LECTURE NOTES

ON

AVIONICS AND INSTRUMENTATION SYSTEM

IV B. Tech II semester (JNTUH-R15)

Ms M. Mary Thraza
Assistant Professor

AERONAUTICAL ENGINEERING

INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
DUNDIGAL, HYDERABAD - 500 043

UNIT - I
AVIONICS –INTRODUCTION-AVIONICS STANDARDS

1.1 Importance and Role of Avionics

‘Avionics’ is a word derived from the combination of aviation and electronics. It was first used in the USA in the early 1950s and has since gained wide scale usage and acceptance although it must be said that it may still be necessary to explain what it means to the lay person on occasions.

The term ‘avionic system’ or ‘avionic sub-system’ is used in this book to mean any system in the aircraft which is dependent on electronics for its operation, although the system may contain electro-mechanical elements.

For example, a Fly- by-Wire (FBW) flight control system depends on electronic digital computers for its effective operation, but there are also other equally essential elements in the sys- tem.

These include solid state rate gyroscopes and accelerometers to measure the angular and linear motion of the aircraft and air data sensors to measure the height, airspeed and incidence.

There are also the pilot’s control stick and rudder sensor assemblies and electro-hydraulic servo actuators to control the angular positions of the control surfaces.

The avionics industry is a major multi-billion dollar industry world-wide and the avionics equipment on a modern military or civil aircraft can account for around 30% of the total cost of the aircraft.

This figure for the avionics content is more like 40% in the case of a maritime patrol/anti-submarine aircraft (or helicopter) and can be over 75% of the total cost in the case of an airborne early warning aircraft (AWACS).

Modern general aviation aircraft also have significant avionics content. For example, colour head down displays, GPS satellite navigation systems, radio communications equipment.

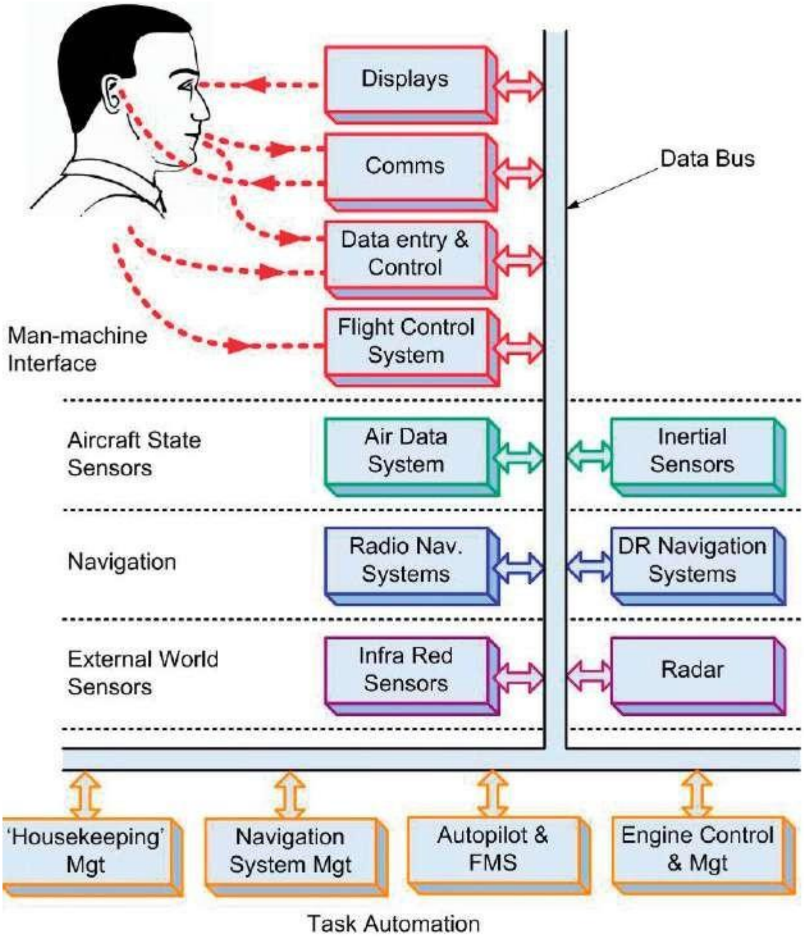
Avionics can account for 10% of their total cost.

It should be noted that unmanned aircraft (UMAs) are totally dependant on the avionic systems.

These comprise displays, communications, data entry and control and flight control. The *Display Systems* provide the visual interface between the pilot and the aircraft systems and comprise
head up displays (HUDs),
helmet mounted displays (HMDs)
and head down displays (HDDs).

Most combat aircraft are now equipped with a HUD. A small but growing number of civil aircraft have HUDs installed.

The HMD is also an essential system in modern combat aircraft and helicopters. The prime advantages of the HUD and HMD are that they project the display information into the pilot's field of view so that the pilot can be head up and can concentrate on the outside world.



Core avionic systems

Systems Which Interface Directly with the Pilot

The *Data Entry and Control Systems* are essential for the crew to interact with the avionic systems.
Such systems range from keyboards and touch panels to the use of direct voice input (DVI) control, exploiting speech recognition technology, and voice warning systems exploiting speech synthesisers.

The *Flight Control Systems* exploit electronic system technology in two areas, namely auto-stabilisation (or stability augmentation) systems and FBW flight control systems.

Most swept wing jet aircraft exhibit a lightly damped short period os- cillatory motion about the yaw and roll axes at certain height and speed conditions, known as ‘Dutch roll’, and require at least a yaw auto-stabiliser system to damp and suppress this motion; a roll auto-stabiliser system may also be required. The short period motion about the pitch axis can also be insufficiently damped and a pitch auto-stabiliser system is necessary.

Most combat aircraft and many civil aircraft in fact require three axis auto-stabilisation systems to achieve acceptable control and handling characteristics across the flight envelope.

FBW flight control enables a lighter, higher performance aircraft to be produced compared with an equivalent conventional design by allowing the aircraft to be designed with a reduced or even negative natural aerodynamic stability.

It does this by providing continuous automatic stabilisation of the aircraft by computer control of the control surfaces from appropriate motion sensors. The system can be designed to give the pilot a manoeuvre command control which provides excellent control and handling characteristics across the flight envelope.

‘Care free manoeuvring’ char- acteristics can also be achieved by automatically limiting the pilot’s commands ac- cording to the aircraft’s state. A very high integrity, failure survival system is of course essential for FBW flight control.

Aircraft State Sensor Systems

These comprise the air data systems and the inertial sensor systems.

The *Air Data Systems* provide accurate information on the air data quantities, that is the altitude, calibrated airspeed, vertical speed, true airspeed, Mach number and airstream incidence angle.

The *Inertial Sensor Systems* provide the information on aircraft attitude and the direction in which it is heading which is essential information for the pilot in executing a manoeuvre or flying in conditions of poor visibility, flying in clouds or at night.

Accurate attitude and heading information are also required by a number of avionic sub-systems which are essential for the aircraft’s mission – for example, the autopilot and the navigation system and weapon aiming in the case of a military aircraft.

Navigation Systems

Accurate navigation information, that is the aircraft’s position, ground speed and track angle (direction of motion of the aircraft relative to true North) is clearly essential for the aircraft’s mission, whether civil or military.

Navigation systems can be divided into dead reckoning (DR) systems and position fixing systems; both types are required in the aircraft.

The *Dead Reckoning Navigation Systems* derive the vehicle’s present position by estimating the distance travelled from a known position from a knowledge of the speed and direction of motion of the vehicle.

They have the major advantages of being completely self contained and independent of external systems.

The main types of DR navigation systems used in aircraft are:

(a) Inertial navigation systems. The most accurate and widely used systems.
(b) Doppler/heading reference systems. These are widely used in helicopters.
(c) Air data/heading reference systems These systems are mainly used as a reversionary navigation system being of lower accuracy than (a) or (b).

A characteristic of all DR navigation systems is that the position error builds up with time and it is, therefore, necessary to correct the DR position error and update the system from position fixes derived from a suitable position fixing system.

The *Position Fixing Systems* used are now mainly radio navigation systems based on satellite or ground based transmitters.

A suitable receiver in the aircraft with a supporting computer is then used to derive the aircraft's position from the signals received from the transmitters.

Task Automation Systems

These comprise the systems which reduce the crew workload and enable minimum crew operation by automating and managing as many tasks as appropriate so that the crew role is a supervisory management one. The tasks and roles of these are very briefly summarised below.

Navigation Management comprises the operation of all the radio navigation aid systems and the combination of the data from all the navigation sources, such as GPS and the INS systems, to provide the best possible estimate of the aircraft position, ground speed and track.

The *Autopilots and Flight Management Systems* have been grouped together. Because of the very close degree of integration between these systems on modern civil aircraft. It should be noted, however, that the Autopilot is a 'stand alone' system and not all aircraft are equipped with an FMS.

The autopilot relieves the pilot of the need to fly the aircraft continually with the consequent tedium and fatigue and so enables the pilot to concentrate on other tasks associated with the mission. Apart from basic modes, such as height hold and heading hold, a suitably designed high integrity autopilot system can also provide a very precise control of the aircraft flight path for such applications as automatic landing in poor or even zero visibility conditions

- Flight planning.
- Navigation management.
- Engine control to maintain the planned speed or Mach number.
- Control of the aircraft flight path to follow the optimised planned route. Control of the vertical flight profile.
- Ensuring the aircraft is at the planned 3D position at the planned time slot; often referred to as 4D navigation. This is very important for air traffic control.
- Flight envelope monitoring.
- Minimising fuel consumption.
- Fuel management. This embraces fuel flow and fuel quantity measurement and control of fuel transfer from the appropriate fuel tanks to minimise changes in the aircraft trim.
- Electrical power supply system management. Hydraulic power supply system management. Cabin/cockpit pressurisation systems.
- Environmental control
  system. Warning systems.

- Maintenance and monitoring systems.

1.2 The Avionic Environment

Avionic systems equipment is very different in many ways from ground based equipment carrying out similar functions. The reasons for these differences are briefly explained in view of their fundamental importance.

1. The importance of achieving minimum weight.
2. The adverse operating environment particularly in military aircraft in terms of operating temperature range, acceleration, shock, vibration, humidity range and electro-magnetic interference.
3. The importance of very high reliability, safety and integrity.

Minimum Weight

There is a gearing effect on unnecessary weight which is of the order of 10:1. For example a weight saving of 10 kg enables an increase in the payload capability of the order of 100 kg.
The process of the effect of additional weight is a vicious circle. An increase in the aircraft weight due to, say, an increase in the weight of the avionics equipment, requires the aircraft structure to be increased in strength, and therefore made heavier, in order to withstand the increased loads during manoeuvres.
Assum- ing the same maximum normal acceleration, or *g*, and the same safety margins on maximum stress levels are maintained

Environmental Requirements

The environment in which avionic equipment has to operate can be a very severe and adverse one in military aircraft; the civil aircraft environment is generally much more benign but is still an exacting one.

Considering just the military cockpit environment alone, such as that experienced by the HUD and HDD. The operating temperature range is usually specified from $-40°C$ to $+70°C$. Clearly, the pilot will not survive at these extremes but if the aircraft is left out in the Arctic cold or soaking in the Middle-East sun, for example, the equipment may well reach such temperatures. A typical specification can demand full performance at 20,000 ft within two minutes of take-off at any temperature within the range.

Reliability

The over-riding importance of avionic equipment reliability can be appreciated in view of the essential roles of this equipment in the operation of the aircraft

Every possible care is taken in the design of avionic equipment to achieve maximum reliability.

A typical RST cycle requires the equipment to operate satisfactorily through the cycle described below.

- Soaking in an environmental chamber at a temperature of $+70°C$ for a given period.
- Rapidly cooling the equipment to $-55°C$ in 20 minutes and soaking at that

temperature for a given period.
- Subjecting the equipment to vibration, for example 0.5g amplitude at 50 Hz, for periods during the hot and cold soaking phases.

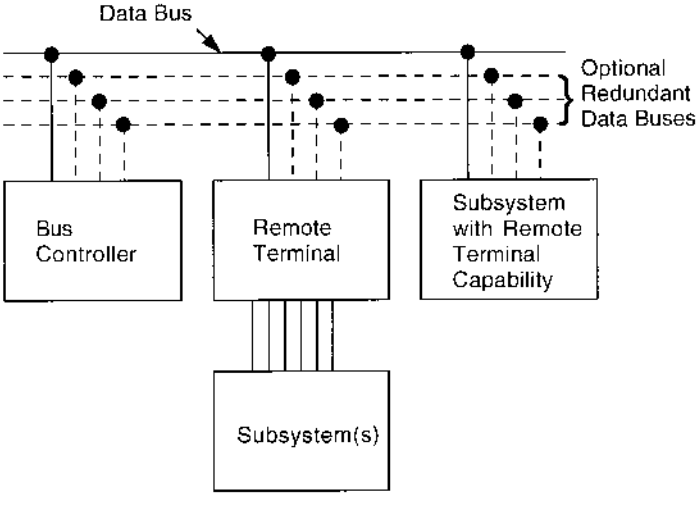
Data Bus Systems

Data bus systems are the essential enabling technologies of avionic systems integration in both federated and integrated modular avionics architectures.

They can be broadly divided into electrical data bus systems where the data are transmitted as electrical pulses by wires, and optical data bus systems where the data are transmitted as light pulses by optical fibres

Electrical Data Bus Systems

There are several electrical serial digital data bus systems in use in avionics systems. These systems can be broadly divided into two categories in terms of their data rate transmission capabilities, namely, data bus systems operating with a maximum throughput of 1 to 2 Mbits/s and high speed data bus systems with a throughput of 50 Mbits/s to 100 Mbits/s.



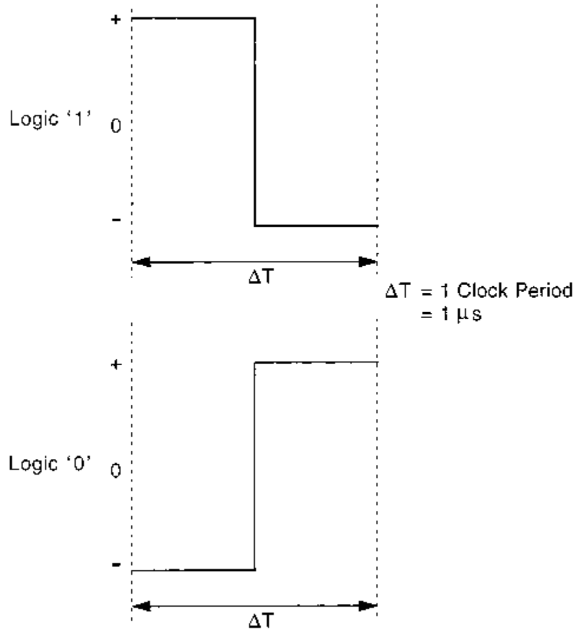
Typical multiplex data bus system architecture.

The version which has been adapted for airborne applications is known as the ‘Avionics Full Duplex Switched Ethernet’, which has been shortened to ‘AFDX Ethernet’ network. It meets the civil aircraft avionic system requirements in all aspects and its commercially sourced components make it a very competitive system.

Its adoption in military aircraft avionic systems would appear to be a likely future development because of its cost advantages.

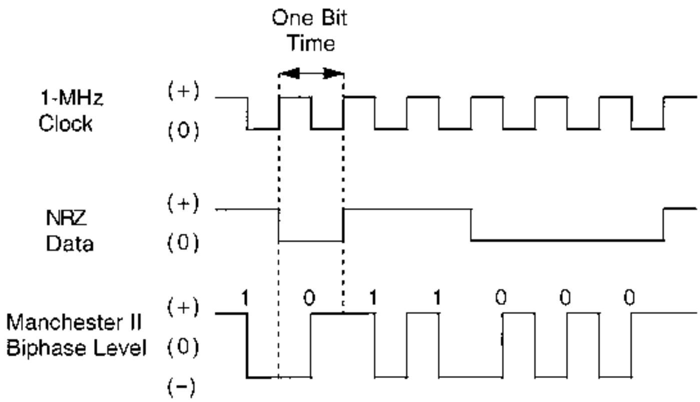
MIL STD 1553 Bus System

MIL STD 1553B is a US military standard which defines a TDM multiple-source–multiple-sink data bus system which is in very wide scale use in military aircraft in many countries. It is also used in naval surface ships, submarines, and land vehicles such as main battlefield tanks. The system is a half duplex system, that is operation of a data transfer can take place in either direction over a single line, but not in both directions on that line simultaneously.



Logic '1' and logic '0'.

the bus controller. Each sub-system is connected to the bus through a unit called a remote terminal (RT). Data can only be transmitted from one RT and received by another RT.



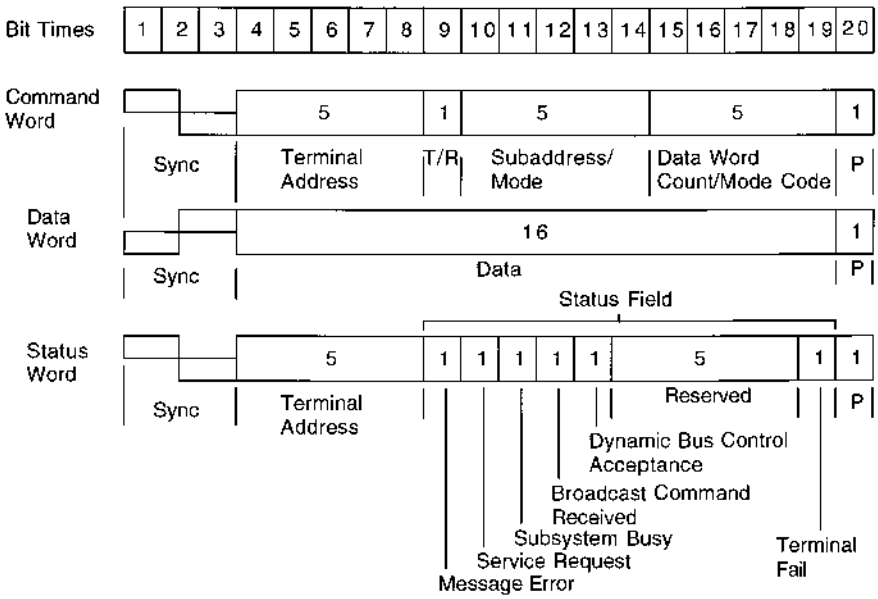
Data encoding

A maximum of 31 terminals can be connected to the bus. The bus operation is asynchronous, each terminal having an independent clock source for transmission.

Decoding is achieved in receiving terminals using clock information derived from the messages.

A command word comprises six separate fields. These are briefly explained below:

- The SYNC signal field is an invalid Manchester waveform so that it cannot be 'confused' with any data bits.
- The RT address field occupies 5 bits, each RT being assigned a unique 5 bit address. Decimal address 31(11111) is not assigned as a unique address and is a broadcast address.
- The T/R bit is 0 if the RT is to receive, and 1 if the RT is to transmit.
- The sub-address/mode field, comprising 5 bits, is used for either an RT sub-address or mode control. The sub-address is used to route data to and from a location in the RT. A code of all zeros (00000) in the sub-address/mode field indicates that the contents of the word counts/mode field are to be decoded as a five bit mode command.



Word formats.

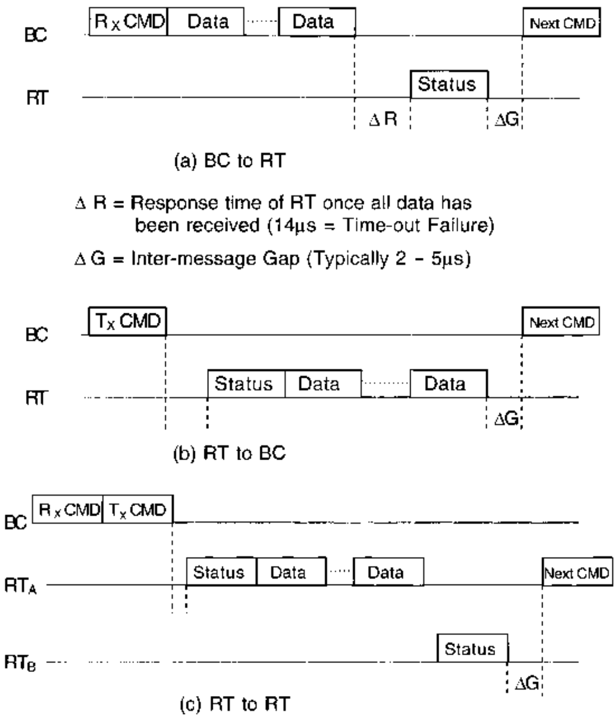
- The data word count/mode code field, comprising 5 bits, is generally used for data transfers. The word count field indicates the number of data words to be transferred in any one message block, the maximum number being 32 (indicated by all zeros).
- The parity bit is 1 if there is an odd number of bits in fields 1–19.

A status word is the first word of a response by an RT to a BC command. It provides:

(a) A summary of the status/health of the RT.
(b) The word count of the data words to be transmitted in response to a command.

The fields are briefly described below:

- The SYNC signal field is the same as with a command word.
- The RT address field (5 bits) confirms the correct RT is responding.
- The status field comprises 11 bits. The message error bit is set if the previous command was not correctly understood. The instrumentation bit =0 to distinguish the word from a command word.
- The parity bit is set by the RT in the same sense as a command word.



Transfer formats.

There are ten possible transfer formats, but the three most commonly used formats are:

- BC to RT
- RT to BC
- RT to RT

- *Message data validation* – the terminal is designed to detect improperly coded signals, data drop-outs or excessively noisy signals.
- *Word validation* – the terminal checks that each word conforms to the following minimum criteria:
  - Word begins with a valid SYNC field.
  - Bits are in valid Manchester II code.
  - Information field has 16 bit plus parity.
  - Word parity is odd.

  When a word fails to conform to the above criteria the word is considered invalid.

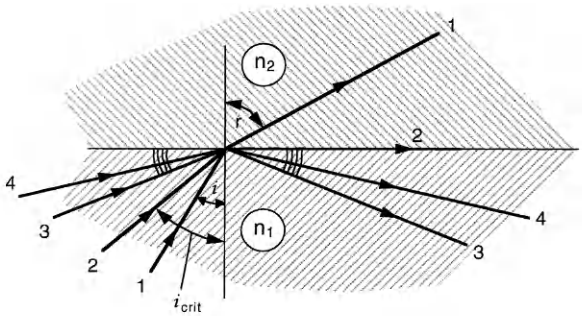
- *Transmission continuity* – the terminal checks the message is contiguous as shown in the formats. Improperly timed data SYNCs are con- sidered a continuity error.
- *Excessive transmission* – the terminal includes a signal time-out which pre- cludes a signal transmission greater than 1 ms plus or minus 0.34 ms.

The data word is deemed valid when the data meet the above criteria and are received

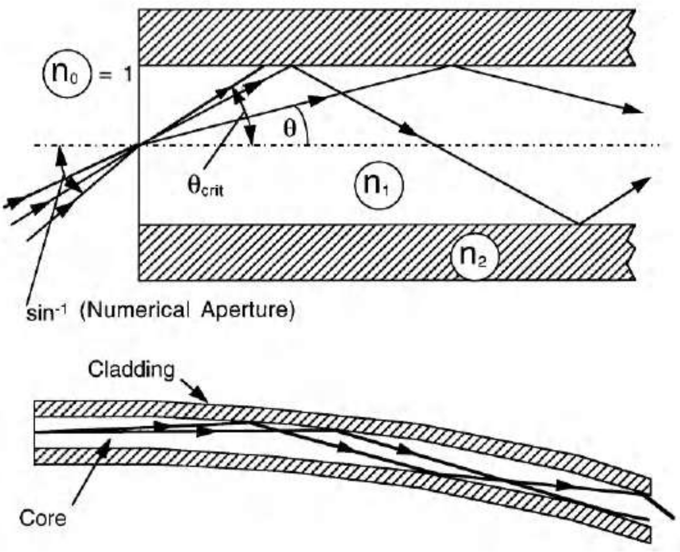
Optical Data Bus Systems

Most readers are probably familiar to some extent with the use of optical fibres to transmit light signals. A brief explanation is set out below for those readers who need to refresh themselves on the subject and also to make clear the difference between multi-mode and single mode optical fibres and their respective applications. The transmission of light signals along any optical fibre depends on the optical property of total internal reflection.. Ray 1 is refracted in passing through the second medium, the relationship between the angle the incident ray makes with the normal, i, and the angle the refracted ray makes with the normal, r, being given by Snell's
law:

$$\frac{\sin i}{\sin r} = \frac{n_2}{n_1}$$

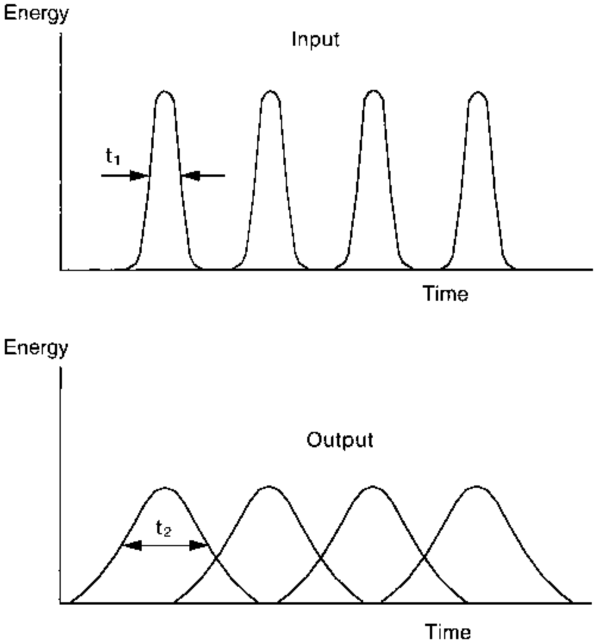


Total internal reflection.



Multi-mode optical fibre.

At the critical incidence angle, i_{crit}, ray 2 is refracted through an angle of 90° and does not pass through the second medium ($i_{crit} = \sin^{-1} n_2/n_1$). All rays with incident angles greater than i_{crit} such as rays 3 and 4 are thus reflected back into the first medium. This condition is known as total internal reflection and is effectively a loss free process.

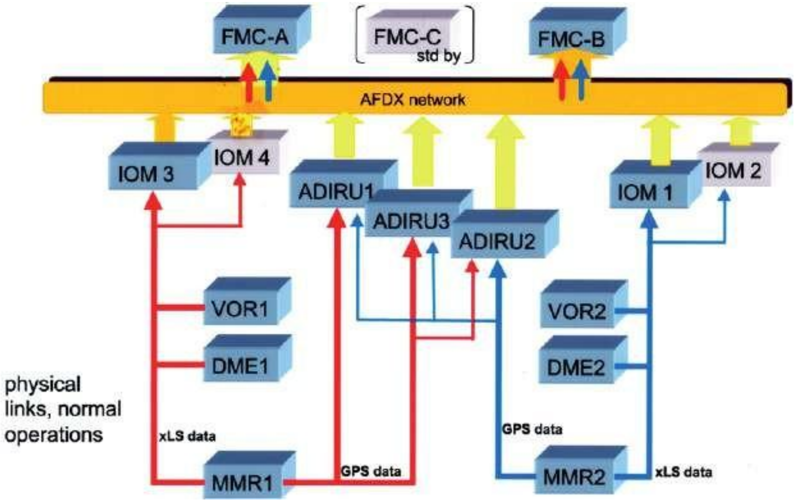


Pulse broadening.

