technique

The pseudo-doppler technique is a phase based DF method that produces a bearing estimate on the received signal by measuring the doppler shift induced on the signal by sampling around the elements of a circular array. The original method used a single antenna that physically moved in a circle but the modern approach uses a multi-antenna circular array with each antenna sampled in succession.

Watson-Watt / Adcock antenna array

The Watson-Watt technique uses two Adcock antenna pairs to perform an amplitude comparison on the incoming signal. An Adcock antenna pair is a pair of monopole or dipole antennas that takes the vector difference of the received signal at each antenna so that there is only one output from the pair of antennas. Two of these pairs are co-located but perpendicularly oriented to produce what can be referred to as the N-S (North-South) and E-W (East-West) signals that will then be passed to the receiver. In the receiver, the bearing angle can then be computed by taking the arctangent of the ratio of the N-S to E-W signal.

Usage

Radio navigation

A portable, battery operated GT-302 Accumatic automatic direction finder for marine use

- *Radio direction finding*, or *RDF* was once the primary aviation navigational aid. Beacons were used to mark "airways" intersections and to define departure and approach procedures. Since the signal transmitted contains no information about bearing or distance, these beacons are referred to as *non-directional beacons*, or *NDB* in the aviation world. Starting in the 1950s, these beacons are generally being replaced by the VOR system, in which the bearing to the navigational aid is measured from the signal itself, therefore no specialized antenna with moving parts is required. Due to relatively low purchase, maintenance and calibration cost, NDB's are still used to mark locations of smaller aerodromes and important helicopter landing sites.

- Similar beacons located in coastal areas are also used for maritime radio navigation, as almost every ship is (was) equipped with a direction finder (Appleyard 1988). Very few maritime radionavigation beacons remain active today (2008) as ships have abandoned navigation via RDF in favor of GPS navigation.

- In the United Kingdom a radio direction finding service is available on 121.5 MHz and 243.0 MHz to aircraft pilots who are in distress or are experiencing difficulties. The service is based on a number of radio DF units located at civil and military airports and certain HM Coastguard stations. These stations can obtain a "fix" of the aircraft and transmit it by radio to the pilot.

Location of illegal, secret or hostile transmitters - SIGINT

In WW2 Considerable effort was expended on identifying secret transmitters in the UK by direction finding. The work was undertaken by the Radio Security Service (RSS). Initially three U Adcock HF DF stations were set up in 1939 by the General Post Office but with the declaration of war, MI5 and MI8 developed this into a larger network. One of the problems with providing coverage of an area the size of UK was installing sufficient DF stations to cover the entire area to receive skywave signals reflected back from the ionised layers in the upper atmosphere. Even with the expanded network, some areas were not adequately covered and for this reason up to 1700 voluntary interceptors (radio amateurs) were recruited to detect illicit transmissions by ground wave. In addition to the fixed stations RSS ran a fleet of mobile DF vehicles around the UK. If a transmitter was identified by the fixed DF stations or voluntary interceptors, the mobile units were sent to the area to home in on the source. The mobile units were H Adcock systems

By 1941 only a couple of illicit transmitters had been identified in the UK but they were German agents that had been 'turned' and were transmiting under MI5 control. But many illicit transmissions had been logged emanating from German agents in occupied and neutral countries in Europe. The traffic became a valuable source of intelligence and the control of RSS was subsequently passed to MI6 who were responsible for secret intelligence originating from outside the UK. The direction finding and interception operation increased in volume and importance until 1945.

The HF Adcock stations consisted of four 10m vertical antennas surrounding a small wooden operators hut containing a receiver and a radiogoniometer which was adjusted to obtain the bearing. MF stations were also used which used four guyed 30m lattice tower antennas. In 1941 RSS began experimenting with Spaced Loop direction finders developed by the Marconi company and the UK NationalPhysical Laboratories. These consisted of two parallel loops 1 to 2m square on the ends of a rotatable 3 to 8m beam. The angle of the beam was combined with results from a radiogoniometer to provide a bearing. The bearing obtained was considerably sharper than that obtained with the U Adcock system but there were ambiguities which prevented the installation of 7 proposed S.L df systems. The SL systems involved the operator being located in a metal tank underground directly below the antennas. 7 underground tanks were installed but only two SL systems were installed at Wymondham, Norfolk and Weaverthorp in Yorkshire. Problems were encountered resulting in the remaining 5 underground tanks being fitted with U adcock systems. The rotating SL antenna was turned by hand by the operator in the underground tank which meant successive measurements were a lot slower than turning the dial of a goniometer. Another experimental SL station was built near Aberdeen in 1942 for the Air Ministry and this involved a semi-underground concrete bunker but this too was abandoned because of operating difficulties. By 1944 a mobile version of the spaced loop had been developed and was used by RSS in France following the D-Day invasion of Normandy. The US military used a shore based version of the Spaced Loop DF in WW2 called "DAB". The loops were placed at the ends of a beam, all of which was located inside a wooden hut with the electronics in a large cabinet with CRT display at the centre of the beam and everything being supported on a central axis. The beam was rotated manually by the operator.

The Royal Navy introduced a variation on the shore based HF DF stations in 1944 to track U-boats in the North Atlantic. They built groups of 5 df stations so that bearings from individual stations in the group could be combined and a mean taken. Four such groups were built in Britain at Ford End, Essex, Gonhavern, Cornwall, Anstruther and Bowermadden in the Scottish Highlands. Groups were also built in Iceland, Nova Scotia and Jamaica. The anticipated improvements were not realised but later statistical work improved the system and the Goonhavern and Ford End groups continued to be used during the Cold War.

Arguably the most comprehensive book on wireless direction finding was written by Roland Keen who was head of the engineering department of RSS at Hanslope Park during WW2. The DF systems mentioned here are described in detail in his exhaustive treatment of the subject in the 1947 edition of his book Wireless Direction Finding.

At the end of WW2 a number of RSS DF stations continued to operate into the cold war under the control of GCHQ the British SIGINT organisation.

Most direction finding effort within the UK now (2009) is directed towards locating unauthorised 'pirate' FM broadcast radio transmissions. A network of remotely operated VHF direction finders are used mainly located around the major cities. The transmissions from mobile telephone handsets are also located by a form of direction finding using the comparative signal strength at the surrounding local 'cell' receivers. This technique is often offered as evidence in UK criminal prosecutions and, almost certainly, for SIGINT purposes.

Emergency aid

There are many forms of radio transmitters designed to transmit as a beacon in the event of an emergency, which are widely deployed on civil aircraft. Modern emergency beacons transmit a unique identification signal that can aid in finding the exact location of the transmitter.

Avalanche rescue

Avalanche transceivers operate on a standard 457 kHz, and are designed to help locate people and equipment buried by avalanches. Since the power of the beacon is so low the directionality of the radio signal is dominated by small scale field effects and can be quite complicated to locate.

Wildlife tracking

Location of radio-tagged animals by triangulation is a widely applied research technique for studying the movement of animals. The technique was first used in the early 1960s, when the technology used in radio transmitters and batteries made them small enough to attach to wild animals, and is now widely deployed for a variety of wildlife studies. Most tracking of wild animals that have been affixed with radio transmitter equipment is done by a field researcher using a handheld radio direction finding device. When the researcher wants to locate a particular animal, the location of the animal can be triangulated by determining the direction to the transmitter from several locations.

Reconnaissance

Phased arrays and other advanced antenna techniques are utilized to track launches of rocket systems and their resulting trajectories. These systems can be used for defensive purposes and also to gain intelligence on operation of missiles belonging to other nations. These same techniques are used for detection and tracking of conventional aircraft.

Sport

Events hosted by groups and organizations that involve the use of radio direction finding skills to locate transmitters at unknown locations have been popular since the end of World War II. Many of these events were first promoted in order to practice the use of radio direction finding techniques for disaster response and civil defense purposes, or to practice locating the source of radio frequency interference. The most popular form of the sport, worldwide, is known as Amateur Radio Direction Finding or by its international acronym ARDF. Another form of the activity, known as "transmitter hunting", "mobile T-hunting" or "fox hunting" takes place in a larger geographic area, such as the metropolitan area of a large city, and most participants travel in motor vehicles while attempting to locate one or more radio transmitters with radio direction finding techniques.

Chapter 4

Error Analysis for the Global Positioning System

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

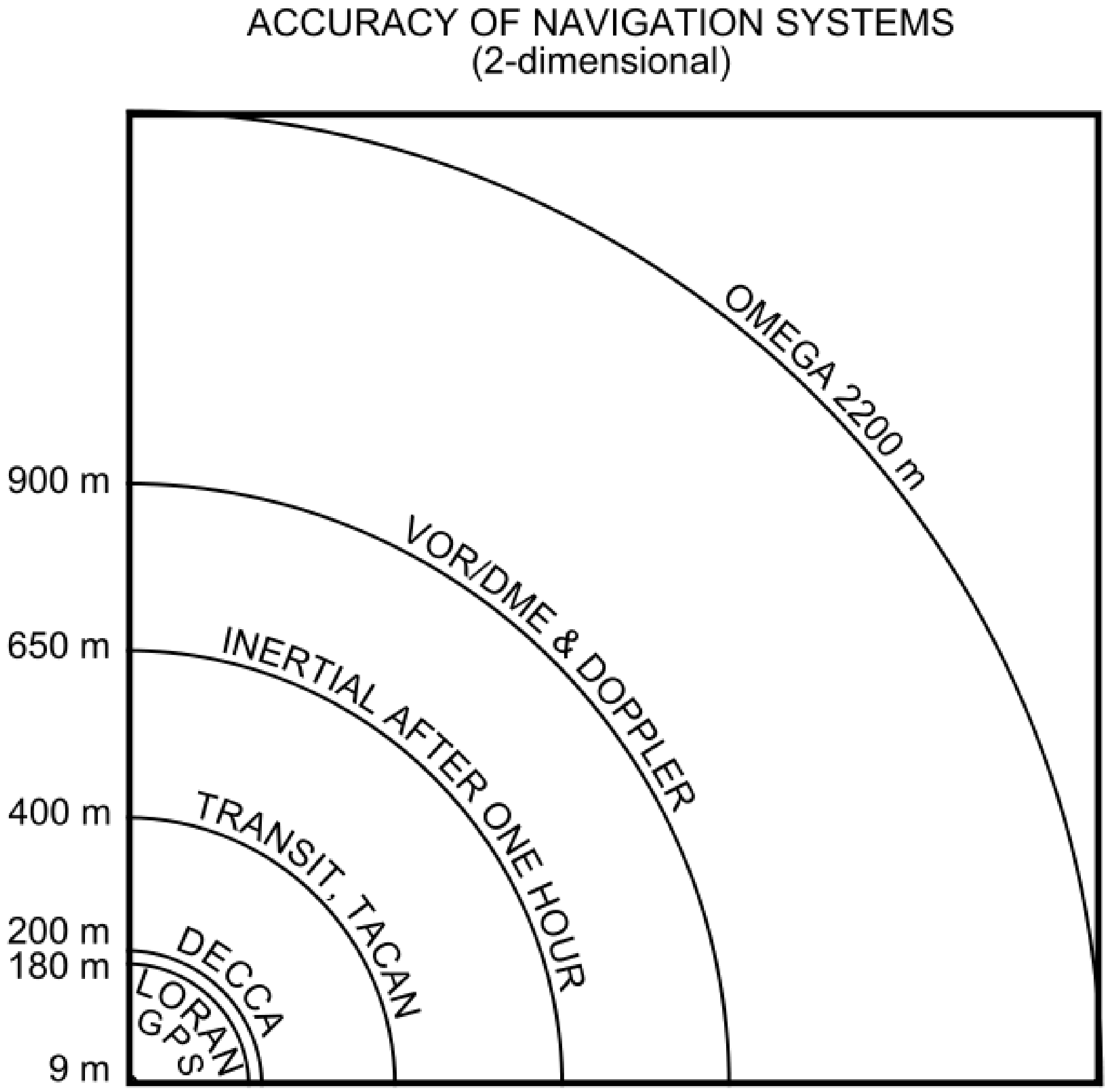
The **analysis of errors in the information reported by the Global Positioning System**, a space-based satellite system for navigation, is important to estimating the accuracy of position estimates and correcting for the errors. The Global Positioning System (GPS) was created by the United States Department of Defense (DOD) in the 1970s. It has come to be widely used for navigation both by the U.S. military and the general public.

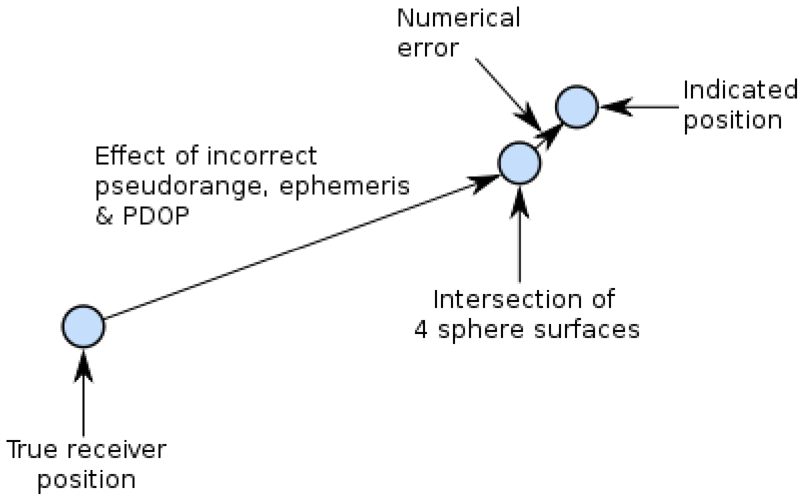
The positioning data provided directly by the satellites is extremely precise but there are many factors that can create make the errors in the data non-trivial. In situations where high accuracy is necessary, understanding and compensating for these sources of error is important. Sources of error include atmospheric distortion (predominantly in the ionosphere), satellite clock inaccuracies, and the travel delays of the satellite signals.

Overview

Sources of User Equivalent Range Errors (UERE)

Source	Effect (m)
Signal arrival C/A	±3
Signal arrival P(Y)	±0.3
Ionospheric effects	±5
Ephemeris errors	±2.5
Satellite clock errors	±2
Multipath distortion	±1
Tropospheric effects	±0.5
σ_RC/A	±6.7
σ_RP(Y)	±6.0


ACCURACY OF NAVIGATION SYSTEMS
(2-dimensional)
900 m
650 m
400 m
200 m
180 m
9 m
OMEGA 2200 m
VOR/DME & DOPPLER
INERTIAL AFTER ONE HOUR
TRANSIT, TACAN
DECCA
LORAN
GPS


Numerical error
Indicated position
Effect of incorrect pseudorange, ephemeris & PDOP
Intersection of 4 sphere surfaces
True receiver position

Geometric Error Diagram Showing Typical Relation of Indicated Receiver Position, Intersection of Sphere Surfaces, and True Receiver Position in Terms of Pseudorange Errors, PDOP, and Numerical Errors

User equivalent range errors (UERE) are shown in the table. There is also a numerical error with an estimated value, σ_{num}, of about 1 meter. The standard deviations, σ_R, for the coarse/acquisition and precise codes are also shown in the table. These standard deviations are computed by taking the square root of the sum of the squares of the individual components (i.e., RSS for root sum squares). To get the standard deviation of receiver position estimate, these range errors must be multiplied by the appropriate dilution of precision terms and then RSS'ed with the numerical error. Electronics errors are one of several accuracy-degrading effects outlined in the table above. When taken together, autonomous civilian GPS horizontal position fixes are typically accurate to about 15 meters (50 ft). These effects also reduce the more precise P(Y) code's accuracy. However, the advancement of technology means that today, civilian GPS fixes under a clear view of the sky are on average accurate to about 5 meters (16 ft) horizontally.

The term user equivalent range error (UERE) refers to the error of a component in the distance from receiver to a satellite. These UERE errors are given as ± errors thereby implying that they are unbiased or zero mean errors. These UERE errors are therefore used in computing standard deviations. The standard deviation of the error in receiver position, σ_{rc}, is computed by multiplying PDOP (Position Dilution Of Precision) by σ_R, the standard deviation of the user equivalent range errors. $\sigma_R$is computed by taking the square root of the sum of the squares of the individual component standard deviations.

PDOP is computed as a function of receiver and satellite positions. A detailed description of how to calculate PDOP is given in the section, geometric dilution of precision computation (GDOP).

$\sigma_R$for the C/A code is given by:

$$\sigma_R = \sqrt{3^2 + 5^2 + 2.5^2 + 2^2 + 1^2 + 0.5^2}\,\mathrm{m} \ = \ 6.7\,\mathrm{m}$$

The standard deviation of the error in estimated receiver position σ_{rc}, again for the C/A code is given by:

$$\sigma_{rc} = \sqrt{PDOP^2 \times \sigma_R^2 + \sigma_{num}^2} = \sqrt{PDOP^2 \times 6.7^2 + 1^2}\,\mathrm{m}$$

The error diagram on the left shows the inter relationship of indicated receiver position, true receiver position, and the intersection of the four sphere surfaces.

Signal arrival time measurement

The position calculated by a GPS receiver requires the current time, the position of the satellite and the measured delay of the received signal. The position accuracy is primarily dependent on the satellite position and signal delay.

To measure the delay, the receiver compares the bit sequence received from the satellite with an internally generated version. By comparing the rising and trailing edges of the bit transitions, modern electronics can measure signal offset to within about one percent of a bit pulse width, $\frac{0.01}{(1.023 \times 10^6/\mathrm{s})}$, or approximately 10 nanoseconds for the C/A code. Since GPS signals propagate at the speed of light, this represents an error of about 3 meters.

This component of position accuracy can be improved by a factor of 10 using the higher-chiprate P(Y) signal. Assuming the same one percent of bit pulse width accuracy, the high-frequency P(Y) signal results in an accuracy of $\frac{(0.01 \times 300{,}000{,}000\ \mathrm{m/s})}{(10.23 \times 10^6/\mathrm{s})}$ or about 30 centimeters.

Atmospheric effects

Inconsistencies of atmospheric conditions affect the speed of the GPS signals as they pass through the Earth's atmosphere, especially the ionosphere. Correcting these errors is a significant challenge to improving GPS position accuracy. These effects are smallest when the satellite is directly overhead and become greater for satellites nearer the horizon since the path through the atmosphere is longer. Once the receiver's approximate location is known, a mathematical model can be used to estimate and compensate for these errors.

Ionospheric delay of a microwave signal depends on its frequency. This phenomenon is known as dispersion and can be calculated from measurements of delays for two or more frequency bands, allowing delays at other frequencies to be estimated. Some military and expensive survey-grade civilian receivers calculate atmospheric dispersion from the different delays in the L1 and L2 frequencies, and apply a more precise correction. This can be done in civilian receivers without decrypting the P(Y) signal carried on L2, by tracking the carrier wave instead of the modulated code. To facilitate this on lower cost receivers, a new civilian code signal on L2, called L2C, was added to the Block IIR-M satellites, which was first launched in 2005. It allows a direct comparison of the L1 and L2 signals using the coded signal instead of the carrier wave.

The effects of the ionosphere generally change slowly, and can be averaged over time. Those for any particular geographical area can be easily calculated by comparing the GPS-measured position to a known surveyed location. This correction is also valid for other receivers in the same general location. Several systems send this information over radio or other links to allow L1-only receivers to make ionospheric corrections. The ionospheric data are transmitted via satellite in Satellite Based Augmentation Systems (SBAS) such as Wide Area Augmentation System (WAAS) (available in North America and Hawaii), EGNOS (Europe and Asia) or Multi-functional Satellite Augmentation System (MSAS) (Japan), which transmits it on the GPS frequency using a special pseudo-random noise sequence (PRN), so only one receiver and antenna are required.

Humidity also causes a variable delay, resulting in errors similar to ionospheric delay, but occurring in the troposphere. This effect both is more localized and changes more quickly than ionospheric effects, and is not frequency dependent. These traits make precise measurement and compensation of humidity errors more difficult than ionospheric effects.

Changes in receiver altitude also change the delay, due to the signal passing through less of the atmosphere at higher elevations. Since the GPS receiver computes its approximate altitude this error is relatively simple to correct, either by applying a function regression or correlating margin of atmospheric error to ambient pressure using a barometric altimeter.

Multipath effects

GPS signals can also be affected by multipath issues, where the radio signals reflect off surrounding terrain; buildings, canyon walls, hard ground, etc. These delayed signals can cause inaccuracy. A variety of techniques, most notably narrow correlator spacing, have been developed to mitigate multipath errors. For long delay multipath, the receiver itself can recognize the wayward signal and discard it. To address shorter delay multipath from the signal reflecting off the ground, specialized antennas (e.g., a choke ring antenna) may be used to reduce the signal power as received by the antenna. Short delay reflections are harder to filter out because they interfere with the true signal, causing effects almost indistinguishable from routine fluctuations in atmospheric delay.

Multipath effects are much less severe in moving vehicles. When the GPS antenna is moving, the false solutions using reflected signals quickly fail to converge and only the direct signals result in stable solutions.

Ephemeris and clock errors

While the ephemeris data is transmitted every 30 seconds, the information itself may be up to two hours old. If a fast time to first fix (TTFF) is needed, it is possible to upload a valid ephemeris to a receiver, and in addition to setting the time, a position fix can be obtained in under ten seconds. It is feasible to put such ephemeris data on the web so it can be loaded into mobile GPS devices.

The satellite's atomic clocks experience noise and clock drift errors. The navigation message contains corrections for these errors and estimates of the accuracy of the atomic clock. However, they are based on observations and may not indicate the clock's current state.

These problems tend to be very small, but may add up to a few meters (tens of feet) of inaccuracy.

For very precise positioning (e.g., in geodesy), these effects can be eliminated by differential GPS: the simultaneous use of two or more receivers at several survey points. In the 1990s when receivers were quite expensive, some methods of *quasi-differential* GPS were developed, using only *one* receiver but reoccupation of measuring points. At the TU Vienna the method was named *qGPS* and adequate software of post processing was developed.

Geometric dilution of precision computation (GDOP)

The concept of geometric dilution of precision was introduced in the section, *error sources and analysis*. Computations were provided to show how PDOP was used and how it affected the receiver position error standard deviation.

When visible GPS satellites are close together in the sky (i.e., small angular separation), the DOP values are high; when far apart, the DOP values are low. Conceptually, satellites that are close together cannot provide as much information as satellites that are widely separated. Low DOP values represent a better GPS positional accuracy due to the wider angular separation between the satellites used to calculate GPS receiver position. HDOP, VDOP, PDOP and TDOP are respectively Horizontal, Vertical, Position (3-D) and Time Dilution of Precision.

Figure 3.1 Dilution of Precision of Navstar GPS data from the U.S. Coast Guard provide a graphical indication of how geometry affect accuracy.

We now take on the task of how to compute the dilution of precision terms. As a first step in computing DOP, consider the unit vector from the receiver to satellite i with components $\frac{(x_i - x)}{R_i}$, $\frac{(y_i - y)}{R_i}$, and $\frac{(z_i - z)}{R_i}$ where the distance from receiver to the satellite, R_i, is given by:

$$R_i = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}$$

where $x, y, and\ z$ denote the position of the receiver and $x_i, y_i, and\ z_i$ denote the position of satellite i. These x, y, and z components may be components in a North, East, Down coordinate system a South, East, Up coordinate system or other convenient system. Formulate the matrix *A* as:

$$A = \begin{bmatrix} \frac{(x_1-x)}{R_1} & \frac{(y_1-y)}{R_1} & \frac{(z_1-z)}{R_1} & c \\ \frac{(x_2-x)}{R_2} & \frac{(y_2-y)}{R_2} & \frac{(z_2-z)}{R_2} & c \\ \frac{(x_3-x)}{R_3} & \frac{(y_3-y)}{R_3} & \frac{(z_3-z)}{R_3} & c \\ \frac{(x_4-x)}{R_4} & \frac{(y_4-y)}{R_4} & \frac{(z_4-z)}{R_4} & c \end{bmatrix}$$

The first three elements of each row of A are the components of a unit vector from the receiver to the indicated satellite. The elements in the fourth column are c where c denotes the speed of light. Formulate the matrix, Q, as

$$Q = \left(A^T A\right)^{-1}$$

This computation is in accordance with Chapter 11 of The global positioning system by Parkinson and Spilker where the weighting matrix, P, has been set to the identity matrix. The elements of the Q matrix are designated as:

$$Q = \begin{bmatrix} d_x^2 & d_{xy}^2 & d_{xz}^2 & d_{xt}^2 \\ d_{xy}^2 & d_y^2 & d_{yz}^2 & d_{yt}^2 \\ d_{xz}^2 & d_{yz}^2 & d_z^2 & d_{zt}^2 \\ d_{xt}^2 & d_{yt}^2 & d_{zt}^2 & d_t^2 \end{bmatrix}$$

The Greek letter σ is used quite often where we have used d. However the elements of the Q matrix do not represent variances and covariances as they are defined in probability and statistics. Instead they are strictly geometric terms. Therefore d as in dilution of precision is used. PDOP, TDOP and GDOP are given by

$$PDOP = \sqrt{d_x^2 + d_y^2 + d_z^2},$$
$$TDOP = \sqrt{d_t^2} = |d_t|, \text{ and}$$
$$GDOP = \sqrt{PDOP^2 + TDOP^2}$$

in agreement with "Section 1.4.9 of PRINCIPLES OF SATELLITE POSITIONING".

The horizontal dilution of precision, $HDOP = \sqrt{d_x^2 + d_y^2}$, and the vertical dilution of precision, $VDOP = \sqrt{d_z^2}$, are both dependent on the coordinate system used. To correspond to the local horizon plane and the local vertical, x, y, and z should denote positions in either a North, East, Down coordinate system or a South, East, Up coordinate system.

Derivation of DOP equations

The equations for computing the geometric dilution of precision terms have been described in the previous section. This section describes the derivation of these equations. The method used here is similar to that used in "Global Positioning System (preview) by Parkinson and Spiker"

Consider the position error vector, $\mathbf{e}$, defined as the vector from the intersection of the four sphere surfaces corresponding to the pseudoranges to the true position of the receiver. $\mathbf{e} = e_x\hat{x} + e_y\hat{y} + e_z\hat{z}$ where bold denotes a vector and $\hat{x}$, $\hat{y}$, and $\hat{z}$ denote unit vectors along the x, y, and z axes respectively. Let e_t denote the time error, the true time minus the receiver indicated time. Assume that the mean value of the three components of $\mathbf{e}$ and e_t are zero.

$$A \begin{bmatrix} e_x \\ e_y \\ e_z \\ e_t \end{bmatrix} = \begin{bmatrix} \frac{(x_1-x)}{R_1} & \frac{(y_1-y)}{R_1} & \frac{(z_1-z)}{R_1} & c \\ \frac{(x_2-x)}{R_2} & \frac{(y_2-y)}{R_2} & \frac{(z_2-z)}{R_2} & c \\ \frac{(x_3-x)}{R_3} & \frac{(y_3-y)}{R_3} & \frac{(z_3-z)}{R_3} & c \\ \frac{(x_4-x)}{R_4} & \frac{(y_4-y)}{R_4} & \frac{(z_4-z)}{R_4} & c \end{bmatrix} \begin{bmatrix} e_x \\ e_y \\ e_z \\ e_t \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad (1)$$

where $e_1,\ e_2,\ e_3,\ and\ e_4$ are the errors in pseudoranges 1 through 4 respectively. This equation comes from linearizing the equation relating pseudoranges to receiver position, satellite positions, and receiver clock errors as shown in. Multiplying both sides by A^{-1} there results

$$\begin{bmatrix} e_x \\ e_y \\ e_z \\ e_t \end{bmatrix} = A^{-1} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad (2).$$

Transposing both sides:

$$\begin{bmatrix} e_x & e_y & e_z & e_t \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & e_3 & e_4 \end{bmatrix} \left(A^{-1}\right)^T \quad (3).$$

Post multiplying the matrices on both sides of equation (2) by the corresponding matrices in equation (3), there results

$$\begin{bmatrix} e_x \\ e_y \\ e_z \\ e_t \end{bmatrix} \begin{bmatrix} e_x & e_y & e_z & e_t \end{bmatrix} = A^{-1} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \begin{bmatrix} e_1 & e_2 & e_3 & e_4 \end{bmatrix} \left(A^{-1}\right)^T \quad (4).$$

Taking the expected value of both sides and taking the non-random matrices outside the expectation operator, E, there results:

$$E\left(\begin{bmatrix} e_x \\ e_y \\ e_z \\ e_t \end{bmatrix}\begin{bmatrix} e_x & e_y & e_z & e_t \end{bmatrix}\right) = A^{-1}\, E\left(\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}\begin{bmatrix} e_1 & e_2 & e_3 & e_4 \end{bmatrix}\right)\left(A^{-1}\right)^T \quad (5)$$

Assuming the pseudorange errors are uncorrelated and have the same variance, the covariance matrix on the right side can be expressed as a scalar times the identity matrix. Thus

$$\begin{bmatrix} \sigma_x^2 & \sigma_{xy}^2 & \sigma_{xz}^2 & \sigma_{xt}^2 \\ \sigma_{xy}^2 & \sigma_y^2 & \sigma_{yz}^2 & \sigma_{yt}^2 \\ \sigma_{xz}^2 & \sigma_{yz}^2 & \sigma_z^2 & \sigma_{zt}^2 \\ \sigma_{xt}^2 & \sigma_{yt}^2 & \sigma_{zt}^2 & \sigma_t^2 \end{bmatrix} = \sigma_R^2\ A^{-1}\left(A^{-1}\right)^T = \sigma_R^2\ \left(A^T A\right)^{-1} \quad (6)$$

since $A^{-1}\left(A^{-1}\right)^T\left(A^T A\right) = I$

Note: $\left(A^{-1}\right)^T = \left(A^T\right)^{-1}$,since $I = \left(AA^{-1}\right)^T = \left(A^{-1}\right)^T A^T$

Substituting for $\left(A^T A\right)^{-1} = Q$ there follows

$$\begin{bmatrix} \sigma_x^2 & \sigma_{xy}^2 & \sigma_{xz}^2 & \sigma_{xt}^2 \\ \sigma_{xy}^2 & \sigma_y^2 & \sigma_{yz}^2 & \sigma_{yt}^2 \\ \sigma_{xz}^2 & \sigma_{yz}^2 & \sigma_z^2 & \sigma_{zt}^2 \\ \sigma_{xt}^2 & \sigma_{yt}^2 & \sigma_{zt}^2 & \sigma_t^2 \end{bmatrix} = \sigma_R^2 \begin{bmatrix} d_x^2 & d_{xy}^2 & d_{xz}^2 & d_{xt}^2 \\ d_{xy}^2 & d_y^2 & d_{yz}^2 & d_{yt}^2 \\ d_{xz}^2 & d_{yz}^2 & d_z^2 & d_{zt}^2 \\ d_{xt}^2 & d_{yt}^2 & d_{zt}^2 & d_t^2 \end{bmatrix} \quad (7)$$

From equation (7), it follows that the variances of indicated receiver position and time are

$$\sigma_{rc}^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2 = \sigma_R^2\left(d_x^2 + d_y^2 + d_z^2\right) = PDOP^2\sigma_R^2$$ and

$$\sigma_t^2 = \sigma_R^2 d_t^2 = TDOP^2\sigma_R^2$$

The remaining position and time error variance terms follow in a straightforward manner.

Selective availability

GPS includes a (currently disabled) feature called *Selective Availability* (*SA*) that adds intentional, time varying errors of up to 100 meters (328 ft) to the publicly available navigation signals. This was intended to deny an enemy the use of civilian GPS receivers for precision weapon guidance.

SA errors are actually pseudorandom, generated by a cryptographic algorithm from a classified *seed* key available only to authorized users (the U.S. military, its allies and a few other users, mostly government) with a special military GPS receiver. Mere possession of the receiver is insufficient; it still needs the tightly controlled daily key.