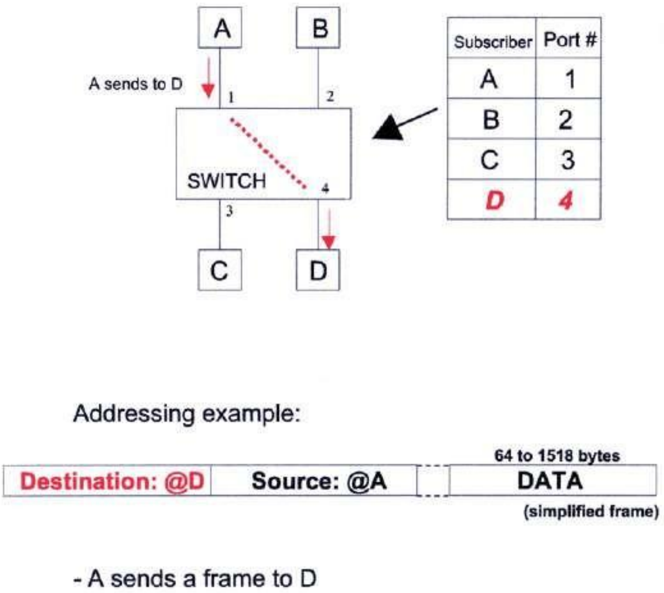
Features of transmission systems.

Parameter	STANAG 3910 Data Bus	LTPB Linear Token Passing Bus	HSRB High Speed Ring Bus
Data rate	20 Mbits/s	50 Mbits/s	50–100 Mbits/s
Encoding technique	Manchester bi-phase	Manchester bi-phase	4B/5B data encoding
Topology	Bus structure	Bus structure	Point to point linked ring
Max message transfer	4096 data words	4096 data words	4096 data words
Number of stations	31	128	128
Bus control philosophy	Central control	Distributed control	Distributed control
Controlling mechanism	1553 bus control	Token passing control	Token passing control
Bus length	Dependent on 1553 network	1000 m	1500 m
Interconnect media	Fibre optics	Fibre optics or electrical	Fibre optics or electrical
Standard	STANAG	SAE	SAE
Country of origin	Europe	USA	USA

To overcome these problems, a number of higher speed transmission systems have been developed both in Europe and the USA.



Interconnection of Navigation System units by the AFDX network (by courtesy of Air- bus).



Switched Ethernet principle.

and communication constraints with the objective of connecting subscribers which exchange large amounts of data with each other to the same switches.

UNIT - II
Displays and Man–Machine Interaction and Communication System

Introduction

The cockpit display systems provide a visual presentation of the information and data from the aircraft sensors and systems to the pilot (and crew) to enable the pilot to fly the aircraft safely and carry out the mission. They are thus vital to the operation of any aircraft as they provide the pilot, whether civil or military, with:

- Primary flight information,
- Navigation information,
- Engine data,
- Airframe data,
- Warning information.

The military pilot has also a wide array of additional information to view, such as:

- Infrared imaging sensors,
- Radar,
- Tactical mission data,
- Weapon aiming,
- Threat warnings.

The pilot is able to rapidly absorb and process substantial amounts of visual information but it is clear that the information must be displayed in a way which can be readily assimilated, and unnecessary information must be eliminated to ease the pilot's task in high work load situations. A number of developments have taken place to improve the pilot–display interaction and this is a continuing activity as new technology and components become available. Examples of these developments are:

- Head up displays,
- Helmet mounted displays,
- Multi-function colour displays,
- Digitally generated colour moving map displays,
- Synthetic pictorial imagery,
- Displays management using intelligent knowledge based system (IKBS) technology,
- Improved understanding of human factors and involvement of human factors specialists from the initial cockpit design stage.

Equally important and complementary to the cockpit display systems in the 'man machine interaction' are the means provided for the pilot to control the operation of the avionic systems and to enter data. Again, this is a field where continual development is taking place. Multi-function keyboards and multi-function touch panel displays are now widely used. Speech recognition technology has now reached sufficient maturity for 'direct voice input' control to be installed in the new generation of military aircraft. Audio warning systems are now well established in both military and civil aircraft.

The integration and management of all the display surfaces by audio/tactile inputs enables a very significant reduction in the pilot’s workload to be achieved in the new generation of single seat fighter/strike aircraft. Other methods of data entry which are being evaluated include the use of eye trackers.

It is not possible in the space of one chapter to cover all aspects of this subject which can readily fill several books. Attention has, therefore, been concentrated on providing an overview and explanation of the basic principles involved in the following topics:

- Head up displays
- Helmet mounted displays
- Computer aided optical design
- Discussion of HUDs versus HMDs
- Head down displays
- Data fusion
- Intelligent displays management
- Display technology
- Control and data entry

Head Up Displays

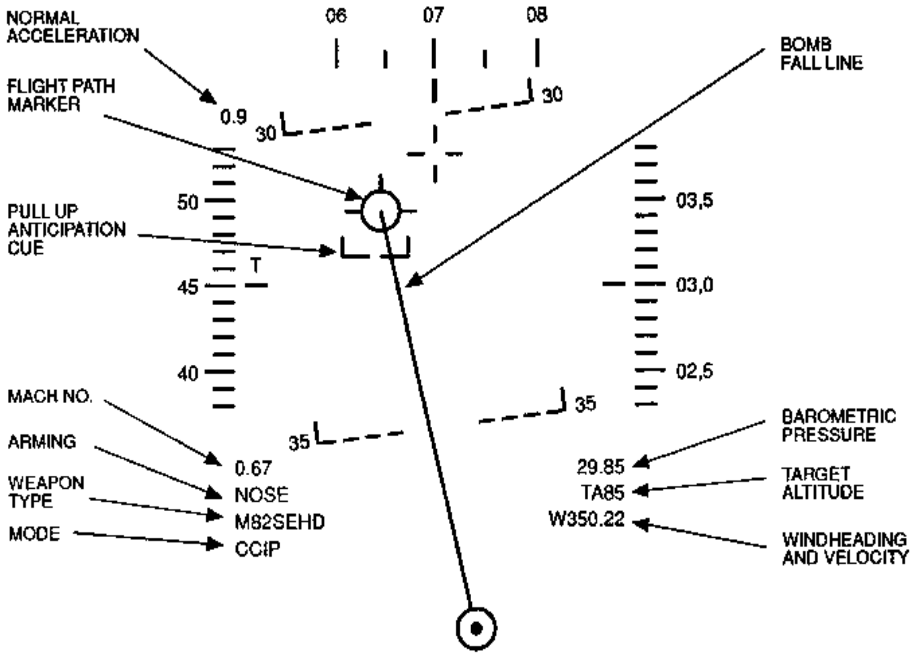
Introduction

Without doubt the most important advance to date in the visual presentation of data to the pilot has been the introduction and progressive development of the Head Up Display or HUD. (The first production HUDs, in fact, went into service in 1962 in the Buccaneer strike aircraft in the UK.)

The HUD has enabled a major improvement in man–machine interaction (MMI) to be achieved as the pilot is able to view and assimilate the essential flight data generated by the sensors and systems in the aircraft whilst head up and maintaining full visual concentration on the outside world scene.

Head-up presentation of primary flight information (by courtesy of BAE Systems

The pilot is thus free to concentrate on the outside world during manoeuvres and does not need to look down at the cockpit instruments or head down displays. It should be noted that there is a transition time of one second or more to re-focus the eyes from viewing distant objects to viewing near objects a metre or less away, such as the cockpit instruments and displays and adapt to the cockpit light environment. In combat situations, it is essential for survival that the pilot is head up and scanning for possible threats from any direction. The very high accuracy which can be achieved by a HUD and computerised weapon aiming system together with the



Typical weapon aiming display (by courtesy of BAE Systems)

HUDs are now being installed in civil aircraft for reasons such as:

1. Inherent advantages of head-up presentation of primary flight information including depiction of the aircraft's flight path vector, resulting in improved situational awareness and increased safety in circumstances such as wind shear or terrain/traffic avoidance manoeuvres.
2. To display automatic landing guidance to enable the pilot to land the aircraft safely in conditions of very low visibility due to fog, as a back up and monitor for the automatic landing system. The display of taxi-way guidance is also being considered.
3. Enhanced vision using a raster mode HUD to project a FLIR video picture of the outside world from a FLIR sensor installed in the aircraft, or, a synthetic picture of the outside world generated from a forward looking millimetric radar sensor

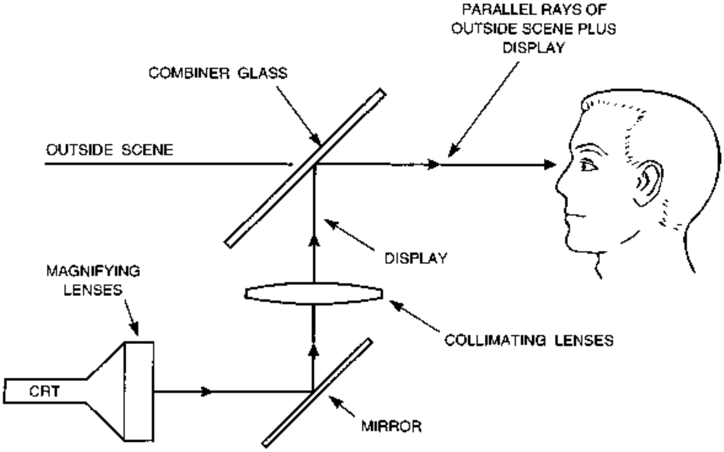
Civil HUD installation – BAE Systems VGS HUD installed in a Boeing 737-800 series airliner (by courtesy of BAE Systems).

in the aircraft. These enhanced vision systems are being actively developed and will enable the pilot to land the aircraft in conditions of very low or zero visibility at airfields not equipped with adequate all weather guidance systems such as ILS (or MLS).

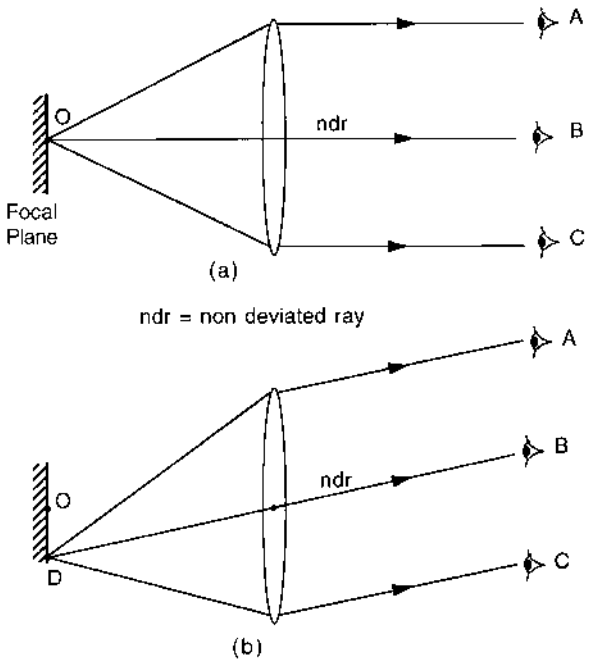
Illustrates a civil HUD installation.

Basic Principles

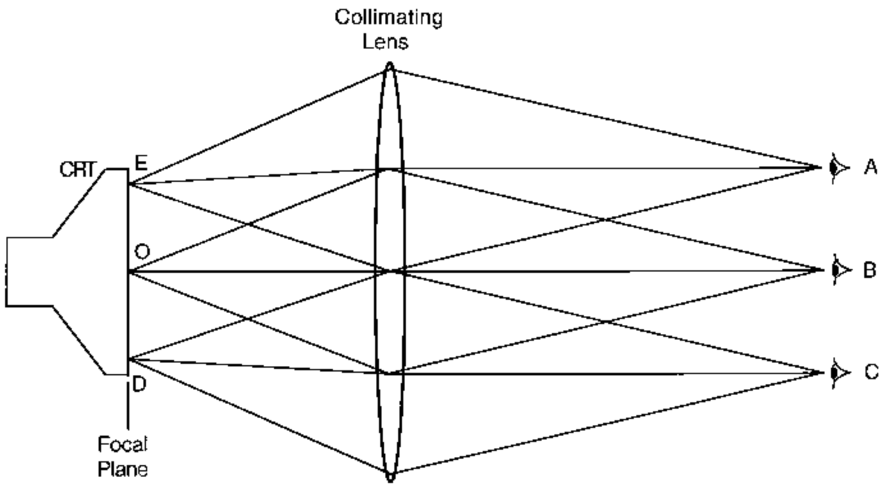
The basic configuration of a HUD is shown schematically in Figure. The pilot views the outside world through the HUD combiner glass (and windscreen). The combiner glass is effectively a 'see through' mirror with a high optical transmission efficiency so that there is little loss of visibility looking through the combiner and windscreen. It is called a combiner as it optically combines the collimated display symbology with the outside world scene viewed through it.



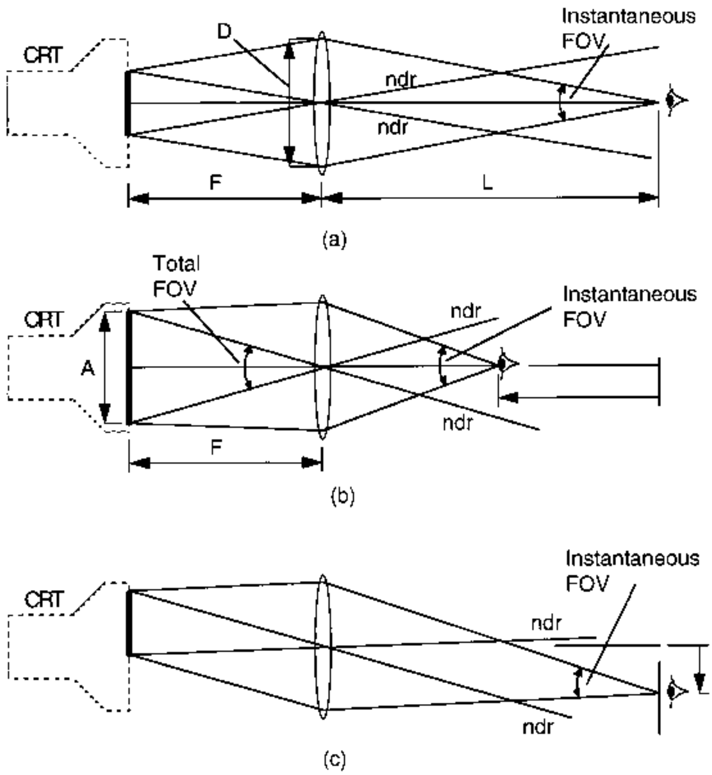
HUD schematic.



Simple optical collimator.

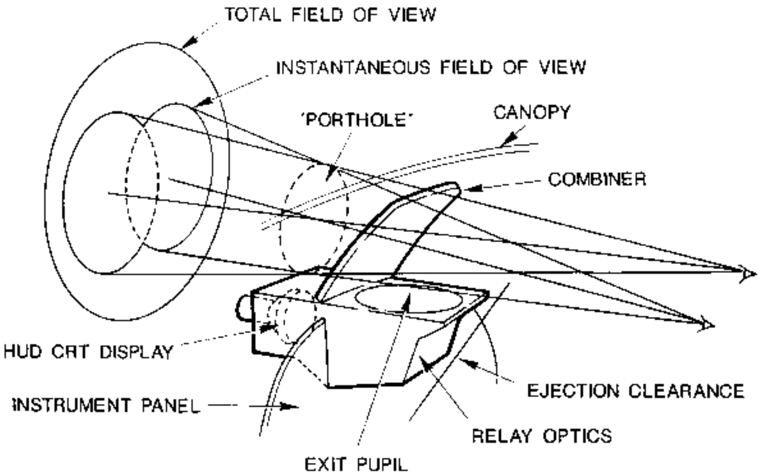


Simple optical collimator ray trace.



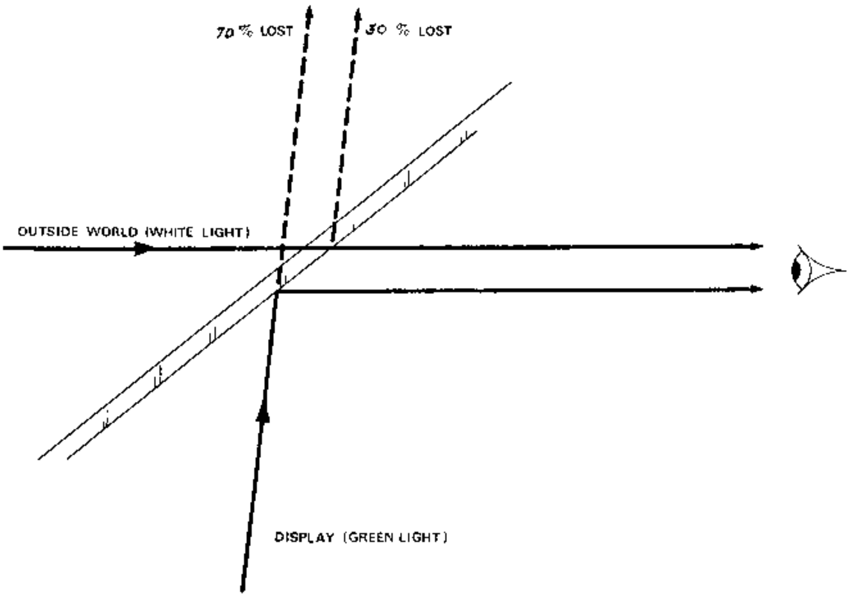
Instantaneous and total FOV.

The analogy can be made of viewing a football match through a knot hole in the fence and this FOV characteristic of a HUD is often referred to as the ‘knot hole effect’.



HUD installation constraints and field of view.

In order to achieve an adequate contrast ratio so that the display can be seen against the sky at high altitude or against sunlit cloud it is necessary to achieve a



Conventional refractive HUD combiner operation.

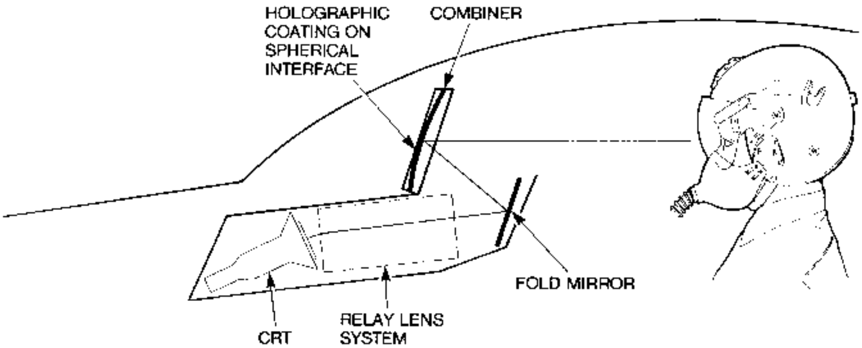
display brightness of 30,000 Cd/m^2 (10,000 ft L) from the CRT. In fact, it is the brightness requirement in particular which assures the use of the CRT as the display source for some considerable time to come, even with the much higher optical efficiencies which can be achieved by exploiting holographic optical elements.

Instantaneous FOV of conventional HUD.

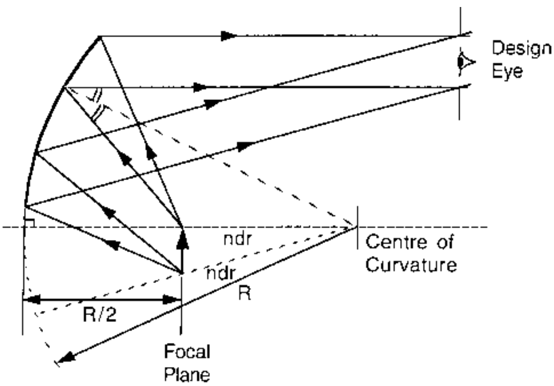
Instantaneous FOV of holographic HUD.

Holographic HUDs

The requirement for a large FOV is driven by the use of the HUD to display a collimated TV picture of the FLIR sensor output to enable the pilot to 'see' through the HUD FOV in conditions of poor visibility, particularly night operations.

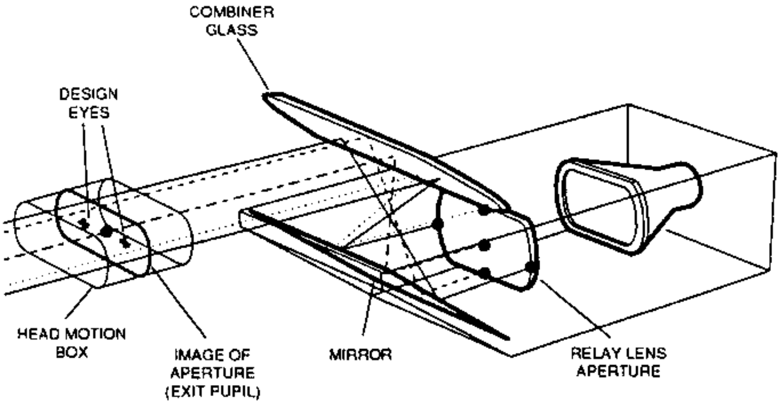


Off-axis holographic combiner HUD configuration.



Collimation by a spherical reflecting surface.

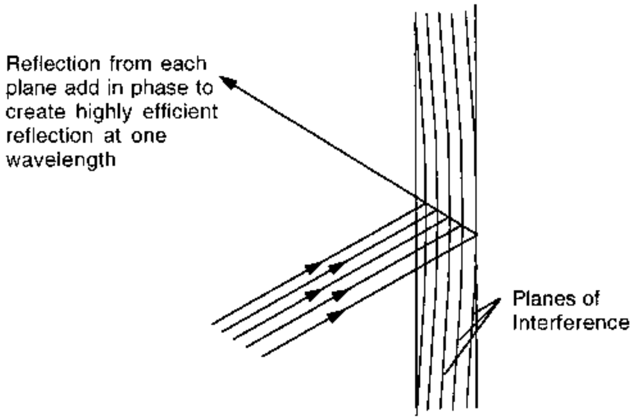
car round Hyde Park Corner with a shattered opaque windscreen with your vision restricted to a hole punched through the window.)



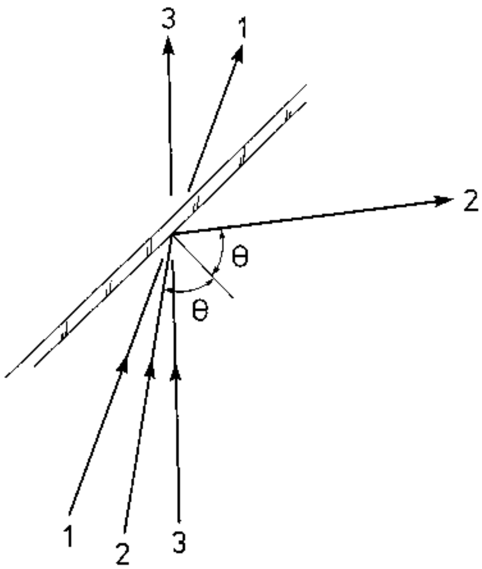
The head motion box concept.

Because the collimating element is located in the combiner, the porthole is considerably nearer to the pilot than a comparable refractive HUD design. The collimating element can also be made much larger than the collimating lens of a refractive HUD, within the same cockpit constraints. The IFOV can thus be increased by a factor of two or more and the instantaneous and total FOVs are generally the same, as the pilot is effectively inside the viewing porthole.

This is where a unique property of holographically generated coatings is used, namely the ability to introduce optical power within the coating so that it can correct the remaining aberration errors. The powered holographic coating produces an effect equivalent to local variations in the curvature of the spherical reflecting surface of the combiner to correct the aberration errors by diffracting.

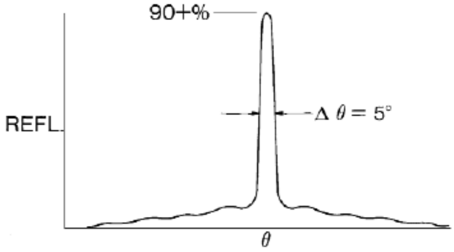


Holographic coating.



Angularly selective reflection of monochromatic rays.

At any point on the surface, the coating will only reflect a given wavelength over a small range of incidence angles. Outside this range of angles, the reflectivity drops off very rapidly and light of that wavelength will be transmitted through the coating.



Holographic coat- ing performance.

Uniformly across the reflector surface.

The process for producing the powered holographic combiner is very briefly outlined below.

The process has three key stages:

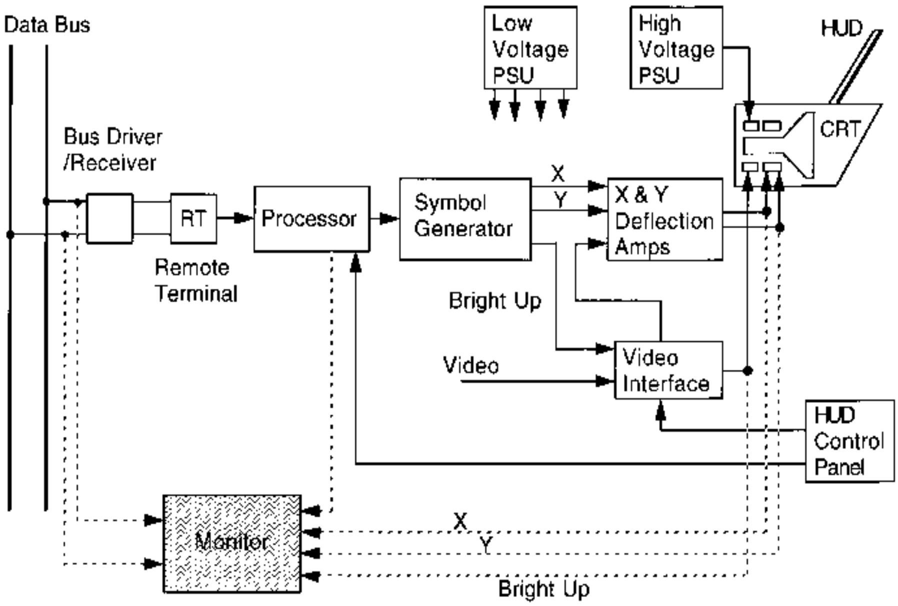
- Design and fabricate the Computer Generated Hologram (CGH).
- Produce master hologram.
- Replicate copies for the holographic combiner elements.