Before it was turned off on May 1, 2000, typical SA errors were about 50 m (164 ft) horizontally and about 100 m (328 ft) vertically. Because SA affects every GPS receiver in a given area almost equally, a fixed station with an accurately known position can measure the SA error values and transmit them to the local GPS receivers so they may correct their position fixes. This is called Differential GPS or *DGPS*. DGPS also corrects for several other important sources of GPS errors, particularly ionospheric delay, so it continues to be widely used even though SA has been turned off. The ineffectiveness of SA in the face of widely available DGPS was a common argument for turning off SA, and this was finally done by order of President Clinton in 2000.

Another restriction on GPS, antispoofing, remains on. This encrypts the *P-code* so that it cannot be mimicked by an enemy transmitter sending false information. Few civilian receivers have ever used the P-code, and the accuracy attainable with the public C/A code is so much better than originally expected (especially with DGPS) that the antispoof policy has relatively little effect on most civilian users. Turning off antispoof would primarily benefit surveyors and some scientists who need extremely precise positions for experiments such as tracking the motion of a tectonic plate.

DGPS services are widely available from both commercial and government sources. The latter include WAAS and the U.S. Coast Guard's network of LF marine navigation beacons. The accuracy of the corrections depends on the distance between the user and the DGPS receiver. As the distance increases, the errors at the two sites will not correlate as well, resulting in less precise differential corrections.

During the 1990-91 Gulf War, the shortage of military GPS units caused many troops and their families to buy readily available civilian units. This significantly impeded the U.S. military's own battlefield use of GPS, so the military made the decision to turn off SA for the duration of the war.

In the 1990s, the FAA started pressuring the military to turn off SA permanently. This would save the FAA millions of dollars every year in maintenance of their own radio navigation systems. The amount of error added was "set to zero" at midnight on May 1, 2000 following an announcement by U.S. President Bill Clinton, allowing users access to the error-free L1 signal. Per the directive, the induced error of SA was changed to add no error to the public signals (C/A code). Clinton's executive order required SA to be set to zero by 2006; it happened in 2000 once the U.S. military developed a new system that provides the ability to deny GPS (and other navigation services) to hostile forces in a specific area of crisis without affecting the rest of the world or its own military systems.

Selective Availability is still a system capability of GPS, and could, in theory, be reintroduced at any time. In practice, in view of the hazards and costs this would induce for U.S. and foreign shipping, it is unlikely to be reintroduced, and various government agencies, including the FAA, have stated that it is not intended to be reintroduced.

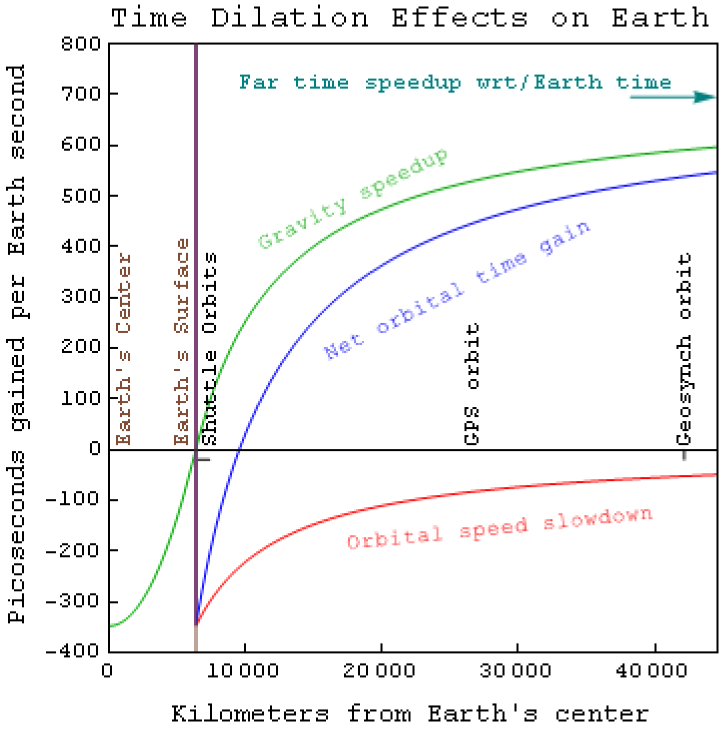
One interesting side effect of the Selective Availability hardware is the capability to add corrections to the outgoing signal of the GPS cesium and rubidium atomic clocks to an accuracy of approximately 2×10^{-13} This represented a significant improvement over the raw accuracy of the clocks.

On 19 September 2007, the United States Department of Defense announced that future GPS III satellites will not be capable of implementing SA, eventually making the policy permanent.

Antispoofing

Another restriction on GPS, antispoofing, remains on. This encrypts the *P-code* so that it cannot be mimicked by a transmitter sending false information. Few civilian receivers have ever used the P-code, and the accuracy attainable with the public C/A code is so much better than originally expected (especially with DGPS) that the antispoof policy has relatively little effect on most civilian users. Turning off antispoof would primarily benefit surveyors and some scientists who need extremely precise positions for experiments such as tracking tectonic plate motion.

Relativity



Satellite clocks are slowed by their orbital speed but sped up by their distance out of the Earth's gravitational well.

A number of sources of error exist due to relativistic effects that would render the system useless if uncorrected. Three relativistic effects are the time dilation, gravitational frequency shift, and eccentricity effects. For example, the relativistic time *slowing* due to the speed of the satellite of about 1 part in 10^{10}, the gravitational time dilation that makes a satellite run about 5 parts in 10^{10} *faster* than an Earth based clock, and the Sagnac effect due to rotation relative to receivers on Earth. These topics are examined below, one at a time.

Special and general relativity

According to the theory of relativity, due to their constant movement and height relative to the Earth-centered, non-rotating approximately inertial reference frame, the clocks on the satellites are affected by their speed. Special relativity predicts that the frequency of the atomic clocks moving at GPS orbital speeds will tick more slowly than stationary ground clocks by a factor of $\frac{v^2}{2c^2} \approx 10^{-10}$, or result in a delay of about 7 μs/day, where the orbital velocity is v = 4 km/s, and c = the speed of light. The time dilation effect has been measured and verified using the GPS.

The effect of gravitational frequency shift on the GPS due to general relativity is that a clock closer to a massive object will be slower than a clock farther away. Applied to the GPS, the receivers are much closer to Earth than the satellites, causing the GPS clocks to be faster by a factor of 5×10^(-10), or about 45.9 μs/day. This gravitational frequency shift is noticeable.

When combining the time dilation and gravitational frequency shift, the discrepancy is about 38 microseconds per day, a difference of 4.465 parts in 10^{10}. Without correction, errors in the initial pseudorange of roughly 10 km/day would accumulate. This initial pseudorange error is corrected in the process of solving the navigation equations. In addition the elliptical, rather than perfectly circular, satellite orbits cause the time dilation and gravitational frequency shift effects to vary with time. This eccentricity effect causes the clock rate difference between a GPS satellite and a receiver to increase or decrease depending on the altitude of the satellite.

To compensate for the discrepancy, the frequency standard on board each satellite is given a rate offset prior to launch, making it run slightly slower than the desired frequency on Earth; specifically, at 10.22999999543 MHz instead of 10.23 MHz. Since the atomic clocks on board the GPS satellites are precisely tuned, it makes the system a practical engineering application of the scientific theory of relativity in a real-world environment. Placing atomic clocks on artificial satellites to test Einstein's general theory was proposed by Friedwardt Winterberg in 1955.

Sagnac distortion

GPS observation processing must also compensate for the Sagnac effect. The GPS time scale is defined in an inertial system but observations are processed in an Earth-centered, Earth-fixed (co-rotating) system, a system in which simultaneity is not uniquely defined. A Lorentz transformation is thus applied to convert from the inertial system to the ECEF system. The resulting signal run time correction has opposite algebraic signs for satellites in the Eastern and Western celestial hemispheres. Ignoring this effect will produce an east-west error on the order of hundreds of nanoseconds, or tens of meters in position.

Natural sources of interference

Since GPS signals at terrestrial receivers tend to be relatively weak, natural radio signals or scattering of the GPS signals can desensitize the receiver, making acquiring and tracking the satellite signals difficult or impossible.

Space weather degrades GPS operation in two ways, direct interference by solar radio burst noise in the same frequency band or by scattering of the GPS radio signal in ionospheric irregularities referred to as scintillation. Both forms of degradation follow the 11 year solar cycle and are a maximum at sunspot maximum although they can occur at anytime. Solar radio bursts are associated with solar flares and Coronal Mass Ejections (CMEs) and their impact can affect reception over the half of the Earth facing the sun. Scintillation occurs most frequently at tropical latitudes where it is a night time phenomenon. It occurs less frequently at high latitudes or mid-latitudes where magnetic storms can lead to scintillation. In addition to producing scintillation, magnetic storms can produce strong ionospheric gradients that degrade the accuracy of SBAS systems.

Artificial sources of interference

In automotive GPS receivers, metallic features in windshields, such as defrosters, or car window tinting films can act as a Faraday cage, degrading reception just inside the car.

Man-made EMI (electromagnetic interference) can also disrupt or jam GPS signals. In one well-documented case it was impossible to receive GPS signals in the entire harbor of Moss Landing, California due to unintentional jamming caused by malfunctioning TV antenna preamplifiers. Intentional jamming is also possible. Generally, stronger signals can interfere with GPS receivers when they are within radio range or line of sight. In 2002 a detailed description of how to build a short-range GPS L1 C/A jammer was published in the online magazine Phrack.

The U.S. government believes that such jammers were used occasionally during the 2001 war in Afghanistan, and the U.S. military claims to have destroyed six GPS jammers during the Iraq War, including one that was destroyed with a GPS-guided bomb. A GPS jammer is relatively easy to detect and locate, making it an attractive target for anti-radiation missiles. The UK Ministry of Defence tested a jamming system in the UK's West Country on 7 and 8 June 2007.

Some countries allow the use of GPS repeaters to allow the reception of GPS signals indoors and in obscured locations; however, under EU and UK laws, the use of these is prohibited as the signals can cause interference to other GPS receivers that receive data from both GPS satellites and the repeater.

Due to the potential for both natural and man-made noise, numerous techniques continue to be developed to deal with the interference. The first is to not rely on GPS as a sole source. According to John Ruley, "IFR pilots should have a fallback plan in case of a GPS malfunction". Receiver Autonomous Integrity Monitoring (RAIM) is a feature included in some receivers, designed to provide a warning to the user if jamming or another problem is detected. The U.S. military has also deployed since 2004 their Selective Availability / Anti-Spoofing Module (SAASM) in the Defense Advanced GPS Receiver (DAGR). In demonstration videos the DAGR was shown to detect jamming and maintain its lock on the encrypted GPS signals during interference which caused civilian receivers to lose lock.

Chapter 5

Garmin G1000

T182T cockpit with Garmin G1000

The **Garmin G1000** is an integrated flight instrument system manufactured by Garmin typically composed of two display units, one serving as a primary flight display, and one as a multi-function display. It serves as a replacement for most conventional flight instruments and avionics.

Components

An aircraft with a basic Garmin G1000 installation contains two LCD displays (one acting as the primary flight display and the other as the multi-function display) as well as an integrated communications panel that fits between the two.

Beyond that, additional features are found on newer and larger G1000 installations, such as in business jets. This includes:

- A third display unit, to act as a co-pilot PFD
- An alphanumeric keyboard
- An integrated flight director/autopilot (without it, the G1000 interfaces with an external autopilot)

Depending on the airplane manufacturer and whether or not a GFC 700 autopilot is installed, the G1000 system will consist of either two GDU 1040 displays (no autopilot), a GDU 1040 PFD/GDU 1043 MFD (GFC 700 autopilot installed), or a GDU 1045 PFD/GDU 1045 MFD (GFC 700 autopilot installed with VNAV).

The GDU 1040 is the standard base bezel with no autopilot/flight director mode selection keys below the heading bug. The GDU 1043 has autopilot/flight director keys for all GFC 700 modes except VNAV. The GDU 1045 is essentially identical to the GDU 1043 except for the addition of an autopilot/flight director mode for VNAV. Depending on how the units are installed, an MFD failure may, or may not, have an impact on autopilot or flight director use. If a GDU 1040 is used as a PFD in an airplane equipped with a GFC 700 autopilot, a failure of the MFD (which houses the autopilot mode selection keys) will leave the autopilot engaged, but the modes cannot be changed because no autopilot keys are present on the PFD. But, if an MFD failure occurs in an airplane with the GFC 700 autopilot and either a GDU 1043 or a GDU 1045 bezel installed as a PFD, the pilot will have full use of the autopilot through the keys on the PFD.

Both the PFD and MFD each have two slots for SD memory cards. The top slot is used to update the Jeppesen aviation database (also known as NavData) every 28 days, and to load software and configuration to the system. The aviation database must be current to use GPS for navigation during IFR instrument approaches. The bottom slot houses the World terrain and Jeppesen obstacle databases. While terrain information rarely changes or needs to be updated, obstacle databases can be updated every 56 days through a subscription service. The top card can be removed from the G1000 system following an update, but the bottom card must stay in both the PFD and MFD to ensure accurate terrain awareness and TAWS-B information.

The following summarizes the important process of updating the memory cards: 1. Go to http://fly.garmin.com and log in. 2. Register or click on the airplane for which you desire to update the cards. 3. Click in your device, such as G1000. 4. Download the desired services onto each of the lower cards, one at a time. You will need to download and install each item twice, one time for each card, because the software on the cards must be identical. Examples of items downloaded are Obstacle, Taxi and Terrain. 5. Search for the Jeppesen Services Update Manager and, if not already installed, install it on your computer. Open the Jeppesen Services Update Manager program on your computer and log into the software. 6. Insert the upper card into the download device that came with your G1000. 7. Download NavData - Jeppesen Aviation Database onto your upper card. 8. Put the two lower cards into the two lower slots of your G1000 device. 9. Put the upper card in the upper left slot of your G1000 device. 10. Turn on the master power switch of your airplane and allow the G1000 device to update. This will update NavData on your left screen. Then turn off the master power switch. 11. Move the upper card to the upper right slot and turn on your master power switch again. This will update NavData on your right screen.

Primary flight display (PFD)

Screenshot of the PFD on the G1000

The primary flight display shows the basic flight instruments, such as the airspeed indicator, the altimeter, the heading indicator, and course deviation indicator. A small map called the "inset map" can be enabled in the corner. The buttons on the PFD are used to set the squawk code on the transponder. The PFD can also be used for entering and activating flight plans. The PFD also has a "reversionary mode" which is capable of displaying all information shown on the MFD (for example, engine gauges and navigational information). This capability is provided in case of an MFD failure.

Multi-function display (MFD)

The MFD usually shows engine instrumentation and a moving map.

The multi-function display typically shows a moving map on the right side, and engine instrumentation on the left. Most of the other screens in the G1000 system are accessed by turning the knob on the lower right corner of the unit. Screens available from the MFD other than the map include the setup menus, information about nearest airports and NAVAIDs, Mode S traffic reports, terrain awareness, XM radio, flight plan programming, and GPS RAIM prediction.

Implementation

The G1000 system consists of several integrated components which sample and exchange data or display information to the pilot.

GDU Display

The GDU display unit acts as the primary source of flight information for the pilot. Each display can interchangeably serve as a primary flight display (PFD) or multi-function display (MFD). The wiring harness within the aircraft specifies which role each display is in by default. All of the displays within an aircraft are interconnected using a high-speed

Ethernet data bus. A G1000 installation may have two GDUs (one PFD and one MFD) or three (one PFD for each pilot and an MFD). There are several different GDU models in service, which have different screen sizes (from 10 inches to 15 inches) and different bezel controls.

In normal operation, the display in front of the pilot is the PFD and will provide aircraft attitude, airspeed, altitude, vertical speed, heading, rate-of-turn, slip-and-skid, navigation, transponder, inset map view (containing map, traffic, and terrain information), and systems annunciation data. The second display, typically positioned to the right of the PFD, operates in MFD mode and provides engine instrumentation and a moving map display. The moving map can be replaced or overlaid by various other types of data, such as satellite weather, checklists, system information, waypoint information, weather sensor data, and traffic awareness information.

Both displays provide redundant information regarding communications and navigation radio frequency settings even though each display is usually only paired with one GIA Integrated Avionics Unit. In the event of a single display failure, the remaining display will adopt a combined "reversionary mode" and automatically become a PFD combined with engine instrumentation data and other functions of the MFD. A red button labeled "reversionary mode" or "display backup," located on the GMA audio panel, is also available to the pilot to select this mode manually if desired.

GMA Audio Panel

The GMA panel provides buttons for selecting what audio sources are heard by each member of the cockpit. It also includes a button for forcing the integrated cockpit into its fail-safe reversionary mode.

GMC/GCU Remote Controllers

The GMC and GCU controllers are panel-mounted modules which provide a more intuitive interface for the pilot than that provided by the GDU. The GMC controls the G1000's autopilot, while the GCU is used to enter navigational data and control the GDU's functions.

GIA Integrated Avionics Unit

The GIA unit is a combined communications and navigation radio, and also serves as the primary data aggregator for the G1000 system. It provides a two-way VHF communications transceiver, a VHF navigation receiver with glideslope, a GPS receiver, and a variety of supporting processors. Each unit is paired with a GDU display, which acts as a controlling unit. The GIA 63W, found on many newer G1000 installations, is an updated version of the older GIA 63 which includes Wide Area Augmentation System support.

GDC Air Data Computer

The GDC computer replaces the internal components of the pitot-static system in traditional aircraft instrumentation. It measures airspeed, altitude, vertical speed, and outside air temperature. This data is then provided to all the displays and integrated avionics units.

GRS Attitude and Heading Reference System (AHRS)

The GRS system uses solid-state sensors to measure aircraft attitude, rate of turn, and slip and skid. This data is then provided to all the integrated avionics units and GDU display units. Unlike many competing systems, the AHRS can be rebooted and recalibrated in flight during turns of up to 20 degrees.

GMU Magnetometer

The GMU magnetometer measures aircraft heading and is a digital version of a traditional compass. It does so through aligning itself with the magnetic flux lines of the earth.

GTX Transponder

Either the GTX 32 or GTX 33 transponder can be used in the G1000 system, although the GTX 33 is far more common. The GTX 32 provides standard mode-C replies to ATC interrogations while the GTX 33 provides mode-S bidirectional communications with ATC and therefore can indicate traffic in the area as well as announce itself spontaneously via "squitting" without prior interrogation.

GEA Engine/Airframe Unit

The GEA unit measures a large variety of engine and airframe parameters, including engine RPM, manifold pressure, oil temperature, cylinder head temperature, exhaust gas temperature, and fuel level in each tank. This data is then provided to the integrated avionics units.

FAA APPROVED

GSD Data Aggregator

The GSD is a data aggregator system included on complex G1000 systems, such as that found on the Embraer Phenom 100. It serves as a point of connection which allows external systems to communicate with the G1000.

Backup systems

As a condition of certification, all aircraft utilizing the G1000 integrated cockpit must have a redundant airspeed indicator, altimeter, attitude indicator, and magnetic compass. In the event of a failure of the G1000 instrumentation, these backup instruments become primary.

In addition, a secondary power source is required to power the G1000 instrumentation for a limited time in the event of a failure of the aircraft's alternator and primary battery.

Certification

The Garmin G1000 is generally certified on new general aviation aircraft, including Beechcraft, Cessna, Diamond, Cirrus Design, Mooney, Piper, Quest Aircraft (the Kodiak 100) , and Tiger. Garmin recently announced its first G1000 retrofit program for the Beechcraft C90 King Air in 2007 followed by the Beechcraft B200 King Air in 2009. The Garmin G1000 became a jet platform in 2007, as the avionics system for the Cessna Citation Mustang Very Light Jet. Versions of the G1000 are also used in the Embraer Phenom 100 and Embraer Phenom 300, HondaJet, and PiperJet.

Competition

The G1000 competes with the Avidyne Entegra and Chelton FlightLogic EFIS glass cockpits. However, there are significant differences with regard to the features, degree of integration, intuitive aspects of the design, and overall product utility. Note that the Chelton system is not typically found in airplanes that include the less expensive G1000 or Avidyne systems.

Advantages and drawbacks

As the system has the GPS, communication, and radio navigation components built directly into the system, it both consolidates components into a centralized location and, for the same reason, becomes potentially more costly to repair or replace. The system has the potential to reduce downtime as key components, such as the AHRS, ADC and PFD, are modular and easily replaced. The system's design also prevents the failure of a single component from "cascading" through other components.

The garmin is compatible with the latest EVS technology. Enhanced vision system use thermal and infrared camera's to see real-time images and help turn obscurants such as bad weather, night time, fog, dust and brownouts into better images that can see 8-10 times farther than the naked eye. This helps people see better and fly safer.

There are some safety concerns with all glass cockpits, such as the failure of the primary flight displays (PFD). These concerns, however, are no more (and perhaps are less) significant than similar considerations with aircraft equipped with traditional instrumentation. The Garmin G1000 system offers a reversionary mode that will present all of the primary flight instrumentation on the remaining display. In addition, there are multiple GPS units, and electronic redundancy incorporated extensively throughout the design of the system.

Another important risk factor is the potential to spend too much time looking inside the cockpit managing the instrumentation, possibly increasing the chance of a collision with other aircraft, obstacles, or terrain.

Training and Training Resources

Flying any glass cockpit aircraft requires transition training to familiarize the pilot with the aircraft's systems. Transition training is most effective when a pilot prepares ahead of time. Most general aviation manufacturers using the G1000 system have FAA Industry Training Standards (FITS) training programs for pilots transitioning into their airplanes. FAA FITS compliant training is recommended for any pilot transitioning to the G1000 or any other glass cockpit prior to the aircraft in instrument meteorological conditions (IMC) or if operating a glass cockpit aircraft for the first time. Glass cockpit aircraft may not be suitable for primary training.

One of the most effective resources for preparing for G1000 transition training include the Garmin simulator software. In addition, some flight schools now have G1000 flight training devices (FTDs) that provide realistic simulation.

All of the most current Garmin G1000 Pilot's Guides are available in PDF format for free downloading from Garmin.

Chapter 6

Helmet Mounted Display

Helmet mounted display

The Integrated Helmet and Display Sight System (IHADSS)

A **helmet mounted display** (HMD) is a device used in some modern aircraft, especially combat aircraft. HMDs project information similar to that of head-up displays (HUD) on an aircrew's visor or reticle, thereby allowing him to obtain situational awareness and/or cue weapons systems to the direction his head is pointing. Some applications refer to these devices as Helmet Mounted Sight and Display (HMSD) or Helmet Mounted Sights (HMS).

Requirement

Aviation HMD designs serve two basic purposes:

- using the pilot's eye (actually head angle) as a pointing device to direct air-to-air and air-to-ground weapons seekers or other sensors (e.g., radar, FLIR) to a target merely by pointing his head at the target and actuating a switch via HOTAS controls. In close combat prior to HMDs, the pilot had to align the aircraft to shoot at a target. HMDs allow the pilot to simply point his head at a target, designate it to weapon and shoot.

- displaying targeting and aircraft performance information (such as airspeed, altitude, target range, weapon seeker status, "G", etc.) to the pilot while "heads-up", eliminating the need to look inside the cockpit.

- displaying sensor video for the purpose of: verification that the chosen sensor has been cued to the right target or location without requiring the pilot to look inside the cockpit, and viewing outside terrain using sensor video in degraded visual conditions.

HMD systems, combined with High Off-Boresight (HOBS) weapons, results in the ability for aircrew to attack and destroy nearly any target seen by the pilot. These systems allow targets to be designated with minimal aircraft maneuvering, minimizing the time spent in the threat environment, and allowing greater lethality, survivability, and pilot situational awareness.

History

The first aircraft with simple HMD devices appeared for experimental purpose in the mid seventies to aid in targeting heat seeking missiles. These rudimentary devices were better described as Helmet Mounted Sights. Mirage F1AZ of the SAAF [South African Airforce] used a locally developed helmet mounted sight. This enables the pilot to make bore attacks, without having to manoeuvre until the optimum firing position. South Africa subsequently emerged as one of the pioneers and leaders in helmet mounted sight technology. The SAAF was also the first Air Force to fly the helmet sight operationally. The US Navy's Visual Target Acquisition System (VTAS), made by Honeywell Corporation was a simple mechanical "ring and bead" style sight fitted to the front of the pilot's helmet that was flown in the 1974-78 ACEVAL/AIMVAL on U.S. F-14 and F-15 fighters

VTAS received praise for its effectiveness in targeting off-boresight missiles, but the U.S. did not pursue fielding it except for integration into late model Navy F-4 Phantoms equipped with the AIM-9 Sidewinder. HMDs were also introduced in helicopters during this time.

The first operational jet fighters with HMD (Mirage F1AZ) were fielded by the South African Air Force. After the South African system had been proven in combat, playing a role in downing Soviet aircraft over Angola, the Soviets embarked on a crash program to counter the technology. As a result, the MiG-29 was fielded in 1985 with an HMD and a high off-boresight weapon (AA-11 Archer/R-73), giving them an advantage in close in maneuvering engagements.

Several nations responded with programs to counter the MiG-29/HMD/AA-11 (and later Su-27) combination once its effectiveness was known, principally through access to former East German MiG-29s that were operated by the unified German Air Force.

The first successful non-Soviet HMD was the Israeli Air Force Elbit DASH series, fielded in conjunction with the Python 4, in the early 1990s. American and European fighter HMDs lagged behind, not becoming widely used until the late 1990s and early 2000s. The US-UK-Germany responded initially with a combined ASRAAM effort. Technical difficulties led to the US abandoning ASRAAM, instead funding development of the AIM-9X and the Joint Helmet Mounted Cueing System in 1990.

Technology

While conceptually simple, implementation of aircraft HMDs is quite complex. There are many variables:

- precision - the angular error between the line-of-sight and the derived cue. The position of the *helmet* is what is used to point the missile, it thus must be calibrated and fit securely on the pilot's head. The line between the pilot's eye and the reticle on the visor is known as the line of sight (LOS) between the aircraft and the intended target. The user's eye must stay aligned with the sight – in other words, current HMDs cannot sense where the eye is looking, but can place a "pipper" between the eye and the target.
- latency or slew rate - how much lag there is between the helmet and the cue.
- field of regard - the angular range over which the sight can still produce a suitably accurate measurement.
- weight and balance - total helmet weight and its center of gravity, which are particularly important under high "*g*" maneuvers. Weight is the largest problem faced by fighter aircraft HMD designers. This is much less a concern for helicopter applications, making elaborate helicopter HMDs common.
- safety and cockpit compatibility, including ejection seat compatibility.
- optical characteristics – calibration, sharpness, distant focus (or 'Collimation', a technique used to present the images at a distant focus, which improves the readability of images), monocular vs. binocular imagery, eye dominance, and binocular rivalry.
- durability and ability to handle day to day wear and tear.
- cost, including integration and training.
- fit and interfacing the aviator's head to the aircraft – head anthropometry and facial anatomy make helmet fitting a crucial factor in the aviator's ability to interface with the aircraft systems. Misalignment or helmet shift can cause an inaccurate picture.

Head Position Sensing

HMD designs must sense the elevation, azimuth and tilt of the pilot's head relative to the airframe with sufficient precision even under high "*g*" and during rapid head movement. Two basic methods are used in current HMD technology - optical and electromagnetic.

Optical tracking

Optical systems employ infrared emitters on the helmet (or cockpit) infrared detectors in the cockpit (or helmet), to measure the pilot's head position. The main limitations are restricted fields of regard and sensitivity to sunlight or other heat sources. The MiG-29/AA-11 Archer system uses this technology.

Electromagnetic tracking

Electromagnetic sensing designs use coils (in the helmet) placed in an alternating field (generated in the cockpit) to produce alternating electrical voltages based on the movement of the helmet in multiple axes. This technique requires precise magnetic mapping of the cockpit to account for ferrous and conductive materials in the seat, cockpit sills and canopy to reduce angular errors in the measurement.

Optics

Modern HMDs typically employ a compact CRT embedded in the helmet, and suitable optics to display symbology on to the pilot's visor or reticle, focused at infinity. Advanced HMDs can also project FLIR or NVG imagery.

Major Systems

Systems are presented in rough chronological order of initial operating capability.

Integrated Helmet And Display Sight System (IHADSS)

IHADSS

In 1984, the U.S. Army fielded the AH-64 Apache and with it the Integrated Helmet and Display Sighting System (IHADSS), a new helmet concept in which the role of the helmet was expanded to provided a visually coupled interface between the aviator and the aircraft. The Honeywell M142 IHADSS is fitted with a 40° by 30° field of view, video-with-symbology monocular display. IR emitters allow a slewable IR imaging sensor, mounted on the nose of the aircraft, to be slaved to the aviator's head movements. The display also enables Nap-of-the-earth night navigation. IHADSS is also used on the Italian Agusta A129 Mangusta.

ZSh-5 / Shchel-3UM

The Russian designed Shchel-3UM HMD design is fit to the ZSh-5 series helmet, and is used on the MiG-29 and Su-27 in conjunction with the AA-11 Archer. The HMD/Archer combination gave the MiG-29 and Su-27 a significantly improved close combat capability and quickly became the most widely deployed HMD in the world.

Display And Sight Helmet (DASH)

The Elbit Systems DASH III was the first modern Western HMD to achieve operational service. Evolution of the DASH began during the mid 1980s, when the IAF issued a requirement for F-15 and F-16 aircraft. The first design entered production around 1986, and the current GEN III helmet entered production during the early to mid 1990s. The current production variant is deployed on IDF F-15, and F-16 aircraft. Additionally, it has been certified on the F/A-18 and F-5. The DASH III has been exported and integrated into various legacy aircraft, including the MiG-21. It also forms the baseline technology for the US JHMCS.

The DASH GEN III is a wholly embedded design, where the complete optical and position sensing coil package is built within the helmet (either USAF standard HGU-55/P or the Israeli standard HGU-22/P) using a spherical visor to provide a collimated image to the pilot. A quick-disconnect wire powers the display and carries video drive signals to the helmet's Cathode Ray Tube (CRT). DASH is closely integrated with the aircraft's weapon system, via a MIL-STD-1553B bus.

Joint Helmet Mounted Cueing System (JHMCS)

JHMCS

After the U.S. withdrawal from ASRAAM, the U.S. pursued and fielded JHMCS in conjunction with the Raytheon AIM-9X, in November 2003 with the 12th and 19th Fighter Squadrons at Elmendorf AFB, Alaska. The Navy conducted RDT&E on the F/A-18C as lead platform for JHMCS, but fielded it first on the F/A-18 Super Hornet E and F aircraft in 2003. The USAF is also integrating JHMCS into its F-15E and F-16 aircraft.

JHMCS is a derivative of the DASH III and the Kaiser Agile Eye HMDs, and was developed by Vision Systems International (VSI), a joint venture company formed by

Rockwell Collins, Elbit and Kaiser Electronics (Kaiser is no longer affiliated with VSI; it is now equally owned by Rockwell Collins and Elbit's US subsidiary Elbit Systems of America). Boeing integrated the system into the F/A-18 and began low-rate initial production delivery in fiscal year 2002. JHMCS is employed in the F/A-18C/D/E/F, F-15C/D/E, and F-16 Block 40/50 with a design that is 95% common to all platforms. This may also be integrated into the system of the F-22.

Unlike the DASH, which is integrated into the helmet itself, JHMCS assemblies attach to modified HGU-55/P, HGU-56/P or HGU-68/P helmets. JHMCS employs a newer, faster digital processing package, but retains the same type of electromagnetic position sensing as the DASH. The CRT package is more capable, but remains limited to monochrome presentation of calligraphic symbology. JHMCS provides support for raster scanned imagery to display FLIR/IRST pictures for night operations and provides collimated symbology and imagery to the pilot. The integration of the night-vision goggles with the JHMCS was a key requirement of the program.

When combined with the AIM-9X, an advanced short-range dogfight weapon that employs a Focal Plane Array seeker and a thrust vectoring tail control package, JHMCS allows effective target designation up to 80 degrees either side of the aircraft's nose.

Aselsan AVCI

Aselsan of Turkey has developed a similar system to the US made JHMCS called the AVCI Helmet Integrated Cueing System. It is envisaged that the Aselsan JHMCS will be incorporated with Turkey's F-16 CCIP modernization program. The system will also be utilized into the T-129 Turkish Attack Helicopter.

Topsight/TopNight

The French thrust vectoring Matra MBDA MICA missile for its Rafale and late model Mirage 2000 fighters was accompanied by the Topsight HMD by Sextant Avionique. TopSight provides a 20 degree FoV for the pilot's right eye, and calligraphic symbology generated from target and aircraft parameters. Electromagnetic position sensing is employed. The Topsight helmet uses an integral embedded design, and its contoured shape is designed to provide the pilot with a wholly unobstructed field of view.

TopNight, a Topsight derivative, is designed specifically for adverse weather and night air to ground operations, employing more complex optics to project infrared imagery overlaid with symbology.

Eurofighter Helmet Mounted Symbology System

The Eurofighter Typhoon utilizes the Helmet Mounted Symbology System (HMSS) developed by BAE Systems and Pilkington Optronics. It is capable of displaying both raster imagery and calligraphic symbology, with provisions for embedded NVGs. As with the DASH helmet, the system employs integrated position sensing to ensure that symbols representing outside-world entities move in line with the pilot's head movements.

Helmet Mounted Display System

Helmet-Mounted Display System for the F-35 Joint Strike Fighter

Vision Systems International (VSI; the Elbit Systems/Rockwell Collins joint venture) along with Helmet Integrated Systems, Ltd. developed the Helmet Mounted Display System (HMDS) for the F-35 Joint Strike Fighter aircraft. In addition to standard HMD capabilities offered by other systems, HMDS fully utilizes the advanced avionics architecture of the F-35 and provides the pilot video with imagery in day or night conditions. As a result, the F-35 is the first tactical fighter jet in 50 years to fly without a HUD.

JedEyes TM

JedEyes TM is a new system recently introduced by Elbit Systems especially to meet Apache and other rotary wing platform requirements. The system is designed for day, night and brownout flight environments. JedEyes TM has a 70 x 40 degree FOV and 2250x1200 pixels resolution.

Cobra

Sweden’s JAS 39 Gripen fighter utilizes the Cobra HMD, developed by BAE Systems, Denel Optronics of South Africa, and Saab. It has been exported to the South African Air Force.

Future technology

- VSI is developing the QuadEyeTM Night Vision Cueing & Display (NVCD) for the US Navy and US Air Force, and is also producing the DASH Generation IV HMD.

- Eye tracking – Eye trackers measure the point of gaze relative to the direction of the head, allowing a computer to sense where the user is looking. These systems are not currently used in aircraft.

- Direct retinal projection – Systems that project information directly onto the wearer’s retina with a low-powered laser (virtual retinal display) are also in experimentation.

Chapter 7

Weather Radar

Weather radar in Norman, Oklahoma with rainshaft

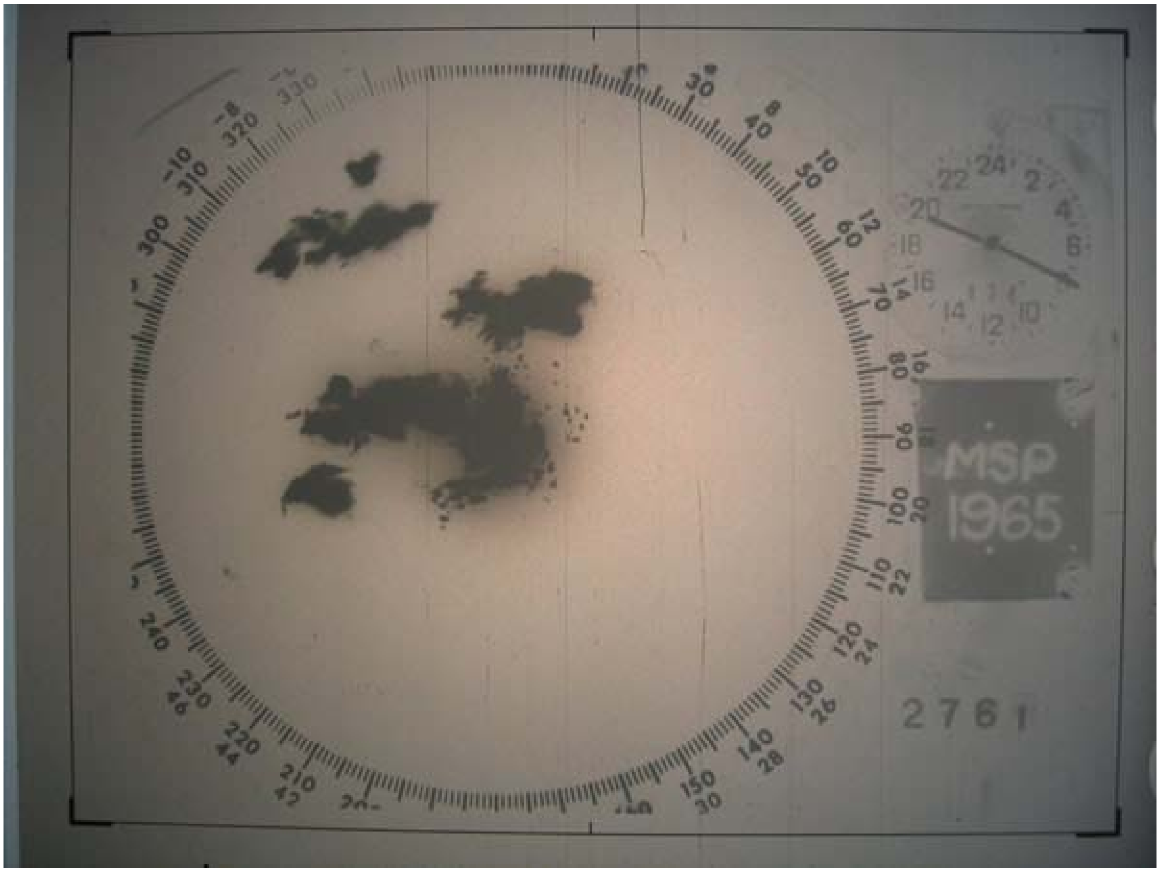
Weather (WF44) radar dish

University of Oklahoma OU-PRIME C-band, polarimetric, weather radar during construction

A **weather radar**, or **weather surveillance radar** (**WSR**), is a type of radar used to locate precipitation, calculate its motion, estimate its type (rain, snow, hail, etc.), and forecast its future position and intensity.

Modern weather radars are mostly pulse-Doppler radars, capable of detecting the motion of rain droplets in addition to intensity of the precipitation. Both types of data can be analyzed to determine the structure of storms and their potential to cause severe weather.

History



1960s radar technology detected tornado producing supercells over the Twin Cities metropolitan area.

During World War II, military radar operators noticed noise in returned echoes due to weather elements like rain, snow, and sleet. Just after the war, military scientists returned to civilian life or continued in the Armed Forces and pursued their work in developing a use for those echoes. In the United States, David Atlas, for the Air Force group at first, and later for MIT, developed the first operational weather radars. In Canada, J.S. Marshall and R.H. Douglas formed the "Stormy Weather Group" in Montreal. Marshall and his doctoral student Walter Palmer are well known for their work on the drop size distribution in mid-latitude rain that led to understanding of the Z-R relation, which correlates a given radar reflectivity with the rate at which water is falling on the ground. In the United Kingdom, research continued to study the radar echo patterns and weather elements such as stratiform rain and convective clouds, and experiments were done to evaluate the potential of different wavelengths from 1 to 10 centimetres.

In 1953, Donald Staggs, an electrical engineer working for the Illinois State Water Survey, made the first recorded radar observation of a "hook echo" associated with a tornadic thunderstorm.

Between 1950 and 1980, reflectivity radars, which measure position and intensity of precipitation, were built by weather services around the world. The early meteorologists had to watch a cathode ray tube. During the 1970s, radars began to be standardized and organized into networks. The first devices to capture radar images were developed. The number of scanned angles was increased to get a three-dimensional view of the precipitation, so that horizontal cross-sections (CAPPI) and vertical ones could be performed. Studies of the organization of thunderstorms were then possible for the Alberta Hail Project in Canada and National Severe Storms Laboratory (NSSL) in the US in particular.

The NSSL, created in 1964, began experimentation on dual polarization signals and on Doppler effect uses. In May 1973, a tornado devastated Union City, Oklahoma, just west of Oklahoma City. For the first time, a Dopplerized 10-cm wavelength radar from NSSL documented the entire life cycle of the tornado. The researchers discovered a mesoscale rotation in the cloud aloft before the tornado touched the ground : the tornadic vortex signature. NSSL's research helped convince the National Weather Service that Doppler radar was a crucial forecasting tool. The Super Outbreak of tornadoes on April 3–4, 1974 and their devastating destruction might have helped to get funding for further developments.

Between 1980 and 2000, weather radar networks became the norm in North America, Europe, Japan and other developed countries. Conventional radars were replaced by Doppler radars, which in addition to position and intensity of could track the relative velocity of the particles in the air. In the United States, the construction of a network consisting of 10 cm (4 in) wavelength radars, called NEXRAD or WSR-88D (Weather Service Radar 1988 Doppler), was started in 1988 following NSSL's research. In Canada, Environment Canada constructed the King City station, with a five centimeter research Doppler radar, by 1985;McGill University dopplerized its radar (J. S. Marshall Radar Observatory) in 1993. This led to a complete Canadian Doppler network between 1998 and 2004. France and other European countries switched to Doppler network by the end of the 1990s to early 2000s. Meanwhile, rapid advances in computer technology led to algorithms to detect signs of severe weather and a plethora of "products" for media outlets and researchers.

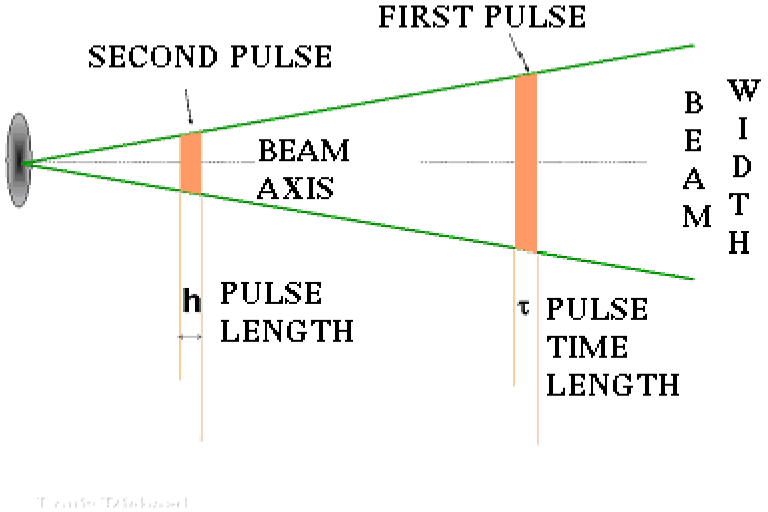
After 2000, research on dual polarization technology has moved into operational use, increasing the amount of information available on precipitation type (e.g. rain vs. snow). "Dual polarization" means that microwave radiation which is polarized both horizontally and vertically (with respect to the ground) is emitted. Wide-scale deployment is expected by the end of the decade in some countries such as the United States, France, and Canada.

Since 2003, the U.S. National Oceanic and Atmospheric Administration has been experimenting with phased-array radar as a replacement for conventional parabolic antenna to provide more time resolution in atmospheric sounding. This would be very important in severe thunderstorms as their evolution can be better evaluated with more timely data.

Also in 2003, the National Science Foundation established the Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere, "CASA", a multidisciplinary, multi-university collaboration of engineers, computer scientists, meteorologists, and sociologists to conduct fundamental research, develop enabling technology, and deploy prototype engineering systems designed to augment existing radar systems by sampling the generally undersampled lower troposphere with inexpensive, fast scanning, dual polarization, mechanically scanned and phased array radars.

How a weather radar works

Sending radar pulses



A radar beam spreads out as it moves away from the radar station, covering an increasingly large volume.

Weather radars send directional pulses of microwave radiation, on the order of a microsecond long, using a cavity magnetron or klystron tube connected by a waveguide to a parabolic antenna. The wavelengths of 1 to 10 cm (4 in) are approximately ten times the diameter of the droplets or ice particles of interest, because Rayleigh scattering occurs at these frequencies. This means that part of the energy of each pulse will bounce off these small particles, back in the direction of the radar station.

Shorter wavelengths are useful for smaller particles, but the signal is more quickly attenuated. Thus 10 cm (4 in) (S-band) radar is preferred but is more expensive than a 5 cm (2 in) C-band system. 3 cm (1 in) X-band radar is used only for very short distance purposes, and 1 cm (0 in) Ka-band weather radar is used only for research on small-particle phenomena such as drizzle and fog.

Radar pulses spread out as they move away from the radar station. This means that the region of air any given pulse is moving through is larger for areas farther away from the station, and smaller for nearby areas, decreasing resolution at far distances. At the end of a 150–200 km sounding range, the volume of air scanned by a single pulse might be on the order of a cubic kilometer. This is called the *pulse volume*

The volume of air that a given pulse takes up at any point in time may be approximately calculated by the formula $v = hr^2\theta^2$, where v is the volume enclosed by the pulse, h is pulse width (in e.g. meters, calculated from the duration in seconds of the pulse times the speed of light), r is the distance from the radar that the pulse has already traveled (in e.g. meters), and $\theta$is the beam width (in radians). This formula assumes the beam is symmetrically circular, "r" is much greater than "h" so "r" taken at the beginning or at the end of the pulse is almost the same, and the shape of the volume is a cone frustum of depth "h".

Listening for return signals

Between each pulse, the radar station serves as a receiver and listens for return signals from particles in the air. The duration of the "listen" cycle is on the order of a millisecond, which is a thousand times longer than the pulse duration. The length of this phase is determined by the need for the microwave radiation (which travels at the speed of light) to propagate from the detector, to the weather target, and back again, a distance which could be several hundred kilometers. The horizontal distance from station to target is calculated simply from the amount of time that lapses from the initiation of the pulse to the detection of the return signal. (The time is converted into distance by multiplying by the speed of light). If pulses are emitted too frequently, the returns from one pulse will be confused with the returns from previous pulses, resulting in incorrect distance calculations.

Determining height

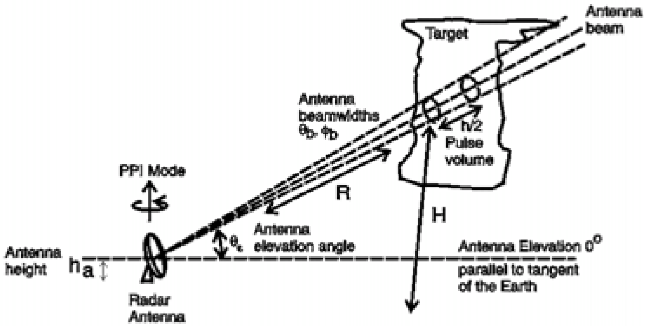
BEAM HEIGHT WITH DISTANCE (AGL)

$$H=\left(\sqrt{r^2+(k_e a_e)^2+2r\,k_e a_e \sin(\theta_e)}\right)-k_e a_e+h_a$$

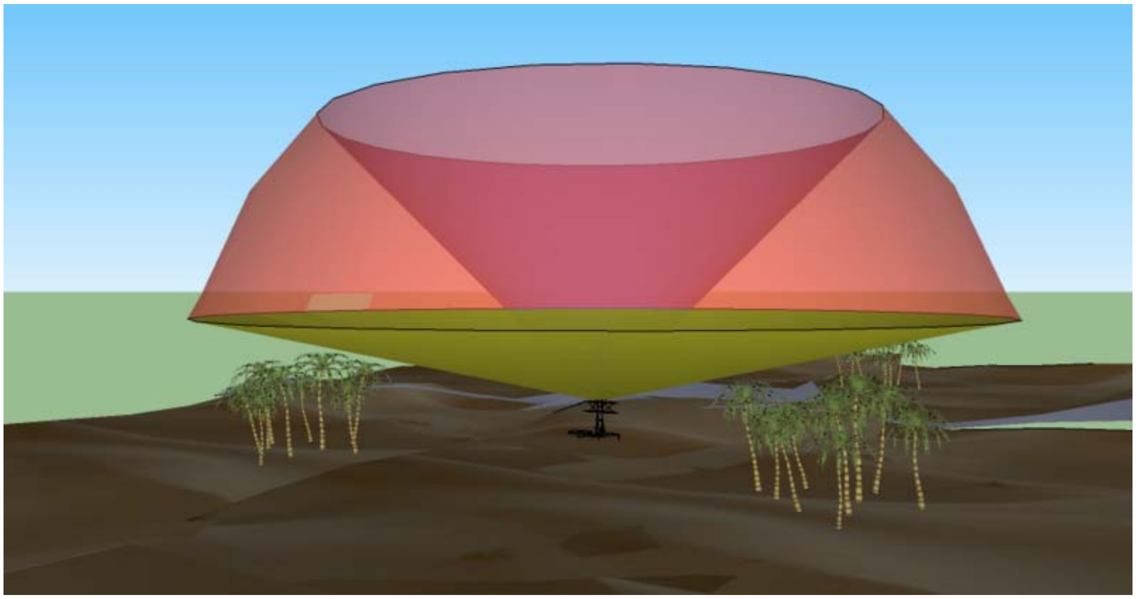
r* : distance** k_e : ***4/3 (Standard refraction coefficient)

a_e : ***Earth radius*** θ_e : ***Elevation angle***

h_a : ***Height of radar above ground***



Echos height above ground



Scanned volume by using multiple elevation angles

Assuming the Earth is round, with knowledge of the variation of the index of refraction through air and the distance to the target, one can calculate the height above ground of the target. The image to the left shows the calculation of the height depending on the elevation angle of the antenna and other considerations.

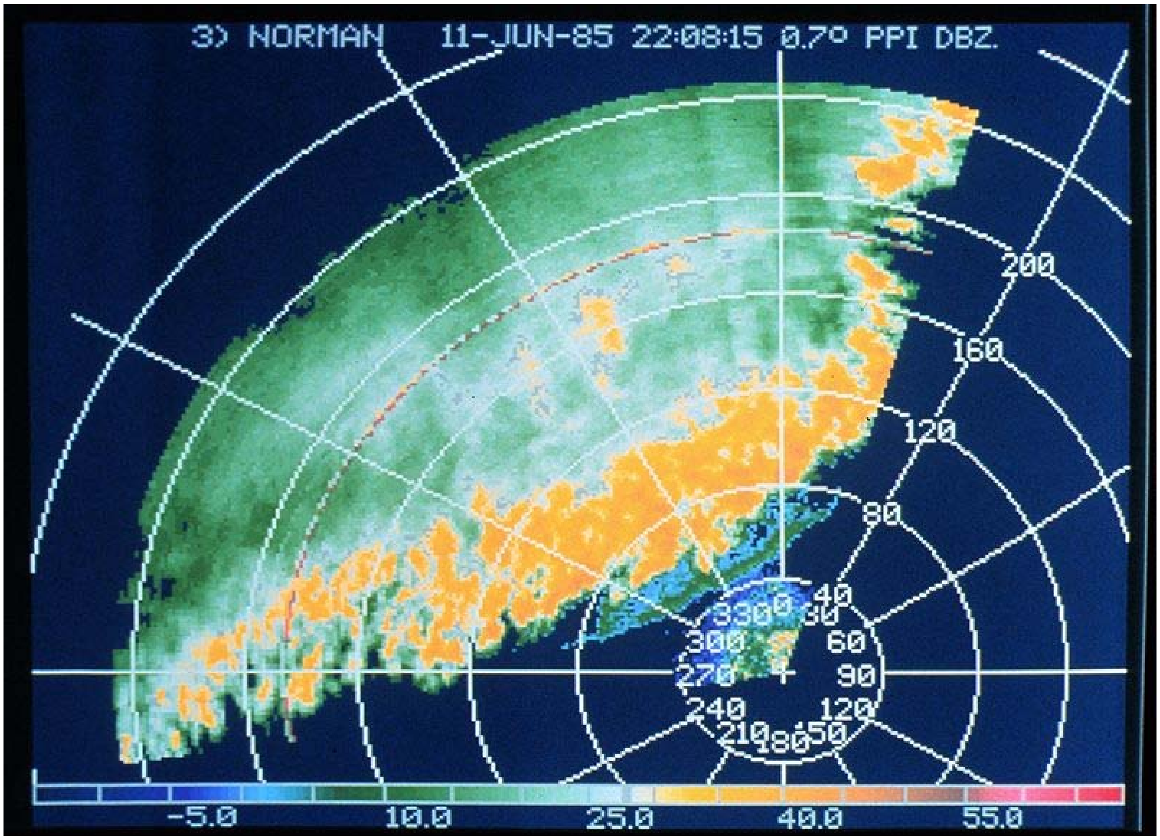
Each Weather radar network use a series of typical angles that will be set according to the needs. After each scanning rotation, the antenna elevation is changed for the next sounding. This scenario will be repeated on many angles to scan all the volume of air around the radar within the maximum range. Usually, this scanning strategy is completed within 5 to 10 minutes to have data within 15 km (9 mi) above ground and 250 km (155 mi) distance of the radar. For instance in Canada, the 5 cm (2 in) weather radars use angles ranging from 0.3 to 25 degrees. The image to the right shows an hypothetical volume scanned when multiple angles are used.

Due to the Earth curvature and change of index of refraction with height, the radar cannot "see" below the height above ground of the minimal angle (shown in green) or closer to the radar than the maximal one (show as a red cone in the center).

Main types of radar outputs

All data from radar scans are displayed according to the need of the users. Different outputs have been developed through time to reach this. Here is a list of common and specialized outputs available.

Plan position indicator



Thunderstorm line viewed in reflectivity (dBZ) on a PPI

Since data are obtained one angle at a time, the first way of displaying them has been the Plan Position Indicator (PPI) which is only the layout of radar return on a two dimensional image. One has to remember that the data coming from different distances to the radar are at different heights above ground.

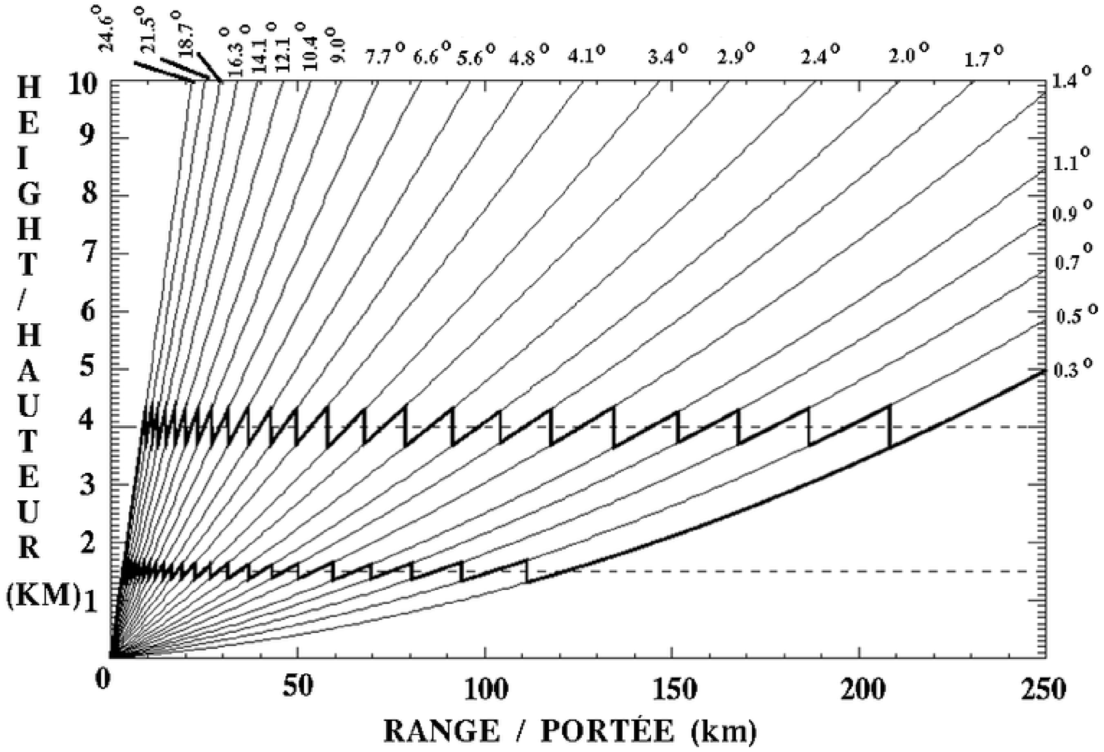
This is very important as a high rain rate seen near the radar is relatively close to what reach the ground but what is seen from 160 km (99 mi) (100 miles) away is about 1.5 km (1 mi) above ground and could be far different from the amount reaching the surface. It is thus difficult to compare weather echoes at different distance from the radar.

PPIs are afflicted with ground echoes near the radar as a supplemental problem. These can be misinterpreted as real echoes. So other products and further treatments of data have been developed to supplement its shortcomings.

USAGE: Reflectivity, Doppler and polarimetric data can use PPI.

N.B.: In the case of Doppler data, two points of view are possible: relative to the surface or the storm. When looking at the general motion of the rain to extract wind at different altitudes, it is better to use data relative to the radar. But when looking for rotation or wind shear under a thunderstorm, it is better to use the storm relative images that subtract the general motion of precipitation leaving the user to view the air motion as if he would be sitting on the cloud.

Constant Altitude Plan Position Indicator



Typical angles scanned in Canada. The zigzag represent data angles used to make CAPPIs at 1.5 & 4 km (2 mi) of altitude

To avoid some of the problems on PPIs, the CAPPI or Constant Altitude Plan Position Indicator has been developed by researchers in Canada. It is basically a horizontal cross-section through radar data. This way, one can compare precipitation on an equal footing at difference distance from the radar and avoid ground echoes. Although data are taken at a certain height above ground, a relation can be inferred between ground stations reports and the radar data.

CAPPIs call for a large number of angles from near the horizontal to near the vertical of the radar to have a cut that is as close as possible at all distance to the height needed. But even then, after a certain distance, there isn't any angle available and the CAPPI becomes the PPI of the lowest angle. The zigzag line on the angles diagram above shows the data used to produce a 1.5 and 4 km (2 mi) height CAPPIs. Notice that the section after 120 km (75 mi) is using the same data.

Usage

Since the CAPPI uses the closest angle to the desired height at each point from the radar, the data can originate from slightly different altitudes, as seen on the image, in different points of the radar coverage. It is therefore crucial to have a large enough number of sounding angles to minimize this height change. Furthermore, the type of data must be changing relatively gradually with height to produced an image that is not noisy.

Reflectivity data being relatively smooth with height, CAPPIs are mostly used for displaying them. Velocity data, on the other hand, can change rapidly in direction with height and CAPPIs of them are not common. It seems that only McGill University is producing regularly Doppler CAPPIs with the 24 angles available on their radar. However, some researchers have published papers using velocity CAPPIs to study tropical cyclones and development of NEXRAD products. Finally, polarimetric data are recent and often noisy. There doesn't seem to have regular use of CAPPI for them although the *SIGMET* company offer a software capable to produce those type of images.

Real time examples

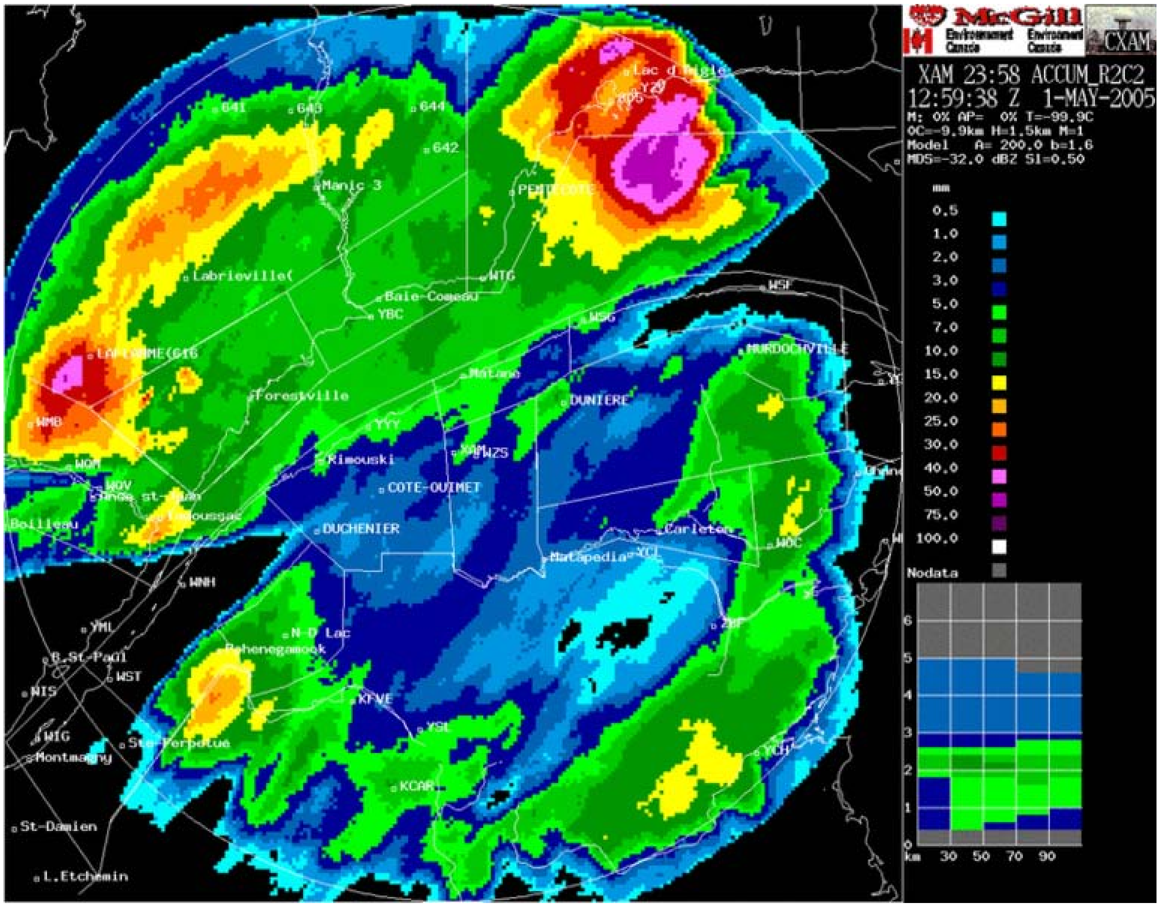
- McGill University
- Environment Canada

Vertical composite

Another solution to the PPI problems is to produce images of the maximum reflectivity in a layer above ground. This solution is usually taken when the number of angles available is small or variable. The American National Weather Service is using such Composite as their scanning scheme can vary from 4 to 14 angles, according to their need, which would make very coarse CAPPIs. The Composite make sure that no strong echo is missed in the layer and a treatment using Doppler velocities eliminate the ground echoes. Comparing base and composite products, one can locate virga and updrafts zones.

Real time example: NWS Burlington radar, one can compare the BASE and COMPOSITE products

Accumulations



24 hours rain accumulation on the Val d'Irène radar in Eastern Canada. Notice the zones without data in the East and Southwest caused by radar beam blocking from mountains.

One of the main use of radar is to be able to assess the amount of precipitations fallen over large basins for hydrological purpose. For instance, river flood control, sewer management and dam construction are all areas where planners want accumulation data. It ideally completes surface stations data which they can use for calibration.

To produce radar accumulations, we have to estimate the rain rate over a point by the average value over that point between one PPI, or CAPPI, and the next; then multiply by the time between those images. If one wants for a longer period of time, one has to add up all the accumulations from images during that time.

Echotops

Aviation is a heavy user of radar data. One map particularly important in this field is the Echotops for flight planning and avoidance of dangerous weather. Most country weather radars are scanning enough angles to have a 3D set of data over the area of coverage. It is relatively easy to estimate the maximum altitude at which precipitation is found within the volume. However, those are not the tops of clouds as they extend to higher altitudes than the precipitation.

Vertical cross sections

To know the vertical structure of clouds, in particular thunderstorms or the level of the melting layer, a vertical cross sections product of the radar data is available to meteorologist. This is done by displaying only the data along a line, from coordinates A to B, taken from the different angles scanned.