The CGH and construction optics create the complex wavefront required to produce the master hologram. The CGH permits control of the power of the diffraction grating over the combiner thus allowing correction of some level of optical distortion resulting in a simplified relay lens system.

Wide FOV holographic HUD installed in Eurofighter Typhoon (by courtesy of BAE Systems).

HUD Electronics

The basic functional elements of a modern HUD electronic system are shown in above figure. These functional elements may be packaged so that the complete HUD system is contained in a single unit, as in the Typhoon HUD above.



HUD electronics.

The display processor processes this input data to derive the appropriate display formats, carrying out tasks such as axis conversion, parameter conversion and format management.

In addition the processor also controls the following functions:

- Self test,
- Brightness control (especially important at low brightness levels),
- Mode sequencing,
- Calibration,
- Power supply control.

Helmet Mounted Displays
Introduction

The advantages of head-up visual presentation of flight and mission data by means of a collimated display have been explained in the preceding section. The HUD, however, only presents the information in the pilot's forward field of view, which even with a wide FOV holographic HUD is limited to about 30° in azimuth and 20° to 25° in elevation. Significant increases in this FOV are not practicable because of the cockpit geometry constraints.

Helmet Design Factors

It is important that the main functions of the conventional aircrew helmet are appreciated as it is essential that the integration of a helmet mounted display system with the helmet does not degrade these functions in any way. The basic functions are:

1. To protect the pilot's head and eyes from injury when ejecting at high airspeeds. For example, the visor must stay securely locked in the down position when subjected to the blast pressure experienced at indicated airspeeds of 650 knots. The helmet must also function as a crash helmet and protect the pilot's head as much as possible in a crash landing.

2. To interface with the oxygen mask attached to the helmet. Combat aircraft use a special pressurised breathing system for high *g* manoeuvring.

3. To provide the pilot with an aural and speech interface with the communications radio equipment. The helmet incorporates a pair of headphones which are coupled to the outputs of the appropriate communications channel selected by the pilot. The helmet and earpieces are also specifically designed to attenuate the cockpit background acoustic noise as much as possible. A speech interface is provided by a throat microphone incorporated in the oxygen mask.

4. In addition to the clear protective visor, the helmet must also incorporate a dark visor to attenuate the glare from bright sunlight.
5. The helmet must also be compatible with NBC (nuclear–biological–chemical) protective clothing and enable an NBC mask to be worn.

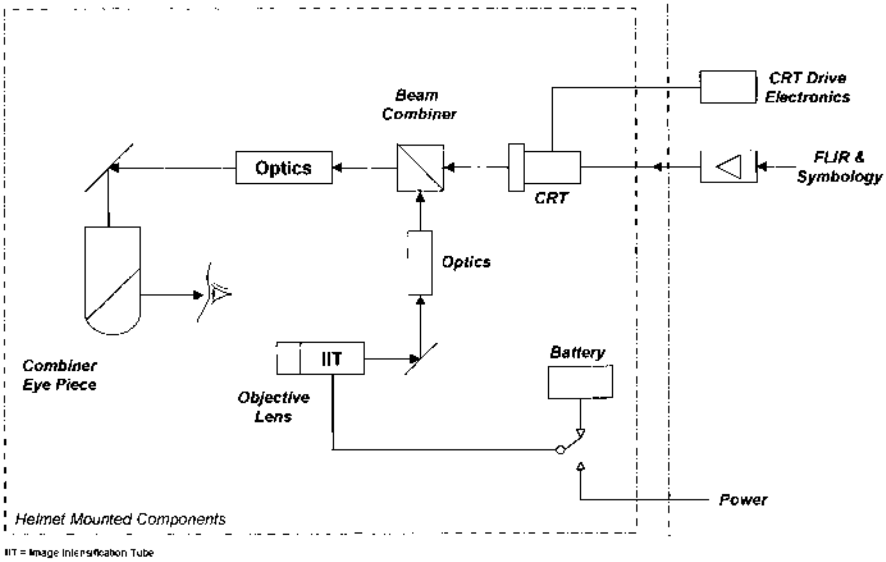
Helmet mounted sight and off-boresight missile launch (by courtesy of BAE Systems)

Helmet Mounted Sights

- A helmet mounted sight (HMS) in conjunction with a head tracker system provides a very effective means for the pilot to designate a target.
- The pilot moves his head to look and sight on the target using the collimated aiming cross on the helmet sight.
- The angular co-ordinates of the target sight line relative to the airframe are then inferred from the measurements made by the head tracker system of the attitude of the pilot’s head.
- In air to air combat, the angular co-ordinates of the target line of sight (LOS) can be supplied to the missiles carried by the aircraft.
- The missile seeker heads can then be slewed to the target LOS to enable the seeker heads to acquire and lock on to the target. (A typical seeker head needs to be pointed to within about 2° of the target to achieve automatic lock on.)
- Missile lock on is displayed to the pilot on the HMS and an audio signal is also given. The pilot can then launch the missiles.

Helmet Mounted Displays

- As mentioned in the introduction, the HMD can function as a ‘HUD on the hel- met’ and provide the display for an integrated night/poor visibility viewing system with flight and weapon aiming symbology overlaying the projected image from the sensor.
- Although monocular HMDs have been built which are capable of displaying all the information normally displayed on a HUD, there can be problems with what is known as ‘monocular rivalry’.
- This is because the brain is trying to process different images from each eye and rivalry can occur between the eye with a display and that without.
- The problems become more acute at night when the eye without the display sees very little, and the effects have been experienced when night-flying with a monocular system in a helicopter.

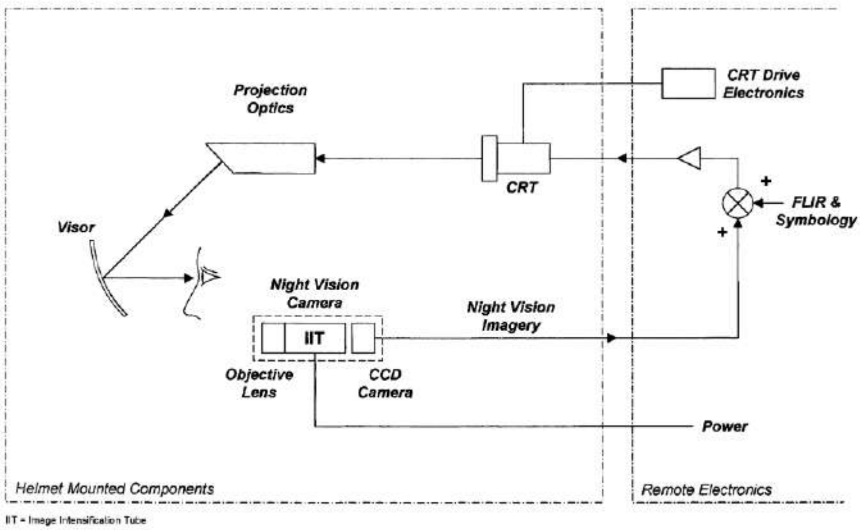


Optical mixing of IIT and CRT imagery.

It should be noted that special cockpit lighting is necessary as conventional cock- pit lighting will saturate the image intensifiers. Special green lighting and comple- mentary filtering are required.

Most night flying currently undertaken by combat aircraft is carried out using NVGs which are mounted on a bracket on the front of a standard aircrew helmet. The weight of the NVGs and forward mounting creates an appreciable out of balance moment on the pilot’s head and precludes the pilot from undertaking manoeuvres involving significant *g*. The NVGs must also be removed before ejecting because of the *g* experienced during ejection. (A quick release mechanism is incorporated in the NVG attachment.)

Helicopter HMD with integrated NVGs (by courtesy of BAE Systems). This HMD forms part of the mission systems avionics for the Tiger 2 attack helicopter for the German army.



Electronic combination of IIT and CRT.

The performance of current night vision cameras is, however, inherently lower in terms of resolution than that provided by the best NVGs.

Eurofighter Typhoon binocular visor projected HMD (by courtesy of BAE Systems). Right side illustration shows optical ray trace. Displays images are collimated by reflection from the spherical visor. Note use of 'brow' mirror to relay display images.

and electrical parts can be replaced quickly and sent off for repair without the pilot losing his personalised helmet.

It should be noted, however, that a helmet with an integrated HMD system will be a relatively expensive piece of kit and will need to be treated with more care and respect by the pilots than the current aircrew helmet.

Head Tracking Systems

The need to measure the orientation of the pilot's head to determine the angular co-ordinates of the pilot's line of sight with respect to the airframe axes has already been explained. It should be noted that the problem of measuring the angular orientation of an object which can both translate and rotate can be a fundamental requirement in other applications such as robotics. In fact the solutions for head tracking systems can have general applications.

Space does not permit a detailed review of the various head tracking systems which have been developed. Most of the physical effects have been tried and exploited such as optical, magnetic and acoustic methods.

Optical tracking systems work in a number of ways, for example,

(a) Pattern recognition using a CCD camera.
(b) Detection of LEDs mounted on the helmet.
(c) Sophisticated measurement of laser generated fringing patterns.

Magnetic tracking systems measure the field strength at the helmet from a magnetic field radiator located at a suitable position in the cockpit. There are two types of magnetic head tracker system:

(a) An AC system using an alternating magnetic field with a frequency of around 10 kHz.
(b) A DC system using a DC magnetic field which is switched at low frequency.

Both systems are sensitive to metal in the vicinity of the helmet sensor. This causes errors in the helmet attitude measurement which are dependent on the cockpit installation. These errors need to be measured and mapped for the particular cockpit and the corrections stored in the tracker computer. The computer then corrects the tracker outputs for the in situ errors. The AC system is more sensitive to these errors than the DC system which is claimed to be 10 times less sensitive to metal than AC systems.

Head Down Displays

Introduction

Electronic technology has exhibited an exponential growth in performance over four decades and is still advancing. By the early 1980s, it became viable to effect a revolution in civil flight-decks and military cockpits by replacing the majority of the traditional dial type instruments with multi-function colour CRTdisplays.

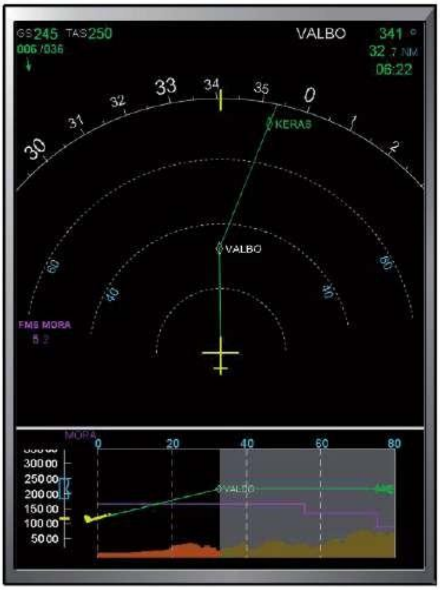
'Wall to wall' colour displays have transformed the civil flight-deck – from large 'jumbo' jets to small commuter aircraft.

Primary flight display (by courtesy of Airbus).

Civil Cockpit Head Down Displays

The displays are duplicated for the Captain and Second Pilot and being multi-function it is possible to reconfigure the displayed information in the event of the failure of a particular display surface.

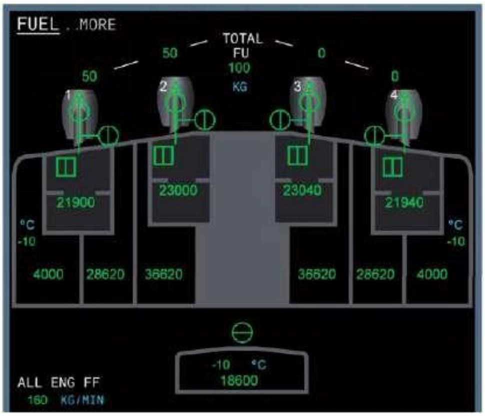
The electronic Primary Flight Display (PFD) replaces six electro-mechanical instruments: altimeter, vertical speed indicator, artificial horizon/attitude director indicator, heading/compass indicator and Mach meter. PFD formats follow the classic 'T' layout of the conventional primary flight instruments, as mentioned in Chapter 7. All the primary flight information is shown on the PFD thereby reducing the pilot scan, the use of colour enabling the information to be easily separated and emphasised where appropriate.



Navigation (or horizontal situation) displays (by courtesy of Airbus).

Engine/Warning display (by courtesy of Airbus).

world visual cues are not present (e.g. flying in cloud or at night) as the pilot can become disorientated. The normal display of attitude on the HUD is not suitable for this purpose, being too fine a scale which consequently moves too rapidly.



Systems Display (by courtesy of Airbus).

ILS or VOR formats, a map mode showing the aircraft's position relative to specific waypoints and a North up mode showing the flight plan. Weather radar displays may be superimposed over the map. The vertical flight profile can also be displayed.

Military Head Down Displays

Video head down displays now include FLIR, LLTV and maps. All the HUD functions may be repeated overlaid on the video pictures. Fuel and engine data, navigation waypoints and a host of 'housekeeping' functions (e.g. hydraulics, pressurisation) may be displayed. A stores management display is also required showing the weapons carried and their status.

It is usual to have a bezel around the display with keys. Sometimes key functions are dedicated or they may be 'soft keys' where the function is written beside the key on the display. So called tactile surfaces are being introduced using such techniques as infrared beams across the surface of the display or even surface acoustic waves where the pressure applied can also be measured to give *X*, *Y* and *Z* co-ordinates.

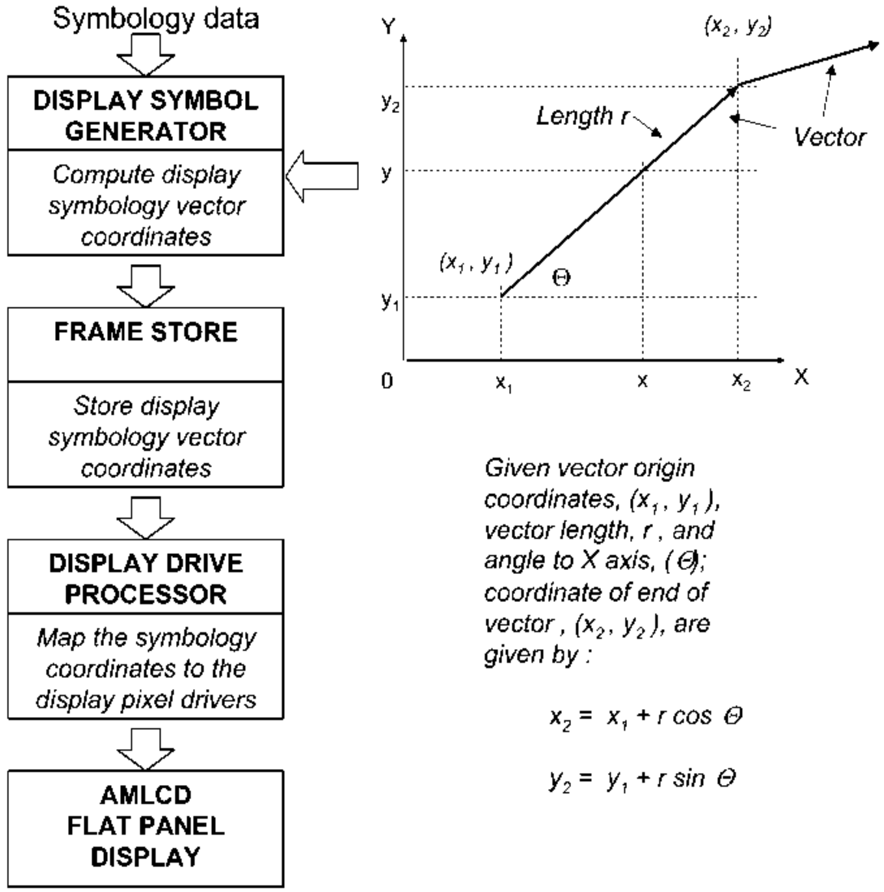
Lockheed Martin 'Lightning 2' Joint Strike Fighter cockpit (by courtesy of Lockheed Martin)

The other two displays are smaller 5×5 inch monochrome displays comprising a Systems Status Display (SSD) displaying systems status data and a Systems Control Display (SCD) displaying systems control data. Both displays have tactile data entry overlay.

The advanced cockpits for the new generation of fighter/strike aircraft have just two large colour displays as the primary head down displays.

Display Symbology Generation

Symbology such as lines, circles, curves, dials, scales, alpha-numeric characters, tabular information, map display features is drawn as a set of straight line segment approximation, or vectors (like a 'join the dots' children's pictures)



Symbology generation and display.

The system electronics use 'commercial off the shelf' graphics processing chips, frame stores and display drive processors (developed for video games, PCs, etc.).

Intelligent Displays Management

The exploitation of intelligent knowledge based systems (IKBS) technology, frequently referred to as 'expert systems', to assist the pilot in carrying out the mission is the subject of a number of very active research programmes, particularly in the United States. One of the US programmes is the 'pilot's associate program' which aims to aid the pilot of a single seat fighter/attack aircraft in a similar way to the way in which the second crew member of a two crew aircraft assists the pilot. The prime aim is to reduce the pilot work load in high work load situations.

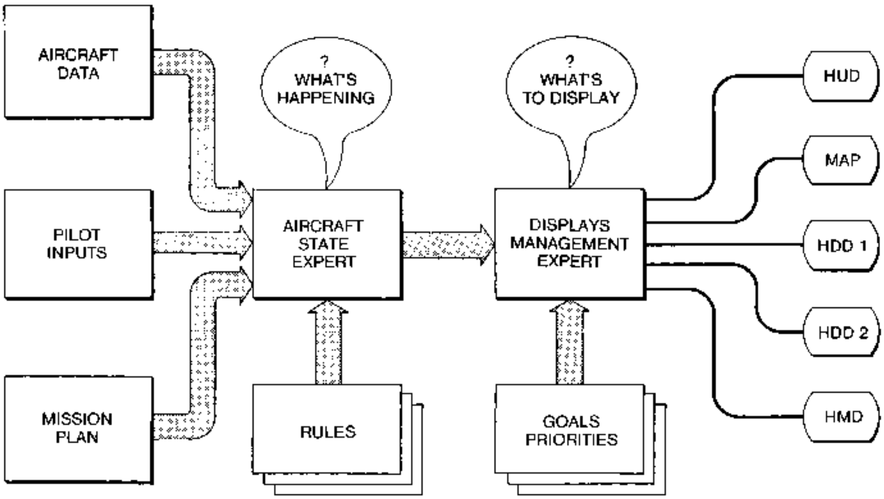
Space constraints do not permit more than a very brief introduction to this topic, which is of major importance to the next generation of military aircraft as these will use a single pilot to carry out the tasks which up to now have required a pilot and a navigator/weapons systems officer.

The exploitation of IKBS technology on the civil flight deck will follow as the technology becomes established in military applications.

A subset of all the proposed expert systems on an aircraft is an intelligent displays management system to manage the information which is visually presented to the pilot in high work load situations.

It is the unexpected or uncontrollable that leads to an excessive work load, examples being:

- The 'bounce' – interception by a counter attacking aircraft with very little warning.
- Evasion of ground threat – SAM (surface–air missile).
- Bird strike when flying at low altitude.
- Engine failure.
- Weather abort or weather diversion emergency.



Intelligent displays management (by courtesy of BAE Systems) which deduces 'what is happening' from the aircraft data, pilot inputs, threat warning systems and the mission plan by the application of an appropriate set of rules. The aircraft state expert in turn controls the 'displays management expert' which determines the information displayed on the various display surfaces: HUD, map, head down displays or HMD according to an appropriate set of goals and priorities.

Displays Technology

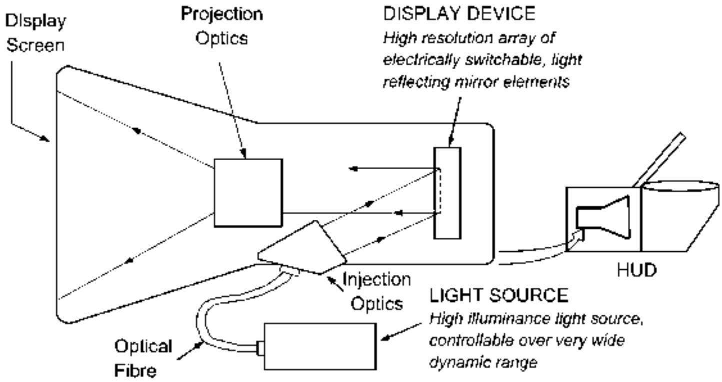
This section deals with two recent developments in HUD and HMD displays technology which will have a major influence on future HUD and HMD performance, size, weight, reliability, and both initial cost and cost of ownership.

Replacing the HUD CRT

The vast majority of HUDs currently in service worldwide use a CRT as the display source, and CRT based HUDs will remain in service for many years to come.

The continuing development, however, of projected display systems has now enabled a competitive, higher reliability replacement of the HUD display source to be produced. New HUD systems exploiting these developments are now entering service and the CRT based HUD will be eventually superseded.

It is also possible to update existing CRT based HUDs with in effect a 'new lamps for old' replacement of the CRT together with updated electronics.

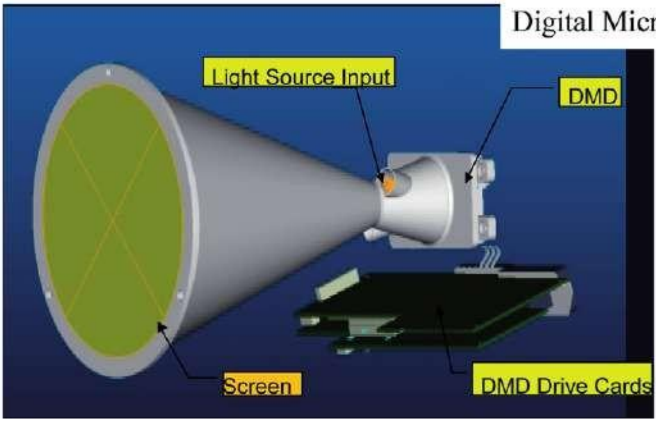


Block diagram of a HUD projected display unit.

The luminance of the light source is controllable over a very wide dynamic range to meet all ambient light conditions from very bright sunlight to night operation.

The only viable light source, until recently, was the laser and production HUDs using a laser light source are now entering service. These offer a five to tenfold increase in HUD reliability with MTBFs in excess of 20,000 hours with maturity.

The Texas Instruments Digital Micromirror



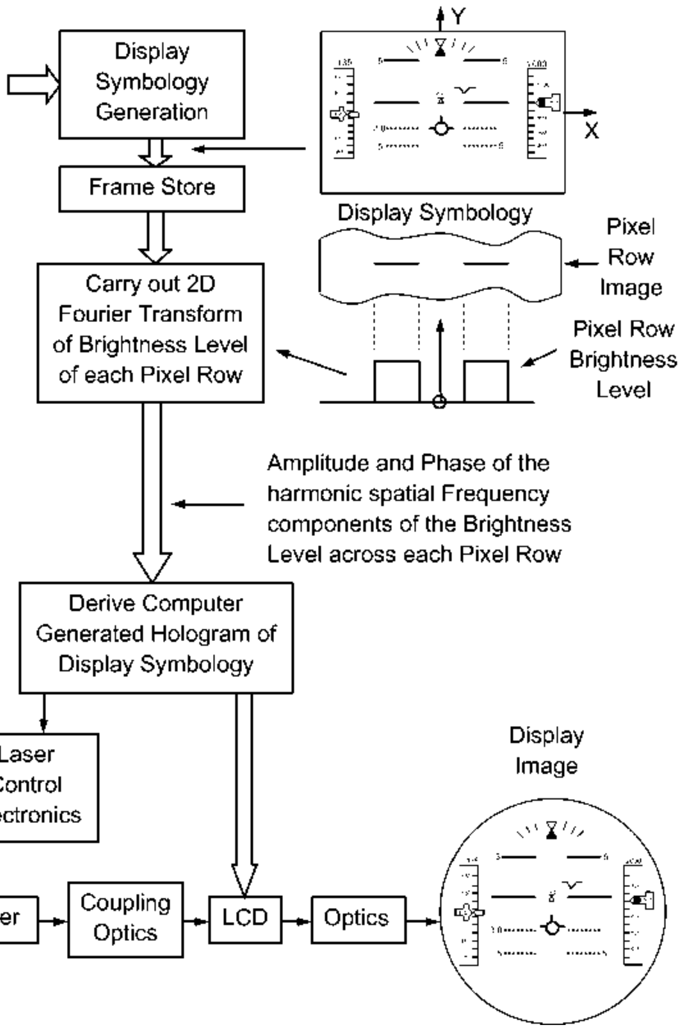
‘Digital Light Engine’ projector display unit (by courtesy of BAE Systems).

demonstrated very high reliability in a wide variety of civil and military display applications. They are also widely used in very large screen projection systems to display high definition video.

The Texas Instruments DMD™ is basically an array of over a million hinged micro-mirrors which can be individually deflected mechanically through (12°). Light modulation in a dark field projection system is achieved by tilting the micro-mirror to reflect light from the illuminating source on to, or away from the display screen. The micro-mirrors are matrix addressed and deflected by the electrostatic forces created by applying a voltage to the appropriate mirror electrodes; each micro-mirror produces a display pixel.

HMD/HUD Optical System Technology

The preceding sections have covered the basic principles, functions and design of HUDs and HMDs which are currently in service, and which will remain in service for many years to come.



Block diagram of active CGH method of generating a fault tolerant display.

Recent developments in holographic waveguide technology, however, will have a profound impact on future HMD and HUD design. Exploitation of this technology offers a major improvement in terms of mass, cost, volume, simplicity and optical performance.

UNIT - III
INERTIAL SENSORS AND GLOBAL POSITIONING SYSTEM

Introduction

Gyroscopes (hereafter abbreviated to gyros) and accelerometers are known as inertial sensors. This is because they exploit the property of inertia, namely the resistance to a change in momentum, to sense angular motion in the case of the gyro and changes in linear motion in the case of the accelerometer.
They are fundamental to the control and guidance of an aircraft. For example, in a FBW aircraft the rate gyros and accelerometers provide the aircraft motion feedback which enables a man- oeuvre command control to be achieved and an aerodynamically unstable aircraft to be stabilised by the flight control system (as explained in Chapter 4).

Gyros and accelerometers are also the essential elements of the spatial reference system or attitude/heading reference system (AHRS) and the inertial navigation system (INS). They largely determine the performance and accuracy of these systems and account for a major part of the system cost.

The AHRS and INS share common technology and operating principles. This chapter covers gyros and accelerometers and attitude derivation from strap-down gyros and accelerometers. It provides the basic background to inertial navigation and AHRS systems, which are covered in Chapter 6 ‘Navigation Systems’.