Range Height Indicator

When a weather radar is scanning in only one direction vertically, it obtains high resolution data along a vertical cut of the atmosphere. The output of this sounding is called a *Range Height Indicator* (RHI) which is excellent for viewing the detailed vertical structure of a storm. This is different from the vertical cross section mentioned above by the fact that the radar is making a vertical cut along specific directions and does not scan over the entire 360 degrees around the site. This kind of sounding and product is only available on research radars.

Radar networks

Berrimah Radar in Darwin, Northern Territory Australia

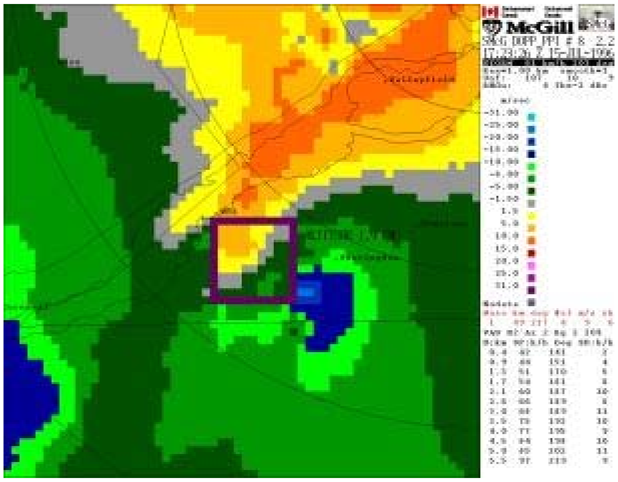
Over the past few decades, radar networks have been extended to allow the production of composite views covering large areas. For instance, all major countries (e.g., the United States, Canada, much of Europe) produce images that include all of their radars. This is not as trivial a task.

In fact, such a network can consist of different types of radar with different characteristics like beam width, wavelength and calibration. These differences have to be taken into account when matching data across the network, particularly to decide what data to use when two radars cover the same point. If one uses the stronger echo but it comes from the most distant radar, one uses returns that are from higher altitude coming from rain or snow that might evaporate before reaching the ground (virga). If one uses data from the closest radar, it might be attenuated passing through a thunderstorm. Composite images of precipitations using a network of radars are made with all those limitations in mind.

Here are some national radar networks :

- Environment Canada
- National Weather Service in United States
- Czech Republic
- South African Republic
- Deutscher Wetterdienst in Germany
- Bureau of Meteorology, Australia
- Smhi, Scandinavia and Baltic sea
- POLRAD - Poland

Automatic algorithms



The square in this Doppler image has been automatically placed by the radar program to spot the position of a mesocyclone. Notice the inbound/outbound doublet (blue/yellow) with the zero velocity line (gray) parallel to the radial to the radar (up right). It is noteworthy to mention that the change in wind direction here occurs over less than 10 km (6 mi) (6 miles).

To help meteorologists to spot dangerous weather, mathematical algorithms have been introduced in the weather radar treatment programs. These are particularly important in the analyzing the Doppler velocity data as they are more complex. The polarization data will even need more algorithms.

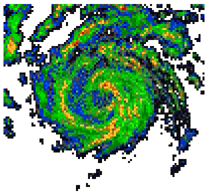
Main algorithms for reflectivity:

- Vertically Integrated Liquid (VIL) is an estimate of the total mass of precipitation in the clouds.
- *VIL Density* is VIL divided by the height of the cloud top. It is a clue to the possibility of large hail in thunderstorms.
- *Potential wind gust*, which can estimate the winds under a cloud (a downdraft) using the VIL and the height of the echotops (radar estimated top of the cloud) for a given storm cell.
- Hail algorithms that estimates the presence and its potential size.

Main algorithms for Doppler velocities:

- Mesocyclone detection: it is triggered by a velocity change over a small circular area. The algorithm is searching for a "*doublet*" of inbound/outbound velocities with the zero line of velocities, between the two, along a radial line from the radar. Usually the mesocyclone detection must be found on two or more stacked progressive tilts of the beam to be significative of rotation into a thunderstom cloud.
- TVS or Tornado Vortex Signature algorithm is essentially a mesocyclone with a large velocity threshold found through many scanning angles. This algorithm is used in NEXRAD to indicate the possibility of a tornado formation.
- Wind shear in low levels. This algorithm detects variation of wind velocities from point to point in the data and looking for a *doublet* of inbound/outbound velocities with the zero line perpendicular to the radar beam. The wind shear is associated with downdraft, (downburst and microburst), gust fronts and turbulence under thunderstorms.
- VAD Wind Profile (VWP) is a display that estimates the direction and speed of the horizontal wind at various upper levels of the atmosphere, using the technique explained in the Doppler section.

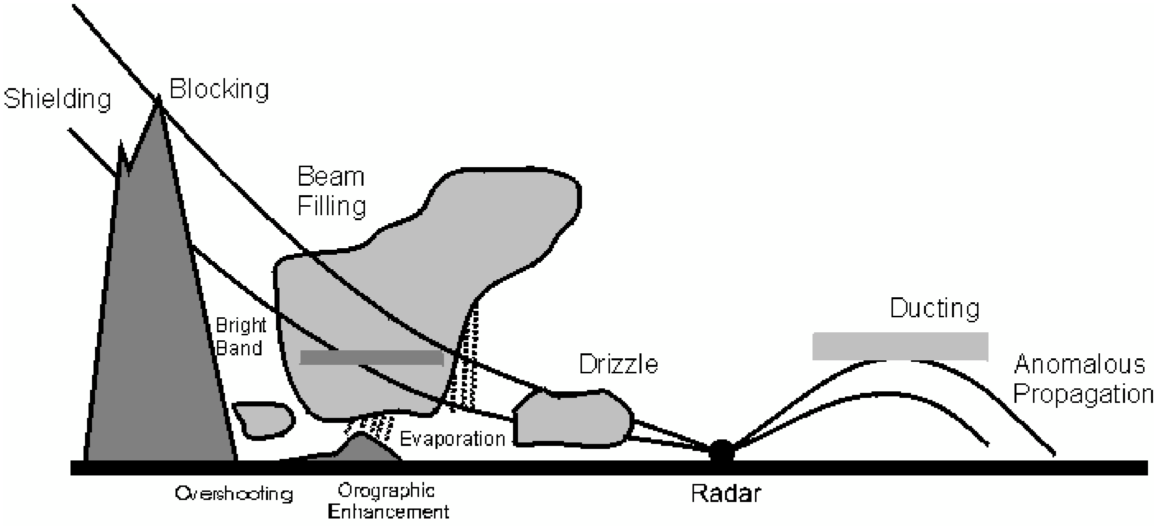
Images



PPI reflectivity loop(in dBZ) showing the evolution of a hurricane

The animation of radar products can show the evolution of reflectivity and velocity patterns. The user can extract information on the dynamics of the meteorological phenomena, including the ability to extrapolate the motion and observe development or dissipation. This can also reveal non-meteorological artifacts (false echoes) that will be discussed later.

Limitations and artifacts



Radar data interpretation depends on many hypotheses about the atmosphere and the weather targets. They are:

- International Standard Atmosphere.
- Targets small enough that they obey the Rayleigh scattering so the return is proportional to the precipitation rate.
- The volume scanned by the beam is full of ***meteorological*** targets (rain, snow, etc..), all of the same variety and in a uniform concentration.
- No attenuation
- No amplification
- Return from side lobes of the beam are negligible.
- The beam is close to a Gaussian function curve with power decreasing to half at half the width.

- The outgoing and returning waves are both polarized similarly.
- There is no return from multiple reflections.

One has to keep in mind that these hypotheses are not necessarily met in many circumstances. One has to be able to recognize the truth from the false echoes.

Anomalous propagation (non-standard atmosphere)

The first assumption is that the radar beam is moving through air that cools down at a certain rate with height. The position of the echoes depend heavily on this hypothesis. However, the real atmosphere can vary greatly from the norm.

Super refraction

It is very common to have temperature inversions forming near the ground, for instance air cooling at night while remaining warm aloft. As the index of refraction of air decreases faster than normal the radar beam bends toward the ground instead of continuing upward. Eventually, it will hit the ground and be reflected back toward the radar. The processing program will then wrongly place the return echoes at the height and distance it would have been in normal conditions.

This type of false return is relatively easy to spot on a time loop if it is due to night cooling or marine inversion as one sees very strong echoes developing over an area, spreading in size laterally but not moving and varying greatly in intensity. However, inversion of temperature exists ahead of warm fronts and the abnormal propagation echoes are then mixed with real rain.

The extreme of this problem is when the inversion is very strong and shallow, the radar beam reflects many times toward the ground as it has to follow a waveguide path. This will create multiple bands of strong echoes on the radar images.

This situation can be found with inversions of temperature aloft or rapid decrease of moisture with height. In the former case, it could be difficult to notice.

Under refraction

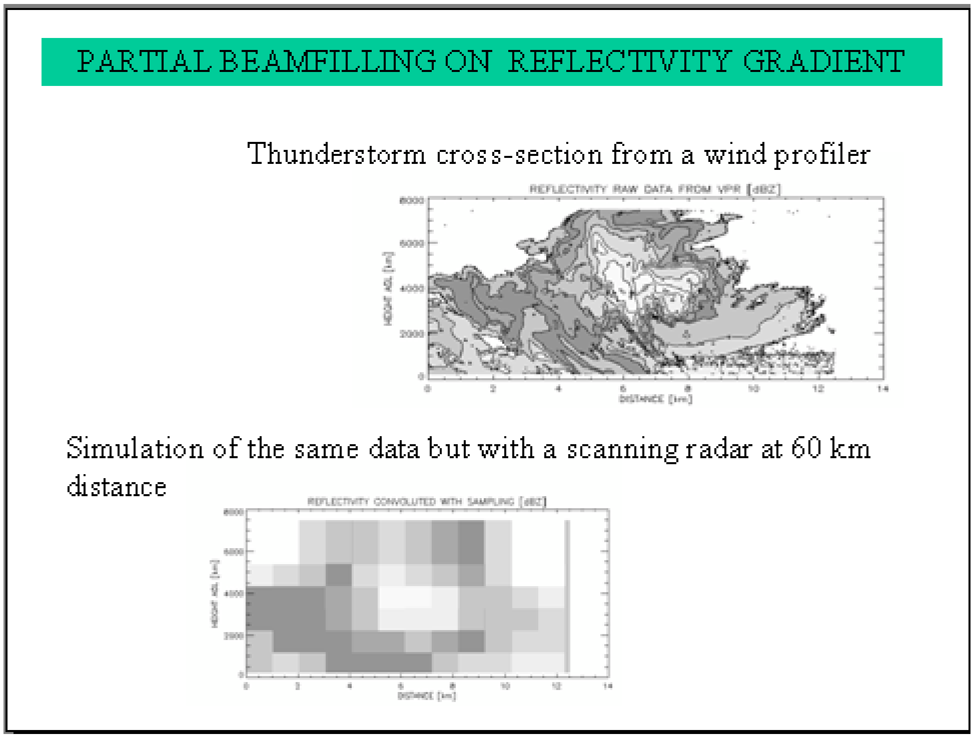
On the other hand, if the air is unstable and cools faster than the standard atmosphere with height, the beam ends up higher than expected. This places the precipitation at a much higher altitude than it actually is. This situation is very difficult to spot.

Non-Rayleigh targets

If we want to reliably estimate the precipitation rate, the targets have to be 10 times smaller than the radar wave according to Rayleigh scattering. This is because the water molecule has to be excited by the radar wave to give a return. This is relatively true for rain or snow as 5 or 10 cm (4 in) radars are used.

However, for very large hydrometeors, since the wavelength is on the order of stone, the return levels off according to Mie theory. A return of more than 55 dBZ is likely to come from hail but won't vary proportionally to the size. On the other hand, very small targets, like cloud droplets, are too small to be excited and don't give a recordable return on common weather radars.

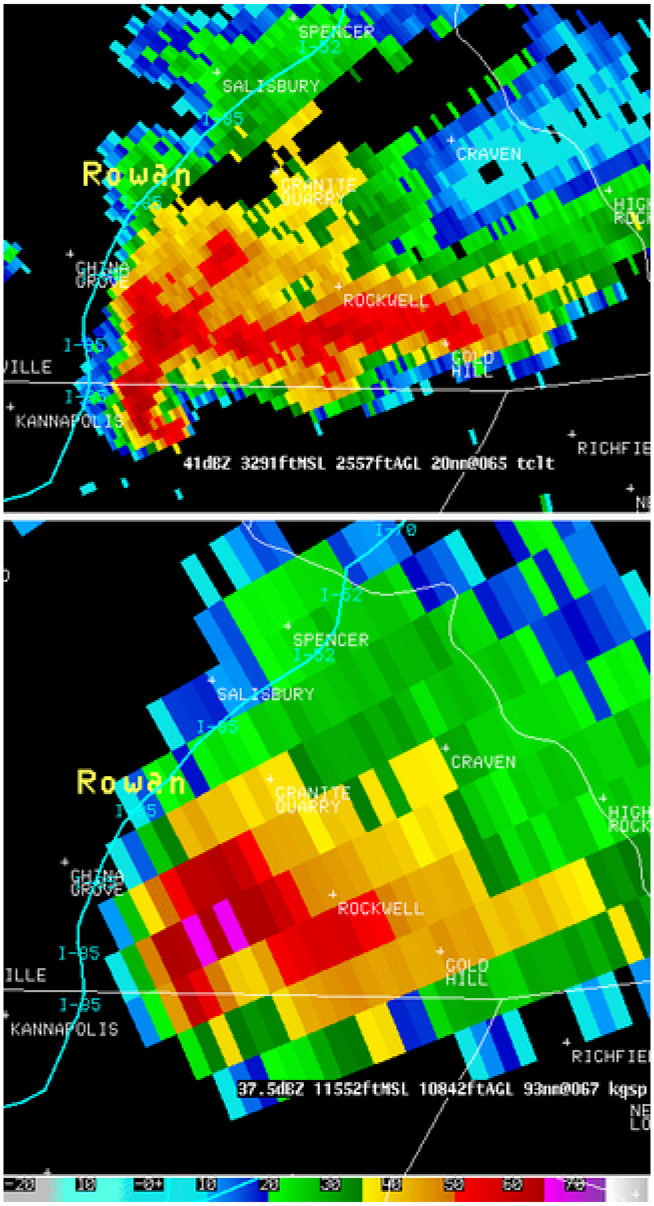
Resolution and partially filled scanned volume


PARTIAL BEAMFILLING ON REFLECTIVITY GRADIENT

Thunderstorm cross-section from a wind profiler

Simulation of the same data but with a scanning radar at 60 km distance

Profiler high resolution view of a thunderstorm (top) and by a weather radar (bottom)



A supercell thunderstorm seen from two radars almost colocated. The top image is from a TDWR and the bottom one from a NEXRAD.

As demonstrated at the start, radar beams have a physical dimension and data are sampled every degree, not continuously, along each angle of elevation. This results in an averaging of the values of the returns for reflectivity, velocities and polarization data on the resolution volume scanned.

In the figure to the left, at the top is a view of a thunderstorm taken by a wind profiler as it was passing overhead. This is like a vertical cross section through the cloud with 150 m vertical and 30 m horizontal resolution. We can see that the reflectivity has large variations in a short distance. Now compare this with a simulated view of what a regular weather radar would see at 60 km (37 mi) (40 miles) at the bottom. Everything has been smoothed out. Not only the coarser resolution of the radar blur the image but the sounding incorporate area that are echo free, thus extending the thunderstorm beyond its real boundaries.

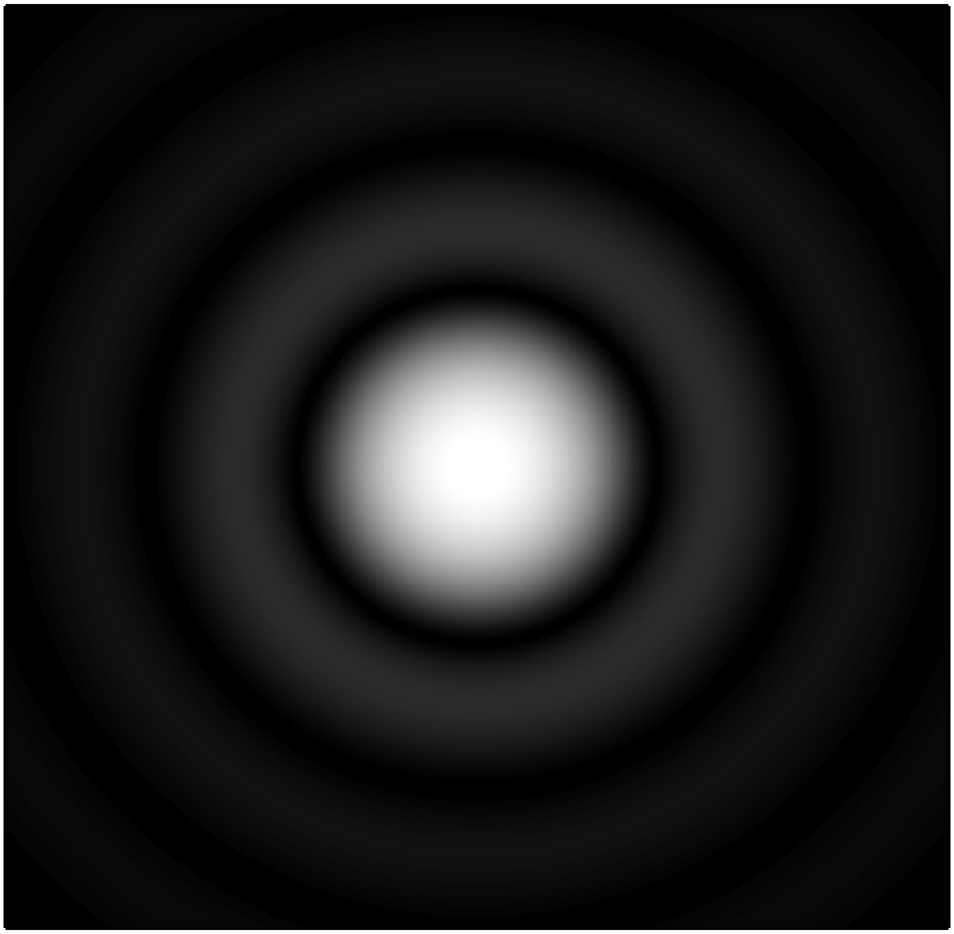
This shows how the output of weather radar is only an approximation of the reality. The image to the right compare real data from two radars almost colocated. The TWDR has about half the beamwidth of the other and one can see twice more details than with the NEXRAD.

Naturally, resolution can be improved by newer equipment but some things cannot. As mentioned previously, the volume scanned increases with distance so the possibility that the beam is only partially filled increases too. This leads to underestimation of the precipitation rate at larger distances and fools the user into thinking that rain is lighter as it moves away.

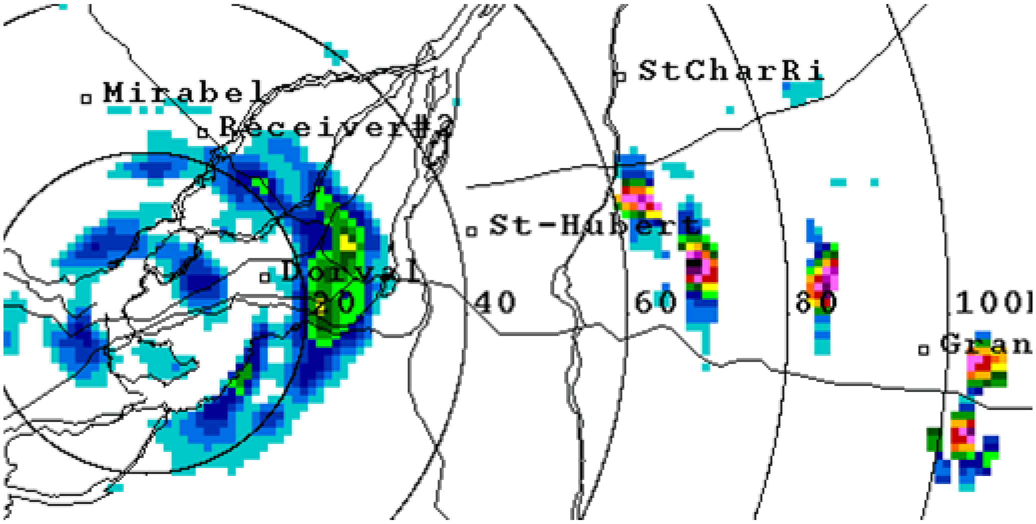
Beam geometry

The radar beam has a distribution of energy similar to the diffraction pattern of a light passing through a slit. This is because the wave is transmitted to the parabolic antenna through a slit in the wave-guide at the focal point. Most of the energy is at the center of the beam and decreases along a curve close to a Gaussian function on each side as mentioned before. However, there are secondary peaks of emission that will sample the targets at off-angles from the center. All is done to minimize the power sent by those lobes but they are never zero.

When a secondary lobe hits a very reflective target, like a mountain or a strong thunderstorm, some of the energy is sent back to the radar. This energy is relatively weak but arrives at the same time the central peak is illuminating a different azimuth. The echo is thus misplaced by the processing program. This has the effect of actually broadening the real weather echo making a smearing of weaker values on each side of it. This causes the user to overestimate the extent of the real echoes.



Diffraction by a circular slit simulating the energy viewed by weather targets



The strong echoes are returns of the central peak of the radar from a series of small hills (yellow and reds pixels). The weaker echoes on each sides of them are from secondary lobes (blue and green)

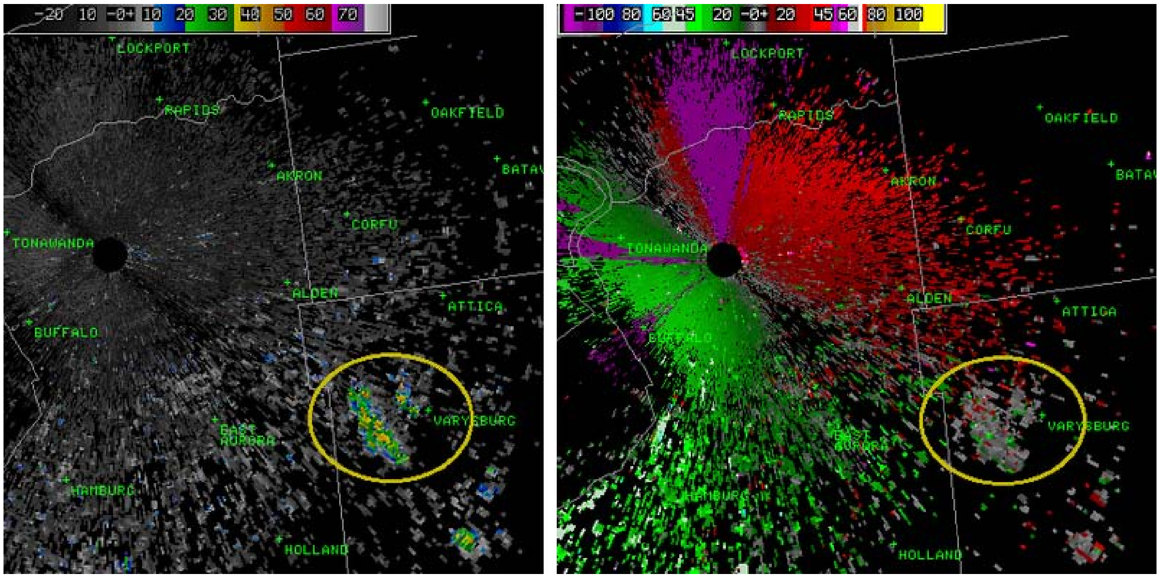
Non weather targets

There is more than rain and snow in the sky. Other objects can be misinterpreted as rain or snow by weather radars. Insects and arthropods are swept along by the prevailing winds, while birds follow their own course. As such, fine line patterns within weather radar imagery, associated with converging winds, are dominated by insect returns. Bird migration, which tends to occur overnight within the lowest 7,000 feet (2,100 m) of the Earth's atmosphere, contaminates wind profiles gathered by weather radar, particularly the WSR-88D, by increasing the environmental wind returns by 15 knots (28 km/h) to 30 knots (56 km/h) Other objects within radar imagery include::

- Thin metal strips (chaff) dropped by military aircraft to fool enemies.
- Solid obstacles such as mountains, buildings, and aircraft.
- Ground and sea clutter.
- Reflections from buildings if the radar is close enough to a city (called urban spikes).

Each of them has their own characteristics that make it possible to distinguish them to the trained eye but they may fool a layman. It is possible to eliminate some of them with post-treatment of data using reflectivity, Doppler, and polarization data.

Wind farms

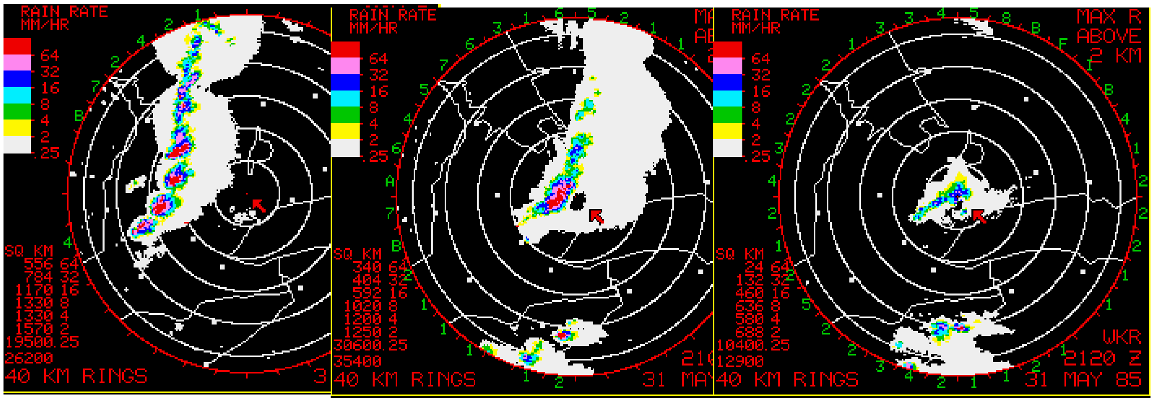


Reflectivity (left) and radial velocities (right) southeast of a NEXRAD weather radar. Echoes in circles are from a wind farm.

The rotating blades of windmills on modern wind farms can return the radar beam to the radar if they are in its path. Since the blade are moving, the echoes will have a velocity and can be mistaken for real precipitation. The closer the wind farm is to the radar, the more important is this artifact as the combined signal from many towers is stronger. If the conditions are right, the radar can even see a doublet of toward and away velocities that can generate false positives for the tornado vortex signature algorithm on weather radar, as happened in 2009 in Dodge City, Kansas.

Finally, the windmills are blocking a part of the radar beam and thus attenuating the return from precipations in the lee of them, leading to underestimation.

Attenuation



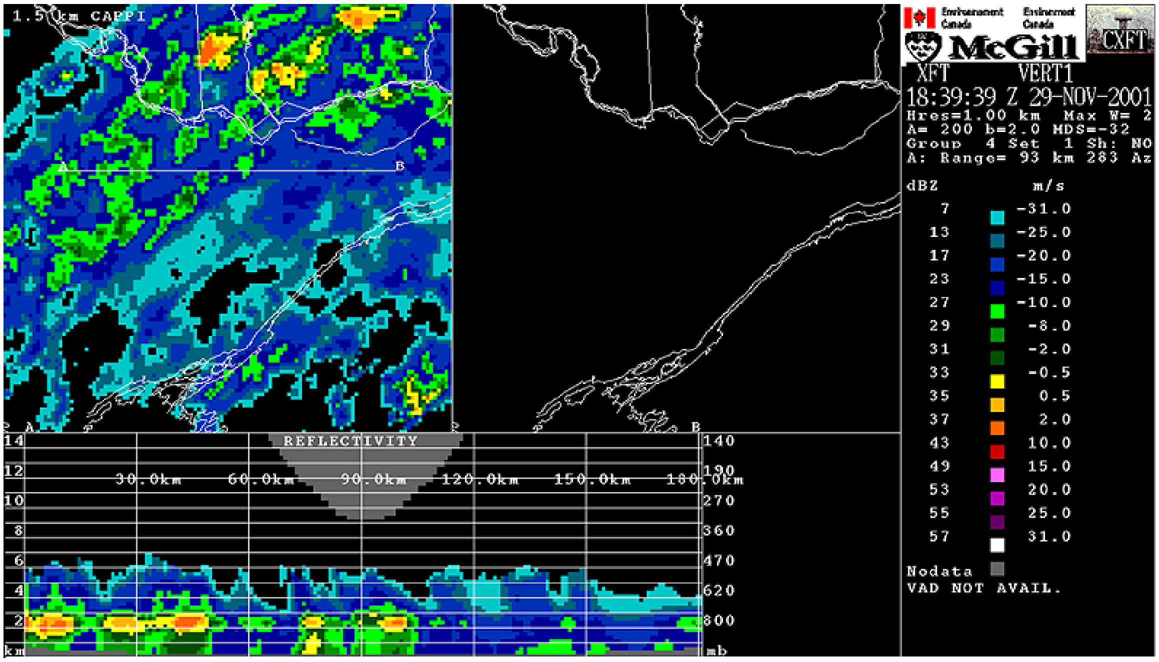
Example of strong attenuation when a line of thunderstorms move over (from left to right images) a 5 cm (2 in) wavelength weather radar (red arrow). Source: Environment Canada

Micro-waves used in weather radars can be absorbed by rain, depending on the wavelength used. For the 10 centimeter radars, this attenuation is negligible. That is the reason why countries with high water content storms are using 10 centimeter wavelength like in the United States with NEXRAD. The cost of a larger antenna, klystron and other related equipments is offset by this benefit.

For a 5 centimeter radar, absorption becomes important in very heavy rain and this attenuation leads to underestimation of echoes in and beyond a strong thunderstorms line. Canada and other northern countries use this less costly kind of radars as their precipitations are usually less intense. However, users have to remember this effect when interpreting data. The images above show how a strong line of echoes seems to vanish as it moves over the radar. To compensate for this behaviour, radar sites are often chosen to somewhat overlap in coverage to give different points of view of the same storms.

Shorter wavelengths are even more attenuated and are only useful on short range radar. Many television stations in the United States have 3-centimeter radars to cover their audience area. Knowing their limitations and using them with the local NEXRAD can supplement the data available to a meteorologist.

Bright band



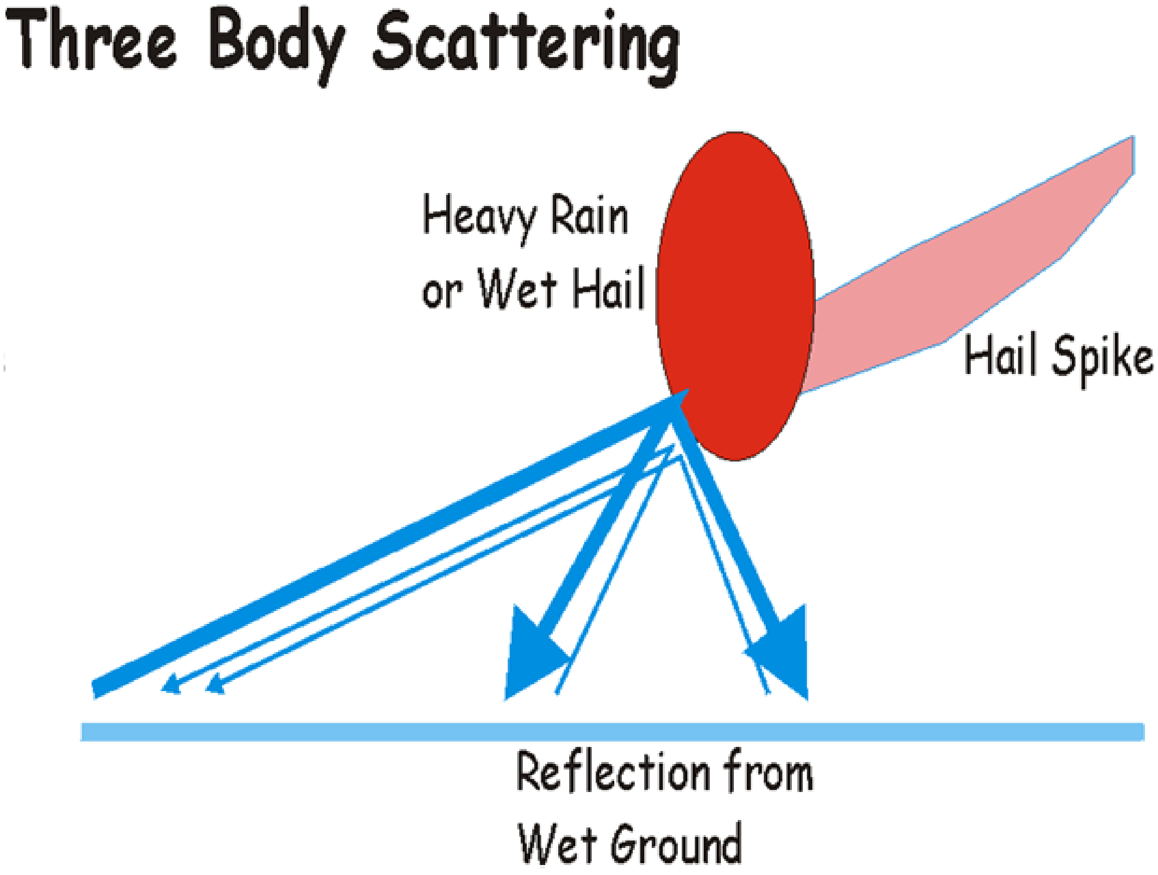
1.5 km altitude CAPPI at the top with strong contamination from the brightband (yellows). The vertical cut at the bottom show that this strong return is only above ground.

As we have seen previously, the reflectivity depends on the diameter of the target and its capacity to reflect. Snow flakes are large but weakly reflective while rain drops are small but highly reflective.

When snow falls through a layer above freezing temperature, it melts and eventually becomes rain. Using the reflectivity equation, one can demonstrate that the returns from the snow before melting and the rain after, are not too different as the change in dielectric constant compensate for the change in size. However, during the melting process, the radar wave "sees" something akin to very large droplets as snow flakes become coated with water.

This gives enhanced returns that can be mistaken for stronger precipitations. On a PPI, this will show up as an intense ring of precipitations at the altitude where the beam crosses the melting level while on a series of CAPPIs, only the ones near that level will have stronger echoes. A good way to confirm a bright band is to make a vertical cross section through the data like in the picture above.

Multiple reflections

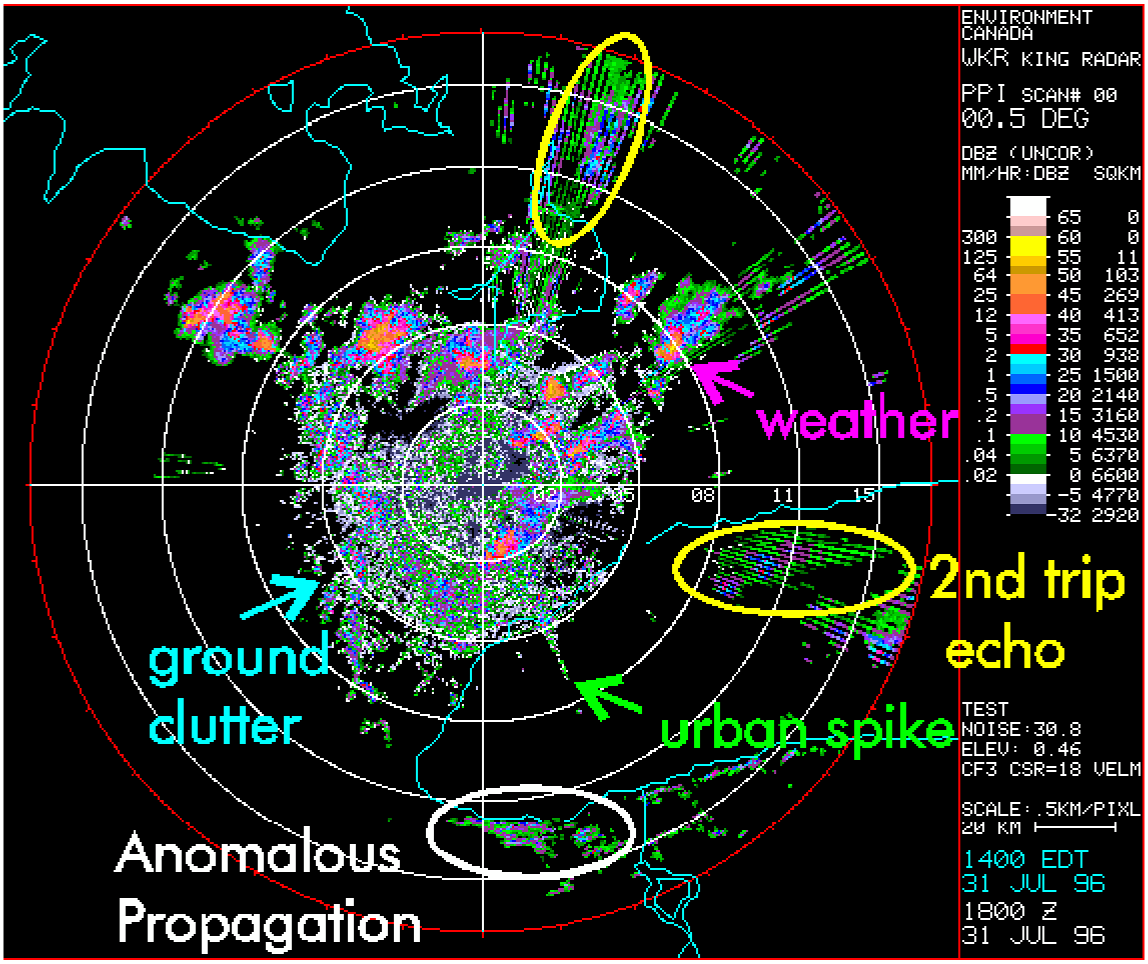


It is assumed that the beam hits the weather targets and returns directly to the radar. In fact, there is energy reemitted in all directions. Most of it is weak, and multiple reflections diminish it even further so what can eventually return to the radar from such an event is negligible. In some cases though, this cannot be.

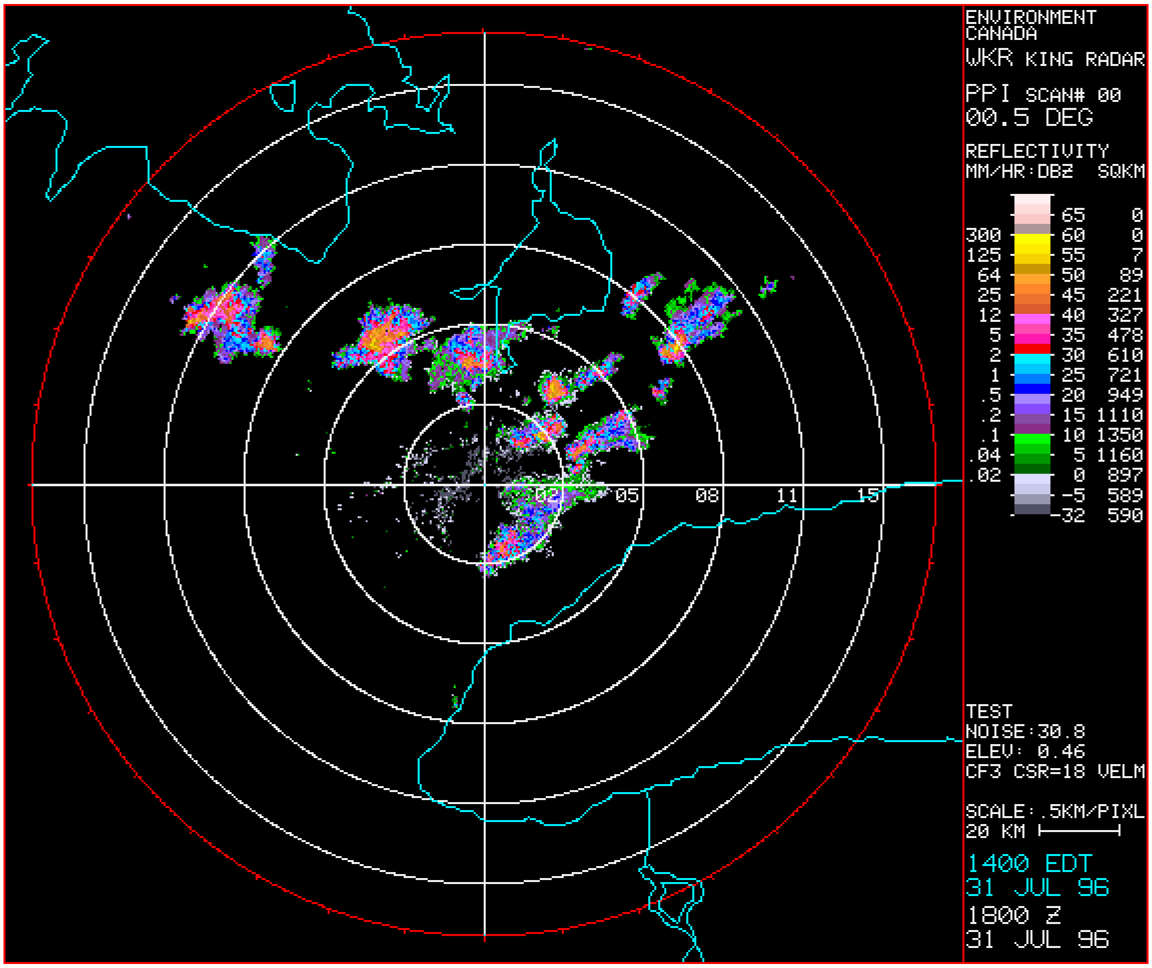
For instance, when the beam hits hail, the energy spread toward the wet ground will be reflected back to the hail and then to the radar. The resulting echo is weak but noticeable. Due to the extra path length it has to go through, it arrives later at the antenna and is placed further than its source. This gives a kind of triangle of false weaker reflections placed radially behind the hail.

Solutions for now and the future

Filtering



Radar image of reflectivity with many non-weather echoes



The same image but cleaned using the Doppler velocities

These two images show what can be achieved already to clean up radar data. The output on the left is made with the raw returns and it is difficult to spot the real weather. Since rain and snow clouds are usually moving, one can use the Doppler velocities to eliminate a good part of the clutter (ground echoes, reflections from buildings seen as urban spikes, anomalous propagation, etc..). The image on the right has been filtered using this property in a somewhat complex technique.

However, not all non-meteorological targets remain still; one can think of birds for instance. Others, like the bright band, depend on the structure of the precipitations. Polarization offers a direct typing of the echoes which could be used to filter more false data or produce separate images for specialized purposes. This recent development in this field is bound to improve the quality of radar products.

Mesonet

Phased Array Weather Radar in Norman, Oklahoma

Another question is the resolution. As mentioned previously, radar data are an average of the scanned volume by the beam. Resolution can be improved by larger antenna or denser networks. A program by the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA) aims to supplement regular NEXRAD using many low cost X band (3 cm) weather radar mounted on cellular telephone towers.. These radars will subdivide the large area of the NEXRAD into smaller domains to look at altitudes below its lowest angle. These will give details not currently available.

Using 3-cm wavelength radars, the antenna of each radar is small (about 1 meter diameter) but the resolution is similar at short distance to that of NEXRAD. The attenuation is significant due to the wavelength used but each point in the coverage area is seen by many radars, each viewing from a different direction and compensating for data lost from others.

Electronic sounding

Timeliness is also a point needing improvement. With 5 to 10 minutes time between complete scans of weather radar, a lot of things can be missed in the development of a thunderstorm. A Phased-array radar is being tested at the National Severe Storms Lab in Norman, Oklahoma, to speed up the gathering of data.

Specialized applications

Weather radar on the wing of Pilatus PC-12

Avionics weather radar

Aircraft application of radar systems include weather radar, collision avoidance, target tracking, ground proximity, and other systems. For commercial Weather Radar Systems, ARINC 708 is the primary weather radar system using an airborne pulse-Doppler radar.

Antennas

Unlike ground weather radar, which is set at a fixed angle, airborne weather radar is being utilized from the nose or wing of an aircraft. Not only will the aircraft be moving up, down, left, and right, but it will be rolling as well. To compensate for this, the antenna is linked and calibrated to the vertical gyro located on the aircraft. By doing this, the pilot is able to set a pitch or angle to the antenna that will enable the stabilizer to keep the antenna pointed in the right direction under moderate maneuvers. The small servo motors will not be able to keep up with too abrupt maneuvers, but it will try. In doing this the pilot is able to adjust the radar so that it will point towards the weather system of interest. If the airplane is at a low altitude, the pilot would want to set the radar at a high angle above the horizon line so that ground clutter is not all that is being displayed on the plan position indicator (PPI). Similarly, if the airplane is at a very high altitude the pilot would want to set the radar at a low or negative angle. The goal here is to point the radar towards the clouds wherever they may be in respect to the aircraft, and if the pilot changes the direction of the airplane the stabilizer will adjust itself accordingly so that the pilot doesn't have to fly with one hand and adjust the radar with the other. The stabilizer is meant to help display the information that the pilot requests, while the airplane is in motion.

Receivers/Transmitters

There are two major systems when talking about the receiver/transmitter: the first is high-powered systems, and the second is low-powered systems; both of which operate in the x-band frequency range (8,000 to 12,500) MHz. High-powered systems operate at power levels between 10,000 and 60,000 watts. These systems consist of magnetrons and vacuum tubes that are fairly expensive (approximately $1,700) and allow for considerable amounts of noise due to irregularities with the system. Thus, these systems are highly dangerous for arcing and are not safe to be used around ground personnel. However, the alternative would be the low-powered systems. These systems operate between 100 to 200 watts, and require a combination of high gain receivers, signal microprocessors, and transistors to operate as effectively as the high-powered systems. The complex microprocessors help to eliminate noise, providing a more accurate and detailed depiction of the sky. Also, since there are fewer irregularities throughout the system, the low-powered radars can be used to detect turbulence via the Doppler Effect. Furthermore, since the low-powered systems operate at considerable less wattage, they are safe from arcing and can be used at virtually all times.

Thunderstorm Tracking

Digital radar systems now have capabilities far beyond that of their predecessors. Digital systems now offer thunderstorm tracking surveillance. This provides users with the ability to acquire detailed information of each storm cloud being tracked. Thunderstorms are first identified by matching precipitation raw data received from the radar pulse to some sort of template preprogrammed into the system. In order for a thunderstorm to be identified, it has to meet strict definitions of intensity and shape that set it apart from any non-convective cloud. Usually, it must show signs of organization in the horizontal and continuity in the vertical : a core or a more intense center to be identified and tracked by digital radar tracking systems. Once the thunderstorm cell is identified, speed, distance covered, direction, and Estimated Time of Arrival (ETA) are all tracked and recorded to be utilized later.

Chapter 8

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.

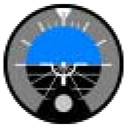
Flight instruments

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

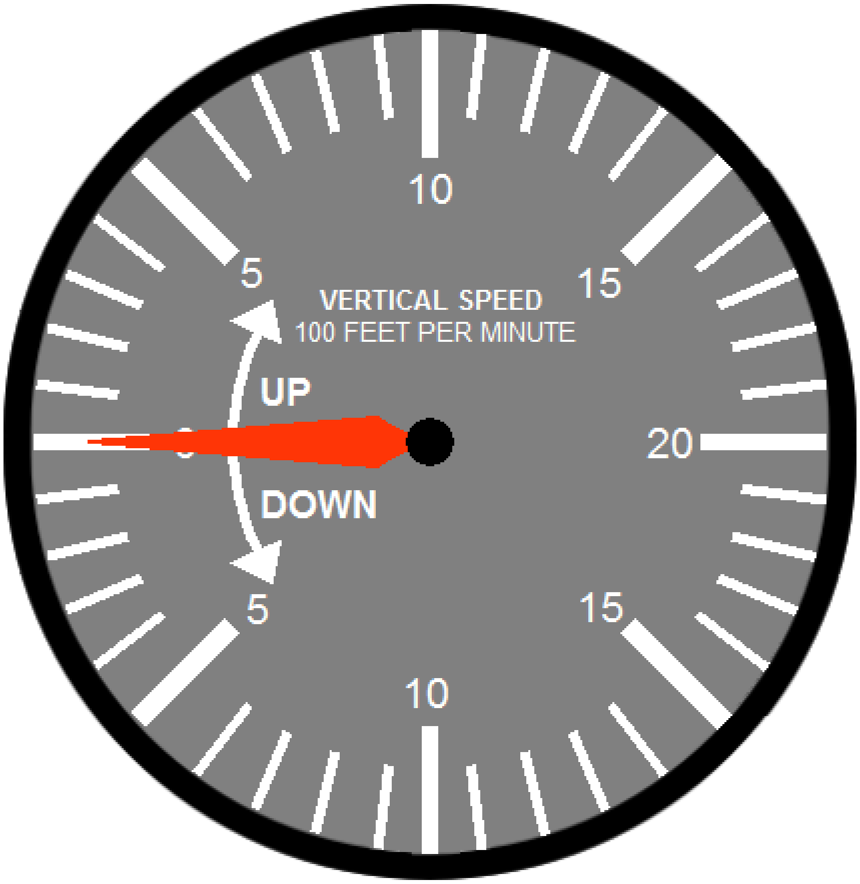
The heading indicator (also known as the directional gyro, or DG; sometimes also called the gyrocompass, though usually not in aviation applications) displays the aircraft's heading with respect to geographical north. Principle of operation is a spinning gyroscope, and is therefore subject to drift errors (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.

Chapter 9

Flight Control Modes (Electronic)

Flight control modes (electronic)

New aircraft designs like this Boeing 777 rely on sophisticated flight computers to aid and protect the aircraft in flight. These are governed by computational laws which assign flight control modes during flight

Aircraft with fly-by-wire flight controls require computer *controlled flight control modes* that are capable of determining the operational mode (computational law) of the aircraft.

A reduction of electronic flight control can be caused by the failure of a computational device, such as the flight control computer or an information providing device, such as the ADIRU.

Electronic flight control systems (EFCS) also provide augmentation in normal flight, such as increased protection of the aircraft from overstress or providing a more comfortable flight for passengers by recognizing and correcting for turbulence and providing yaw damping.

Two aircraft manufacturers produce commercial passenger aircraft with primary flight computers that can perform under different flight control modes (or laws). The most well-known are the *Normal, Alternate, Direct* and *Mechanical Laws* of the Airbus A320-A380.

Boeing's fly-by-wire system is used in the Boeing 777. Boeing also has two other commercial aircraft under development, the 787 and the 747-8, which will use fly-by-wire controls.

These newer generation of aircraft use the lighter weight electronic systems to increase safety and performance while lowering aircraft weight. Since these systems can also protect the aircraft from overstress situations, the designers can therefore reduce over-engineered components, further reducing weight.

Philosophies of design

Aircraft designers have created a set of flight control modes that include redundant electronics to safeguard against system failures. Failures can occur singly or combined to render systems inoperable. Pilots must be able to control the aircraft with some, or even none, of the computational electronics functioning. In the case of Airbus the back-ups are the direct and mechanical modes. Boeing's direct mode removes many of the computational 'limitations'.

In older aircraft, control is through the pilot's control column, rudder pedals, trim wheel or throttles that mechanically move cables, pulleys or hydraulic servo valves. These then move control surfaces or change engine settings.

Many newer aircraft replace these mechanical controls with fly-by-wire systems. These aircraft have flight control computers which operate control surfaces, inform the pilot and provide performance information. In older aircraft the pilot's mechanical controls are resisted by the forces acting on the control surface, but nothing prevents the aircraft from stalling, over-speeding or an excessive bank angle at high speed. Fly-by-wire systems limit control surface movements to ensure that aircraft limits are not exceeded.

Another function of flight control laws is to assess the performance of the aircraft under various conditions, such as takeoff, landing or normal cruise when flight control computers partially or completely fail. Designers build in the ability to by-pass the computers or for the standby systems to operate without the computers.

Flight control laws (Airbus)

A330-200 in flight mode

Airbus aircraft designs after the A300/A310 are almost completely controlled by fly-by-wire equipment. These newer aircraft, including the A320, A330, A340, A350 and A380 operate under Airbus flight control Laws. The flight controls on the Airbus 330, for example, are all electronically controlled and hydraulically activated. Some surfaces, such as the rudder, can also be mechanically controlled. While in normal flight the computers act to prevent excessive forces in the pitch and roll.

Airbus 320-100 Cockpit

Illustration of the Air-data reference system on Airbus A330

The aircraft is controlled by three primary control computers (Captain's, First Officer's and Standby) and two secondary control computers (Captain's and First Officer's). In addition there are two Flight Control Data Computers (FCDC) that read information from the sensors, such as air data (airspeed, altitude). This is fed along with GPS data, into three processing units known as ADIRUs (Air data/inertial reference units) which act both as an air data reference and inertial reference. ADIRUs are part of the air data inertial reference system, which, on the Airbus is linked to eight air data modules: three are linked to pitot tubes and five are linked to static sources. Information from the ADIRU is fed into one of several flight control computers (Primary and secondary flight control). The computers also receive information from the control surfaces of the aircraft and from the pilots aircraft control devices and autopilot. Information from these computers is sent both to the pilot's primary flight display and also to the control surfaces.

There are four named flight control laws, however *Alternate Law* consists of two modes, *Alternate Law 1* and *Alternate Law 2*. Each of these modes have different sub modes: ground mode, flight mode and flare, plus a back-up *Mechanical Law*.

Normal law

Normal Law differs depending on the stage of flight. These include:

- Stationary at the gate
- Taxiing from the gate to a runway or from a runway back to the gate
- Beginning the take-off roll
- Initial climb
- Cruise climb and cruise flight at altitude
- Final descent, flare and landing.

Normal Law is different depending on the stage of flight. During the transition from take-off to cruise there is a 5 second transition, from descent to flare there is a two second delay and from flare to ground there is another 2 second transition in *Normal Law*.

Ground mode

The aircraft behaves as in direct mode: The autotrim feature is turned off and there is a direct response of the elevators to control column (or sidestick on Airbus) inputs. The horizontal stabilizer is set to 4° up but manual settings (e.g. for center of gravity) override this setting. After the wheels leave the ground, a 5 second transition occurs where *Normal Law - flight mode* takes over from ground mode.

Flight mode

The flight mode of *Normal Law* provides five types of protection: Pitch attitude, load factor limitations, high speed, high-AOA and bank angle. Flight mode is operational from take-off to 100 feet above the ground, but can be lost as a result of pilot commands or system failures. Loss of *Normal Law* as a result of a system failure results in *Alternate Law 1* or *2*.

Unlike conventional controls, in *Normal Law flight mode* the sidestick provides a load factor proportional to stick deflection which is independent of aircraft speed. When the stick is neutral and the load factor is 1g the aircraft remains in level flight without the pilot changing the elevator trim. The aircraft also maintains a proper pitch angle once a turn has been established, up to 33° bank. The system prevents further trim up when the angle of attack is excessive, the load factor exceeds 1.3g or when the bank angle exceeds 33°.

Alpha protection (α-Prot) prevents stalling and the effects of windshear. The protection engages when the angle of attack is between α-Prot and α-Max and limits the angle of attack commanded by the pilot's sidestick or, if autopilot is engaged, it disengages the autopilot.

High speed protection will automatically recover from an overspeed. There are two speed limitations for high altitude aircraft, V_{MO} (Velocity Maximum Operational) and M_{MO}

(Mach Maximum Operational) the two speeds are the same at approximately 31,000 feet, below which overspeed is determined by V_{MO} and above 31,000 feet by M_{MO}.

Flare mode

A380 in take off

This mode is automatically engaged when the radar altimeter indicates 100 feet above ground. At 50 feet the aircraft trims the nose slightly down. During the flare, *Normal Law* provides high-AOA protection and bank angle protection. The load factor is permitted to be from 2.5g to -1g, or 2.0g to 0g when slats are extended. Pitch attitude is limited to +30 to -15° which is reduced to 25° as the aircraft slows.

Alternate law

The are four reconfiguration modes for the Airbus fly-by-wire aircraft, two *Alternate Law (1 and 2)*, *Direct Law* and *Mechanical Law*. The ground mode and flare modes for *Alternate Law* are identical to those modes for *Normal Law*.

Alternate law 1 (ALT1) mode combines a *Normal Law* lateral mode with the load factor, bank angle protections retained. High angle of attack protection may be lost and low energy (level flight stall) protection is lost. High speed and high angle of attack protections enter alternative law mode.

ALT1 maybe entered if there are faults in the horizontal stabilizer, an elevator, yaw-damper actuation, slat or flap sensor, or a single air data reference fault.

Alternate law 2 (ALT2) loses *Normal Law* lateral mode (replaced by roll direct mode and yaw alternate mode) along with pitch attitude protection, bank angle protection and low energy protection. Load factor protection is retained. High angle of attack and high speed protections are retained unless the reason for *Alternate 2 Law* mode is the failure of two air-data references or if the two remaining air data references disagree.

ALT2 mode is entered when 2 engines flame out (on dual engine aircraft), faults in two inertial or air-data references, with the autopilot being lost, except with an ADR disagree. This mode may also be entered with an all spoilers fault, certain ailerons fault, or pedal transducers fault.

Direct law

Direct mode (DIR) loses normal lateral mode and all protections, the aircraft assumes *Alternate Law* yaw mode and *Direct Law* roll mode. Elevator can then only be controlled by the manual trim. Control surface motion is directly related to the sidestick and rudder pedal motion.

DIR is entered if there is failure of three inertial reference units or the primary flight computers, faults in two elevators, flame out in two engines (on a two engine aircraft) or when the captain's primary flight computer is inoperable.

Mechanical law

In the *Mechanical Law* back-up mode, pitch is controlled by the mechanical trim system and lateral direction is controlled by the rudder pedals operating the rudder mechanically.

Boeing 777 Primary Flight Control System

The cockpit of the 777 is similar to 747-400, a fly-by-wire control simulating mechanical control

The fly-by-wire electronic flight control system of the Boeing 777 differs from the Airbus EFCS. The design principle is to provide a system that responds similarly to a mechanically controlled system. Because the system is controlled electronically the flight control system can provide flight envelope protection.

The electronic system is subdivided between 2 levels, the 4 actuator control electronics (ACE) and the 3 primary flight computers (PFC). The ACEs control actuators (from those on pilot controls to control surface controls and the PFC). The role of the PFC is to calculate the control laws and provide feedback forces, pilot information and warnings.

Standard Protections and augmentations

The flight control system on the 777 is designed to restrict control authority beyond certain range by increasing the back pressure once the desired limit is reached. This is done via electronically controlled backdrive actuators (controlled by ACE). The protections and augmentations are: bank angle protection, turn compensation, stall protection, over-speed protection, pitch control, stability augmentation and thrust asymmetry compensation. The design philosophy is: "to inform the pilot that the command being given would put the aircraft outside of its normal operating envelope, but the ability to do so is not precluded."

Normal mode

In *Normal mode* the PFCs transmit actuator commands to the ACEs, which convert them into analog servo commands. Full functionality is provided, including all enhanced performance, envelope protection and ride quality features.

Secondary mode

Boeing *Secondary mode* is comparable to the Airbus *Alternate Law*, with the PFCs supplying commands to the ACEs. However, EFCS functionality is reduced, including loss of flight envelope protection. Like the Airbus system, this state is entered when a number of failures occur in the EFCS or interfacing systems (e.g. ADIRU or SAARU).

Direct mode

In *Direct mode* each ACE decodes pilot commands directly from the pilot controller transducers. This mode can be manually or automatically entered. Automatic entry occurs when there is a failure of all PFCs, ACEs, and/or loss of a control data bus.

Chapter 10

Variometer & Machmeter

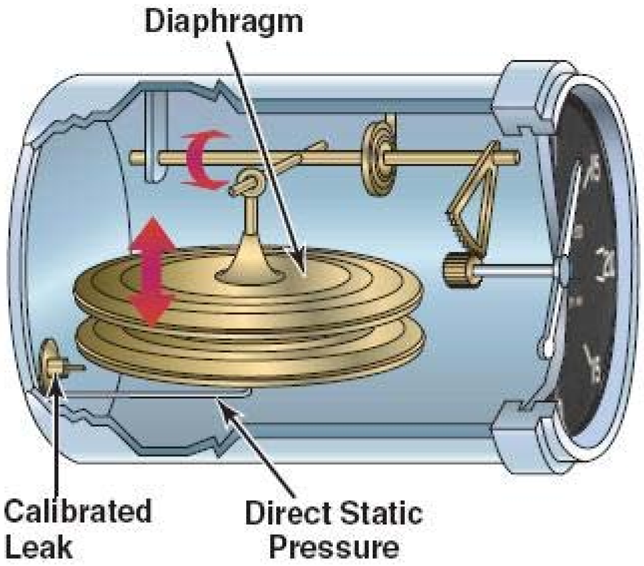
Variometer

A **variometer** (also known as a **rate of climb and descent Indicator (RCDI)**, **rate-of-climb indicator**, **vertical speed indicator (VSI)**, or **vertical velocity indicator (VVI)**) is one of the flight instruments in an aircraft used to inform the pilot of the near instantaneous (rather than averaged) rate of descent or climb. It can be calibrated in knots, feet per minute (101.333 ft/min = 1 kn) or metres per second, depending on country and type of aircraft.

In powered flight the pilot makes frequent use of the **VSI** to ascertain that level flight is being maintained, especially during turning manoeuvres. In gliding, the instrument is used almost continuously during normal flight, often with an audible output, to inform the pilot of rising or sinking air. It is usual for gliders to be equipped with more than one type of variometer. The simpler type does not need an external source of power and can therefore be relied upon to function regardless of whether a battery or power source has been fitted. The electronic type with audio needs a power source to be operative during the flight. The instrument is of little interest during launching and landing, with the exception of aerotow, where the pilot will usually want to avoid releasing in sink.

Variometer for Paragliders, Hang Gliders and Ballooneers

Description



Schematic drawing of the internals of a classic aircraft variometer

Variometers measure the rate of change of altitude by detecting the change in air pressure (static pressure) as altitude changes. A simple variometer can be constructed by adding a large reservoir (a thermos bottle) to augment the storage capacity of a common aircraft rate-of-climb instrument. In its simplest electronic form, the instrument consists of an air bottle connected to the external atmosphere through a sensitive air flow meter. As the aircraft changes altitude, the atmospheric pressure outside the aircraft changes and air flows into or out of the air bottle to equalise the pressure inside the bottle and outside the aircraft. The rate and direction of flowing air is measured by the cooling of one of two self-heating thermistors and the difference between the thermistor resistances will cause a voltage difference; this is amplified and displayed to the pilot. The faster the aircraft is ascending (or descending), the faster the air flows. Air flowing out of the bottle indicates that the altitude of the aircraft is increasing. Air flowing into the bottle indicates that the aircraft is descending.

Newer variometer designs directly measure the static pressure of the atmosphere using a pressure sensor and detect changes in altitude directly from the change in air pressure instead of by measuring air flow. These designs tend to be smaller as they do not need the air bottle. They are more reliable as there is no bottle to be affected by changes in temperature and less chances for leaks to occur in the connecting tubes.

The designs described above, which measure the rate of change of altitude by automatically detecting the change in static pressure as the aircraft changes altitude are referred to as "uncompensated" variometers. The term "vertical speed indicator" or "VSI" is most often used for the instrument when it is installed in a powered aircraft. The term "variometer" is most often used when the instrument is installed in a glider or sailplane.

An "Inertia lead" VSI or ILVSI compensates for relative "*g*" forces experienced in a turn (powered aircraft) and provides appropriate mechanical compensation to remove otherwise erroneous indications of climb or descent.