Gyros and Accelerometers

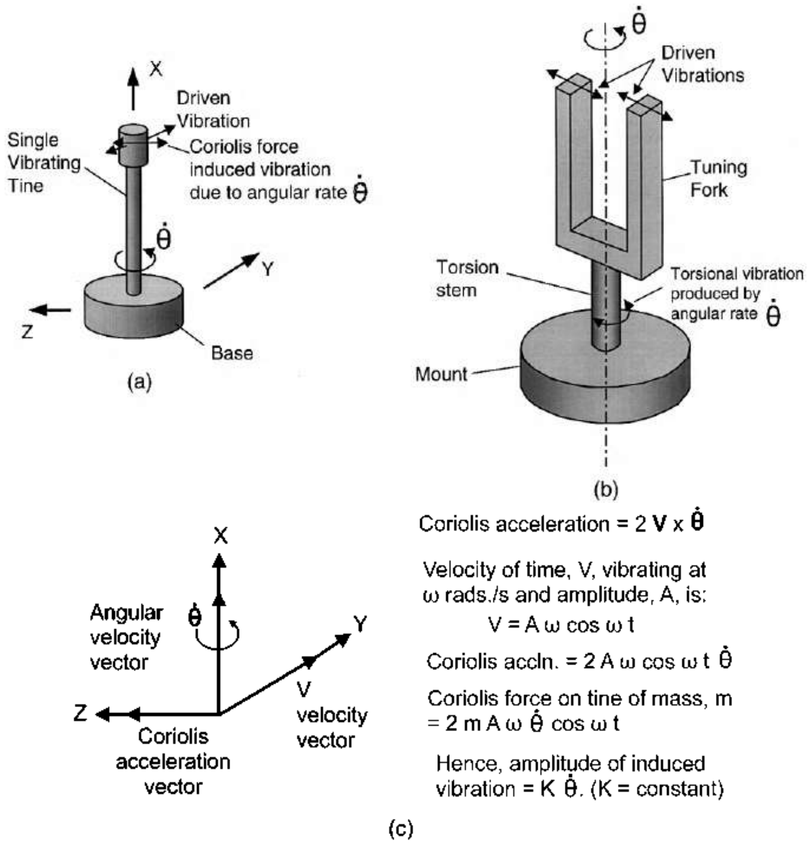
Introduction

The accuracy requirements for gyros and accelerometers can differ by several orders of magnitude depending on the application.

Not surprisingly, the costs can also differ by an order of magnitude or more.

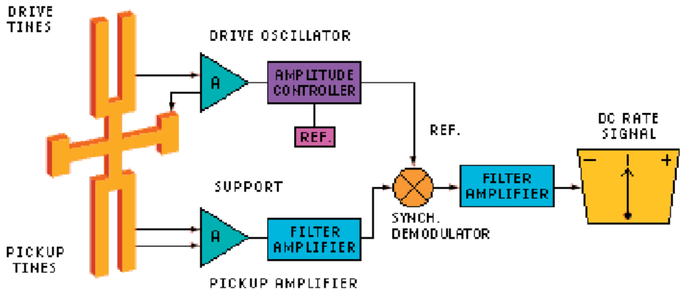
The quest for and attainment of the accuracies required for inertial navigation have involved many billions of dollars expenditure worldwide and the continual exploitation of state of the art technology. For example, the world’s first laser was demonstrated in 1960 and the first experimental ring laser gyro (RLG) was demonstrated in 1963 with an accuracy of a few degrees per hour.

The first production RLG based inertial navigation systems (which requires 0.01°/hour accuracy) went into large scale civil airline service in 1981. Strap-down RLG based IN systems now
dominate the market; a technology revolution

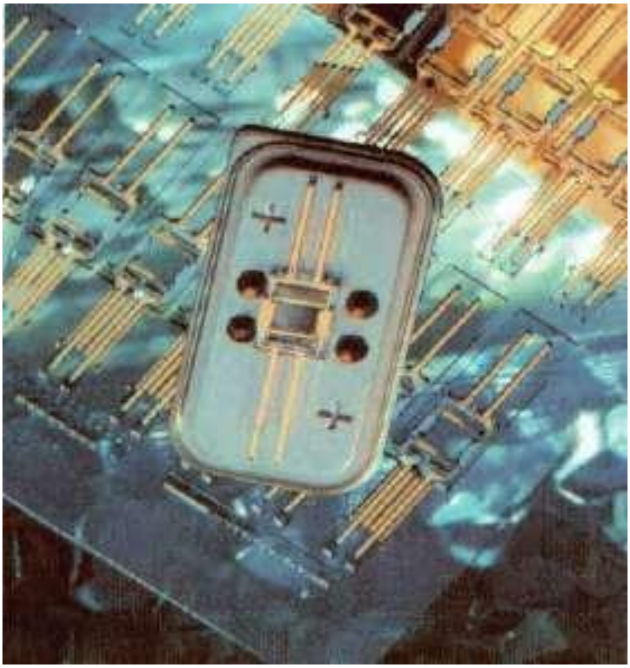


Tuning fork rate gyro.

ment suffers from the unacceptable characteristic that the smallest linear motions applied to its base cause unacceptably large errors. To overcome the effect of base motion, it is necessary to use balanced oscillations in which the oscillations of one mass are counter-balanced by equal and opposite motion of a second equal mass.



Quartz rate sensor (courtesy of Systron Donner Inertial Division).



Quartz rate sensor fabrication wafer with device overlaid (courtesy of Systron Donner Inertial Division).

Optical Gyroscopes

Introduction

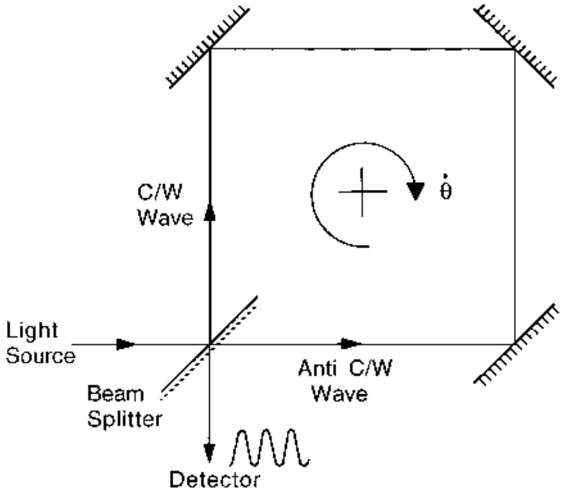
Optical gyroscopes such as the ring laser gyro and the fibre optic gyro measure angular rate of rotation by sensing the resulting difference in the transit times for laser light waves travelling around a closed path in opposite directions.

This time difference is proportional to the input rotation rate and the effect is known as the 'Sagnac effect' after the French physicist G

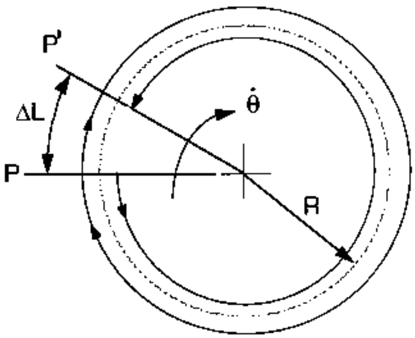
The Sagnac effect time difference, OT , between the clockwise (cw) and anti-clockwise (acw) paths is given by

$$OT = \frac{4A}{c^2}\theta$$

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



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



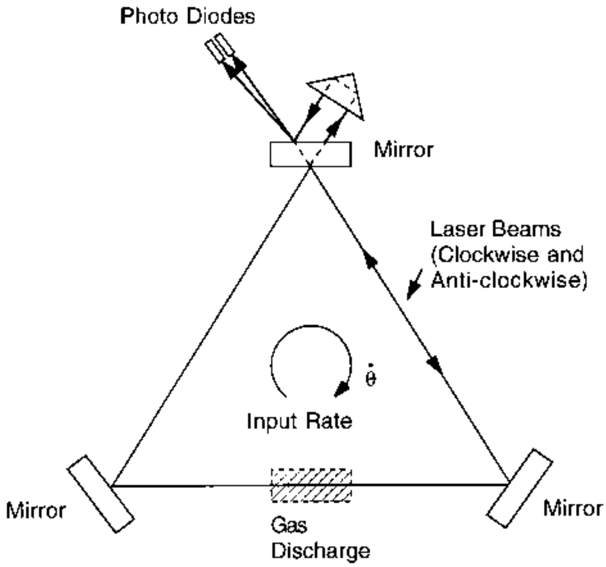
Sagnac effect.

$$OL = cOT$$
$$OL = \frac{4A}{c}\dot{\theta}$$

A rigorous derivation of the above formulae requires the use of the general theory of relativity. A simpler kinematic explanation, however, is given below for the case of a circular path in vacuo.

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

The difference in transit time,



Laser gyro schematic.

The IFOG operates basically as a Sagnac interferometer and measures the phase shift in the fringe pattern produced by the input rotation rate. The accurate measurement of very small Sagnac phase shifts has required the development of special techniques which will be explained. The IFOG was initially developed to an AHRS level of accuracy and the first large production orders for IFOG based AHRS were placed in 1991. Development to inertial accuracy has since taken place in many organisations and IFOG based IN systems are now available.

The resonator type of fibre optic gyro, or RFOG, operates in a similar manner to the RLG but with the resonant ring of optical fibre driven by an external laser diode. Two acousto-optical modulators are used to shift the laser frequency injected into the cw and acw paths respectively with the modulation frequencies being controlled by servo loops to maintain the resonance peak condition. The difference in the resonant frequencies of the two paths is then directly proportional to the input rotation rate as with the RLG. Development of the RFOG appears to be on the 'back-burner' at the present time, as the IFOG type is giving an excellent all round performance.

Attention has thus been concentrated on the RLG and the IFOG.

The Interferometric Fibre Optic Gyro

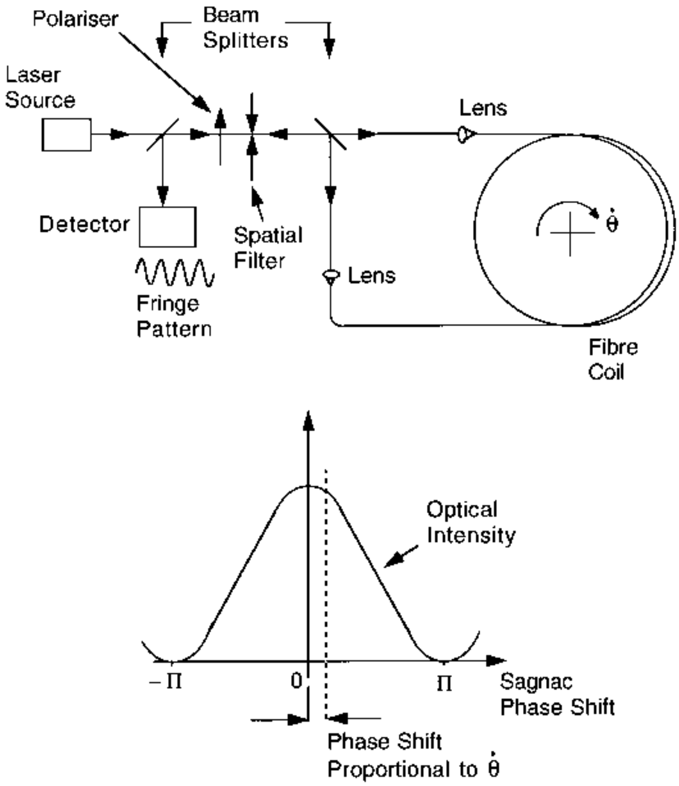
The implementation of the interferometric type of FOG is best explained in a series of stages starting with the simple basic system.

Light from the laser diode source is passed through a first beam splitter and a single optical mode is selected. The light passes through a second beam splitter and propagates in both directions around the fibre coil. In the absence of rotation, as already explained, the transit times are identical so that when the light arrives back at the second beam splitter, perfect constructive interference occurs with accompanying fringe pattern. The gyro output signal is obtained by directing the returning light via the first beam splitter to a photo-detector.

As explained, an input rotation rate about an axis normal to the plane of the coil results in a difference in the transit times between the clockwise and anti-clockwise beams as given by equation,
viz.

$$OT = \frac{4A}{c^2}\theta$$

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



Basic fibre optic gyro.

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

hence

$$OT = \frac{LD}{c^2}\theta$$

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

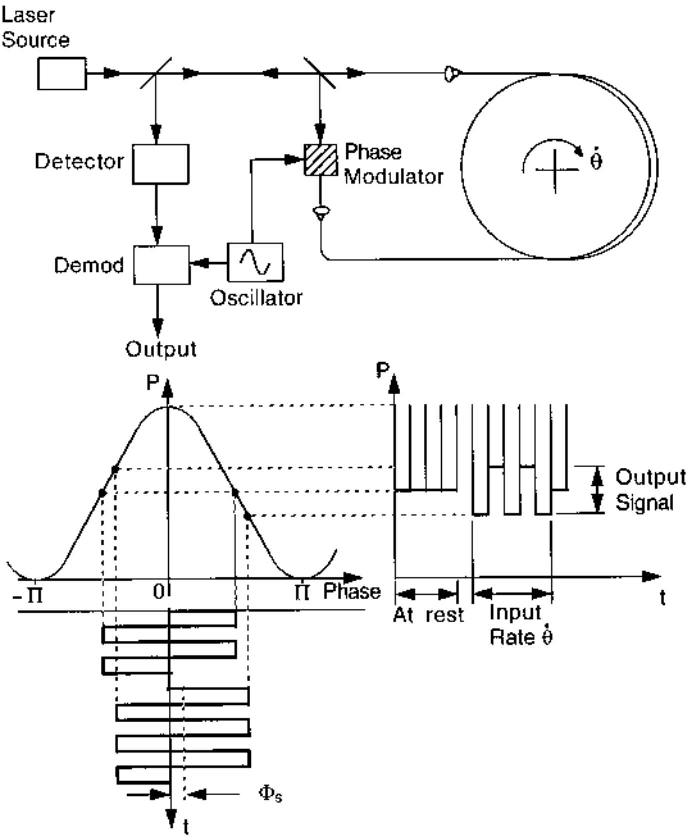
$$8_s = \omega OT$$

This phase shift between the clockwise and anti-clockwise travelling light waves results in a reduction in the intensity of the light at the detector. This change in intensity is very small for useful rotation rates and has required the development of special techniques.

For example, a typical FOG with an optical path length $L = 200$ m, coil diameter, $D = 70$ mm and source laser wavelength $= 1.3$ microns, when subjected to an input rate of 10°/sec will have a Sagnac phase shift of

$$8_s = \frac{2\pi \times 200 \times 70 \times 10^{-2}}{1.3 \times 10^{-6} \times 3 \times 10^{-8}} \times 10 = 2.26^{\circ}$$

Resolution of Earth's rate (15°/hour) requires the measurement of a phase shift of 3.38 arc seconds. Resolution of 0.01°/hour seems barely credible at first sight but is achievable as will be shown.



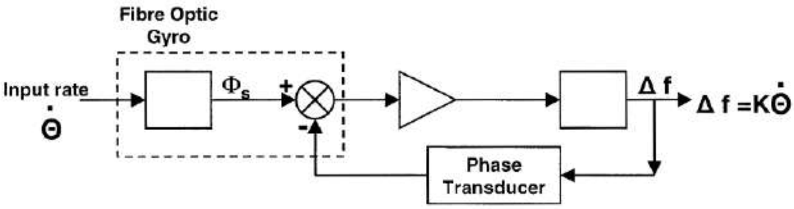
Phase modulation – open-loop IFOG.

The light intensity output signal from the photo-detector is converted to a digital value by an A to D converter followed by digital demodulation and integration. The loop is closed by driving the integrated optics phase modulator with a voltage ramp whose slope is proportional to the input rate. This ramp produces a non-reciprocal phase shift between the light waves so as to restore and maintain the sensor at the zero rotation condition.

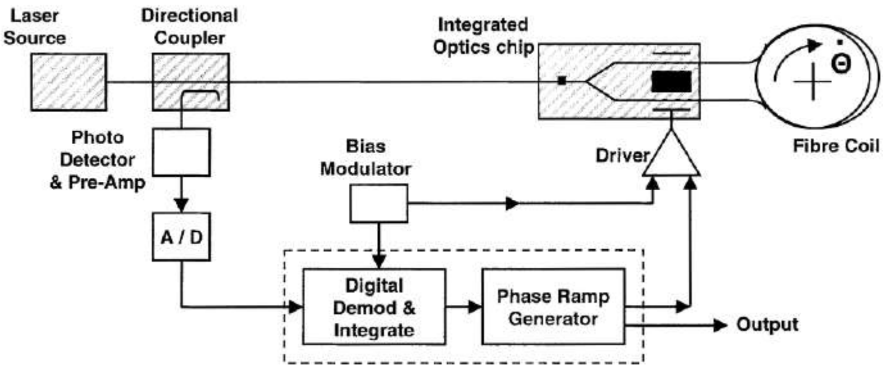
The advantages of the closed-loop system are:

(a) Output is independent of light source intensity variations as the system is always operated at null.
(b) Output is independent of the gains of individual components in the measurement system as long as a very high open-loop gain is maintained.
(c) Output linearity and stability depend only on the phase transducer.

The voltage ramp driving the phase modulator generates a frequency difference to null the phase difference produced by the input rate.



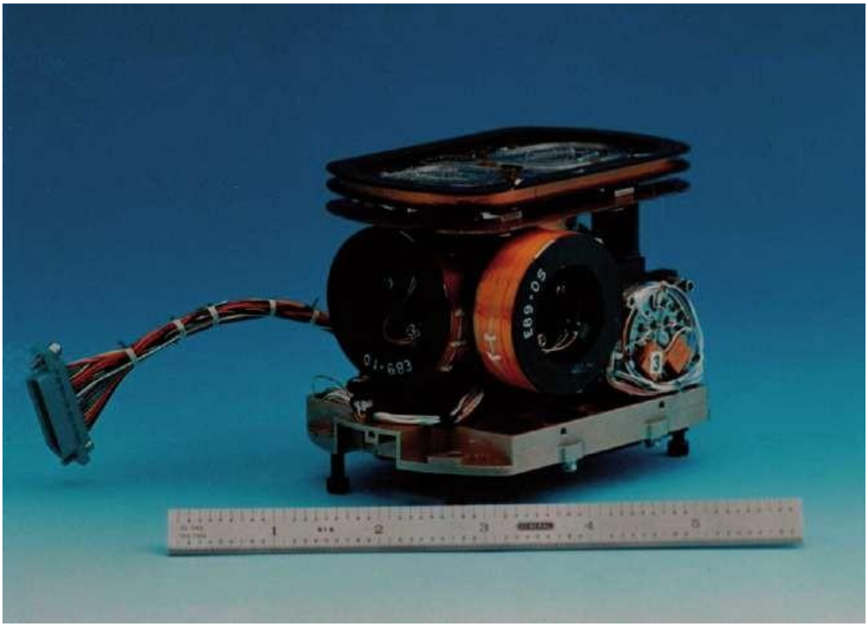
(a) Concept Diagram



(b)

Closed-loop interferometric FOG.

Frequency difference and phase shift are directly related. The phase shift, *8*, that light experiences in propagating along a single mode optical fibre of length, *L*, and refractive index, *n*, is given by



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

i.e.

$$Of = K_1\dot{\theta}$$

where $K_1 = D/n\lambda$.

The output characteristic is thus the same as the RLG and the IFOG behaves as an integrating rate gyro.

The scale factor stability is dependent on the stability of D, n and λ.

For the example FOG already used $D = 70$ mm, $\lambda = 1.3$ mm, $n = 1.5$. Substituting these values, one output pulse corresponds to an angular increment of 5.75 arc seconds.

Equation (5.11) can be shown to be the same as the RLG output equation by multiplying the numerator and denominator of (5.11) by $\pi D/4$, viz.

$$Of = \frac{D}{n\lambda}\,\frac{\pi D/4}{\pi D/4}\dot{\theta} = \frac{4A}{\lambda(n\pi D)}\dot{\theta}$$

The equivalent optical perimeter ($n\pi D$) corresponds to the RLG path length, L. The inherent simplicity of the FOG and the use of *Integrated Optic Chips* where the various electro-optical elements are integrated on to a substrate leads to low manufacturing costs and the IFOG is a highly competitive gyro in terms of cost, reliability, ruggedness and performance.

Integrated three axis versions are being produced with the three FOGs sharing the same laser diode source.

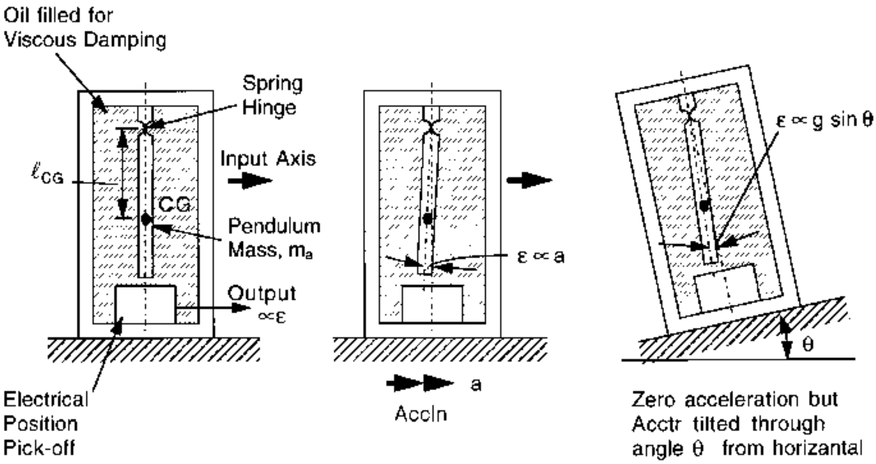
Accelerometers

Introduction – Specific Force Measurement

The acceleration of a vehicle can be determined by measuring the force required to constrain a suspended mass so that it has the same acceleration as the vehicle on which it is suspended, using Newton's law: force = mass × acceleration.

The measurement is complicated by the fundamental fact that it is impossible to distinguish between the force acting on the suspended mass due to the Earth's gravitational attraction and the force required to overcome the inertia and accelerate the mass so that it has the same acceleration as the vehicle. The vehicle acceleration, **a**, being produced by the vector sum of the external forces acting on the vehicle, namely, the propulsive thrust, **T**, lift, **L**, drag, **D**, and the gravitational force, $m\mathbf{g}$, acting on the aircraft mass, *m*.

(The bold print denote the quantities are vector quantities.)



Simple spring restrained pendulous accelerometer.

$$\mathbf{T} + \mathbf{L} + \mathbf{D} + m\mathbf{g} = m\mathbf{a}$$

$$\mathbf{a} = \frac{(\mathbf{T} + \mathbf{L} + \mathbf{D})}{m} + \mathbf{g}$$

The vector sum of the external forces excluding the gravitational force divided by the aircraft mass, that is $(\mathbf{T} + \mathbf{L} + \mathbf{D})/m$, is known as the 'specific force'. The force, $\mathbf{F_a}$, required to constrain the suspended mass, m_a, is thus given by

$$\mathbf{F_a} + m_a\mathbf{g} = m_a\mathbf{a}$$

$$\frac{\mathbf{F_a}}{m_a} + \mathbf{g} = \mathbf{a} = \frac{(\mathbf{T} + \mathbf{L} + \mathbf{D})}{m} + \mathbf{g}$$

Hence

$$\mathbf{F_a} = m_a$$

the accelerometer will thus measure the specific force component along its input axis and *not* the vehicle acceleration component.

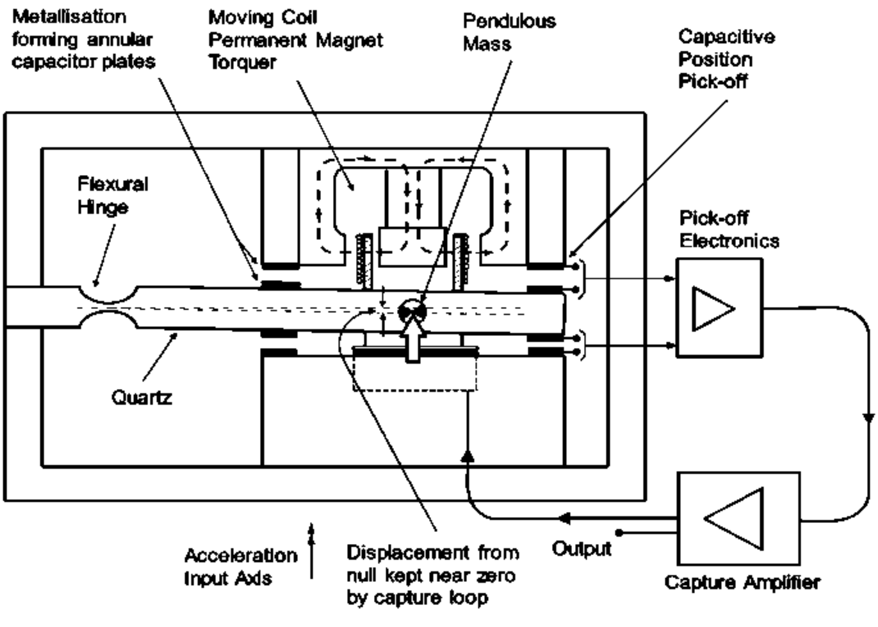
Simple Spring Restrained Pendulous Accelerometer

A simple spring restrained pendulous accelerometer is shown schematically. This comprises an unbalanced pendulous mass which is restrained by the spring hinge so that it can only move in one direction, that is along the input axis. The spring hinge exerts a restoring torque which is proportional to the angular de- flection from the null position. When the case is accelerated the pendulum deflects from the null position until the spring torque is equal to the moment required to ac- celerate the centre of mass of the pendulum at the same acceleration as the vehicle.

The transfer function of the simple accelerometer described is a simple quadratic lag filter of the type

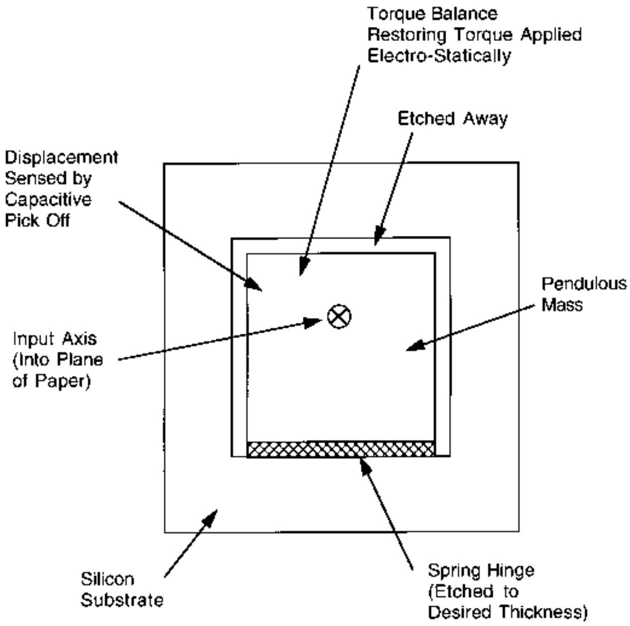
$$\frac{\text{Output}}{\text{Input Accln.}} = \frac{K_0}{D^2 + 2\zeta\,\omega_0 D + \omega_0{}^2}$$

K_0 is the accelerometer scale factor, and the undamped natural frequency ω_0



Torque balance pendulous accelerometer schematic.

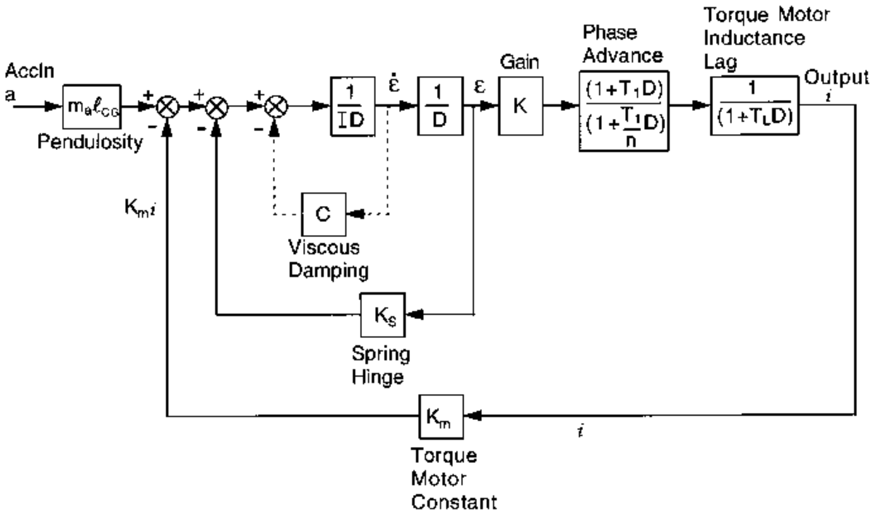
(Strictly speaking, the specific force is measured.) Pulse torque operation of the capture amplifier is generally used giving a pulse rate output which is directly proportional to the input acceleration. Each pulse represents a velocity increment, for example 0.05 m/s (0.1 knots approx.), providing an inherent integration.



'Solid state' dry accelerometer construction.

Accelerometers used in very severe environments with very high vibration and shock levels are generally oil filled to increase their robustness.

'Solid state' accelerometers fabricated from silicon using semi-conductor manufacturing technology are now being widely used for lower accuracy applications. The technology offers very small size and low manufacturing costs.



Torque balance accelerometer loop.

given by

$a = a_{max} \sin \omega t$

where a_{max} is the peak value of the vibration acceleration, a, and ω is the angular frequency.

The displacement of the pendulum from null is given by

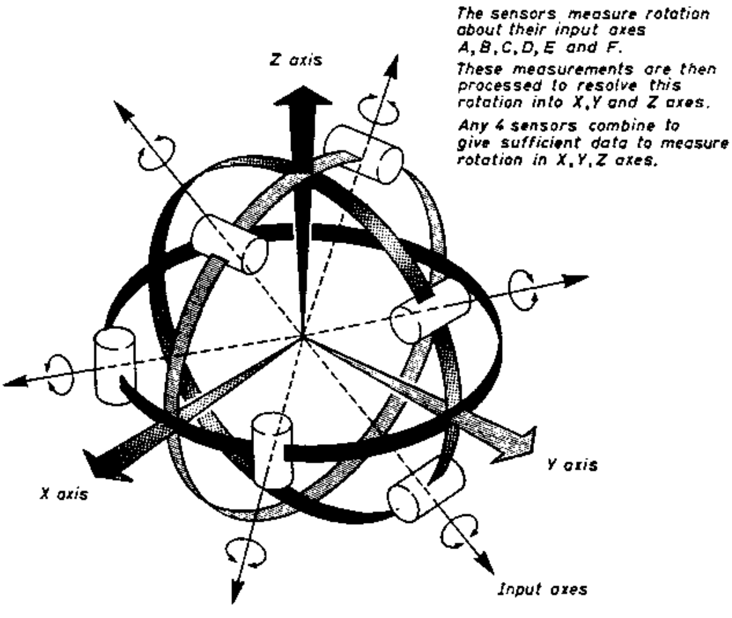
$$\varepsilon = \frac{a_{max}}{K} \sin(\omega t + 8)$$

where K is dependent on the loop gain at frequency ω and 8 is the phase shift between ε and a.

Sensed cross-axis acceleration is $\varepsilon \cdot a_{max} \sin \omega t$

$$= \frac{a^2_{max}}{K} \sin \omega t \cdot \sin(\omega t + \varphi)$$

The product term $[\sin \omega t \,.\, \sin(\omega t + \varphi)]$ has an average value of $\cos 8/2$ so that a proportion of the input vibration is rectified. This error source contributes to the 'g^2 error sensitivity' of the accelerometer as it is proportional to the square of the vibration acceleration.



Skewed axes sensors – dodecahedron configuration.

Skewed Axes Sensor Configurations

One of the attractive features of a strap-down system is that economical failure absorption configurations are available. One such configuration is an arrangement of six single axis rate gyros and six accelerometers, with their input axes skewed with respect to the principal axes

The gyro outputs denoted by letters A to F are functions of the body rates.

$$
\begin{aligned}
A &= p \sin \alpha + r \cos \alpha \\
B &= -p \sin \alpha + r \cos \alpha \\
C &= p \cos \alpha + q \sin \alpha
\end{aligned}
$$

$$
\begin{aligned}
D &= p \cos \alpha - q \sin \alpha \\
E &= q \cos \alpha + r \sin \alpha \\
F &= q \cos \alpha - r \sin \alpha
\end{aligned}
$$

where $\alpha = 31.716747^\circ$, the angle between the gyro input axes and the body axes. There are thus six equations in three unknowns which can give two independent measures of each body rate. Sophisticated algorithms can be used to isolate a failed gyro and re-combine the data from the remaining sensors in an optimum way.

There are other configurations of skewed sensors which are also used. For example, three orthogonal sensors with their input axes aligned with the aircraft principal axes and a fourth sensor with its input axis skewed at 45° to the principal axes.

Two such independent sensor sub-assemblies can tolerate three gyro failures.

GPS – Global Positioning System

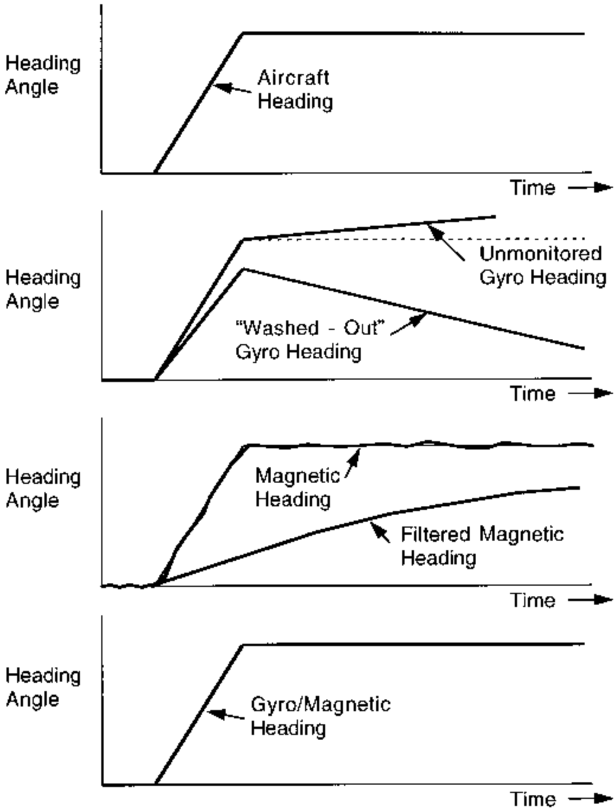
Introduction

GPS is basically a radio navigation system which derives the user's position from the radio signals transmitted from a number of orbiting satellites.

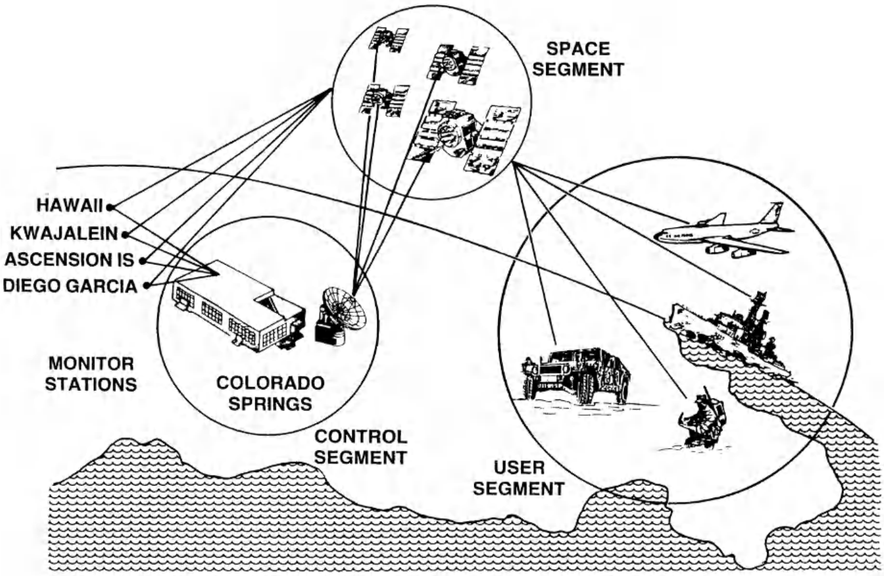
The fundamental difference between GPS and earlier radio navigation systems, such as LORAN-C (now no longer in use), is simply the geometry of propagation from ground based transmitters compared with space borne transmitters. An orbiting satellite transmitter can provide line of sight propagation over vast areas of the world. This avoids the inevitable trade-offs of less accuracy for greater range which are inherent with systems using ground based transmitters. The satellite signals also penetrate the ionosphere rather than being reflected by it so that the difficulties encountered with sky waves are avoided.

GPS provides a superior navigation capability to all previous radio navigation systems. For these reasons and also space constraints, coverage of radio navigation systems has been confined to GPS.

Satellite navigation can be said to have started with the successful launching by the Russians of the world's first orbiting satellite, SPUTNIK 1 in October 1957. The development of the first satellite navigation system TRANSIT 1, was triggered by observations made on the radio signals transmitted from SPUTNIK 1 and was initiated at the end of 1958. TRANSIT 1 resulted in a worldwide navigation system which has been in continuous operation since 1964.



Complementary filtered gyro/magnetic heading response to constant rate of turn.



The GPS system.

A very similar satellite navigation system called GLONASS has been developed by the Russians. Although designed for military applications (like GPS), it is available but is not being used elsewhere at the present time (2010).

GPS System Description

The overall GPS system comprises three segments, namely the space segment, the control segment and the user segment and is shown schematically

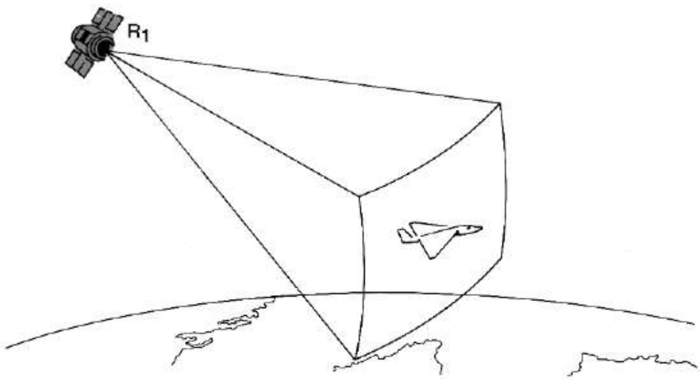
Space Segment. This comprises 24 GPS satellites placed in six orbital planes at 55° to the equator in geo-synchronous orbits at 20,000 km above the Earth. The orbit tracks over the Earth, forming an 'egg beater' type pattern.

Twenty-one satellites are required for full worldwide coverage and three satellites act as orbiting spares.

Basic Principles of GPS

The basic principle of position determination using the GPS system is to measure the spherical ranges of the user from a minimum of four GPS satellites. The orbital positions of these satellites relative to the Earth are known to extremely high accuracy and each satellite transmits its orbital position data.

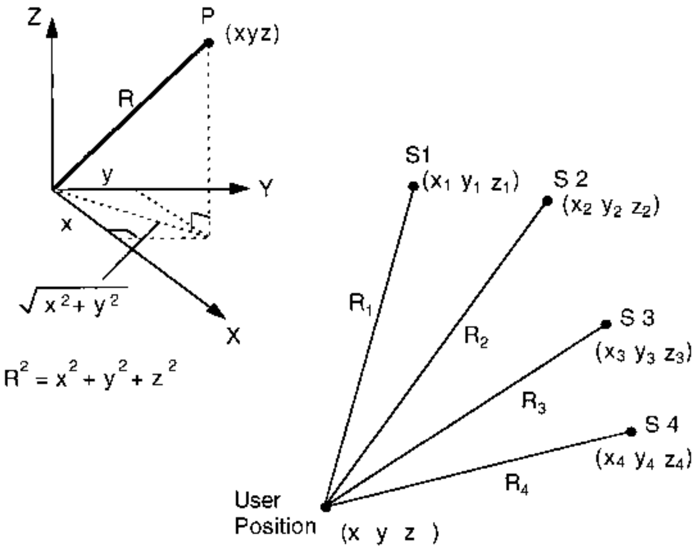
Each satellite transmits a signal which is modulated with the C/A pseudo-random code in a manner which allows the time of transmission to be recovered.



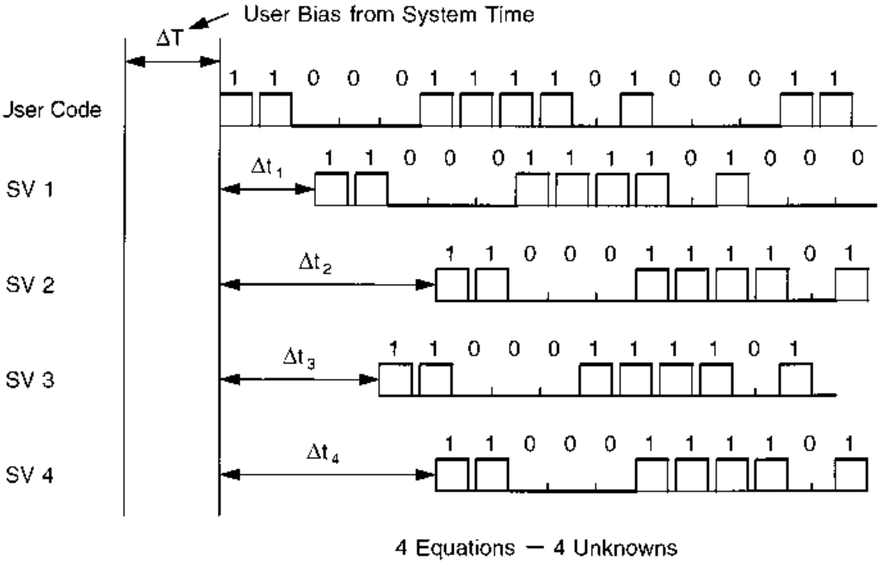
GPS spherical ranging.

The spherical range of the user from the individual transmitting satellite can be determined by measuring the time delay for the satellite transmission to reach the user.

Multiplying the time delay by the velocity of light then gives the spherical range, *R*, of the user from the transmitting satellite. The user's position hence lies on the surface of a sphere of radius, *R*.



User–satellite geometry.



GPS satellite waveforms with perfect satellite clocks.

$$R_{1p} = cOt_1$$

$$R_{2p} = cOt_2$$

$$R_{3p} = cOt_3$$

$$R_{4p} = cOt_4$$

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

$$T = c_o T$$

Hence, from basic 3D co-ordinate geometry

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

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

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

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

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

.

Each satellite time is related to GPS time by a mathematical expression and the user corrects the satellite time to the GPS time using the equation

$$t = t_{s/c} - Ot_{s/c}$$

where t is the GPS time in seconds, $t_{s/c}$ is the effective satellite time at signal transmission in seconds and $Ot_{s/c}$ is the time offset between the satellite and GPS master time.

The time offset is $Ot_{s/c}$ computed from the following equation

$$Ot_{s/c} = a_0 + a_1(t - t_{o/c}) + a_2(t - t_{o/c})2 + Ot_r$$

where a_0, a_1, and a_2 are polynomial coefficients representing the phase offset, frequency offset and ageing term of the satellite clock with respect to the GPS master time and Ot_r is the relativistic term (seconds). The parameter t is the GPS time and $t_{o/c}$ is the *epoch time* at which the polynomial coefficients are referenced and generally $t_{o/c}$ is chosen at the mid point of the fit interval. The polynomial coefficients a_0, a_1, a_2 are estimated by the control segment for each satellite clock and periodically uplinked to the satellite.

These coefficients are transmitted together with the satellite orbital position data, termed the *Ephemeris parameters*, to the navigation user equipment as navigation messages. The clock corrections for the four satellites are designated τ_1, τ_2, τ_3, and τ_4.

The spherical ranges R_1, R_2, R_3 and R_4 are thus given by

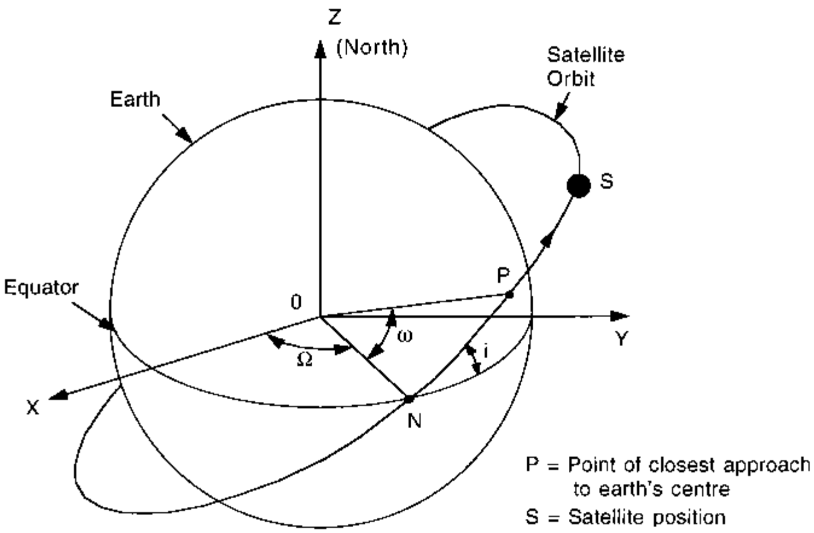
$$R_1 = c(Ot_1 + OT - \tau_1)$$

$$R_2 = c(Ot_2 + OT - \tau_2)$$

$$R_3 = c(Ot_3 + OT - \tau_3)$$

$$R_4 = c(Ot_4 + OT - \tau_4)$$

All satellite clocks are mathematically synchronised to the GPS master time by means of these clock correction terms. The error in synchronisation will grow, however, if the polynomial coefficients a_0, a_1 and a_2 are not updated periodically.



Satellite-Earth geometry. Ephemeris parameter definition.

The navigation equations are basically non-linear as can be seen in equations (6.56), (6.57), (6.58) and (6.59), but can be linearised about nominal values for their solution by applying Taylor's series approximations.

Integration of GPS and INS

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

- Calibration and correction of INS errors – the GPS enables very accurate calibration and correction of the INS errors in flight by means of a Kalman filter.
- The INS can smooth out the step change in the GPS position output which can occur when switching to another satellite because of the change in inherent errors.
- Jamming resistance – like any radio system, GPS can be jammed, albeit over a local area, although it can be given a high degree of resistance to jamming. The INS, having had its errors previously corrected by the Kalman filter, is able to provide accurate navigation information when the aircraft is flying over areas subjected to severe jamming.
- Antenna obscuration – GPS is a line of sight system and it is possible for the GPS antenna to be obstructed by the terrain or aircraft structure during manoeuvres.
- Antenna location corrections – the GPS derived position is valid at the antenna and needs to be corrected for reference to the INS location. The INS provides attitude information which together with the lever arm constants enables this

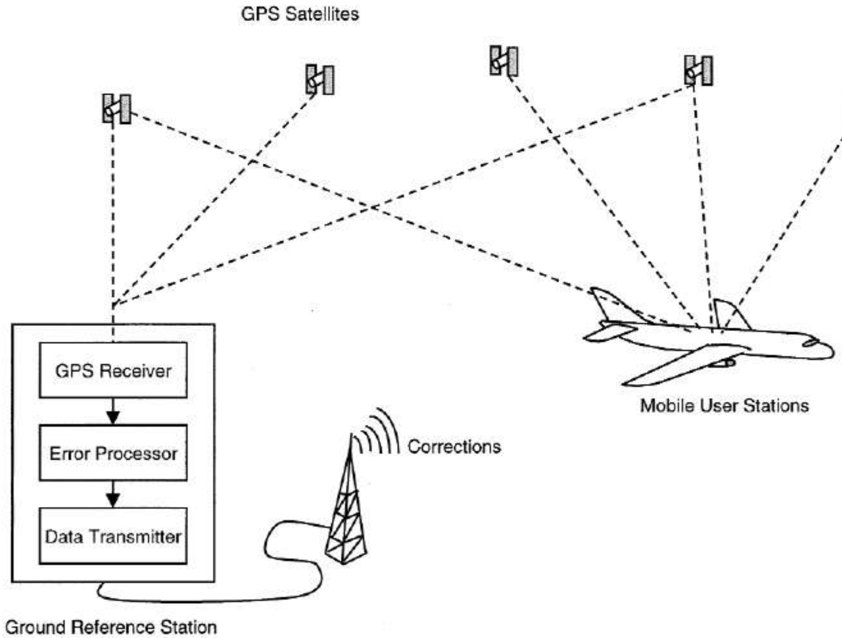
correction to be made.

Differential GPS

Introduction

As explained in the preceding section the horizontal position accuracy available to all GPS users (civil and military) is now 16 m. This was not the case, however, until 2000 when the restriction of 'Selective Availability' was removed.

Concerns about potential enemies using GPS to deliver missiles and other weapons against the US had led to a policy of accuracy denial, generally known as Selective Availability. The GPS ground stations deliberately introduced satellite timing errors to reduce the positioning accuracy available to civil users to a horizontal positioning accuracy of 100 m to a 95% probability level.



The differential GPS concept.

DGPS can be defined as:

The positioning of a mobile station in real-time by corrected (and possibly Doppler or phase smoothed) GPS pseudo ranges.

The corrections are determ- ined at a static 'reference station' and transmitted to the mobile station. A monitor station may be part of the system, as a quality check on the reference station transmissions.

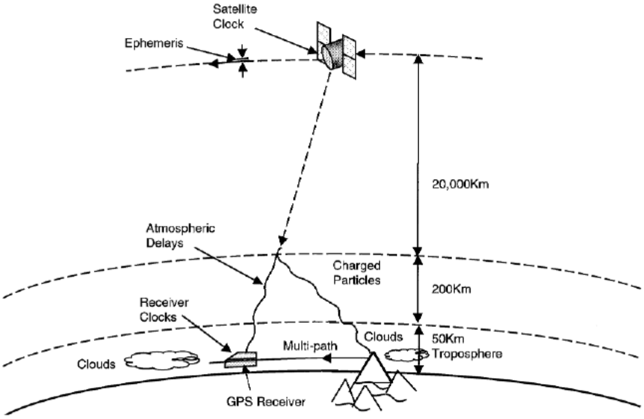
The success of DGPS can be seen from its application to new markets such as locating land vehicles used by the emergency services. Successful trials for automatic landings and taxi-way guidance have also been conducted.

It is now widely used in land and hydrographic surveying applications.

Basic Principles

The basic principle underlying DGPS is the fact that the errors experienced by two receivers simultaneously tracking a satellite at two locations fairly close to each other will largely be common to both receivers.

The basic differential GPS concept is illustrated. The position of the stationary GPS Reference Station is known to very high accuracy so that the satellite ranges can be very accurately determined, knowing the satellite ephemeris data.



GPS error sources.

The errors in the pseudo-range measurements can then be derived and the required corrections computed and transmitted to the user's receiver over a radio link. The errors present in a GPS system are illustrated

GPS satellite clocks. GPS satellites are equipped with very accurate atomic clocks and corrections are made via the Ground Stations, as explained in the preceding section.

Summary of GPS error sources.

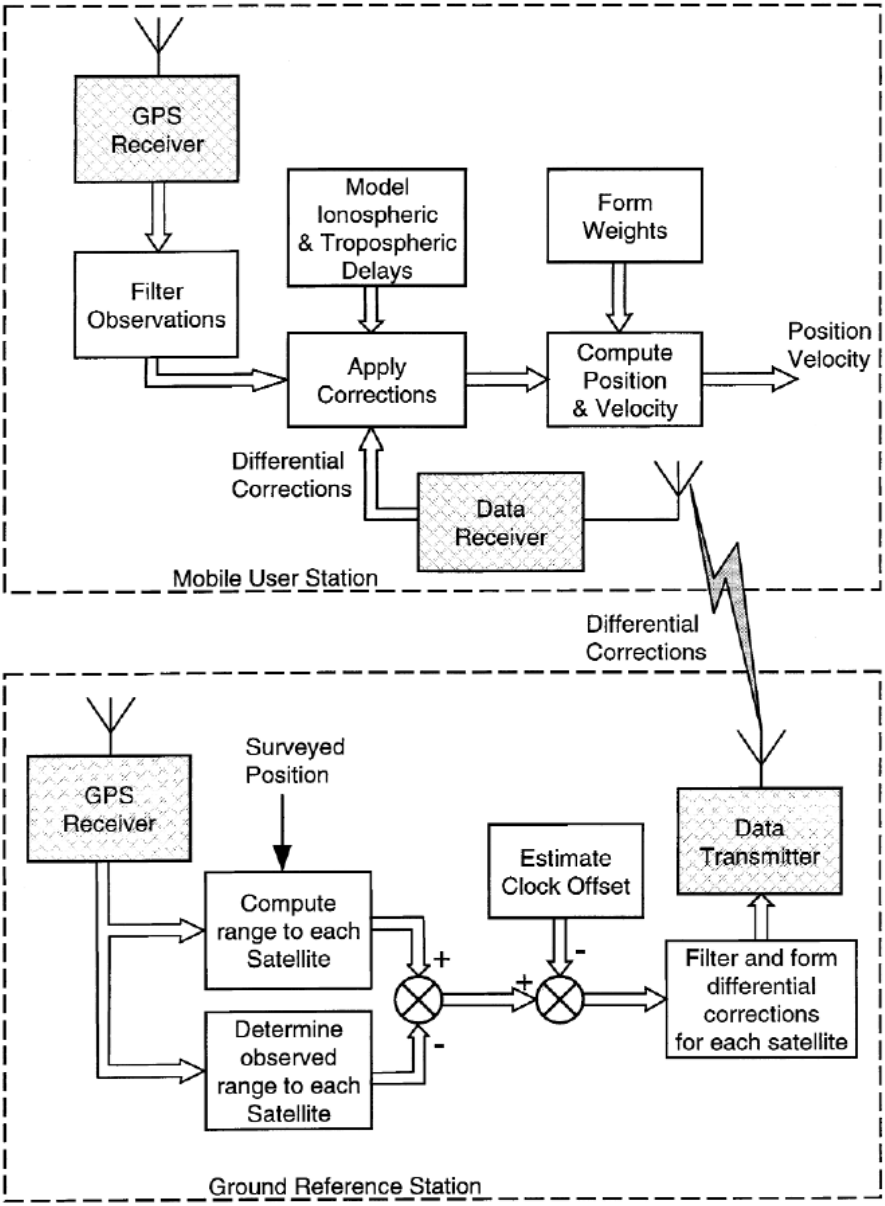
Typical error budget (in metres)		
Per satellite accuracy	Standard GPS	Differential GPS
Satellite clocks	1.5	0
Orbit errors	2.5	0
Ionosphere	5.0	0.4
Troposphere	0.5	0.2
Selective availability*	30.0	0
Receiver noise	0.3	0.3
Multi-path (reflections)	0.6	0.6
Typical positioning accuracy		
Horizontal	50.0	1.3
Vertical	78.0	2.0

3D	93.0	2.8

**Note*: Selective Availability error is shown to demonstrate the effectiveness of the DGPS technique, although the Selective Availability restriction has now been removed.