Panel mounted variometer for gliders

Purpose

Human beings, unlike birds and other flying animals, are not able directly to sense climb and sink rates. Before the invention of the variometer, sailplane pilots found it very hard to soar. Although they could readily detect abrupt *changes* in vertical speed ("in the seat of the pants"), their senses did not allow them to distinguish lift from sink, or strong lift from weak lift. The *actual* climb/sink rate could not even be guessed at, unless there was some clear fixed visual reference nearby. Being near a fixed reference means being near to a hillside, or to the ground. Except when hill-soaring (exploiting the lift close to the up-wind side of a hill), these are generally very unprofitable positions for glider pilots to be in. The most useful forms of lift (thermal and wave lift) are found at higher altitudes and it is very hard for a pilot to detect or exploit them without the use of a variometer. After the variometer was invented in 1929 by Alexander Lippisch and Robert Kronfeld, the sport of gliding moved into a new realm.

The vertical speed indicator from a Robinson R22

Total energy compensation

As the sport developed, however, it was found that these very simple "uncompensated" instruments had their limitations. The information that glider pilots really need to soar is the vertical speed of the glider in isolation of stick thermals, i.e., in isolation of changes in altitude due exclusively to changes in speed.

When the pilot chooses to pull up to enter a thermal or to dive to exit a sink area, an uncompensated variometer will include the change in altitude due to the change in velocity in its read-out, thus marring the airmass' climb or sink rate. Therefore an uncompensated variometer can only accurately indicate the vertical speed of the airmass when flying at constant speed.

The action of diving or pulling up affects the speed of the sailplane. A sailplane can exchange height for speed or speed for height, i.e. potential energy for kinetic energy or kinetic energy for potential energy. In fact, in still air, the sum of potential energy and kinetic energy, i.e., the *Total Energy*, remains constant (neglecting energy loss due to drag), hence the name *Total Energy compensation*.

Most modern sailplanes are equipped with *Total Energy compensated* variometers.

Total energy compensation in theory

While the driving principle is:

1. Potential Energy + Kinetic Energy = Total Energy

the compensation to cancel stick thermals is:

2. Potential Energy **Gained** = Kinetic Energy **Lost** (*stick thermal*)

i.e.:

3. $\Delta E_{pot} = -\Delta E_{kin}$

or

4. $\Delta E_{pot} + \Delta E_{kin} = 0$

Since

5. Potential Energy is proportional to Height ($E_{pot} \propto h$)

and

6. Kinetic Energy is proportional to Velocity squared ($E_{kin} \propto v^2$),

then from (3):

7. $\Delta h \propto -\Delta(v^2)$

where

8. Δh is the compensation to apply to the uncompensated variometer reading.

The Vertical Speed Indicator in this Van's RV-4 light aircraft is on the top row, on the right.

Total energy compensation in practice

In most sailplanes, total energy compensation is achieved by connecting the variometer to the atmosphere via a "total energy probe", that produces vacuum proportional to the square of the glider's air speed - in effect, a negative pitot. Alternatively, the subtraction may be done electronically by the flight computer based on indicated airspeed (pitot).

Very few powered aircraft have total energy variometers. The pilot of a powered aircraft is more interested in the true rate of change of altitude, as he often wants to hold a constant altitude or maintain a steady climb or descent.

The total energy probe used to be shaped as a classical venturi (two small funnels connected back-to-back by their narrow ends), or nowadays the Irving Tube - a slot or pair of holes on the back side of a quarter inch vertical tube. The geometry of the total energy probe is such that air flow generates suction (reduced pressure).

To maximise the precision of this compensation effect, the total energy probe needs to be in undisturbed airflow ahead of the aircraft nose or tail fin (the "Braunschweig tube", the long cantilevered tube with a kink in the end that can be seen projecting from the leading edge of the tail fin on most modern sailplanes.)

Netto variometer

A second type of compensated variometer is the **Netto** or **airmass** variometer. In addition to TE compensation, the Netto variometer adjusts for the intrinsic sink rate of the glider at a given speed (the polar curve) adjusted for the wing loading due to water ballast. The Netto variometer will always read zero in still air. This provides the pilot with the accurate measurement of air mass vertical movement critical for final glides.

The **Relative Netto Variometer** indicates the vertical speed the glider would achieve IF it flies at thermalling speed - independent of current air speed and attitude. This reading is calculated as the Netto reading minus the glider's minimum sink.

When the glider circles to thermal, the pilot needs to know the glider's vertical speed instead of that of the air mass. The **Relative Netto Variometer** (or sometimes the **super Netto**) includes a g-sensor to detect thermalling.

When thermalling, the sensor will detect acceleration (gravity plus centrifugal) above 1 g and tell the relative netto variometer to stop subtracting the sailplane's wing load-adjusted polar sink rate for the duration. Some earlier nettos used a manual switch instead of the g sensor.

Electronic variometers

In modern gliders, most electronic variometers generate a sound whose pitch and rhythm depends on the instrument reading. Typically the audio tone increases in frequency as the variometer shows a higher rate of climb and decreases in frequency towards a deep groan as the variometer shows a faster rate of descent. When the variometer is showing a climb, the tone is often chopped, while during a descent the tone is not chopped and the rate of chopping may be increased as the climb rate increases. The vario is typically silent in still air or in lift which is weaker than the typical sink rate of the glider at minimum sink. This audio signal allows the pilot to concentrate on the external view instead of having to watch the instruments, thus improving safety and also giving the pilot more opportunity to search for promising looking clouds and other signs of lift. A variometer that produces this type of audible tone is known as an "audio variometer".

Advanced electronic variometers in gliders can present other information to the pilot from GPS receivers. The display can thus show the bearing, distance and height required to reach an objective. In cruise mode (used in straight flight), the vario can also give an audible indication of the correct speed to fly depending on whether the air is rising or sinking. The pilot merely has to input the estimated MacCready setting, which is the expected rate of climb in the next acceptable thermal.

There is an increasing trend for advanced variometers in gliders which tend towards flight computers and present other information such as controlled airspace, lists of turnpoints and even collision warnings. Some will also store positional GPS data during the flight for later analysis.

Radio controlled soaring

Variometers are also used in radio controlled gliders. Typically it takes the form of a radio transmitter in the plane, and a receiver held by the pilot on the ground. Depending on the design, the receiver may give the pilot the current altitude of the plane (an altimeter) and some sort of display that indicates if the plane is gaining or losing altitude—often via a tone just like in full scale gliders. Other forms of telemetry may also be provided by the system, giving things such as airspeed and battery voltage. Varios used in radio controlled planes may or may not feature total energy compensation (the better/more expensive ones generally do.)

Variometers are strictly optional for R/C glider use—a skilled pilot can generally determine if their plane is going up or down via visual cues alone, and so the use of a variometer is often seen as a `crutch', as a replacement for skill, and many pilots prefer not to use them at all, as the tone can be distracting, and the (usually small) amount of weight added to the plane does affect performance. The use of variometers is permitted in some R/C soaring contests and prohibited in others.

Perhaps the most popular brands of R/C variometers are the Picolario and the WsTech CS Voice.

Machmeter

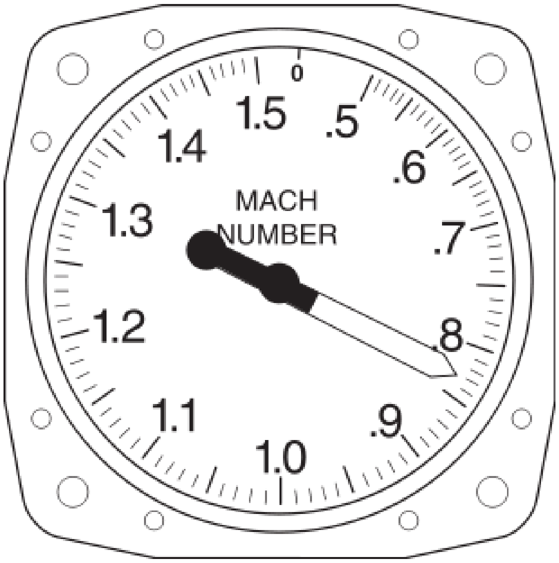


Illustration showing the face of a Machmeter reading a Mach number of 0.83

A **Machmeter** is an aircraft pitot-static system flight instrument that shows the ratio of the true airspeed to the speed of sound, a dimensionless quantity called Mach number. This is shown on a Machmeter as a decimal fraction. An aircraft flying at the speed of sound is flying at a Mach number of one, expressed as *Mach 1*.

Use

As an aircraft in transonic flight approaches the speed of sound, it first reaches its critical mach number, where air flowing over low-pressure areas of its surface locally reaches the speed of sound, forming shock waves. The indicated airspeed for this condition changes with ambient pressure, which in turn changes with altitude. Therefore, indicated airspeed is not entirely adequate to warn the pilot of the impending problems. Mach number is more useful, and most high-speed aircraft are limited to a maximum operating Mach number, also known as $\mathbf{M_{MO}}$.

For example, if the $\mathbf{M_{MO}}$ is Mach 0.83, then at 30,000 feet (9,144 m) where the speed of sound under standard conditions is 590 knots (1,093 km/h; 679 mph), the true airspeed at M_{MO} is 489 knots (906 km/h; 563 mph). The speed of sound increases with air temperature, so at Mach 0.83 at 10,000 feet (3,048 m) where the air is much warmer than at 30,000 feet (9,144 m), the true airspeed at M_{MO} would be 530 knots (982 km/h; 610 mph).

Operation

Some older mechanical Machmeters use an altitude aneroid and an airspeed capsule which together convert pitot-static pressure into Mach number. Modern electronic Machmeters use information from an air data computer system.

Calibration

In subsonic flow the Mach meter can be calibrated according to:

$$M = \sqrt{5\left[\left(\frac{q_c}{P} + 1\right)^{\frac{2}{7}} - 1\right]}$$

where:

M is Mach number
q_c is impact pressure and
P is static pressure
and assuming the ratio of specific heats is 1.4

When a shock wave forms across the pitot tube the required formula is derived from the Rayleigh Supersonic Pitot equation, and is solved iteratively:

$$M = 0.88128485\sqrt{\left[\left(\frac{q_c}{P}+1\right)\left(1-\frac{1}{[7M^2]}\right)^{\frac{5}{2}}\right]}$$

where:

q_c is now impact pressure measured behind a normal shock.

Note that the inputs required are impact pressure (or total pressure) and static pressure. Air temperature input is not required

Chapter 11

Attitude Indicator & Heading Indicator

Attitude Indicator

Attitude indicator with integrated localizer and glideslope indicators, earth below (brown) and sky above (blue) it's perfectly level and not turning, possibly in a slight dive/descent.

An **attitude indicator** (AI), also known as **gyro horizon** or **artificial horizon**, is an instrument used in an aircraft to inform the pilot of the orientation of the aircraft relative to earth. It indicates pitch (fore and aft tilt) and bank or **roll** (side to side tilt) and is a primary instrument for flight in instrument meteorological conditions. Attitude indicators also have significant application under visual flight rules, though some light aircraft do not have them installed.

Use

The essential components of the indicator are:

- "miniature wings", horizontal lines with a dot between them representing the actual wings and nose of the aircraft.

- the center horizon bar separating the two halves of the display, with the top half usually blue in color to represent sky and the bottom half usually dark to represent earth.
- degree marks representing the bank angle. They run along the rim of the dial. On a typical indicator, the first 3 marks on both sides of the center mark are 10 degrees apart. The next is 60 degrees and the mark in the middle of the dial is 90 degrees.

If the symbolic aircraft dot is above the horizon line (blue background) the aircraft is nose up. If the symbolic aircraft dot is below the horizon line (brown background) the aircraft is nose down. When the dot and wings are on the horizon line, the aircraft is in level flight. Because it is the horizon that moves up and down and turns, while the symbolic aircraft is fixed relative to the rest of the instrument panel, trainees get confused; a standard corrective given by flight instructors is "Fly the little airplane, not the horizon."

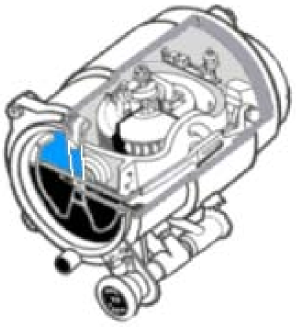
A 45 degree turn is approximated by placing the indicator equidistant between the 30 and 60 degree marks.

The pitch angle is relative to the ground, which is not as helpful as knowing the angle of attack of the wing, a much more critical measure of performance. The pilot must infer the total performance by using other instruments such as the airspeed indicator, altimeter, vertical speed indicator, and power instruments, e.g. an engine tachometer.

Most Russian-built aircraft have a somewhat different design. The background display is colored as in a Western instrument, but moves up and down only to indicate pitch. A symbol representing the aircraft (which is fixed in a Western instrument) rolls left or right to indicate bank angle.

It was proposed that a hybrid version of the Western and Russian artificial horizon systems to be developed that would be more intuitive than either.

Operation



Schematics drawing of the insides of a classic attitude indicator

Attitude indicators use a gyroscope (powered via vacuum pump or electrical motor) to establish an inertial platform. The gyroscope is geared to a display that has two dimensions of freedom, simultaneously displaying pitch and bank. The display may be colored to indicate the horizon as the division between the two colored segments (typically blue for sky and brown for ground), and is intended to be intuitive to use. The actual bank angle is calibrated around the circumference of the instrument. The pitch angle is indicated by a series of calibration lines, each representing 5° or 10° of pitch depending on design. The Artificial Horizon has turning errors when turning through 090 and 270 degrees, and it has no turning errors when turning through 180 and 000 degrees. For example, when turning through 090 degrees the Artificial Horizon will show nose up and bank angle too low. When turning through 180 degrees it will show nose up and bank angle correct.

Some attitude indicators can only tolerate a specific range of bank angles. If the aircraft rolls too steeply — while performing aerobatics, for example — the attitude indicator can "tumble" and become temporarily unusable. For this reason, some attitude indicators are fitted with a "cage" (a device to restore the gyroscope to an erect position). Most modern attitude indicators slowly re-erect back to level after a tumble. Others do not tumble at all.

AHRS are able to provide three-axis information that can be shared with multiple devices in the aircraft, such as "glass cockpit" primary flight displays (PFDs). AHRS have been proven to be highly reliable and are in wide use in commercial and business aircraft. Recent advances in MEMS manufacturing have brought the price of FAA-certified AHRS down to less than $15,000, making them practical for general aviation aircraft.

With most AHRS systems, if an aircraft's AIs have failed there will be a standby AI located in the center of the instrument panel, where other standby basic instruments such as the airspeed indicator and the attitude indicator are also available. These mostly mechanical standby instruments may be available even if the electronic flight instruments fail, though the standby attitude indicator is electrically driven and will, after a short time, fail if electrical power to it fails.

Heading Indicator

A heading indicator in a small aircraft

The **heading indicator** (also called an **HI**) is a flight instrument used in an aircraft to inform the pilot of the aircraft's heading. It is sometimes referred to by its older names, the **directional gyro** or **DG**, and also (UK usage) **direction indicator** or **DI**.

Use

The primary means of establishing the heading in most small aircraft is the magnetic compass, which, however, suffers from several types of errors, including that created by the 'dip' or downward slope of the earth's magnetic field. Dip error causes the magnetic compass to read incorrectly whenever the aircraft is in a bank, or during acceleration, making it difficult to use in any flight condition other than perfectly straight and level. To remedy this, the pilot will typically maneuver the airplane with reference to the heading indicator, as the gyroscopic heading indicator (HI) is unaffected by dip and acceleration errors. The pilot will periodically reset the HI to the heading shown on the magnetic compass.

Operation

The DG works using a gyroscope, tied to the aircraft horizontal, to establish an inertial platform. As such any configuration of the aircraft horizontal which does not match the local earth horizontal results in gimbal error (essentially leading to a variation in the predictable 'apparent' wander known in this instance as drift.) The HI is arranged so that only the horizontal axis is used to drive the display, which consists of a circular compass card calibrated in degrees. The gyroscope is spun either electrically, or using air from a vacuum pump (sometimes a pressure pump in high altitude aircraft) driven from the aircraft's engine. Because the earth rotates (ω, 15° per hour), and because of small accumulated errors caused by friction and imperfect balancing of the gyro, the HI will drift over time, and must be reset from the compass periodically. The apparent drift is predicted by **ω sin Latitude** and will thus be greatest over the poles. Another sort of apparent drift exist in the form of transport wander, where aircraft movement will essentially add or subtract to the effect of the earth's rotation upon a gyroscope. To counter for the effect of earth rate drift a latitude nut can be set (on the ground only) which induces a (hopefully equal and opposite) real wander in the gyroscope. Normal procedure is to realign the direction indicator once each ten to fifteen minutes during routine in-flight checks. Failure to do this is a common source of navigation errors among new pilots.

Variations

Some more expensive heading indicators are 'slaved' to a sensor (called a 'flux gate'). The flux gate continuously senses the earth's magnetic field, and a servo mechanism constantly corrects the heading indicator. These 'slaved gyros' reduce pilot workload by eliminating the need for manual realignment every ten to fifteen minutes.

The prediction of drift in degrees per hour, is as follows:

SOURCE	Drift Rate (Degrees per Hour)	Sign (Northern Hemisphere)	Sign (Southern Hemisphere)
Earth Rate	15 * sin Mean Operating Latitude	- (causing an under-read)	+ (causing an over-read)
Latitude Nut	15 * sin Latitude of Setting	+	-
Transport Wander EAST	(East Groundspeed Component (or sine track angle x groundspeed or change in longitude/flight time in hours) * tan Mean Operating Latitude)/60	-	+
Transport Wander WEST	(West Groundspeed Component (or sine track angle x groundspeed or change in longitude/flight time in hours) * tan Mean Operating Latitude)/60	+	-
Real/Random Wander	As given in the Aircraft Operating Manual	As given	As given

Although it is possible to predict the drift, there will be minor variations from this basic model, accounted for by gimbal error (operating the aircraft away from the local horizontal), among others. A common source of error here is the improper setting of the latitude nut (to the opposite hemisphere for example). The table however allows one to gauge whether an indicator is behaving as expected, and as such, is compared with the realignment corrections made with reference to the magnetic compass. Transport wander is an undesirable consequence of apparent drift.

Chapter 12

Turn Coordinator & Turn Indicator

Turn Coordinator

Image showing the face of a turn coordinator during a standard rate coordinated right turn.

The **turn coordinator** (TC) is a flight instrument which displays to a pilot information about the rate of yaw (turn), roll, and the coordination of the turn. The turn coordinator was developed to replace the older turn and bank indicator, which displayed rate and quality of turn but not rate of roll.

History

The turn coordinator was initially developed as a single instrument used by autopilots to control the roll axis of an aircraft. The turn coordinator without autopilot sensing became popular in general aviation airplanes largely because of the newer look. The turn coordinator rarely appeared in jets or large airplanes.

Use

The indicator includes a miniature airplane as seen from behind. When the miniature airplane wings are level, the yaw rate plus the roll rate is zero. When the wings are not level, the amount and direction of tilt indicates the rate of turn plus the rate of roll. The indicator includes hash marks to indicate "wings level" flight and standard rate turns of 3° per second.

Coordinated turn

Skid

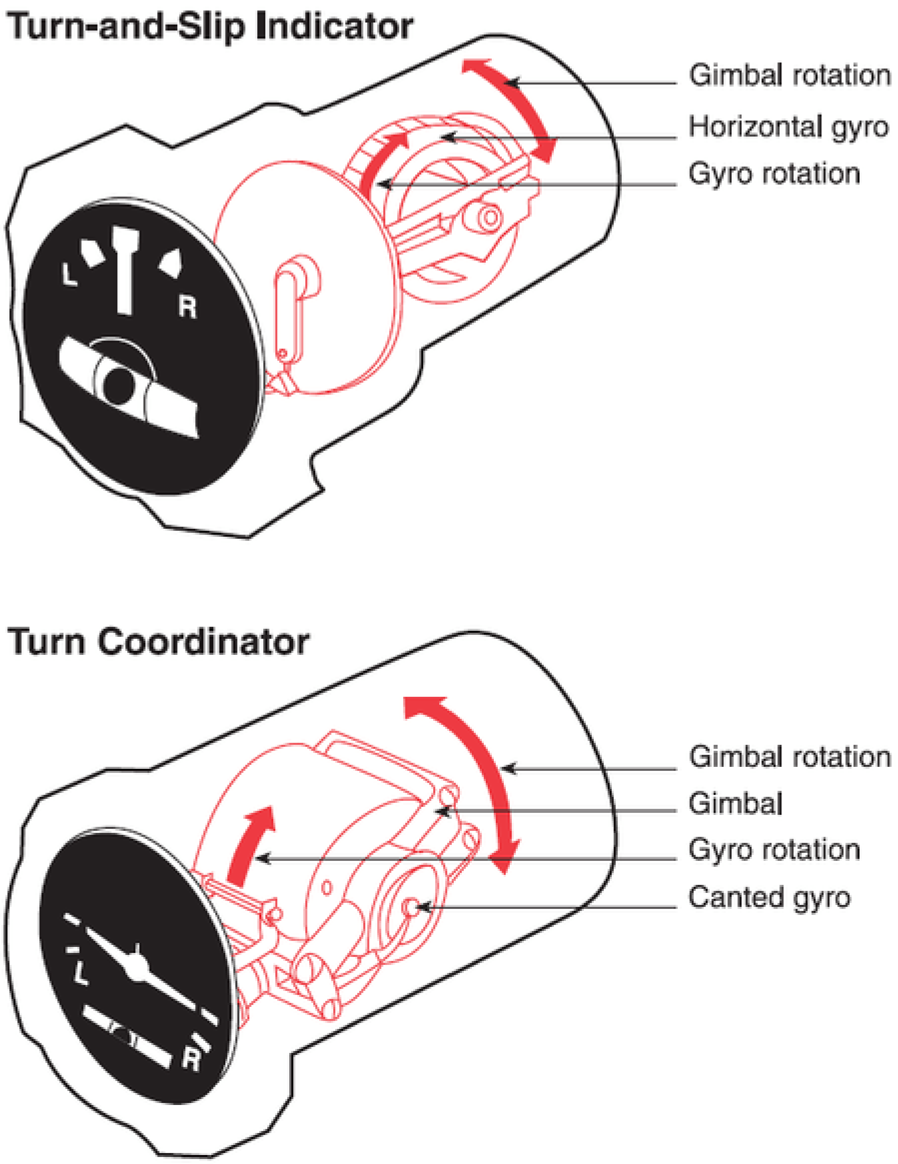
Slip

The ball inside the curved tube indicates the degree of coordination between two primary flight controls (aileron and rudder). The position of the ball indicates coordinated flight, slip or skid.

A pilot corrects for uncoordinated flight with inputs to the rudder. During a coordinated turn, the ball will remain centered. Ball deflection in the direction of the turn indicates insufficient rudder input resulting in a slip. Ball deflection opposite the direction of the turn indicates excessive rudder input resulting in a skid. The mnemonic device *"step on the ball"* aids the pilot to remember which rudder pedal requires additional pressure to return the aircraft to coordinated flight.

Unlike an attitude indicator, the turn coordinator indicates only yaw rate and roll. The attitude indicator indicates pitch and roll. To avoid confusion, some turn coordinators are marked *"No pitch information"* on the face.

Operation



Graphic of a turn and bank indicator and a turn coordinator

The turn coordinator is a gyroscopic instrument. An internal gyroscope spins at approximately 20,000 rpm with the spin axis perpendicular to the longitudinal axis of the airplane and the free axis tilted up 30° from it. As the aircraft rotates about the yaw or roll axis, the principle of gyroscopic inertia causes the gyro to resist the change in its rotational axis about the free axis. This resisting force works against a spring; thus, a slow rate of turn deflects the gyro slightly while a higher rate of roll or yaw deflects it more. The gimbal is linked to the indicator dial on which is the rear view of the miniature airplane.

The quality of turn is indicated by an inclinometer ball located below the miniature airplane. The inclinometer consists of a glass tube filled with kerosene, and a dense ball. The tube is curved such that its center is the lowest point. Normally, the ball will then sit in the center position of the tube, which represents a 'coordinated' turn. This position is marked by two vertical wires on the tube. The ball is said to be 'centered' when it sits perfectly evenly between the two wires.

Variations

The turn coordinator differs from the older turn and bank indicator in that the turn coordinator has the gyro mounted at a 30° tilt. This allows the turn coordinator to respond to roll as well as turn. The TC indication represents a sum of the roll rate and the yaw rate so it responds more quickly at the beginning and end of a turn than a turn and bank indicator. Pilots who are unfamiliar with this principle sometimes have difficulty using the turn coordinator properly, as they may see a roll indication and interpret it as a rate of turn.

The turn coordinator should be used as a performance instrument when the attitude indicator has failed. Called "partial panel" operations, this can be unnecessarily difficult or even impossible if either the pilot does not understand that the instrument is showing roll rates at some times and turn rates at others, or the internal dashpot is worn out. In the latter case the instrument is said to be underdamped; in turbulence it will indicate large full-scale deflections to the left and right, all of which are roll rate responses. In this condition it may not be possible for the pilot to maintain control of the aircraft in partial-panel operations in instrument meteorological conditions. For this and other reasons many highly experienced pilots prefer the "older" turn and bank indicator design.

Turn Indicator

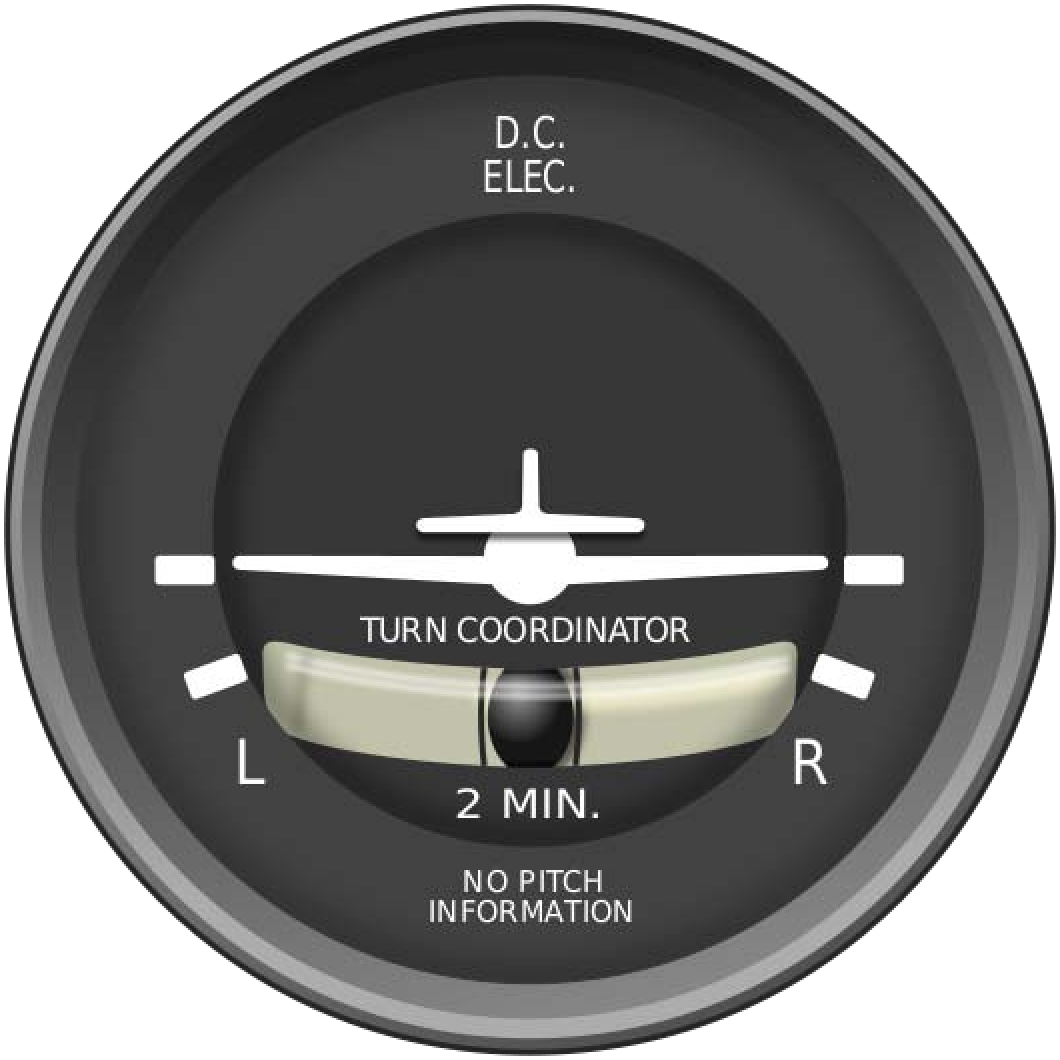
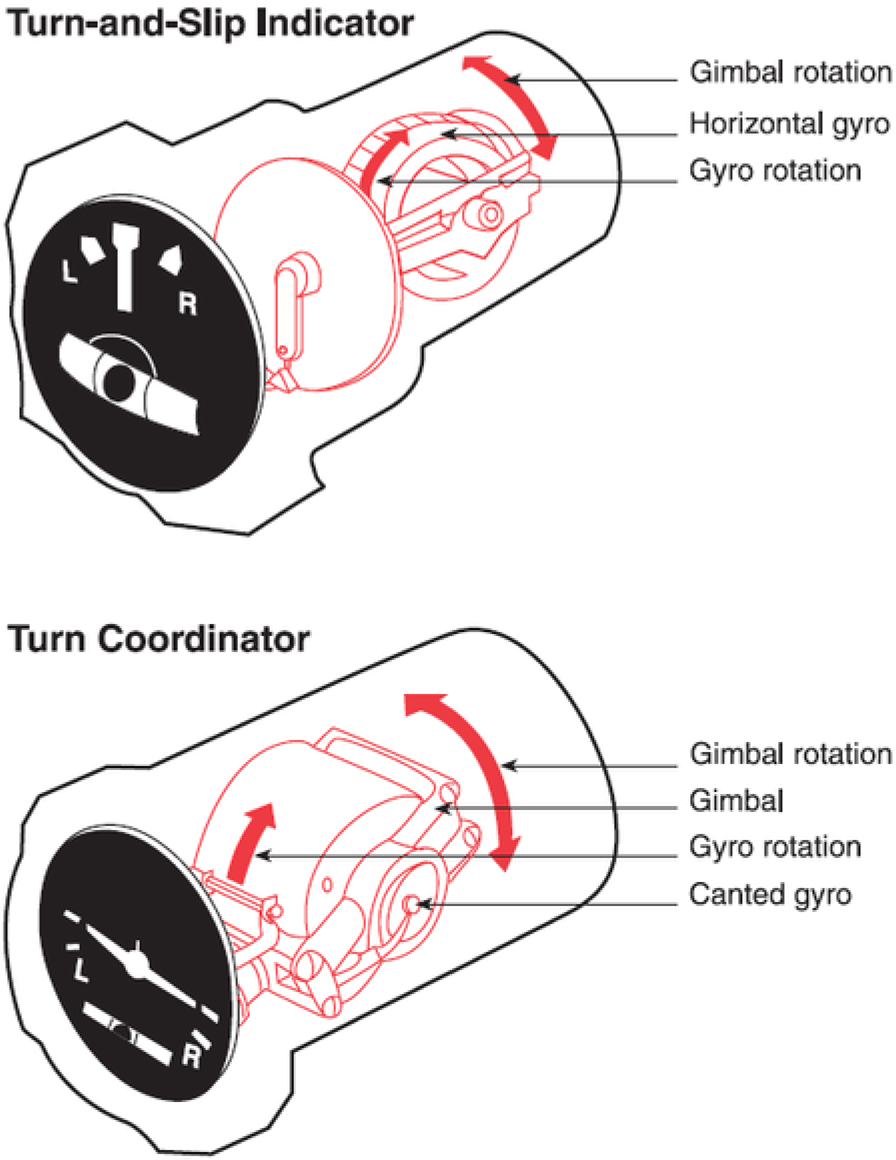


Diagram of a turn indicator

A **turn indicator** is an aircraft flight instrument that shows the rate of turn. It is used by the pilot to maintain control when flying under Instrument flight rules.