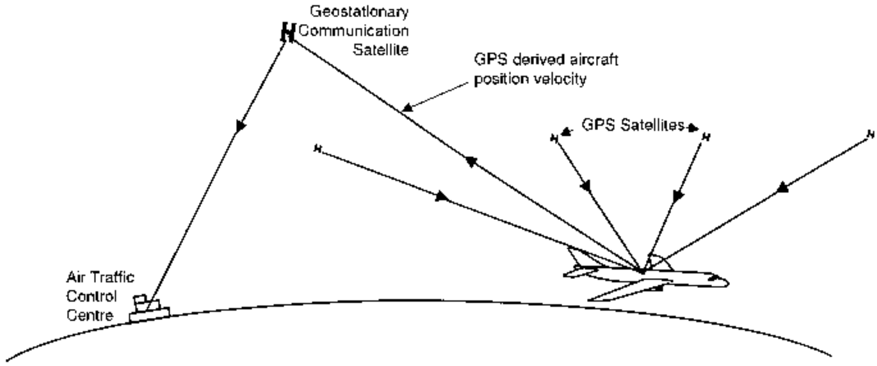
A predicted correction factor for the Earth's atmosphere path is made in the receiver but this is based on a statistical model and there are inevitable residual errors present.

Multi-path errors. The GPS satellite signal is received by the direct line of sight (LOS) path, but the signal may also be received as the result of reflections off local obstructions.



Simplified functional diagram of a generic differential GPS system.

The satellite pseudo-range measurement in the user's GPS receiver is carried out by the correlation of the received satellite signal with the receiver generated replica of the known satellite C/A code using a code tracking loop. The accuracy of the time measurement by the C/A code tracking loop is inevitably ultimately limited by noise sources such as multi-path reception.



Remote air traffic control.

GPS receivers also employ carrier tracking loops to monitor the Doppler shifted carrier component of the satellite navigation signals. The Doppler shift is proportional to the radial velocity of the receiver relative to the satellite. The velocity of the user vehicle is computed from these Doppler shifts.

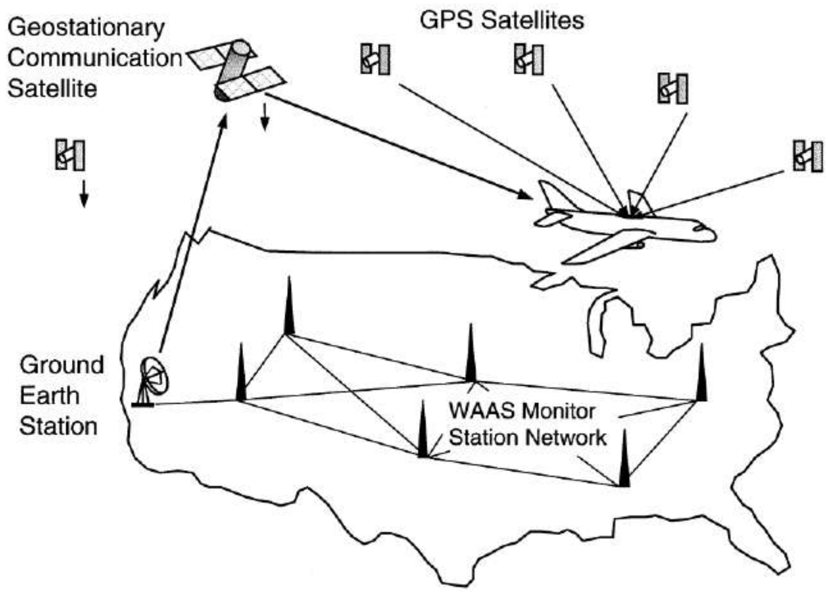
The Ground Reference GPS Station in the DGPS system uses these Doppler frequency shift measurements to improve the accuracy of the pseudo-range measurements. Integrating the Doppler shift over an interval provides a measurement of the change in satellite-to-receiver range.

Future Augmented Satellite Navigation Systems

The advent of satellite navigation systems and satellite communication links has provided new capabilities for aircraft precision navigation, particularly in civil operations.

Providing the integrity and accuracy requirements can be met, satellite navigation systems are able to support all phases of flight including all-weather precision approaches to airports not equipped with ILS (or MLS) installations.

Successful concept proving trials were, in fact, conducted by the UK Air Traffic Control authorities in conjunction with British Airways around 1996. The trials used the on-board GPS receivers and SAT COM radios in a British Airways Boeing 747 airliner to monitor the aircraft flight paths on normal commercial flights to the West Indies.



Wide area augmentation system, WAAS, concept.

Extensive trials of a WAAS evaluation system have been conducted by the FAA over the last decade, with very promising results. It appears very likely that the full system will be built and become operational over the next decade.

UNIT - IV
NAVIGATION RANGING AND LANDING SYSTEMS

Introduction and Basic Principles

The dictionary definition of navigation is a good one.

> *Navigation* – The act, science or art of directing the movement of a ship or aircraft.

Navigation thus involves both control of the aircraft's flight path and the guidance for its mission.

The measurement of the aircraft's attitude with respect to the horizontal plane in terms of the pitch and bank angles and its heading, that is the direction in which it is pointing in the horizontal plane with respect to North, is essential for both control and guidance.

This information is vital for the pilot in order to fly the aircraft safely in all weather conditions, including those when the normal visibility of the horizon and landmarks is poor or not available, for example in haze or fog conditions, flying in cloud and night flying. Attitude and heading information is also essential for the key avionic systems which enable the crew to carry out the aircraft's mission. These systems include the autopilot system (e.g., Attitude and Heading Hold modes, Autoland, etc.) navigation system and the weapon aiming system. The information is also required for pointing radar beams and infrared sensors.

.

In order of increasing accuracy these are:

1. *Air data based DR navigation*. The basic information used comprises the true airspeed (from the air data computer) with wind speed and direction (forecast or estimated) and the aircraft heading from the Attitude Heading Reference System, (AHRS).
2. *Doppler/heading reference systems*. These use a Doppler radar velocity sensor system to measure the aircraft's ground speed and drift angle. The aircraft heading is provided by the AHRS.
3. *Inertial navigation systems*. These derive the aircraft's velocity components by integrating the horizontal components of the aircraft's acceleration with respect to time. These components are computed from the outputs of very high accuracy gyroscopes and accelerometers which measure the aircraft's angular and linear motion.
4. *Doppler inertial navigation systems*. These combine the Doppler and INS outputs, generally by means of a Kalman filter, to achieve increased DR navigation accuracy.

The primary DR navigation system which is also the primary source of very accurate attitude and heading information is the Inertial Navigation System (INS). The term Inertial Reference System (IRS), is also used in civil aircraft. The IRS can have a lower inertial navigation accuracy of up to 4 NM/hour error compared with 1 to 2 NM/hour for a typical INS.

The attitude and heading accuracy, however, is still very high. The terminology INS/IRS is used in this chapter to show they are essentially the same.

The main position fixing navigation systems in current use are briefly summarised below.

(1) Range and Bearing ('R/θ') Radio Navigation Aids

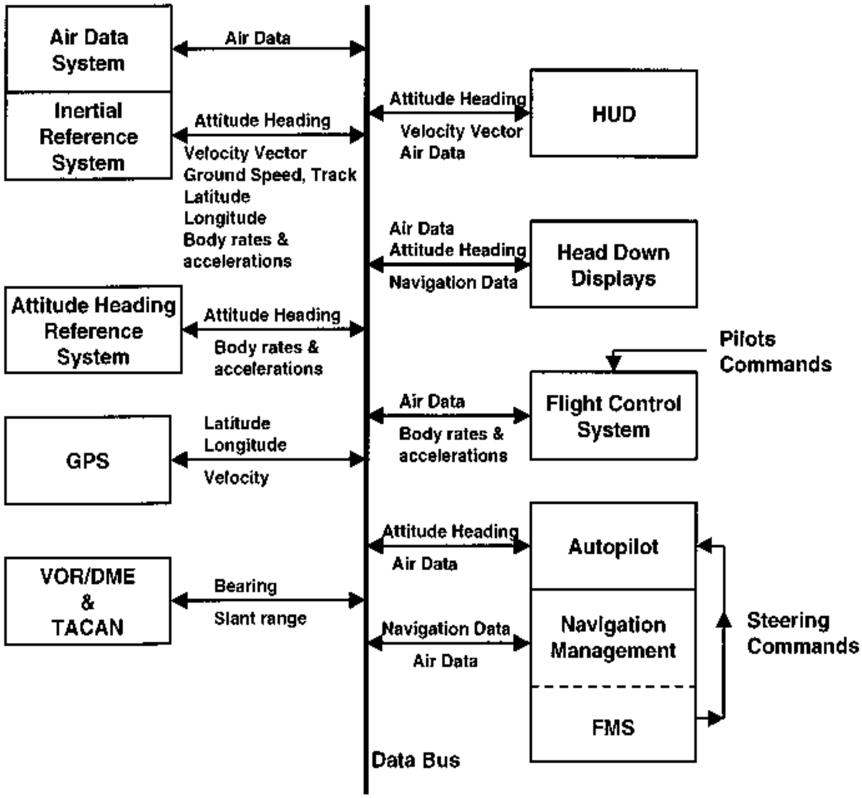
These comprise VOR/DME and TACAN.

VOR (VHF omni-directional range) is an internationally designated short-distance radio navigation aid and is an integral part of Air Traffic Control procedures.

(2) Satellite Navigation Systems – GPS (Global Position System)

GPS is the most important and accurate position fixing system developed to date. It is being used by every type of vehicle – aircraft – ships – land vehicles. Civilian use is now very widespread, for example, GPS receivers are fitted in very many cars, vans and lorries and they are readily affordable for ramblers.

The equipment required by the GPS user is entirely passive and requires a GPS receiver only. Electronic miniaturisation has enabled very compact and light weight GPS receivers to be produced.



Navigation system information flow to user systems.

(3) Terrain Reference Navigation (TRN) Systems

Terrain reference navigation systems derive the vehicle's position by correlating the terrain measurements made by a sensor in the vehicle with the known terrain feature data in the vicinity of the DR estimated position. The terrain feature data is obtained from a stored digital map database.

Basic Navigation Definitions

A very brief review of the terms and quantities used in navigation is set out below. Position on the Earth's surface is generally specified in terms of *latitude* and *longitude* co-ordinates which provide a circular grid over the surface of the Earth. The Earth is basically a sphere – the variation in the radius of the Earth is only about 40 NM in a radius of 3,438 NM at the equator, being slightly flattened at the poles (this variation is taken into account in the navigation computations).

Basic DR Navigation Systems

The basic principles of deriving a DR navigation position estimate are explained below. The following quantities are required:

1. Initial position – latitude/longitude.
2. The northerly and easterly velocity components of the aircraft, V_N and V_E.

Referring to figure it can be seen that the rate of change of latitude is

$$\dot{\lambda} = \frac{V_N}{R}$$

the rate of change of longitude is

$$\dot{\mu} = \frac{V_E}{R\cos\lambda}$$

i.e.

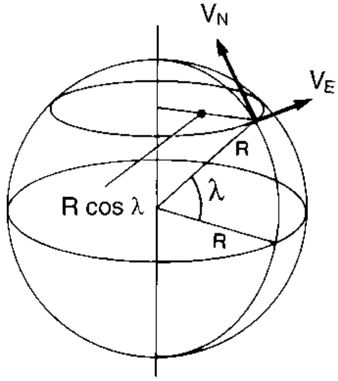
$$\dot{\mu} = \frac{V_E}{R}\sec\lambda$$

The change in latitude over time, t, is thus equal to $1/R\int_0^t V_N\,dt$ and hence the present latitude at time t can be computed given the initial latitude λ_0.

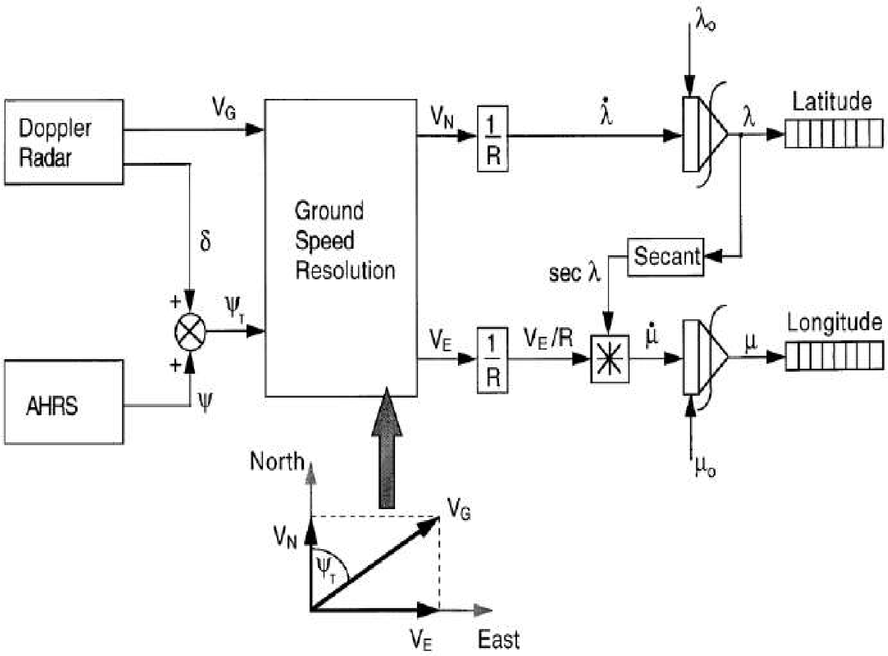
Similarly, the change in longitude is equal to $1/R\int_0^t V_E \sec\lambda\,dt$ and hence the present longitude can be computed given the initial longitude, μ_0, viz.

It can be seen that a mathematical singularity is approached as λ approaches 90° and sec λ approaches infinity. This method of computing the latitude and longitude of the DR position is hence limited to latitudes below 80°.

A different co-ordinate reference frame is used to deal with high latitudes as will be explained later.



Derivation of rates of change of latitude and longitude.



Doppler/heading reference DR navigation system.

Hence

$$V_N = V_G \cos \psi_T$$

$$V_E = V_G \sin \psi_T$$

In the case of an air data based DR navigation system the northerly and easterly velocity components of the aircraft can be derived as follows:

1. The horizontal velocity component V_H of the true airspeed V_T is obtained by resolving V_T through the aircraft pitch angle θ, viz.

$$V_H = V_T \cos\theta$$

2. The northerly and easterly velocity components of the airspeed are then derived by resolving V_H through the aircraft heading angle ψ, viz.

$$\text{Northerly airspeed} = V_H \cos\psi$$

$$\text{Easterly airspeed} = V_H \sin\psi$$

3. The forecast (or estimated) wind speed V_W and direction ψ_W is resolved into its northerly and easterly components, viz.

$$\text{Northerly wind component} = V_W \cos\psi_W$$

$$\text{Easterly wind component} = V_W \sin\psi_W$$

4. The northerly and easterly velocity components of the aircraft are then given by:

$$V_N = V_H \cos\psi + V_W \cos\psi_W \tag{6.5}$$

$$V_E = V_H \sin\psi + V_W \sin\psi_W \tag{6.6}$$

Such a system provides a reversionary DR navigation system in the absence of Doppler (or an INS) and would generally be used in conjunction with a radio navigation system.

Inertial Navigation

Introduction

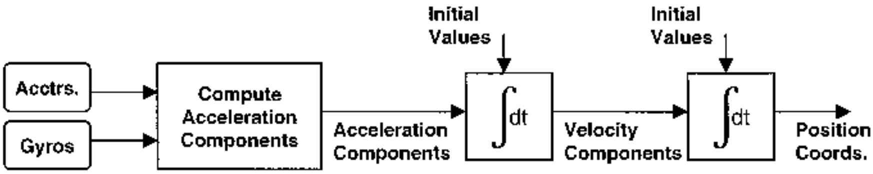
It is instructive to briefly review the reasons for the development of inertial navigation and its importance as an aircraft state sensor.

The attributes of an ideal navigation and guidance system for military applications can be summarised as follows:

- High accuracy
- Self-contained
- Autonomous – does not depend on other systems
- Passive – does not radiate
- Unjammable
- Does not require reference to the ground or outside world.

In the late 1940s these attributes constituted a ‘wish list’ and indicated the development of inertial navigation as the only system which could be capable of meeting all these requirements. It was thus initially developed in the early 1950s for the navigation and guidance of ballistic missiles, strategic bombers, ships and submarines (Ships Inertial Navigation System, SINS).

- Accurate position in whatever coordinates are required – e.g. latitude/longitude, etc.
- Ground speed and track angle.
- Euler angles: heading, pitch and roll to very high accuracy.



Basic principles of inertial navigation.

- Aircraft velocity vector (in conjunction with the air data system).

Accurate velocity vector information together with an accurate vertical reference are essential for accurate weapon aiming and this has led to the INS being installed in military strike aircraft from the early 1960s onwards as a key element of the navigation/weapon aiming system.

Initial Alignment and Gyro Compassing

Inertial navigation can only be as accurate as the initial conditions which are set in. It is therefore essential to know the orientation of the accelerometer measuring axes with respect to the gravitational vector, the direction of true North, the initial position and the initial velocity components to very high accuracy.

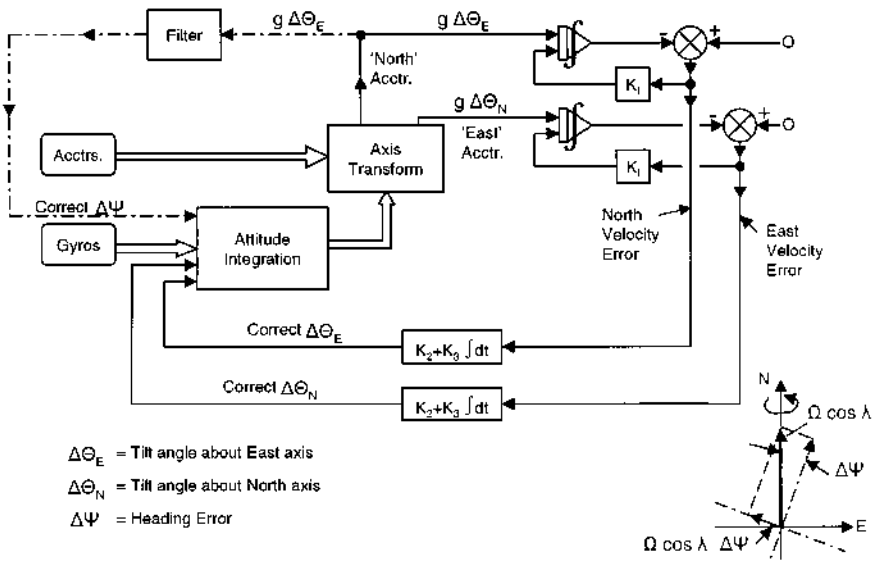
The two basic references used to align an inertial system are the Earth’s gravitational vector and the Earth’s rotation vector.

The initial alignment process is basically the same in a stable platform and strap-down INS.

The difference being that in a stable platform INS, the stable platform is physically rotated to bring it into alignment with the local NED axes by applying precession torques to the vertical and azimuth gyros on the platform.

It is thus easier to visualise (literally). Whereas the strap-down system carries out the axis rotations within the system computer to create, in effect, a virtual stable platform as explained earlier.

The North pointing virtual accelerometer will hence be tilting away from the horizontal at a rate $K \cos \lambda . O\psi$ in the absence of any levelling or gyro compassing



Fine levelling and gyro compassing loops.

loops. (This is assuming the appropriate corrections are made to compensate for $K\cos\lambda$ and $K\sin\lambda$ respectively.)

The gyro compassing loop adjusts the computed heading until the east component of the gyro angular rate measurement in the horizontal plane is nulled, the angular rate of rotation about the East axis being estimated from the summed East axis

$$O\theta_E = \ K\cos\lambda . O\psi dt)$$

The allowable gyro drift rate uncertainty can be determined from the accuracy required of the heading alignment. For example, if an accuracy of 0.1° is required for a latitude of 45°, then the component of the Earth's rate sensed at this latitude with a misalignment of 0.1° is equal to $(0.1/57.3)\sin 45°$ degrees per hour, that is 0.017 degrees per hour.

It can be seen that the magnitude of the component of Earth's rate to be sensed decreases with increasing latitude, so that gyro compassing is effectively restricted to latitudes below 80°.

The major factors which affect alignment accuracy and alignment times are:

(a) Initial tilt.
(b) Aircraft movements, e.g., effect of wind gusts etc.
(c) Accelerometer bias errors and gyro drift rates.
(d) Change of the above quantities (c) with time as the system warms up.
(e) Accelerometer resolution and gyro threshold.

The loop gains in the levelling and gyro compassing loops are generally controlled by means of a Kalman filter to give an optimal alignment process. Typical alignment times are of the order of seven minutes for full accuracy IN performance.

Reduced alignment times are sometimes used and the system corrected to give full IN accuracy by subsequent position fixes using a position fixing navigation system (e.g., GPS).

Effect of Azimuth Gyro Drift

The effect of azimuth gyro drift is to cause position errors which build up with time and are a function of azimuth error, aircraft velocity and latitude.

From the previous section on gyro compassing it can be seen that an azimuth error $O\psi$, will effectively inject a drift of $K \cos \lambda . O\psi$ about the East axis. An error of 0.2° for instance, would generate an effective drift of 0.033°/hour (two minutes of arc/hour approximately).

=

Vertical Navigation Channel

It has been shown in Section 6.2.2 that Schuler tuning effectively constrains the error build up with time in the horizontal channels of the INS. Unfortunately, however, there is no such action in the vertical channel and small errors in computing the vertical acceleration correction terms are integrated with time – in fact the vertical distance error builds up exponentially with time. As explained earlier, the principal correction terms for the vertical accelerometer channel are:

(i) Gravity – gravity varies with altitude according to the inverse square law. The gravitational acceleration at an altitude H is given by

$$g = g_0 \frac{R_0^2}{(R_0 + H)^2}$$

where g_0 = gravitational acceleration at the earth's surface.

Errors in correcting for the variation in gravitational acceleration with height result in a vertical distance error which builds up as cosh function of time. Also, gravity varies over the Earth's surface and has anomalous values in certain places.

(ii) Centrifugal acceleration – a vehicle moving over the Earth's surface is describing a circular trajectory in space and so will experience a centrifugal acceleration $(V_E^2 + V_N^2)/R$ where V_E and V_N are the easterly and northerly velocity components of the vehicle respectively.

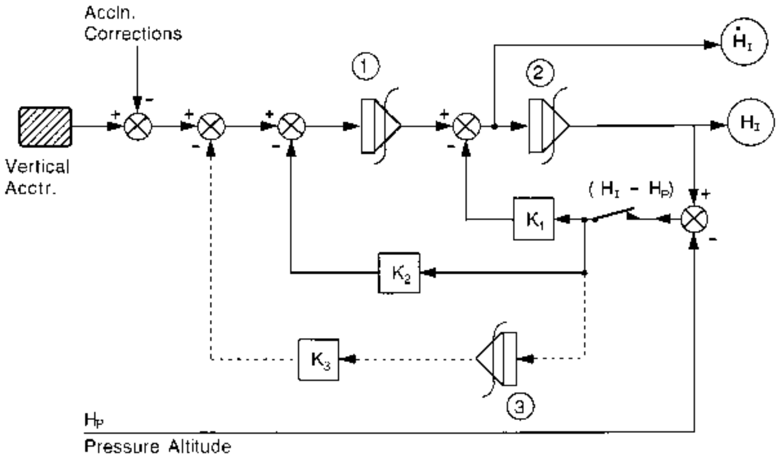
(iii) Coriolis acceleration – the component of the Earth's rotation rate $K \cos \lambda$ about the North axis is combined with the easterly velocity V_E and generates a Cori-
Consider first the simple second-order mixing system shown in bold. The baro-inertial feedback loop is capable of being opened and the gains K_1 and K_2 which determine the damping and response times respectively are also capable of being varied. The system response can be shown to be given by

$$H_I = \frac{D^2}{(D^2 + K_1 D + K_2)} \cdot H + \frac{(K_2 + K_1 D)}{(D^2 + K_1 D + K_2)} \cdot H_P + \frac{1}{(D^2 + K_1 D + K_2)} \cdot Oa_D \tag{6.21}$$

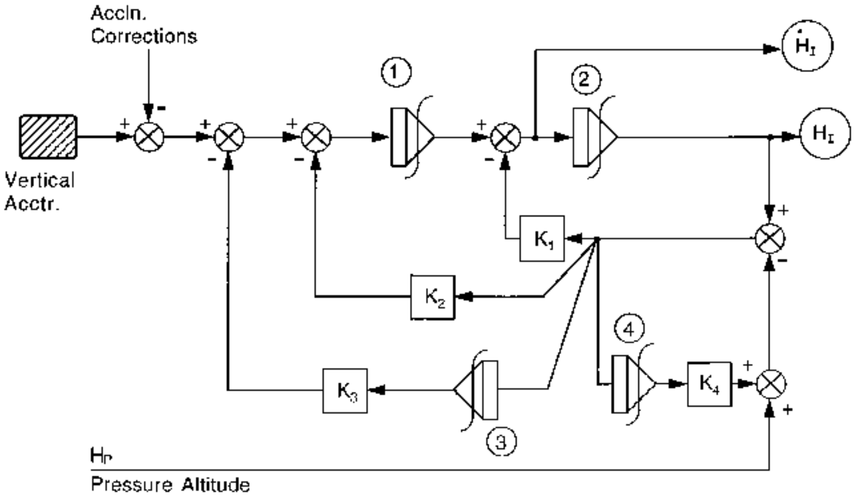
where H is the true height, H_P is the pressure altitude, Oa_D is the uncorrected vertical acceleration error and H_I is the baro/inertial mixed height.

The inertially derived component of H_I , that is

$$\frac{D^2}{(D^2 + K_1 D + K_2)} \cdot H$$



(a)



(b)

(a) Baro-inertial mixing. Simple second-order baro-inertial mixing shown in bold lines. Integral term correction for inertial drift shown in dotted lines. (b) Fourth-order baro-inertial mix- ing (with correction for inertial channel drift and pressure altitude error).

couples the inertially measured height through a second-order high pass filter transfer function which responds to dynamic changes in height without lag, but 'DC blocks' the steady-state value.

The barometrically derived component of H_I, that is

$$\frac{(K_2 + K_1 D)}{(D^2 + K_1 D + K_2)} \cdot H_P$$

couples the barometrically derived height through a low pass filter which smooths the pressure height signal and attenuates the noise present in the signal. The dynamic response is thus slugged by the low pass filter but the steady-state value of the pressure height is not affected. The combination of the two components results in a fast response with a low noise content.

Choice of Navigation Co-ordinates

The choice of a navigation co-ordinate system reduces practically to either:

(a) Spherical co-ordinates (latitude/longitude).
(b) Direction cosines.

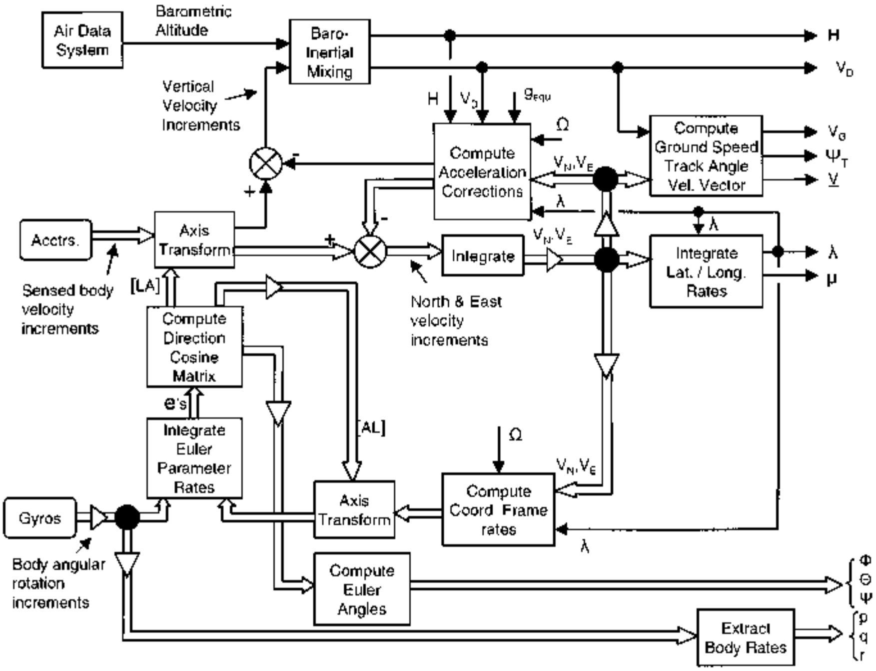
A spherical co-ordinate system has two mathematical 'poles' or singularities where the value of one co-ordinate (the longitude) is indeterminate and near which the high rates of change of that co-ordinate make any system implementation impractical (longitude rate $\dot{\mu} = (V_E/R) \sec \lambda$ becomes infinite when $\lambda = 90°$).
With the latitude/longitude co-ordinate system (also known as the geographical or 'geodetic' system) these poles are the true North and South poles of the Earth.
This co-ordinate system is satisfactory provided there is no requirement for navigation in the polar regions. It also provides the latitude/longitude information for display to the flight crew.

When polar navigational capability is required there are three possible ways to overcome the difficultly, namely:

(a) A spherical co-ordinate system with the poles removed to some other regions.
(b) A 'unipolar' system.
(c) Direction cosines.

The spherical co-ordinate system with transferred poles gives rise to added complications in the ellipsoidal corrections because of the asymmetry which results from the misalignment between the main co-ordinate axis and the axis of the geographic ellipsoid. More seriously, it is a solution of limited usefulness because the singularities still exist.

.



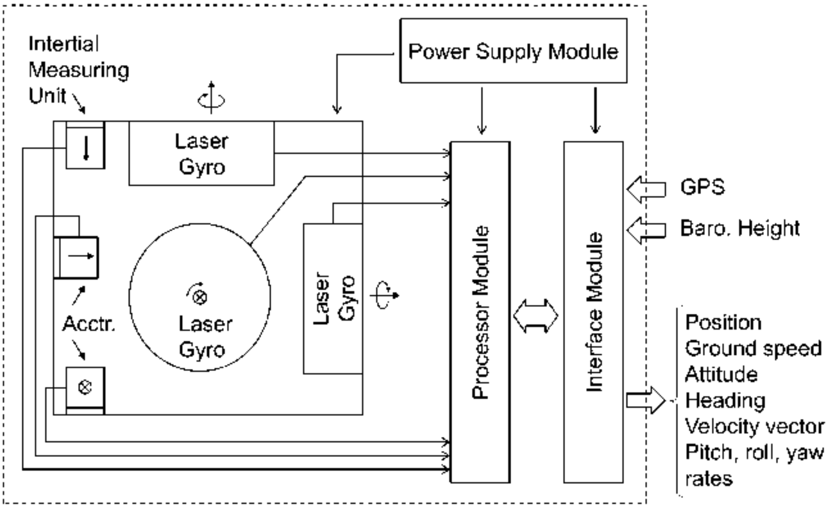
Strap-down INS computing flow diagram.

Strap-down IN System Computing

The basic computing flow diagram for a strap-down INS is shown in. A strap-down IN system carries out the same functions as a stable platform type. INS and many elements and functional areas are common to both systems. There are two crucial areas in the strap-down mechanisation. These are:

1. Attitude integration whereby the vehicle attitude is derived by an integration process from the body incremental angular rotations measured by the gyros.
2. Accelerometer resolution whereby the corrected outputs of the body mounted accelerometers are suitably resolved to produce the horizontal and vertical acceleration components of the aircraft.

The derivation of the aircraft attitude from the strap-down gyro outputs by continuously updating the four



Solid state modular implementation of a strap-down INS.

The derivation of the equivalent (or virtual) horizontal and vertical accelerometer outputs from the body mounted accelerometers is as explained in Chapter 5, Section 5.3.2.2, using the Direction Cosine Matrix derived from the Euler parameters.

The computing system, after deriving the equivalent North, East, Down accelerometer outputs and transforming the co-ordinate frame rates from local NED axes to aircraft body axes, is then the same as for a stable platform mechanisation.

The implementation of a strap-down INS is shown. The aim is to give the reader an appreciation of the small number of solid state modules required. These comprise:

1. The Inertial Measuring Unit (IMU) comprising three orthogonally mounted laser gyros and three orthogonally mounted accelerometers.
2. The Processor Module carrying out all the processing tasks just described, and also monitoring and self test functions.
3. The Interface Module carrying out all the interfacing tasks.
4. The Power Supply Unit.

The complexity resides in the software and the processor and interface micro-chips.

The Honeywell *Laseref VI Micro Inertial Reference System* is an example of a current state of the art INS. The inset illustration and data are published by courtesy of Honeywell.

- Overall size: 6.5″ × 6.4″ × 6.4″
  16.5 cm × 12.3 cm × 12.3 cm)
- Weight: 9.1 lbs (4.14 kg)
- Power consumption: 20 watts
- Reliability > 50,000 hours

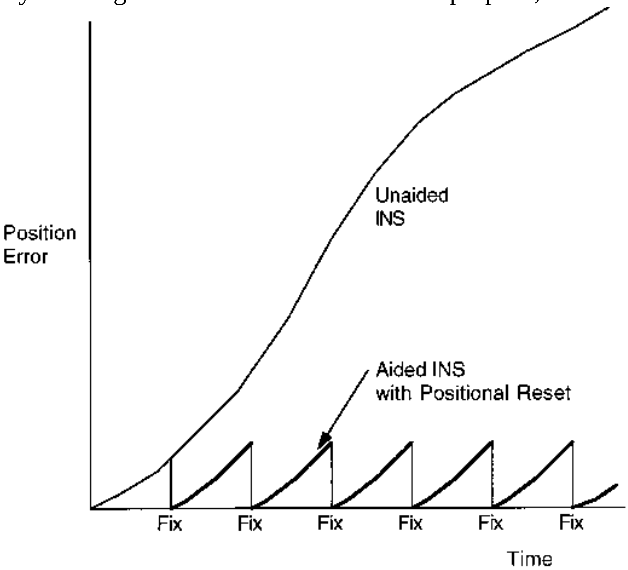
Position accuracy (unaided) is 2 NM/hour.

The inertial sensors comprise the Honeywell GG 1320 AN Digital Laser Gyro and the QA 2000 Q Flex Accelerometer, a closed-loop, torque balance accelerometer. The system incorporates GPS integration. Hybrid mode position and velocity accuracy is 12 metres and 0.25 knots (0.13 m/s). GPS integration also enables IRS alignment in motion and even despatch prior to navigation mode. This feature eliminates delays while waiting for the IRS to align.

Aided IN Systems and Kalman Filters

The time dependent error growth of an IN system makes it necessary that some form of aided navigation system using an alternative navigation source is introduced to correct the INS error build up on long distance flights. For example, as mentioned earlier, a good quality unaided INS of 1 NM/hour accuracy could have a positional error of 5 NM after a five hour flight.

A variety of navigation aids can be used for this purpose, for example GPS.



Aided INS with simple positional reset.

To recap, the various navigation sources comprise:

1. *Position Data*
   - GPS, VOR/DME, TACAN.
   - Terrain reference navigation systems.

- Radar.
- Visual fixes (e.g., use of helmet mounted sight).
- Astro (Stellar) navigation (using automatic star trackers).

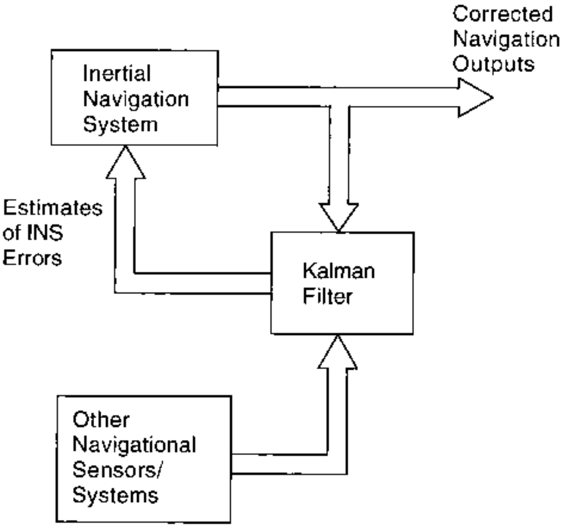
2. *Velocity Data*

- Doppler radar.
- GPS.

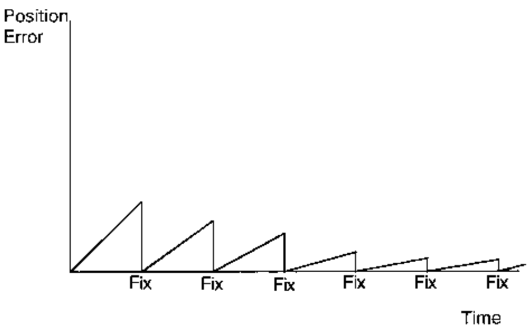
3. *Altitude Data*

- Barometric altitude from the air data computer.
- Radio altimeter.

These sources provide good information on the average at low frequency but are subject to high frequency noise due to causes such as instrument noise, atmospheric effects, antenna oscillation, unlevel ground effects, etc.



Block diagram of aided IN system with Kalman filter.

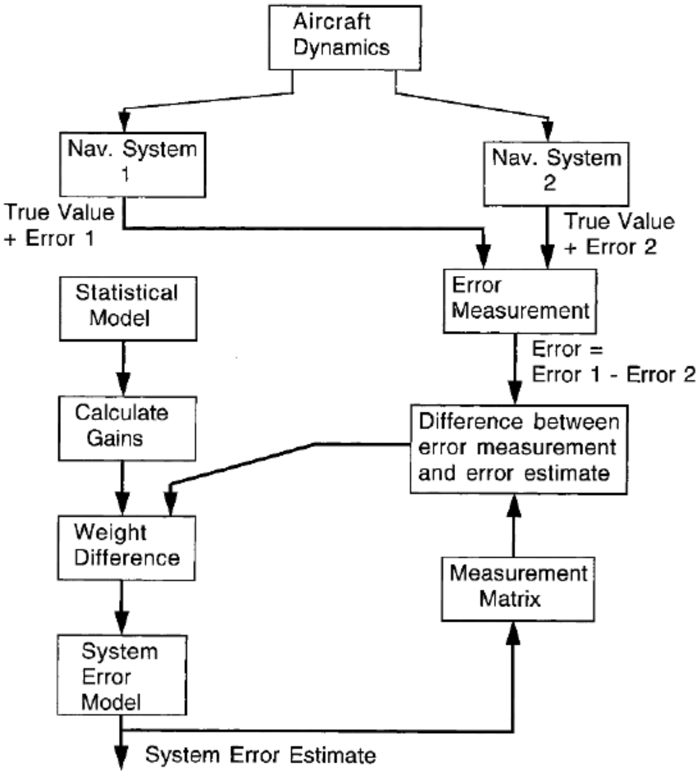


Aided INS with Kalman filter.

however, is poor due to the inherent long term drift characteristics, as already explained.

It should be stressed at this point that a Kalman filter can be used to provide an optimum estimate of the errors in any measuring system and its use is not confined to navigation systems, although it has been particularly effective in this field. Examples of other applications include:

- Radar and infrared automatic tracking systems.
- Fault detection and in monitoring of multiple (redundant) sensors.
- Initial alignment and gyro compassing of an INS.



Application of Kalman filter to mixed navigation systems.

The Kalman filter was first introduced in 1960 by Dr. Richard Kalman (see reference list at the end of this chapter). It is essentially an optimal, recursive data processing algorithm which processes sensor measurements to estimate the quantities of interest (states) of the system using:

1. A knowledge of the system and the measurement device dynamics.
2. A statistical model of the system model uncertainties, noises, measurement errors.
3. Initial condition information.

UNIT - V
SURVEILLANCE SYSTEM AND AUTOPILOT SYSTEM

Introduction

Autopilots and flight management systems (FMS) have been grouped together in one chapter as modern aircraft require a very close degree of integration between the two systems, particularly in their respective control laws.

Much of the background to the understanding of autopilots has been covered in earlier chapters. For instance, the dynamics of aircraft control and fly-by-wire flight control systems, including pitch rate and roll rate manoeuvre command control systems. Chapter 4 also covers the meth- ods used to achieve the very high integrity and safety required to implement FBW flight control through the use of redundancy.

Similar levels of safety and integrity are required in some autopilot modes, for example automatic landing and automatic terrain following (T/F). The digital implementation of flight control systems is also covered.

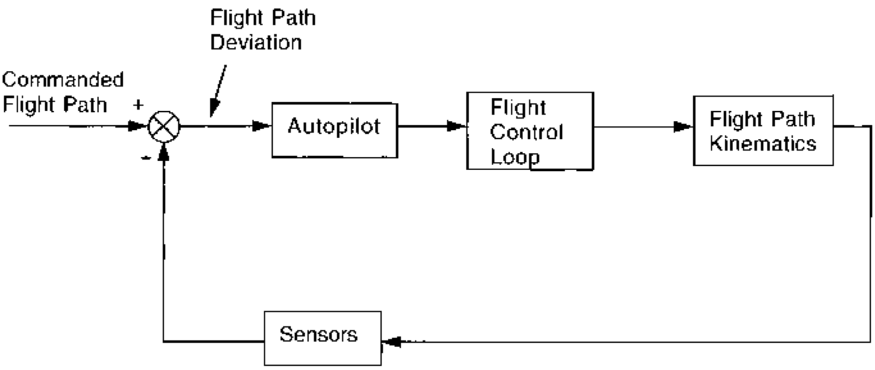
The autopilot is thus an essential equipment for most military and civil aircraft, including helicopters. The advent of the micro-processor has also enabled relatively sophisticated and affordable autopilots to be installed in large numbers of general aviation type aircraft.

The prime role of the flight management system is to assist the pilot in managing the flight in an optimum manner by automating as many of the tasks as appropriate to reduce the pilot workload. The FMS thus performs a number of functions, such as:

- Automatic navigation and guidance including '4D' navigation.
- Presentation of information.
- Management of aircraft systems.
- Efficient management of fuel.
- Reduction of operating costs.

The broad concepts and operation of an FMS are covered in Section 8.3. It should be appreciated that the detailed implementation of an FMS is a complex subject and can involve over 100 man-years of software engineering effort and very extensive (and expensive) flight trials before certification of the system can be obtained from the regulatory authorities. It is only possible, therefore, because of space constraints to give an overview of the subject.

It should also be pointed out that an FMS has an equally important role in a military aircraft. Accurate adherence to an optimum flight path and the ability to keep a rendezvous at a particular position and time for, say, flight re-fuelling or to join up with other co-operating aircraft are clearly very important requirements.



Autopilot loop.

Autopilots

Basic Principles

The basic loop through which the autopilot controls the aircraft's flight path is shown in the block diagram

The autopilot exercises a guidance function in the outer loop and generates commands to the inner flight control loop.

These com- mands are generally attitude commands which operate the aircraft's control surfaces through a closed-loop control system so that the aircraft rotates about the pitch and roll axes until the measured pitch and bank angles are equal to the commanded angles.

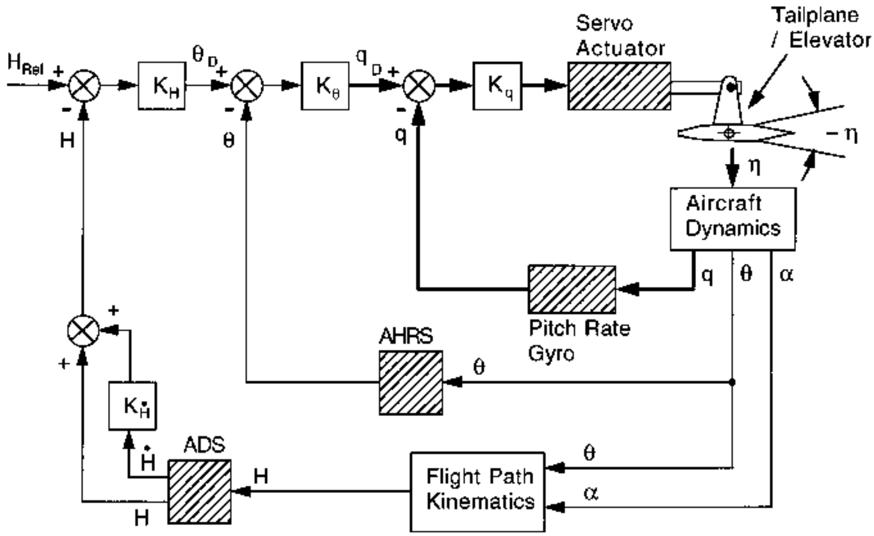
The changes in the aircraft's pitch and bank angles then cause the aircraft flight path to change through the flight path kinematics.

For example, to correct a vertical deviation from the desired flight path, the aircraft's pitch attitude is controlled to increase or decrease the angular inclination of the flight path vector to the horizontal. The resulting vertical velocity component thus causes the aircraft to climb or dive so as to correct the vertical displacement from the desired flight path.

To correct a lateral displacement from the desired flight path requires the aircraft to bank in order to turn and produce a controlled change in the heading so as to correct the error.

The pitch attitude control loop and the heading control loop, with its inner loop commanding the aircraft bank angle, are thus fundamental inner loops in most autopilot control modes.

The outer autopilot loop is thus essentially a slower, longer period control loop compared with the inner flight control loops which are faster, shorter period loops.



Height control (heavy lines show FBW pitch rate command loop). Pitch rate gyro and AHRS are shown separately for clarity. In a modern aircraft, a strap-down AHRS/INS provides both θ and q outputs.

Height Control

Height is controlled by altering the pitch attitude of the aircraft, as just explained. The basic height control autopilot loop is shown in figure.

The pitch rate command inner loop provided by the pitch rate gyro feedback enables a fast and well damped response to be achieved by the pitch attitude command autopilot loop. As mentioned earlier, a FBW pitch rate command flight control system greatly assists the autopilot integration.

The pitch attitude loop bandwidth thus determines the bandwidth of the height loop so that the importance of achieving a fast pitch attitude response can be seen.

The transfer function of the flight path kinematics is derived as follows. Vertical component of the aircraft's velocity vector is

$$\mathbf{V_T}\ \boldsymbol{\sin}\ \boldsymbol{\theta_F} = \dot{\mathbf{H}}$$

Heading Control Autopilot

The function of the heading control mode of the autopilot is to automatically steer the aircraft along a particular set direction. As explained in Chapter 3, Section 3.5, aircraft bank to turn, and assuming a perfectly co-ordinated turn (i.e. no sideslip).

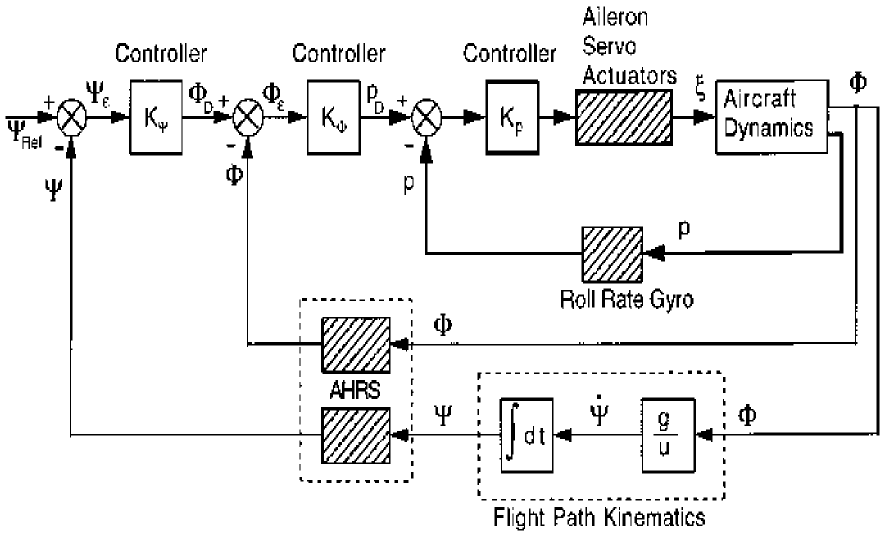
Hence, for small bank angles

$$\dot{W} = \frac{g}{U}\mathit{8}$$

The basic control law used to control the aircraft heading is thus to demand a bank angle, 8_D, which is proportional to the heading error, $W_E = (W_{COM} - W)$, where W_{COM} is the commanded heading angle and W the heading angle.

$$8_D = K_W \cdot W_E$$

where K_ψ is the heading error gain.



Heading control loop (heavy lines show FBW roll rate command loop). Roll rate gyro and AHRS are shown separately for clarity. In a modern aircraft, a strap-down AHRS/INS provides both p and 8 outputs.

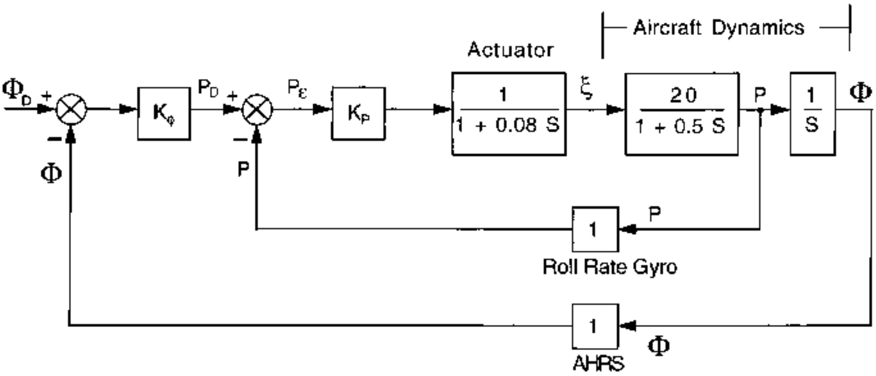
However, a good appreciation of the aircraft and autopilot behaviour can be obtained by assuming pure rolling motion and neglecting the effects of cross-coupling between the roll and yaw axes.

This can also be justified to some extent by assuming effective auto-stabilisation loops which minimise the cross-coupling effects.

$$\dot{8} = p + q \sin 8 \tan\theta + r \cos 8 \tan\theta$$

These assumptions are used in the worked example which follows in order to simplify the aircraft dynamics.

It is also assumed that there is no loss of height when the aircraft banks to correct heading, appropriate changes to the pitch attitude and hence incidence being effected by the height hold autopilot control.



Block diagram of bank angle demand loop.@

Worked Example of Heading Control Autopilot

The aircraft in this example has a roll rate response time constant $T_R = 0.5$ seconds and $L_p/L_\xi = 20°$/second per degree aileron deflection.

The response of the aileron servo actuators approximates to that of a simple first-order filter with a time constant of 0.08 seconds. The aircraft's forward velocity, $U = 250$ m/sec (approximately 500 knots) and is assumed to be constant.

Determine suitable values for the autopilot gains (or gearings) K_p, K_φ and K_ψ .

The overall loop for the heading control is as shown in figure. The block diagram for the bank angle demand loop is shown in figure.

It should be noted that the Laplace operator, s, has been substituted for operator $D(d/dt)$.

Determination of Roll Rate Eerror Gain, K_p

The open-loop transfer function of the innermost loop, namely, the roll rate feedback loop is

$$\frac{p}{p_E} = \frac{K_p 20}{(1 + 0.08s)(1 + 0.5s)} = KG(s) \tag{8.7}$$

where $p_E = p_D - p$.