Types



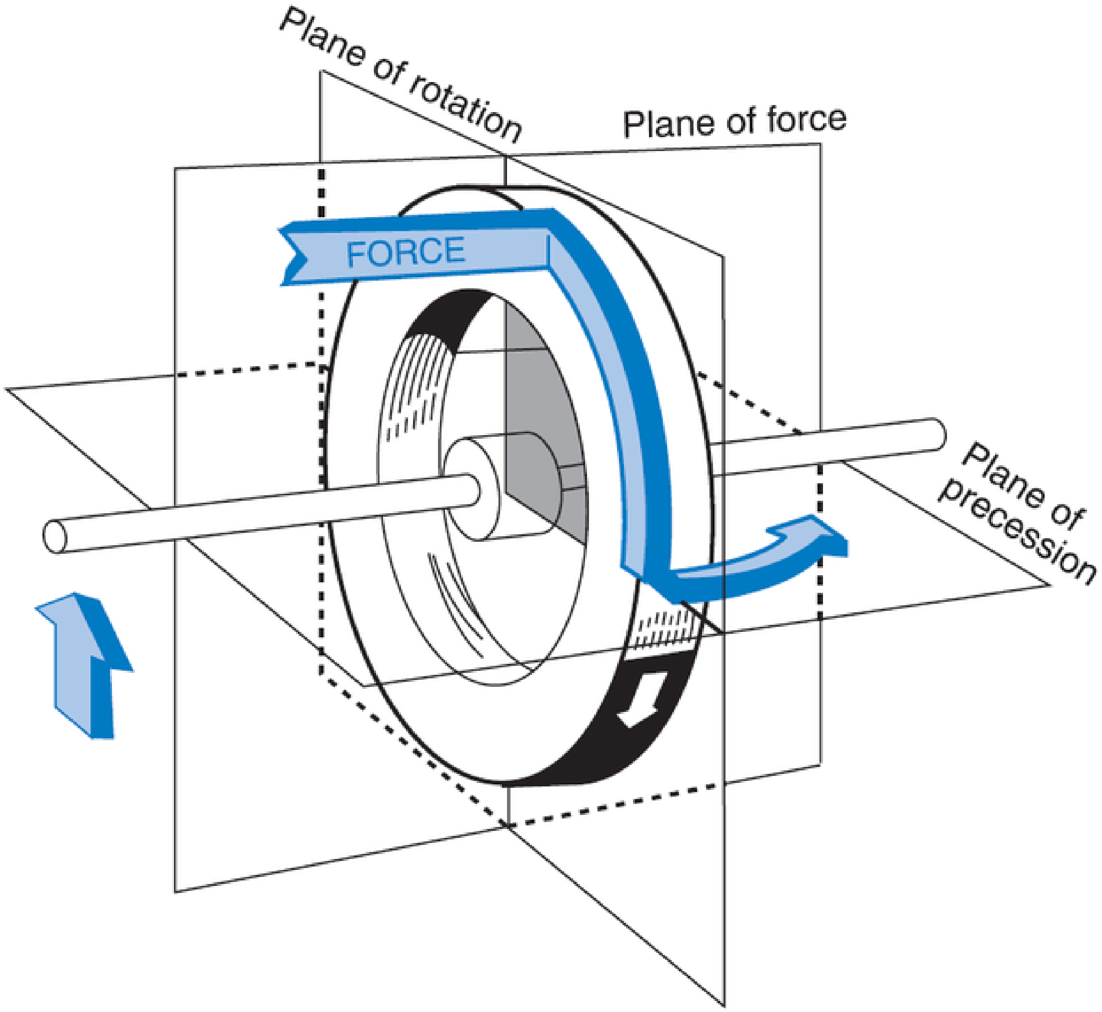
The rate gyro in a turn-and-slip indicator and turn coordinator

The first gyroscopic aircraft instrument was the **turn and bank indicator** (also known as **turn indicator in the needle and ball**, or more recently **turn and slip indicator**).

The major limitation of this instrument is that it senses rotation only about the vertical axis of the aircraft. It tells nothing of the rotation around the longitudinal axis, which in normal flight occurs before the aircraft begins to turn.

The **turn coordinator** was created to overcome this problem by having its gimbal frame angled upward about 30° from the longitudinal axis of the aircraft. This allows it to sense both roll and yaw.

Use and operation



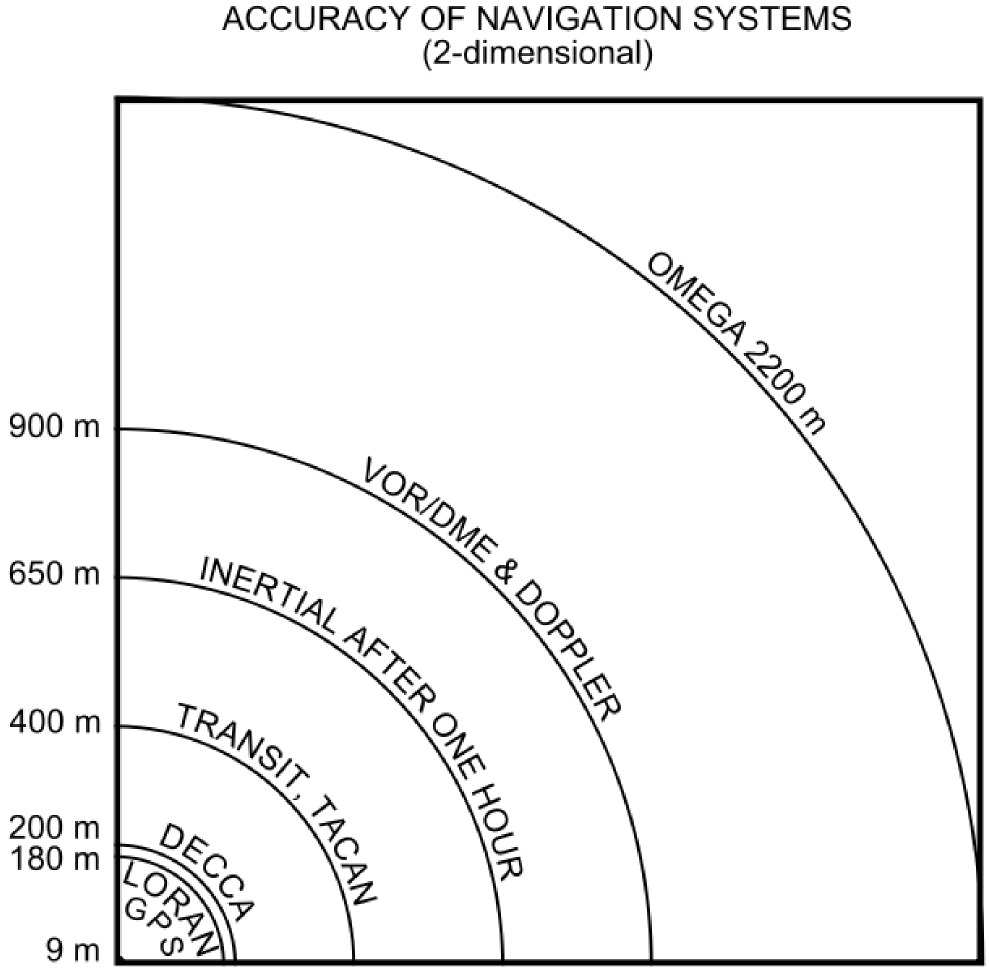
Precession causes a force applied to a spinning wheel to be felt 90° from the point of application in the direction of rotation.

The dial of these instruments is usually marked "2 min. turn", although some turn-and-slip indicators used in faster aircraft like the Concorde are marked "4 min. turn". In gliders, the calibration spring is usually set for a one-minute standard rate turn. A standard-rate turn is being made whenever the needle aligns with a turn marking ("doghouse").

Both instruments incorporate an inclinometer-like indicator, which is either a pendulum or a black glass ball sealed inside a curved glass tube that is partially filled with a liquid similar to the fluid used in a compass. This ball measures the relative strength of the force of gravity and the force of inertia caused by a turn. When the aircraft is flying straight-and-level, there is no inertia acting on the ball, and it remains in the center of the tube between two wires. In a turn made with a bank angle that is too steep, the force of gravity is greater than the inertia and the ball rolls down to the inside of the turn. If the turn is made with too shallow a bank angle, the inertia is greater than gravity and the ball rolls upward to the outside of the turn. The ball does not indicate the amount of bank, neither is it limited to an indication of slip; it only indicates the relationship between the angle of bank and the rate of yaw.

Chapter 13

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 (v, xdv, /dt) 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

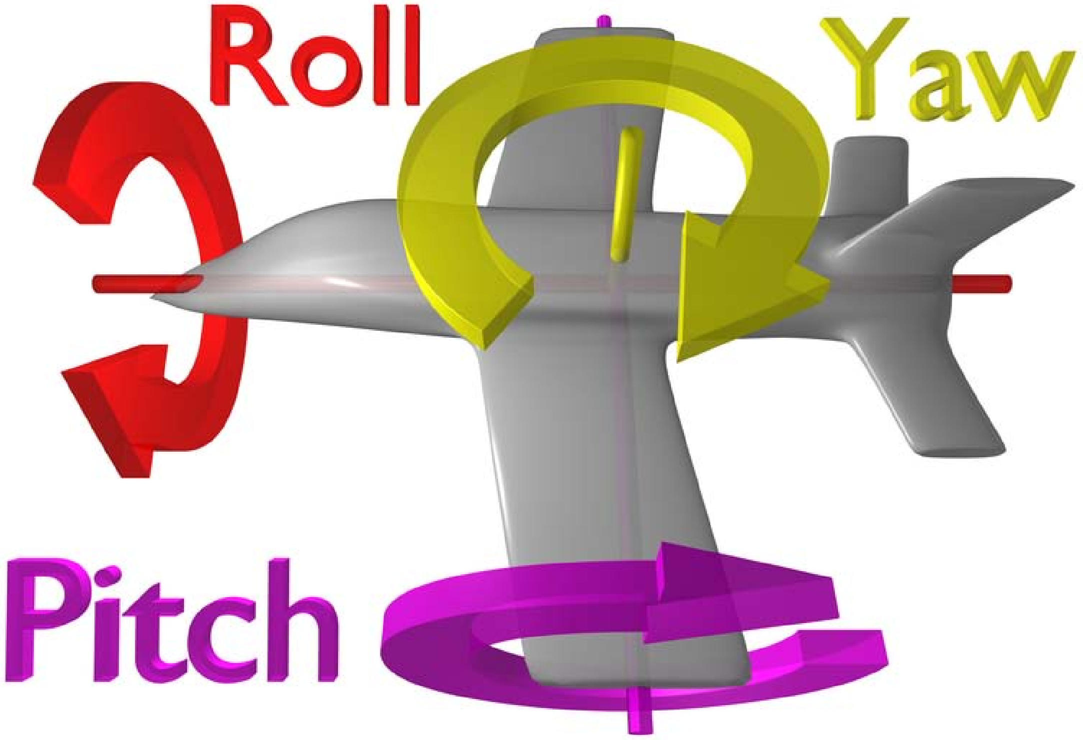
In Feb of 1961 NASA Awarded MIT a contract for preliminary design study of a guidance and navigation system for Apollo. MIT and the Delco Electronics Div. of General Motors Corp. were awarded the joint contract for design and production of the Apollo Guidance and Navigation systems for the Command Module 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.

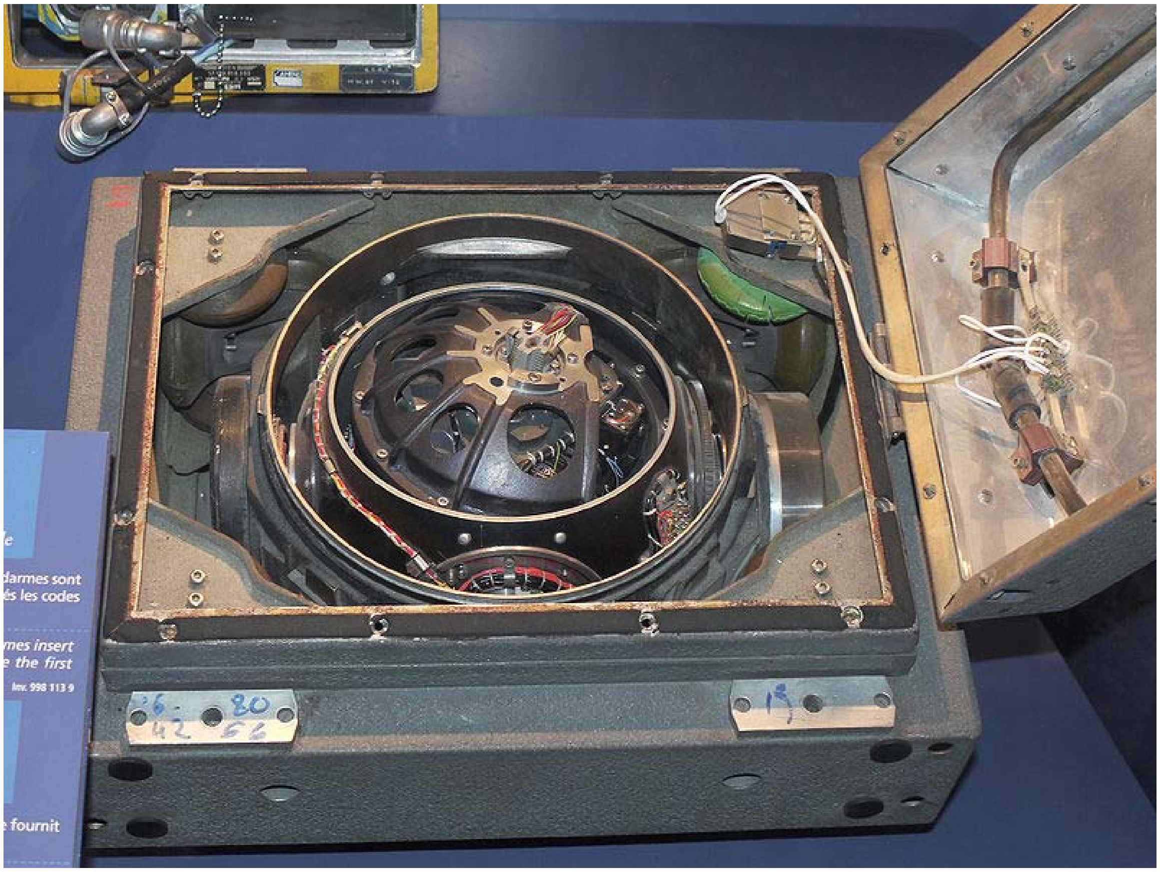
For the space shuttle, an open loop (no feedback) guidance is used to guide the shuttle from lift off until Solid Rocket Booster (SRB) separation. After SRB separation the primary space shuttle guidance is named PEG4 (Powered Explicit Guidance). PEG4 takes into account both the Q system and the predictor-corrector 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 reliablility 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

INSs 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 θ_x, θ_y and θ_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 gimbaled 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 gimbaled 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

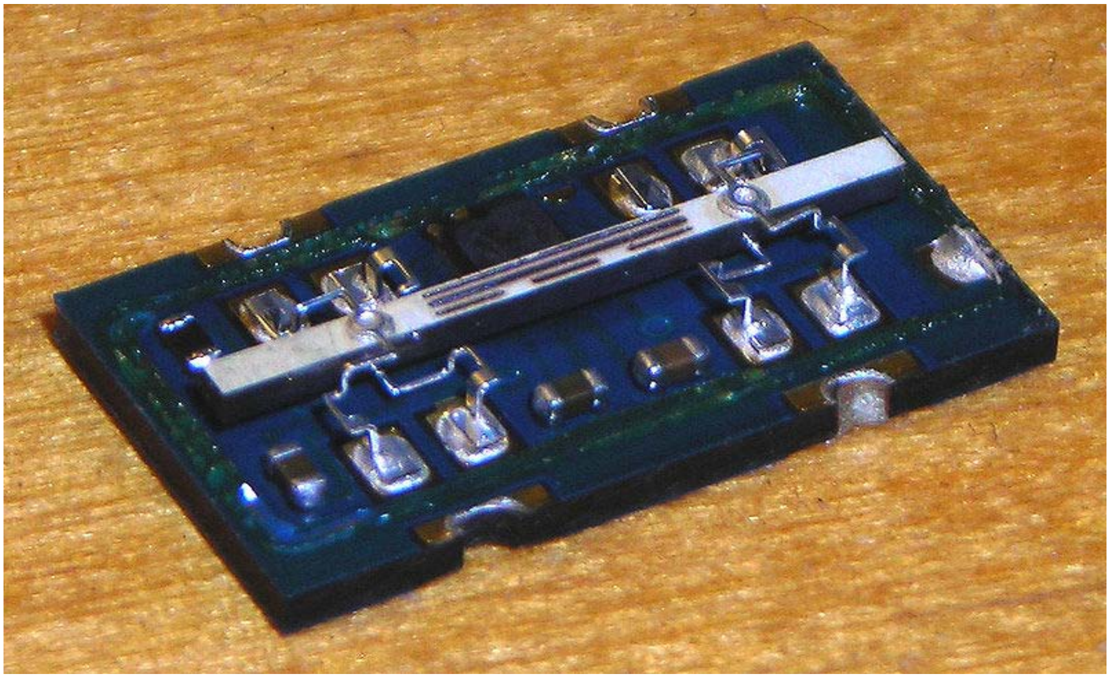
Less-expensive navigation systems, intended for use in automobiles, may use a vibrating structure gyroscope to detect changes in heading, and the odometer pickup to measure distance covered along the vehicle's track. This type of system is much less accurate than a higher-end INS, but it is adequate 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 (≈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

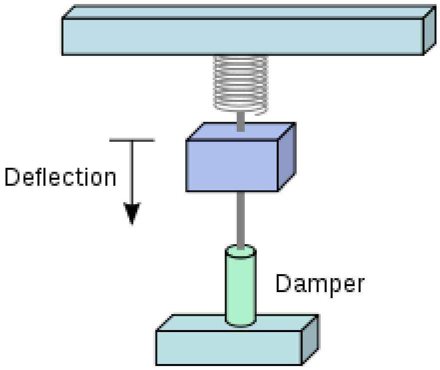
A ring laser gyro splits a beam of laser light into two beams in opposite directions through narrow tunnels in a closed optical circular path around the perimeter of a triangular block of temperature-stable Cervit glass with reflecting mirrors placed in each corner. When the gyro is rotating at some 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



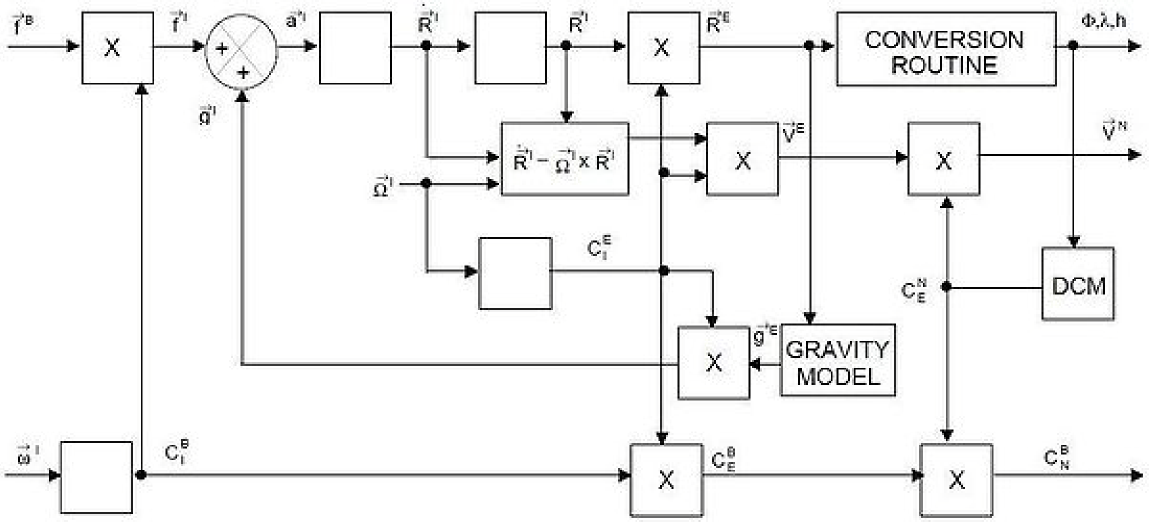
Principle of open loop accelerometer. Acceleration in the upward direction causes the mass to deflect downward.

The basic, open-loop accelerometer consists of a mass attached to a spring. The mass is constrained to move only in-line with the spring. Acceleration causes deflection of the mass and the offset distance is measured. The acceleration is derived from the values of deflection distance, mass, and the spring constant. The system must also be damped to avoid oscillation. A closed-loop accelerometer achieves higher performance by using a feedback loop to cancel the deflection, thus keeping the mass nearly stationary. Whenever the mass deflects, the feedback loop causes an electric coil to apply an equally negative force on the mass, cancelling the motion. Acceleration is derived from the amount of negative force applied. Because the mass barely moves, the non-linearities of the spring and damping system are greatly reduced. In addition, this accelerometer provides for increased bandwidth past the natural frequency of the sensing element.

Both types of accelerometers have been manufactured as integrated micromachinery on silicon chips.

Methodology

In one form, the navigational system of equations acquires linear and angular measurements from the inertial and body frame, respectively and calculates the final attitude and position in the NED frame of reference.



Where: f is specific force, ω is angular rate, a is acceleration, R is position, $\dot{R}$ and V are velocity, Ω is the angular velocity of the earth, g is the acceleration due to gravity, Φ,λ and h are the NED location parameters. Also, super/subscripts of E, I and B are representing variables in the Earth centered, Inertial or Body reference frame, respectively and C is a transformation of reference frames.

Chapter 14

Glass Cockpit

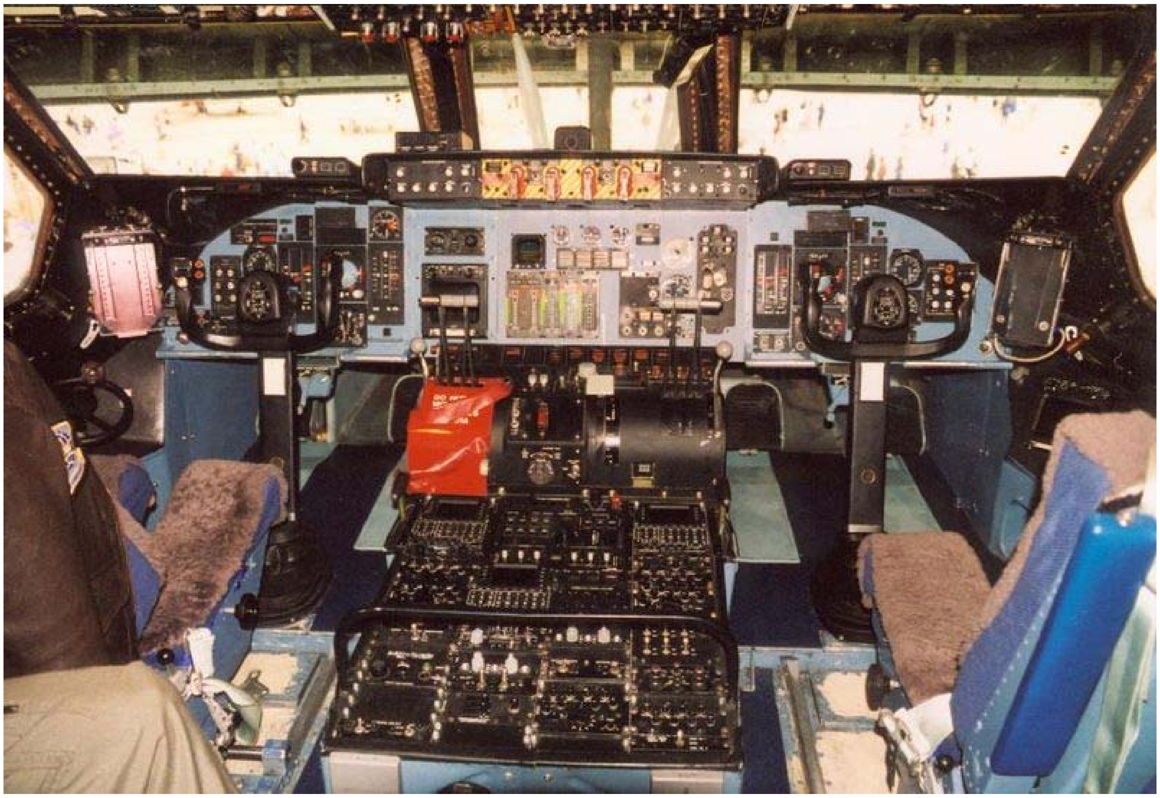
The Airbus A380 glass cockpit featuring "pull out keyboards and 2 wide computer screen on the sides for pilots".

A **glass cockpit** is an aircraft cockpit that features electronic instrument displays. Where a traditional cockpit relies on numerous mechanical gauges to display information, a glass cockpit uses several displays driven by flight management systems, that can be adjusted to display flight information as needed. This simplifies aircraft operation and navigation and allows pilots to focus only on the most pertinent information. They are also popular with airline companies as they usually eliminate the need for a flight engineer. In recent years the technology has become widely available in small aircraft.

As aircraft displays have modernized, the sensors that feed them have modernized as well. Traditional gyroscopic flight instruments have been replaced by electronic Attitude and Heading Reference Systems (AHRS) and Air Data Computers (ADCs), improving reliability and reducing cost and maintenance. GPS receivers are usually integrated into glass cockpits.

Early glass cockpits, found in the McDonnell Douglas MD-80/90, Boeing 737 Classic, 757 and 767-200/-300, and in the Airbus A300-600 and A310, used Electronic Flight Instrument Systems (EFIS) to display attitude and navigational information only, with traditional mechanical gauges retained for airspeed, altitude and vertical speed. Later glass cockpits, found in the Boeing 737NG, 747-400, 767-400, 777, A320 and later Airbuses, Ilyushin Il-96 and Tupolev Tu-204 have completely replaced the mechanical gauges and warning lights in previous generations of aircraft.

History



Instrument panel of a C-5A

New instrument panel for C-5 as part of AMP program

Prior to the 1970s, air transport operations were not considered sufficiently demanding to require advanced equipment like electronic flight displays. Also, computer technology was not at a level where sufficiently light and powerful circuits were available. The increasing complexity of transport aircraft, the advent of digital systems and the growing air traffic congestion around airports began to change that.

The average transport aircraft in the mid-1970s had more than one hundred cockpit instruments and controls, and the primary flight instruments were already crowded with indicators, crossbars, and symbols, and the growing number of cockpit elements were competing for cockpit space and pilot attention. As a result, NASA conducted research on displays that could process the raw aircraft system and flight data into an integrated, easily understood picture of the flight situation, culminating in a series of flights demonstrating a full glass cockpit system.

The success of the NASA-led glass cockpit work is reflected in the total acceptance of electronic flight displays beginning with the introduction of the MD-80 in 1979. Airlines and their passengers alike have benefited. The safety and efficiency of flights has been increased with improved pilot understanding of the aircraft's situation relative to its environment (or "situational awareness").

By the end of the 1990s, Liquid crystal display (LCD) panels were increasingly favored among aircraft manufacturers because of their efficiency, reliability and legibility. Earlier LCD panels suffered from poor legibility at some viewing angles and poor response times, making them unsuitable for aviation. Modern aircraft such as the Boeing 737 Next Generation, 777, 717, 747-400ER, 767-400ER, 747-8, and 787, Airbus A320 family (later versions), A330 (later versions), A340-500/600, A340-300 (later versions), A380 and A350 are fitted with glass cockpits consisting of LCD units.

The glass cockpit has become standard equipment in airliners, business jets, and military aircraft. It was even fitted into NASA's Space Shuttle orbiters *Atlantis*, *Columbia*, *Discovery*, and *Endeavour*, and the current Russian Soyuz TMA model spacecraft that was launched in 2002. By the end of the century glass cockpits began appearing in general aviation aircraft as well. By 2005, even basic trainers like the Piper Cherokee and Cessna 172 were shipping with glass cockpits as options (which nearly all customers chose), and many modern aircraft such as the Diamond Aircraft twin-engine travel and training aircraft DA42, and Cirrus Design SR20 and SR22 are available with glass cockpit only.

Usage

In commercial aviation

The Space Shuttle glass cockpit

Unlike the previous era of glass cockpits—where designers merely copied the look and feel of conventional electromechanical instruments onto cathode ray tubes—the new displays represent a true departure. They look and behave very similarly to other computers, with windows and data that can be manipulated with point-and-click devices. They also add terrain, approach charts, weather, vertical displays, and 3D navigation images.

The improved concepts enable aircraft makers to customize cockpits to a greater degree than previously. All of the manufacturers involved have chosen to do so in one way or another—such as using a trackball, thumb pad or joystick as a pilot-input device in a computer-style environment. Many of the modifications offered by the aircraft manufacturers improve situational awareness and customize the human-machine interface to increase safety.

Modern glass cockpits might include Synthetic Vision (SVS) or Enhanced Vision systems (EVS). Synthetic Vision systems display a realistic 3D depiction of the outside world (similar to a flight simulator), based on a database of terrain and geophysical features in conjunction with the attitude and position information gathered from the aircraft navigational systems. Enhanced Vision systems add real-time information from external sensors, such as an infrared camera.

All new airliners such as the Airbus A380, Boeing 787 and private jets such as Bombardier Global Express and Learjet use glass cockpits.

In general aviation

Garmin G1000 displays in a Cessna 182

Certain general aviation aircraft, such as the 4-seat Diamond Aircraft DA40, DA42 and DA50 and the 4-seat Cirrus Design SR20 and SR22, are available with glass cockpits. Systems such as the Garmin G1000 are now available on many new GA aircraft, including the classic Cessna 172. Some small aircraft can also be modified post-production to replace steam gauges.

Glass cockpits are also popular as a retrofit for older private jets and turboprops such as Dassault Falcons, Raytheon Hawkers, Bombardier Challenger, Cessna Citations, Gulfstreams, King Airs, Learjets, Astras and many others. Aviation service companies work closely with equipment manufacturers to address the needs of the owners of these aircraft.

Safety

As aircraft operation becomes more dependent on glass cockpit systems, flight crews must be trained to deal with possible failures. In one glass-cockpit aircraft, the Airbus A320, fifty incidents of glass-cockpit blackout have occurred. On 25 January 2008

United Airlines Flight 731 experienced a serious glass-cockpit blackout, losing half of the ECAM displays as well as all radios, transponders, TCAS, and attitude indicators. Partially due to good weather and daylight conditions, the pilots were able to land successfully at Newark Airport without radio contact. Airbus has offered an optional fix, which the US NTSB has suggested to the US FAA as mandatory, but the FAA has yet to make it a requirement. A preliminary NTSB factsheet is available. In 2010, the NTSB published a study done on 8,000 general aviation light aircraft. The study found that, although aircraft equipped with glass cockpits had a lower overall accident rate, they also had a larger chance of being involved in a fatal accident. The NTSB Chairman said in response to the study,

"Training is clearly one of the key components to reducing the accident rate of light planes equipped with glass cockpits, and this study clearly demonstrates the life and death importance of appropriate training on these complex systems... While the technological innovations and flight management tools that glass cockpit equipped airplanes bring to the general aviation community should reduce the number of fatal accidents, we have not – unfortunately – seen that happen."

Chapter 15

Autopilot

Autopilot panel of an older Boeing 747 aircraft

An **autopilot** is a mechanical, electrical, or hydraulic system used to guide a vehicle without assistance from a human being. An autopilot can refer specifically to aircraft, self-steering gear for boats, or auto guidance of space craft and missiles. The autopilot of an aircraft is sometimes referred to as "George".

First autopilots

In the early days of aviation, aircraft required the continuous attention of a pilot in order to fly safely. As aircraft range increased allowing flights of many hours, the constant attention led to serious fatigue. An autopilot is designed to perform some of the tasks of the pilot.

The first aircraft autopilot was developed by Sperry Corporation in 1912. The autopilot connected a gyroscopic Heading indicator and attitude indicator to hydraulically operated elevators and rudder (ailerons were not connected as wing dihedral was counted upon to produce the necessary roll stability.) It permitted the aircraft to fly straight and level on a compass course without a pilot's attention, greatly reducing the pilot's workload.

Lawrence Sperry (the son of famous inventor Elmer Sperry) demonstrated it two years later in 1914 at an aviation safety contest held in Paris. At the contest, Lawrence Sperry demonstrated the credibility of the invention were shown by flying the aircraft with his hands away from the controls and visible to onlookers of the contest. This autopilot system was also capable of performing take-off and landing, and the French military command showed immediate interest in the autopilot system. Wiley Post used a Sperry autopilot system to fly alone around the world in less than eight days in 1933.

Further development of the autopilot were performed, such as improved control algorithms and hydraulic servomechanisms. Also, inclusion of additional instrumentation such as the radio-navigation aids made it possible to fly during night and in bad weather. In 1947 a US Air Force C-53 made a transatlantic flight, including takeoff and landing, completely under the control of an autopilot.

In the early 1920s, the Standard Oil tanker *J.A Moffet* became the first ship to use an autopilot.

Modern autopilots

Not all of the passenger aircraft flying today have an autopilot system. Older and smaller general aviation aircraft especially are still hand-flown, while small airliners with fewer than twenty seats may also be without an autopilot as they are used on short-duration flights with two pilots. The installation of autopilots in aircraft with more than twenty seats is generally made mandatory by international aviation regulations. There are three levels of control in autopilots for smaller aircraft. A single-axis autopilot controls an aircraft in the roll axis only; such autopilots are also known colloquially as "wing levellers", reflecting their limitations. A two-axis autopilot controls an aircraft in the pitch axis as well as roll, and may be little more than a "wing leveller" with limited pitch-oscillation-correcting ability; or it may receive inputs from on-board radio navigation systems to provide true automatic flight guidance once the aircraft has taken off until shortly before landing; or its capabilities may lie somewhere between these two extremes. A three-axis autopilot adds control in the yaw axis and is not required in many small aircraft.

Autopilots in modern complex aircraft are three-axis and generally divide a flight into taxi, takeoff, ascent, level, descent, approach and landing phases. Autopilots exist that automate all of these flight phases except the taxiing. An autopilot-controlled landing on a runway and controlling the aircraft on rollout (i.e. keeping it on the centre of the runway) is known as a CAT IIIb landing or Autoland, available on many major airports' runways today, especially at airports subject to adverse weather phenomena such as fog. Landing, rollout and taxi control to the aircraft parking position is known as CAT IIIc. This is not used to date but may be used in the future. An autopilot is often an integral component of a Flight Management System.

Modern autopilots use computer software to control the aircraft. The software reads the aircraft's current position, and controls a Flight Control System to guide the aircraft. In such a system, besides classic flight controls, many autopilots incorporate thrust control capabilities that can control throttles to optimize the air-speed, and move fuel to different tanks to balance the aircraft in an optimal attitude in the air. Although autopilots handle new or dangerous situations inflexibly, they generally fly an aircraft with a lower fuel-consumption than a human pilot.

The autopilot in a modern large aircraft typically reads its position and the aircraft's attitude from an inertial guidance system. Inertial guidance systems accumulate errors over time. They will incorporate error reduction systems such as the carousel system that rotates once a minute so that any errors are dissipated in different directions and have an overall nulling effect. Error in gyroscopes is known as drift. This is due to physical properties within the system, be it mechanical or laser guided, that corrupt positional data. The disagreements between the two are resolved with digital signal processing, most often a six-dimensional Kalman filter. The six dimensions are usually roll, pitch, yaw, altitude, latitude and longitude. Aircraft may fly routes that have a required performance factor, therefore the amount of error or actual performance factor must be monitored in order to fly those particular routes. The longer the flight the more error accumulates within the system. Radio aids such as DME, DME updates and GPS may be used to correct the aircraft position.

Computer system details

The hardware of an autopilot varies from implementation to implementation, but is generally designed with redundancy and reliability as foremost considerations. For example, the Rockwell Collins AFDS-770 Autopilot Flight Director System used on the Boeing 777, uses triplicated FCP-2002 microprocessors which have been formally verified and are fabricated in a radiation resistant process.

Software and hardware in an autopilot is tightly controlled, and extensive test procedures are put in place.

Some autopilots also use design diversity. In this safety feature, critical software processes will not only run on separate computers and possibly even using different architectures, but each computer will run software created by different engineering teams, often being programmed in different programming languages. It is generally considered unlikely that different engineering teams will make the same mistakes. As the software becomes more expensive and complex, design diversity is becoming less common because fewer engineering companies can afford it. The flight control computers on the Space Shuttle uses this design: there are five computers, four of which redundantly run identical software, and a fifth backup running software that was developed independently. The software on the fifth system provides only the basic functions needed to fly the Shuttle, further reducing any possible commonality with the software running on the four primary systems.

Categories

Instrument-aided landings are defined in categories by the International Civil Aviation Organization. These are dependent upon the required visibility level and the degree to which the landing can be conducted automatically without input by the pilot.

CAT I - This category permits pilots to land with a decision height of 200 ft (61 m) and a forward visibility or Runway Visual Range (RVR) of 550 m. Simplex autopilots are sufficient.

CAT II - This category permits pilots to land with a decision height between 200 ft and 100 ft (≈ 30 m) and a RVR of 300 m. Autopilots have a fail passive requirement.

CAT IIIa -This category permits pilots to land with a decision height as low as 50 ft (15 m) and a RVR of 200 m. It needs a fail-passive autopilot. There must be only a 10^{-6} probability of landing outside the prescribed area.

CAT IIIb - As IIIa but with the addition of automatic roll out after touchdown incorporated with the pilot taking control some distance along the runway. This category permits pilots to land with a decision height less than 50 feet or no decision height and a forward visibility of 250 ft (76 m, compare this to aircraft size, some of which are now over 70 m long) or 300 ft (91 m) in the United States. For a landing-without-decision aid, a fail-operational autopilot is needed. For this category some form of runway guidance system is needed: at least fail-passive but it needs to be fail-operational for landing without decision height or for RVR below 100 m.

CAT IIIc - As IIIb but without decision height or visibility minimums, also known as "zero-zero".

Fail-passive autopilot: in case of failure, the aircraft stays in a controllable position and the pilot can take control of it to go around or finish landing. It is usually a dual-channel system.

Fail-operational autopilot: in case of a failure below alert height, the approach, flare and landing can still be completed automatically. It is usually a triple-channel system or dual-dual system.

Radio-controlled models

In radio-controlled modelling, and especially RC aircraft and helicopters, an autopilot is usually a set of extra hardware and software that deals with pre-programming the model's flight.

Chapter 16

Electronic Flight Instrument System

EFIS on an Airbus A380

EFIS on an Eclipse 500

Garmin G1000 on a Diamond DA42

An **Electronic Flight Instrument System** (EFIS) is a flight deck instrument display system in which the display technology used is electronic rather than electromechanical. EFIS normally consists of a primary flight display (PFD), multi-function display (MFD) and Engine Indicating and Crew Alerting System (EICAS) display. Although cathode ray tube (CRT) displays were used at first, liquid crystal displays (LCD) are now more common.

The complex electromechanical attitude director indicator (ADI) and horizontal situation indicator (HSI) were the first candidates for replacement by EFIS. However, there are now few flight deck instruments for which no electronic display is available.

Overview

EFIS installations vary greatly. A light aircraft might be equipped with one display unit, on which are displayed flight and navigation data. A wide-body aircraft is likely to have six or more display units.

Typical EFIS displays and controls can be seen at this B737 technical information web site. The equivalent electromechanical instruments are also shown here.

An EFIS installation will have the following components:

- Displays
- Controls
- Data processors

A basic EFIS might have all these facilities in the one unit.

Display units

Primary Flight Display (PFD)

On the flight deck, the display units are the most obvious parts of an EFIS system, and are the features which give rise to the name "glass cockpit". The display unit taking the place of the ADI is called the primary flight display (PFD). If a separate display replaces the HSI, it is called the navigation display. The PFD displays all information critical to flight, including calibrated airspeed, altitude, heading, attitude, vertical speed and yaw. The PFD is designed to improve a pilot's situational awareness by integrating this information into a single display instead of six different analog instruments, reducing the amount of time necessary to monitor the instruments. PFDs also increase situational awareness by alerting the aircrew to unusual or potentially hazardous conditions — for example, low airspeed, high rate of descent — by changing the color or shape of the display or by providing audio alerts.

The names Electronic Attitude Director Indicator and Electronic Horizontal Situation Indicator are used by some manufacturers. However, a simulated ADI is only the centerpiece of the PFD. Additional information is both superimposed on and arranged around this graphic.

Multi-function displays can render a separate navigation display unnecessary. Another option is to use one large screen to show both the PFD and navigation display.

The PFD and navigation display (and multi-function display, where fitted) are often physically identical. The information displayed is determined by the system interfaces where the display units are fitted. Thus, spares holding is simplified: the one display unit can be fitted in any position.

LCD units generate less heat than CRTs; an advantage in a congested instrument panel. They are also lighter, and occupy a lower volume.

Multi-Function Display (MFD) / Navigation Display (ND)

The MFD (Multi-Function Display) displays navigational and weather information from multiple systems. MFDs are most frequently designed as "chart-centric", where the aircrew can overlay different information over a map or chart. Examples of MFD overlay information include the aircraft's current route plan, weather information from either onboard radar or lightning detection sensors or ground-based sensors, e.g., NEXRAD, restricted airspace and aircraft traffic. The MFD can also be used to view other non-overlay type of data (e.g., current route plan) and calculated overlay-type data, e.g., the glide radius of the aircraft, given current location over terrain, winds, and aircraft speed and altitude.

MFDs can also display information about aircraft systems, such as fuel and electrical systems. As with the PFD, the MFD can change the color or shape of the data to alert the aircrew to hazardous situations.

Engine Indications and Crew Alerting System (EICAS) / Electronic Centralized Aircraft Monitoring (ECAM)

EICAS (Engine Indications and Crew Alerting System) displays information about the aircraft's systems, including its fuel, electrical and propulsion systems (engines). EICAS displays are often designed to mimic traditional round gauges while also supplying digital readouts of the parameters.

EICAS improves situational awareness by allowing the aircrew to view complex information in a graphical format and also by alerting the crew to unusual or hazardous situations. For example, if an engine begins to lose oil pressure, the EICAS might sound an alert, switch the display to the page with the oil system information and outline the low oil pressure data with a red box. Unlike traditional round gauges, many levels of warnings and alarms can be set. Proper care must be taken when designing EICAS to ensure that the aircrew are always provided with the most important information and not overloaded with warnings or alarms.

ECAM is a similar system used by Airbus, which in addition to providing EICAS functions also recommend remedial action.

Control panels

The pilots are provided with controls, with which they select display range and mode (for example, map or compass rose) and enter data (such as selected heading).

Where inputs by the pilot are used by other equipment, data buses broadcast the pilot's selections so that the pilot only needs to enter the selection once. For example, the pilot selects the desired level-off altitude on a control unit. The EFIS repeats this selected altitude on the PFD and by comparing it with the actual altitude (from the air data computer) generates an altitude error display. This same altitude selection is used by the automatic flight control system to level off, and by the altitude alerting system to provide appropriate warnings.

Data processors

The EFIS visual display is produced by the symbol generator. This receives data inputs from the pilot, signals from sensors, and EFIS format selections made by the pilot. The symbol generator can go by other names, such as display processing computer, display electronics unit, etc.

The symbol generator does more than generate symbols. It has (at the least) monitoring facilities, a graphics generator and a display driver. Inputs from sensors and controls arrive via data buses, and are checked for validity. The required computations are performed, and the graphics generator and display driver produce the inputs to the display units.

Monitoring

Like personal computers, flight instrument systems need power-on-self-test facilities and continuous self-monitoring. Flight instrument systems, however, need additional monitoring capabilities:

- Input validation — verify that each sensor is providing valid data
- Data comparison — cross check inputs from duplicated sensors
- Display monitoring — detect failures within the instrument system

Former practice

Traditional (electromechanical) displays were equipped with synchro mechanisms which would transmit, to an instrument comparator, the pitch, roll and heading that were actually being shown on the Captain's and First Officer's instruments. The comparator warned of excessive differences between the Captain and First Officer displays. Even a fault as far *downstream* as a jam in, say, the roll mechanism of an ADI would trigger a comparator warning.

The instrument comparator thus provided both comparator monitoring and display monitoring.

Comparator monitoring

With EFIS, the comparator function is as simple as ever. Is the roll data (bank angle) from sensor 1 the same as the roll data from sensor 2? If not, put a warning caption (such as **CHECK ROLL**) on both PFDs. Comparison monitors will give warnings for airspeeds, pitch, roll and altitude indications. The more advanced EFIS systems, more comparator monitors will be enabled.

Display monitoring

An EFIS display allows no easy re-transmission of what is shown on the display. What is required is a new approach to display monitoring that provides safety equivalent to that of the traditional system. One solution is to keep the display unit as simple as possible, so that it is unable to introduce errors. The display unit either works or does not work. A failure is always obvious, never insidious. Now the monitoring function can be shifted *upstream* to the output of the symbol generator.

In this technique, each symbol generator contains two display monitoring channels. One channel, the internal, samples the output from its own symbol generator to the display unit and computes, for example, what roll attitude should produce that indication. This computed roll attitude is then compared with the roll attitude input to the symbol generator from the INS or AHRS. Any difference has probably been introduced by faulty processing, and triggers a warning on the relevant display.

The external monitoring channel carries out the same check on the symbol generator on the other side of the flight deck: the Captain's symbol generator checks the First Officer's, the First Officer's checks the Captain's. Whichever symbol generator detects a fault, puts up a warning on its own display.

The external monitoring channel also checks sensor inputs (to the symbol generator) for reasonableness. A spurious input, such as a radio height greater than the radio altimeter's maximum, results in a warning.

Human factors

Clutter

At various stages of a flight, a pilot uses different combinations of data. Ideally, only the data in use would be displayed, but an electromechanical instrument has to be in view all the time. To improve display clarity, intricate mechanisms are used on ADIs and HSIs to remove superfluous indications temporarily, e.g., removing the glide slope scale when it is not being used.

With EFIS, some indications, e.g., engine vibration, might not be displayed under normal conditions. If limits are exceeded, then the reading will be displayed. In similar fashion, EFIS is programmed to show the glideslope scale and pointer only during an ILS approach.

If a failure of input data is detected, electromechanical instruments add yet another indicator to the display. Typically, a bar is dropped across the erroneous data. EFIS, on the other hand, removes invalid data from the display and substitutes an appropriate warning.

A de-clutter mode is activated automatically when the pilot's attention is required to be focused on a specific item. For example, if the aircraft is pitched up or down above a specified pitch, usually 30 to 60 degrees, the attitude indicator will de-clutter items from sight until the pitch is brought to an acceptable level. This allows the pilot to focus on the most important matter of aircraft control.

Color

Although color has long been used in traditional instruments, it is restricted to aiding in identification of the data. There is no means of changing the color of any display component.

This restriction has been lifted with EFIS. For example, as an aircraft approaches the glideslope, a blue caption could indicate glide slope is armed; on capture the color might change to green.

On a typical EFIS system, the navigation needles are color coded to reflect the type of navigation being used. Green needles are used for ground based navigation such as VORs, Localizers and ILS systems. Magenta needles are used for GPS navigation.

Advantages

EFIS offers **versatility** by avoiding some of the physical limitations of traditional instruments. Thus, the same display which shows a course deviation indicator, can be switched to show the planned track provided by an area navigation or flight management system. If desired, the weather radar picture can then be superimposed on the displayed route.

The **flexibility** afforded by software modifications, minimises costs when new aircraft equipment and new regulations are introduced. The EFIS system can be updated with new software to extend its capabilities. Such updates introduced in the 1990s included enhanced GPWS, and TCAS.

A degree of **redundancy** is available even with the simple two-screen EFIS installation. Should the PFD fail, transfer switching repositions its vital information to the screen normally occupied by the navigation display.

Advances in EFIS

In the late 1980s, EFIS became standard equipment on most Boeing and Airbus airliners, and many business aircraft adopted EFIS in the 1990s.

Recent advances in computing power and reductions in the cost of liquid-crystal displays and navigational sensors (such as GPS and Attitude and Heading Reference Systems) have finally brought EFIS to general aviation aircraft. Notable examples are the Garmin G1000 and Chelton Flight Systems EFIS-SV.

Several EFIS manufacturers have focused on the experimental aircraft market, producing EFIS and EICAS systems for as little as US$1,000. The low cost is possible for several reasons, including steep drops in sensor prices and a lack of requirements to receive Federal Aviation Administration certification. This latter point restricts their use to experimental aircraft and certain other aircraft categories depending on local regulations. Uncertified EFIS systems are also found in Sport Pilot category aircraft, including factory built, microlight and ultralight aircraft. These systems can be fitted to certified aircraft in some cases as secondary or backup systems depending on local aviation authorities rules and regulations.

Chapter 17

Pitot-Static System

A **pitot-static system** is a system of pressure-sensitive instruments that is most often used in aviation to determine an aircraft's airspeed, Mach number, altitude, and altitude trend. A pitot-static system generally consists of a pitot tube, a static port, and the pitot-static instruments. This equipment is used to measure the forces acting on a vehicle as a function of the temperature, density, pressure and viscosity of the fluid in which it is operating. Other instruments that might be connected are air data computers, flight data recorders, altitude encoders, cabin pressurization controllers, and various airspeed switches. Errors in pitot-static system readings can be extremely dangerous as the information obtained from the pitot static system, such as altitude, is often critical to a successful flight. Several commercial airline disasters have been traced to a failure of the pitot-static system.

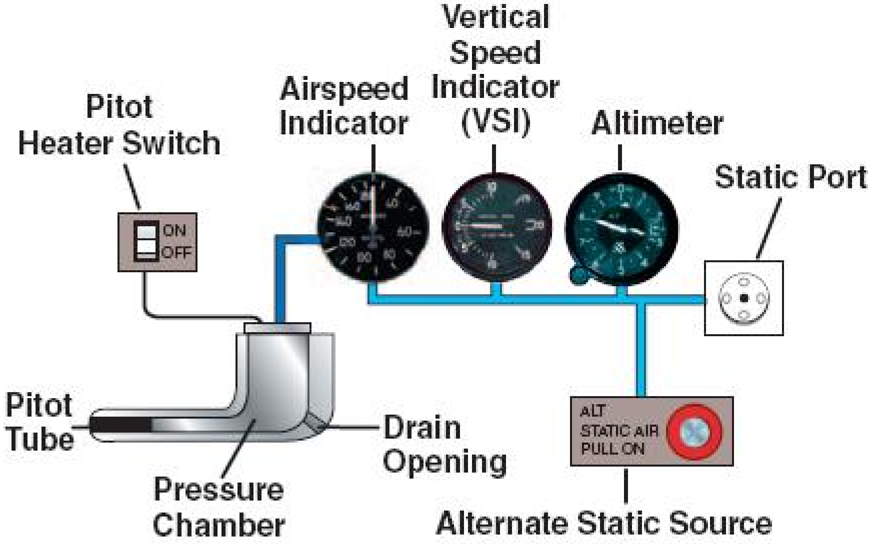
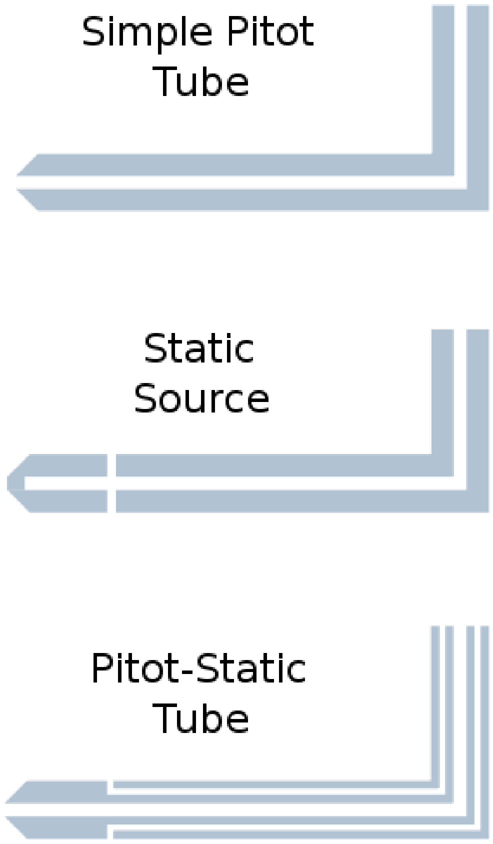


Diagram of a pitot-static system including the pitot tube, pitot-static instruments and static port

Pitot-static pressure



Examples of pitot tube, static tube, and pitot-static tube.

The pitot-static system of instruments uses the principle of air pressure gradient. It works by measuring pressures or pressure differences and using these values to assess the speed and altitude. These pressures can be measured either from the static port (static pressure) or the pitot tube (pitot pressure). The static pressure is used in all measurements, while the pitot pressure is only used to determine airspeed.

Pitot pressure

The pitot pressure is obtained from the pitot tube. The pitot pressure is a measure of ram air pressure (the air pressure created by vehicle motion or the air ramming into the tube), which, under ideal conditions, is equal to stagnation pressure, also called total pressure. The pitot tube is most often located on the wing or front section of an aircraft, facing forward, where its opening is exposed to the relative wind. By situating the pitot tube in such a location, the ram air pressure is more accurately measured since it will be less distorted by the aircraft's structure. When airspeed increases, the ram air pressure is increased, which can be translated by the airspeed indicator.

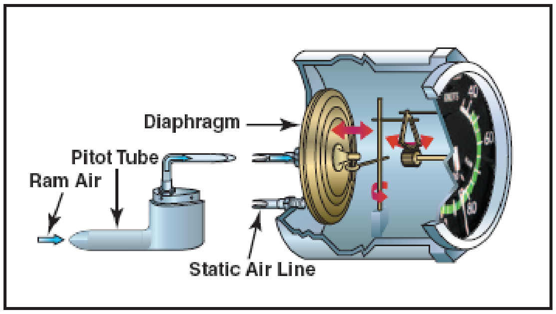
Static pressure

The static pressure is obtained through a static port. The static port is most often a flush-mounted hole on the fuselage of an aircraft, and is located where it can access the air flow in a relatively undisturbed area. Some aircraft may have a single static port, while others may have more than one. In situations where an aircraft has more than one static port, there is usually one located on each side of the fuselage. With this positioning, an average pressure can be taken, which allows for more accurate readings in specific flight situations. An alternative static port may be located inside the cabin of the aircraft as a backup for when the external static port(s) are blocked. A pitot-static tube effectively integrates the static ports into the pitot probe. It incorporates a second coaxial tube (or tubes) with pressure sampling holes on the sides of the probe, outside the direct airflow, to measure the static pressure.When aircraft climb, static pressure will decreases.

Multiple pressure

Some pitot-static systems incorporate single probes that contain multiple pressure-transmitting ports that allow for the sensing of air pressure, angle of attack, and angle of sideslip data. Depending on the design, such air data probes may be referred to as 5-hole or 7-hole air data probes. Differential pressure sensing techniques can be used to produce angle of attack and angle of sideslip indications.

Pitot-static instruments



Airspeed indicator diagram showing pressure sources from both the pitot tube and static port

The pitot-static system obtains pressures for interpretation by the pitot-static instruments. While the explanations below explain traditional, mechanical instruments, many modern aircraft use an air data computer (ADC) to calculate airspeed, rate of climb, altitude and Mach number. In some aircraft, two ADCs receive total and static pressure from independent pitot tubes and static ports, and the aircraft's flight data computer compares the information from both computers and checks one against the other. There are also "standby instruments", which are back-up pneumatic instruments employed in the case of problems with the primary instruments.

Airspeed indicator

The airspeed indicator is connected to both the pitot and static pressure sources. The difference between the pitot pressure and the static pressure is called "impact pressure". The greater the impact pressure, the higher the airspeed reported. A traditional mechanical airspeed indicator contains a pressure diaphragm that is connected to the pitot tube. The case around the diaphragm is airtight and is vented to the static port. The higher the speed, the higher the ram pressure, the more pressure exerted on the diaphragm, and the larger the needle movement through the mechanical linkage.

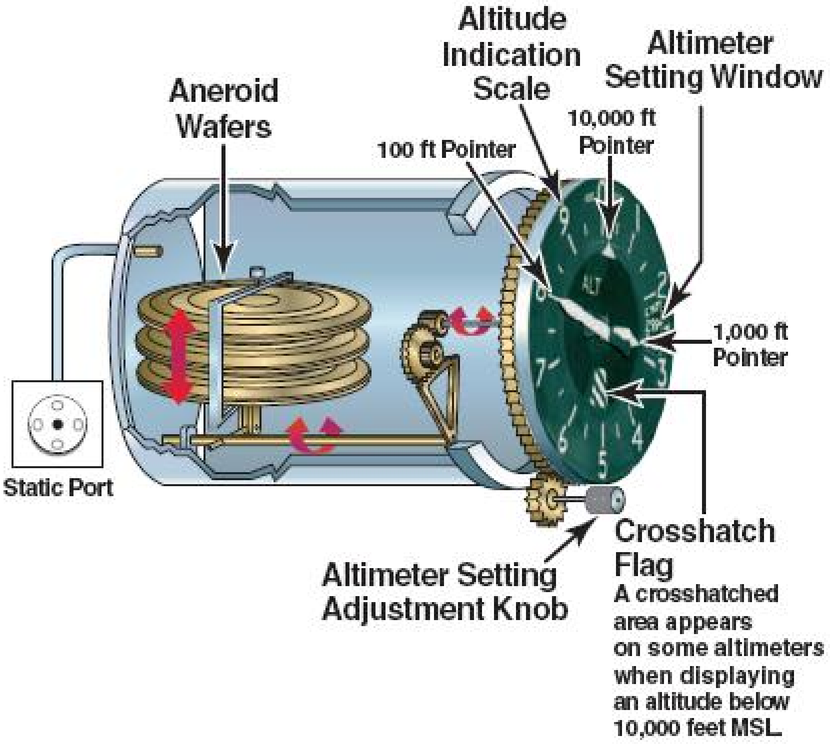


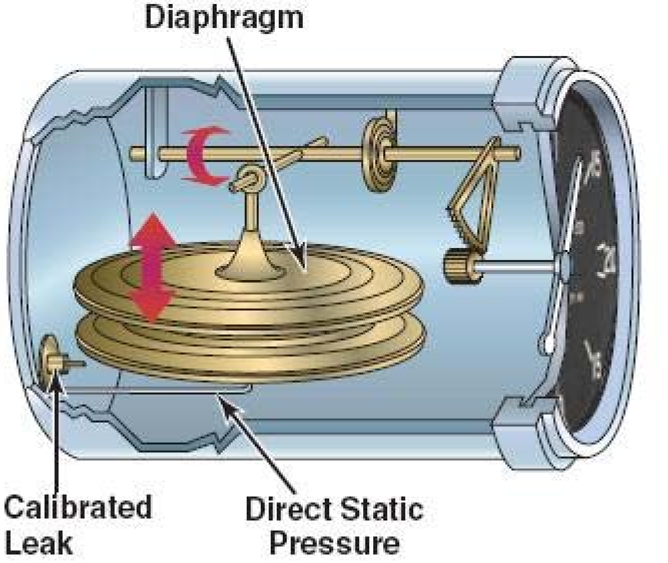
Diagram of an altimeter

Altimeter

The pressure altimeter, also known as the barometric altimeter, is used to determine changes in air pressure that occur as the aircraft's altitude changes. Pressure altimeters must be calibrated prior to flight to register the pressure as an altitude above sea level. The instrument case of the altimeter is airtight and has a vent to the static port. Inside the instrument, there is a sealed aneroid barometer. As pressure in the case decreases, the internal barometer expands, which is mechanically translated into a determination of altitude. The reverse is true when descending from higher to lower altitudes.

Machmeter

Aircraft designed to operate at transonic or supersonic speeds will incorporate a machmeter. The machmeter is used to show the ratio of true airspeed in relation to the speed of sound. Most supersonic aircraft are limited as to the maximum Mach number they can fly, which is known as the "Mach limit". The Mach number is displayed on a machmeter as a decimal fraction.



A vertical airspeed indicator

Vertical airspeed indicator

The variometer, also known as the vertical speed indicator (VSI) or the vertical velocity indicator (VVI), is the pitot-static instrument used to determine whether or not an aircraft is flying in level flight. The vertical airspeed specifically shows the rate of climb or the rate of descent, which is measured in feet per minute or meters per second. The vertical airspeed is measured through a mechanical linkage to a diaphragm located within the instrument. The area surrounding the diaphragm is vented to the static port through a calibrated leak (which also may be known as a "restricted diffuser"). When the aircraft begins to increase altitude, the diaphragm will begin to contract at a rate faster than that of the calibrated leak, causing the needle to show a positive vertical speed. The reverse of this situation is true when an aircraft is descending. The calibrated leak varies from model to model, but the average time for the diaphragm to equalize pressure is between 6 and 9 seconds.

Pitot-static errors

There are several situations that can affect the accuracy of the pitot-static instruments. Some of these involve failures of the pitot-static system itself—which may be classified as "system malfunctions"—while others are the result of faulty instrument placement or other environmental factors—which may be classified as "inherent errors".

System malfunctions

Blocked pitot tube

A blocked pitot tube is a pitot-static problem that will only affect airspeed indicators. A blocked pitot tube will cause the airspeed indicator to register an increase in airspeed when the aircraft climbs, even though actual airspeed is constant. This is caused by the pressure in the pitot system remaining constant when the atmospheric pressure (and static pressure) are decreasing. In reverse, the airspeed indicator will show a decrease in airspeed when the aircraft descends. The pitot tube is susceptible to becoming clogged by ice, water, insects or some other obstruction. For this reason, aviation regulatory agencies such as the U.S. Federal Aviation Administration (FAA) recommend that the pitot tube be checked for obstructions prior to any flight. To prevent icing, many pitot tubes are equipped with a heating element. A heated pitot tube is required in all aircraft certificated for instrument flight except aircraft certificated as Experimental Amateur-Built.

Blocked static port

A blocked static port is a more serious situation because it affects all pitot-static instruments. One of the most common causes of a blocked static port is airframe icing. A blocked static port will cause the altimeter to freeze at a constant value, the altitude at which the static port became blocked. The vertical speed indicator will become frozen at zero and will not change at all, even if vertical airspeed increases or decreases. The airspeed indicator will reverse the error that occurs with a clogged pitot tube and cause the airspeed be read less than it actually is as the aircraft climbs. When the aircraft is descending, the airspeed will be over-reported. In most aircraft with unpressurized cabins, an alternative static source is available and can be toggled from within the cockpit of the airplane.

Inherent errors

Inherent errors may fall into several categories, each affecting different instruments. *Density errors* affect instruments reporting airspeed and altitude. This type of error is caused by variations of pressure and temperature in the atmosphere. A *compressibility error* can arise because the impact pressure will cause the air to compress in the pitot tube. At standard sea level pressure altitude the calibration equation correctly accounts for the compression so there is no compressibility error at sea level. At higher altitudes the compression is not correctly accounted for and will cause the instrument to read greater than equivalent airspeed. A correction may be obtained from a chart.

Compressibility error becomes significant at altitudes above 10,000 feet (3,000 m) and at airspeeds greater than 200 knots (370 km/h). *Hysteresis* is an error that is caused by mechanical properties of the aneroid capsules located within the instruments. These capsules, used to determine pressure differences, have physical properties that resist change by retaining a given shape, even though the external forces may have changed. *Reversal errors* are caused by a false static pressure reading. This false reading may be caused by abnormally large changes in an aircraft's pitch. A large change in pitch will cause a momentary showing of movement in the opposite direction. Reversal errors primarily affect altimeters and vertical speed indicators.

Position errors

Another class of inherent errors is that of position error. A position error is produced by the aircraft's static pressure being different from the air pressure remote from the aircraft. This error is caused by the air flowing past the static port at a speed different from the aircraft's true airspeed. Position errors may provide positive or negative errors, depending on one of several factors. These factors include airspeed, angle of attack, aircraft weight, acceleration, aircraft configuration, and in the case of helicopters, rotor downwash. There are two categories of position errors, which are "fixed errors" and "variable errors". Fixed errors are defined as errors which are specific to a particular make of aircraft. Variable errors are caused by external factors such as deformed panels obstructing the flow of air, or particular situations which may overstress the aircraft.