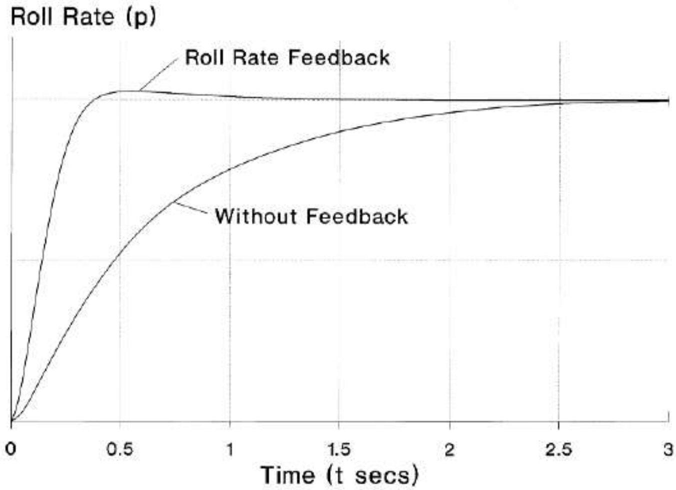
The closed-loop transfer function p/p_D is given by

$$\frac{p}{p_D} = \frac{KG(s)}{1 + KG(s)} \tag{8.8}$$

The stability of the closed loop is determined by the roots of the characteristic equation

$$1 + KG(s) = 0 \tag{8.9}$$

i.e.



Roll rate response.

$$(1 + 0.08s)(1 + 0.5s) + 20K_p = 0$$

$$s^2 + 14.5s + 25\left(1 + 20K_p\right) = 0$$

Comparing with the standard form of second-order equation

$$2\zeta\omega_0 = 14.5$$

$$\omega_0^2 = 25(1 + 20K_p)$$

Choosing a damping ratio, $\zeta = 0.7$, requires the undamped natural frequency, $\omega_0 = 10.4$ rad/sec $(14.5/2 \times 0.7)$. From which $K_p = 0.165°$ aileron angle per degree/second roll rate error.

$$K_p = 0.165 \text{ seconds}$$

$$\frac{p}{p_D} = \frac{0.767}{1 + \frac{2\zeta}{\omega_0}s + \frac{1}{\omega_0^2}s^2}$$

where $\omega_0 = 10.4$ and $\zeta = 0.7$.

Determination of Bank Angle Error Gain, K_φ

Consider now the bank angle demand loop. The open-loop transfer function is

$$\frac{\varphi}{\varphi_E} = \frac{0.767K_\varphi}{s\left(1 + \frac{2\zeta}{\omega_0}s + \frac{1}{\omega_0^2}s^2\right)}$$

where $\omega_0 = 10.4$ and $\zeta = 0.7$.

The value of K_φ is determined using frequency response methods, as follows.

The open-loop frequency response can be obtained by substituting $j\omega$ for s.

$$\frac{\varphi}{\varphi_E}(j\omega) = \frac{K}{j\omega\left(1 - \left(\frac{\omega}{\omega_0}\right)^2 + j2\zeta\frac{\omega}{\omega_0}\right)}$$

$$= KF_1(j\omega) \cdot F_2(j\omega)$$

where $K = 0.767K_\varphi$.

$$F_1(j\omega) = \frac{1}{j\omega}$$

$$F_2(j\omega) = \frac{1}{1 - \left(\frac{\omega}{\omega_0}\right)^2 + j2\zeta\frac{\omega}{\omega_0}}$$

$F_1(j\omega)$ and $F_2(j\omega)$ can be expressed in terms of their respective moduli and phase angles (or arguments)

$$F_1(j\omega) = \left|\frac{1}{\omega}\right|, \quad \angle -90^\circ$$

It is convenient to use a non-dimensional frequency, $u = \omega/\omega_0$, in $F_2(j\omega)$

$$F_2(j\omega) = \frac{1}{1 - u^2 + j2\zeta u}$$

$$F_2(j\omega) = \left|\frac{1}{\sqrt{(1-u^2)^2 + 4\zeta^2u^2}}\right|, \quad \angle \tan^{-1}\frac{2\zeta u}{1-u^2}$$

The evaluation of $KF_1(j\omega)\,F_2(j\omega)$ at specific values of ω is carried out using the logarithms of the moduli of the respective elements of the transfer function.

$$20\log_{10}|KF_1(j\omega) \cdot F_2(j\omega)| = 20\log_{10}K + 20\log_{10}|F_1(j\omega)| + 20\log_{10}|F_2(j\omega)|$$

The log modulus of the overall transfer function is thus the sum of the logs of the moduli of the individual elements.

The overall phase angle is similarly the sum of the individual phase angles of the respective elements

Frequency response.

	$F_1(j\omega)$		$F_2(j\omega)$		$F_1(j\omega)\cdot F_2(j\omega)$	
ω	20 log M	Phase	20 log M	Phase	20 log M	Phase
0.5	6 dB	−90°	0 dB	−3.9°	6 dB	−94°
1	0 dB	−90°	0 dB	−7.8°	0 dB	−98°
2	−6 dB	−90°	0 dB	−15.6°	−6 dB	−106°
3	−9.5 dB	−90°	0 dB	−23.8°	−9.5 dB	−114°
5	−14 dB	−90°	−0.2 dB	−41.2°	−14.2 dB	−131°
7	−16.9 dB	−90°	−0.7 dB	−59.9°	−17.6 dB	−150°
10.4	−20.3 dB	−90°	−2.9 dB	−90°	−23.2 dB	−180°
15	−23.5 dB	−90°	−7 dB	−118.1°	−30.5 dB	−208°
20	−26 dB	−90°	−11.6 dB	−135.1°	−37.6 dB	−225°
30	−29.5 dB	−90°	−18.4 dB	−151.1°	−47.9 dB	−241°

M = Modulus

$$\angle F_1(j\omega)\cdot F_2(j\omega) = \angle F_1(j\omega) + \angle F_2(j\omega)$$

The overall open-loop frequency response in terms of log gain and phase has been calculated from the above formulae.

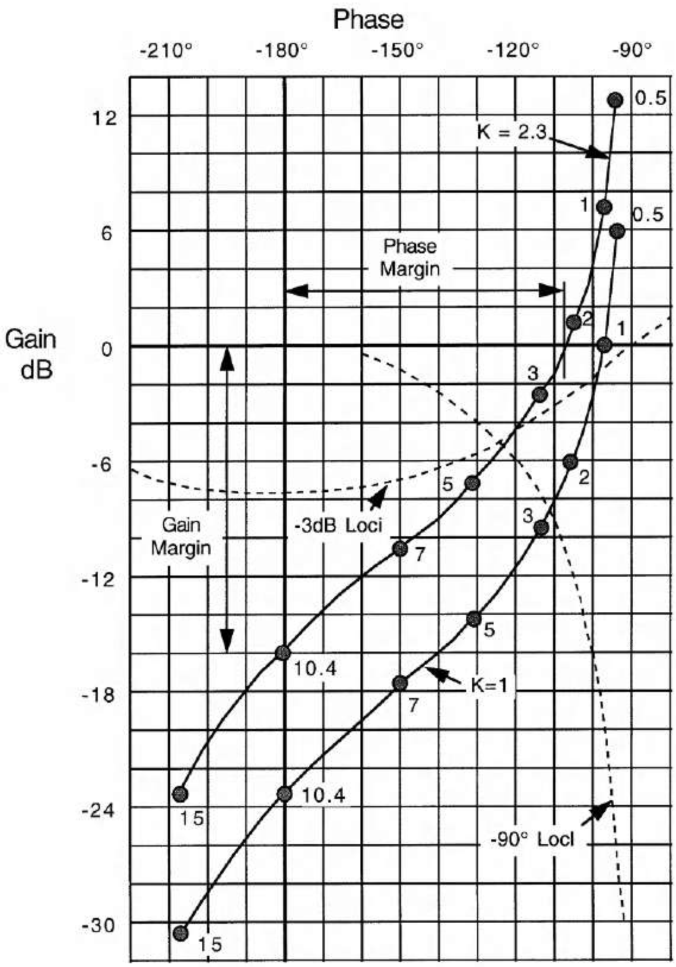
It should be noted that there are standard tables and graphs of the frequency responses of the common transfer function elements. There are also computer programs for evaluating the overall frequency response of a transfer function comprising a number of elements.

The open-loop frequency response is plotted. Which shows the log gain and phase at the specific values of ω chosen in the table. Such a graph is known as a 'Nichol's chart'. Lines of constant gain and phase loci for converting the open- loop frequency response into the closed-loop frequency response are normally superimposed on such a chart. The lines corresponding to −3 dB and −90° are shown on the figure, but the rest have been omitted for clarity.

The Nichol's chart has the following advantages:

1. The logarithmic gain scale enables the response at all the frequencies of interest to be shown on the graph.
2. The effect of shaping the open-loop response can be readily seen. For example, increasing the loop gain, or introducing phase advance or one plus integral of error control, or the effect of additional lags in the system.
3. The closed-loop frequency response can be obtained using the constant gain and phase loci.
4. The gain and phase margins can be read off directly.

The value for the open-loop scalar gain, $K = 1$, gives a very stable loop, with a gain margin of 23.2 dB and a phase margin of 82°. The bandwidth, however, is relatively low. The closed-loop response is 3 dB down at $\omega = 1.2$ approximately (the intersection of the −3 dB loci with the open-loop response). The phase lag is



Bank angle open-loop frequency response (Nichol's chart).

90° at $\omega = 2.8$ approximately (the intersection of the -90° loci with the open-loop response).

Increasing the open-loop scalar gain K to 2.3, that is by 7.2 dB, improves the closed-loop bandwidth considerably. The closed-loop response is 3 dB down at $\omega =$ 3.8 approximately, and there is 90° phase lag at $\omega = 4.2$ approximately. The gain and phase margins are reduced to 16 dB and 71° respectively, but are still very adequate.

The open-loop transfer function is

$$\frac{\varphi}{\varphi_E} = \frac{2.3}{s(1 + 0.1346s + 0.0092456s^2)}$$

The closed-loop stability and damping can be confirmed by finding the roots of the characteristic equation

$$s(1 + 0.1346s + 0.0092456s^2) + 2.3 = 0$$

i.e.

$$s^3 + 14.56s^2 + 108.16s + 246.77 = 0$$

This factorises into

$$(s + 3.63)(s^2 + 10.93s + 68.5) = 0$$

The $(s + 3.63)$ factor produces an exponential subsidence component in the response with a time constant of 1/3.63 seconds, that is 0.28 seconds. The quadratic factor $(s^2 + 10.93s + 68.5)$ produces an exponentially damped sinusoidal component with an undamped natural frequency $\omega_0 = \sqrt{68.5}$, that is 8.28 rad/s (1.32 Hz) and with a damping ratio $\zeta = 0.66$, confirming a reasonably well damped closed-loop response. Hence

$$\frac{8}{8_D} = \frac{1}{(1 + 0.28s)(1 + 0.16s + 0.0146s^2)}$$

The required value of K_φ is obtained from $K = 0.767K_\varphi = 2.3$

$$K_\varphi = 3$$

Determination of Heading Error Gain, K_ψ

Referring to figure. It can be seen that the open-loop transfer function of the heading control loop is

The factors in the denominator have been expressed in a form which enables the cut off frequencies to be readily seen. The scalar gain $K = K_\psi\, g/U$. It can be seen from the above expression that at low frequencies (considerably lower than the first cut off frequency of 3.63 rad/s), both the first and second-order filter components of the transfer function approach unity.

The open-loop gain thus falls at 6 dB/octave at low frequencies.
The value of the frequency at which the open-loop gain is equal to unity (that is 0 dB) is known as the '0 dB cross-over frequency'.
The value of K is chosen to give a 0 dB cross-over frequency such that the rate of change of open-loop gain with frequency is 6 dB/octave for well over an octave in frequency either side of the 0 dB cross-over frequency.
This satisfies Bode's stability criterion.
The 0 dB cross-over frequency is chosen to be 0.7 rad/s.

Hence

$$\frac{K}{0.7} = 1$$

$$K = 0.7 = K_\psi\, 9.81/25$$

$K_\psi = 18$ degrees bank angle per degree heading error.

From the open-loop transfer function it can be seen that the gain starts to fall off at -12 dB/octave ($1/\omega^2$) between $\omega = 3.63$ and $\omega = 8.28$ and then at -24 dB /octave ($1/\omega^4$) above $\omega = 8.28$ rad/s. The phase changes very rapidly over this frequency range, for example $-115°$ at 1 rad/s, $-138°$ at 2 rad/s, $-180°$ at 4.1 rad/s, $-246°$ at 8.28 rad/s, $-310°$ at 15 rad/s and approaches $-360°$ as the frequency increases still higher.

The gain and phase margins with $K_\psi = 18$ are approximately 19 dB and 73° respectively. The closed-loop bandwidth is about 1.4 rad/s when the phase lag is $-90°$. The 3 dB down frequency is just over 1 rad/s.

The loop gain clearly varies with forward velocity, U, and a generous gain margin has been allowed to cope with changes in speed. Gain scheduling with speed is required, however, to cover the whole of the aircraft speed envelope.

A bank angle limiter is also required to prevent too large a bank angle being demanded when there is a significant initial heading error on engaging the autopilot. As stated earlier, the aircraft dynamics have been simplified in this example to give an insight into the design of the inner and outer autopilot loops.

In practice, a full representation of the aircraft dynamics including the yaw/roll cross-coupling effects would be used to establish the autopilot design.

Automatic Landing

Introduction

The previous sections have described the automatic coupled approach phase of an automatic landing using the guidance signals from the ILS (or MLS) system in Cat.I or Cat. II visibility conditions. The pilot, however, takes over control from the autopilot when the decision height is reached and lands the aircraft under manual control.

Attempting to land an aircraft under manual control with decision heights of less than 100 ft, as in Cat. III conditions, is very demanding because of the lack of adequate visual cues and the resulting disorientation which can be experienced. There are only two alternatives for effecting a safe landing in such conditions:

(a) A fully automatic landing system with the autopilot controlling the landing to touchdown. A very high integrity autopilot system is required with failure survival capability provided by redundancy such that the probability of a catastrophic failure is less than 10^{-7}/hour. High integrity autopilot systems capable of carrying out a fully automatic landing in Cat. III conditions are now at a mature stage of development and large numbers of civil jet aircraft operated by major airlines worldwide are now equipped with such systems.

(b) The use of an enhanced vision system with a HUD as described in Chapter 2, using a millimetric wavelength radar sensor in the aircraft to derive a synthetic runway image. This is presented on the HUD together with the primary flight data, including the flight path velocity vector, and provides sufficient information for the pilot to land the aircraft safely under manual control.

Visibility categories.

Category	Minimum Visibility Ceiling	Runway Visual Range
I	200 ft	800 m
II	100 ft	400 m
IIIa	12–35 ft	100–300 m
	Depending on aircraft type and size	
IIIb	12 ft	<100 m
IIIc	0 ft	0 m

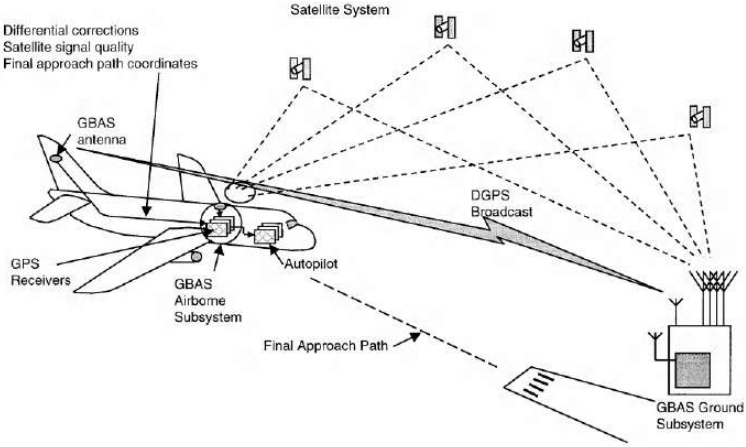
and France should also be acknowledged. At that time, efforts in both countries were directed towards the development of simpler fail passive automatic landing systems. The Sud Aviation SE 210 Caravelle airliner was in fact certified for Cat. III operation with a decision height of 50 ft around the same time as the Trident III. The experience and the design methodologies developed on these pioneering systems, such as redundancy management and techniques such as failure modes and effects analyses have been of great value in subsequent programmes and have been disseminated worldwide.

Satellite Landing Guidance Systems

The navigation position accuracy of 1 m which can be achieved with the differential GPS technique is being exploited in the US for landing guidance with a system called the Ground Based Augmentation System, GBAS. The Ground Based Augmentation System, when installed at an airport, will be able to provide the high integrity and accurate guidance necessary for landing in Cat. III visibility conditions.

The equipment is simpler and less expensive to install and maintain than an Instrument Landing System (ILS) or Microwave Landing System (MLS), so the GBAS life-cycle operation costs are a fraction of the other systems. It is therefore an attractive proposition for the many smaller airports which are not equipped with ILS or MLS.

It is also a more flexible system. For example, the final approach path need not be limited to straight line approaches, but can be curved or stepped, horizontally or vertically.



GBAS landing guidance system.

The flight path kinematics to align the aircraft's flight path with the commanded flight path are the same whether this is defined by a radio/microwave beam or the GBAS

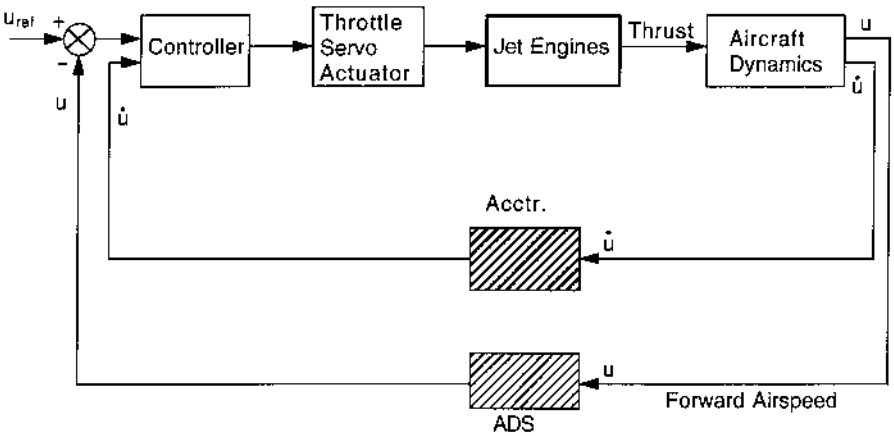
The signals from the GBAS are tailored to provide an 'ILS look alike' guidance for the pilot's displays in terms of their sensitivity. This is to maximise pilot and operator acceptability.

An optional provision is made in the GBAS for additional ranging signals to be provided by ground based transmitters called 'pseudolites' which can be installed to meet high availability requirements.

Speed Control and Auto-Throttle Systems

Control of the aircraft speed is essential for many tasks related to the control of the aircraft flight path, for example the position of the aircraft relative to some reference point.

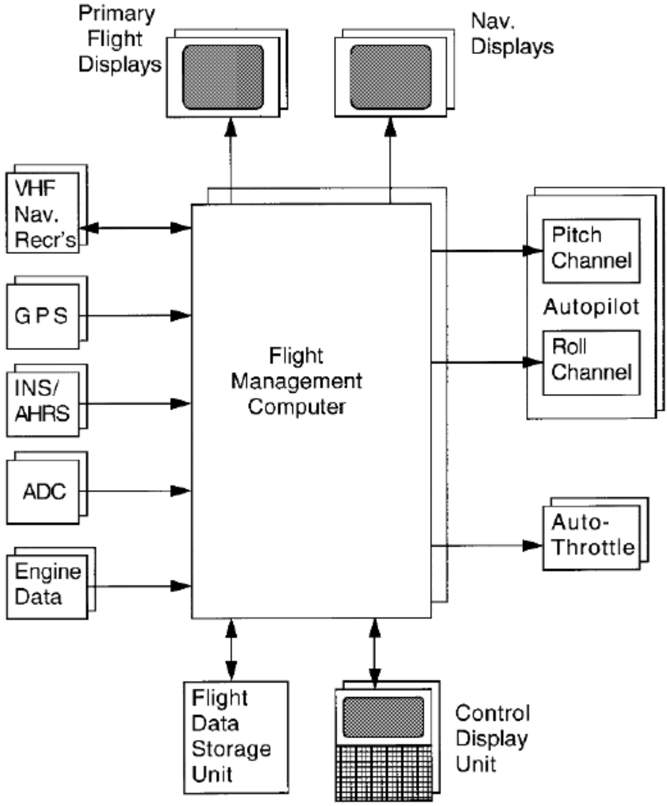
The aircraft speed is controlled by changing the engine thrust by altering the quantity of fuel flowing to the engines by operating the engine throttles. Automatic control of the aircraft's airspeed can be achieved by a closed-loop control system whereby the measured airspeed error is used to control throttle servo actuators which operate the engine throttles. The engine thrust is thus automatically increased or decreased to bring the airspeed error to near zero and minimise the error excursions resulting from disturbances. A typical airspeed control system is shown in the block.



Airspeed control system.

In any closed-loop system, the lags in the individual elements in the loop resulting from energy storage processes (e.g. accelerating inertias) exert a destabilising effect and limit the loop gain and hence the performance of the automatic control system.

The dynamic behaviour of the engines over the range of flight conditions, the throttle actuator response and the aircraft dynamics must thus be taken into account in the design of the speed control system.



Flight management system block diagram.

Flight Management Systems

Introduction

The FMS has become one of the key avionics systems because of the major reduc- tion in pilot work load which is achieved by its use. In the case of military aircraft, they have enabled single crew operation of advanced strike aircraft.

Flight management systems started to come into use in the mid 1980s and are now in very wide scale use, ranging from relatively basic systems in commuter type aircraft to 'all-singing, all-dancing' systems in long range wide body jet airliners. They have enabled two crew operation of the largest civil airliners and are generally a dual FMS installation because of their importance.

Block diagram of a typical flight management system. The benefits they confer are briefly set out below:

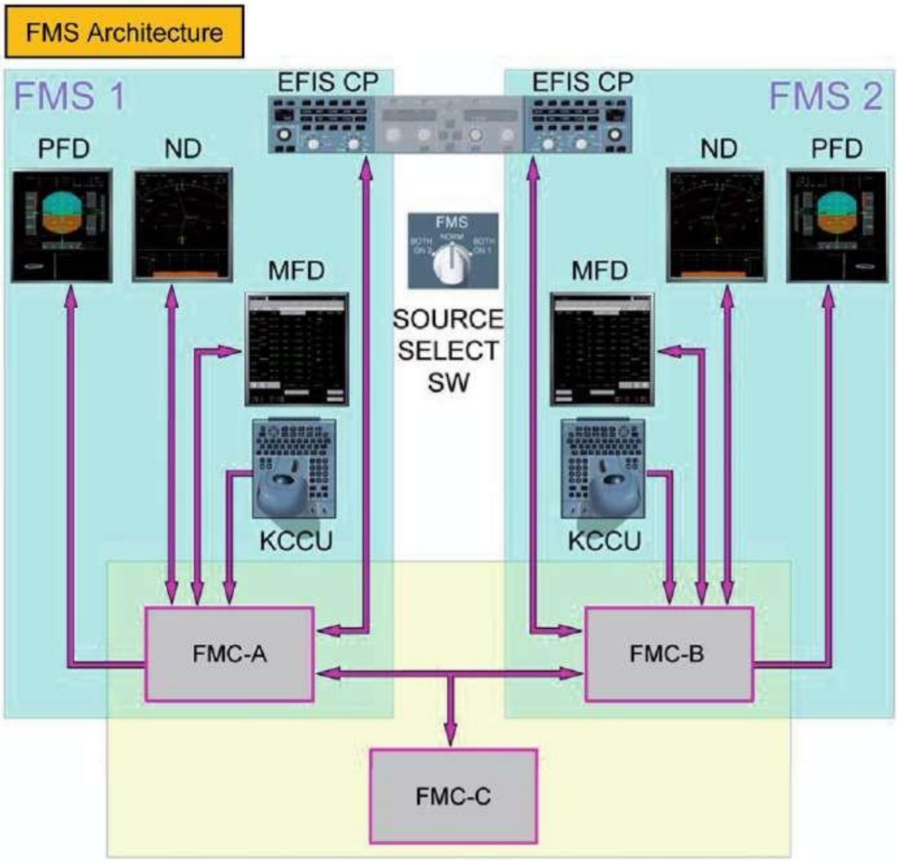
- *Quantifiable economic benefits* – provision of automatic navigation and flight path guidance to optimise the aircraft's performance and hence minimise flight costs.
- *Air traffic* – growth of air traffic density and consequently more stringent ATC requirements, particularly the importance of 4D navigation.

- *Accurate navigation sources* – availability of accurate navigation sources. For example, INS /IRS, GPS, VOR, DME and ILS / MLS.
- *Computing power* – availability of very powerful, reliable, affordable computers.
- *Data bus systems* – ability to interconnect the various sub-systems.

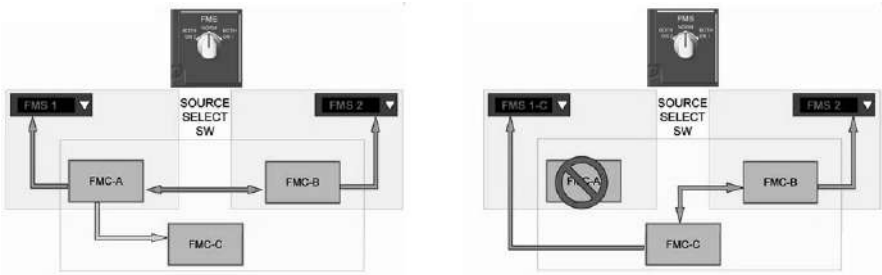
The FMS carries out the following tasks:

1. Flight guidance and lateral and vertical control of the aircraft flight path.
2. Monitoring the aircraft flight envelope and computing the optimum speed for each phase of the flight and ensuring safe margins are maintained with respect to the minimum and maximum speeds over the flight envelope.
3. Automatic control of the engine thrust to control the aircraft speed.

In addition the FMS plays a major role in the flight planning task, provides a computerised flight planning aid to the pilot and enables major revisions to the flight plan to be made in flight, if necessary, to cope with changes in circumstances.



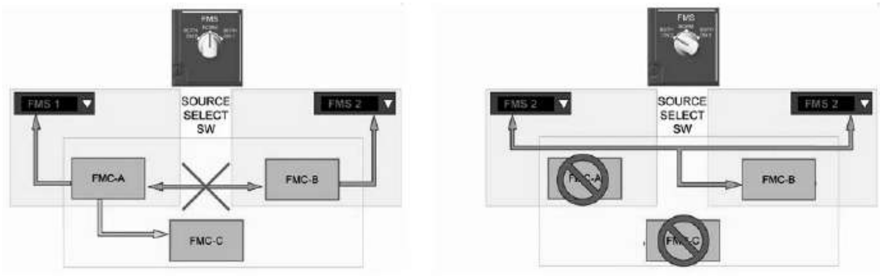
FMS architecture (by courtesy of Airbus).



Dual Mode operation (by courtesy of Airbus).

Dual Mode

Both flight management systems, FMS-1 and FMS-2, are healthy. the configuration in normal operation in the left side illustration and the configura- tion after a single flight management computer failure in the right side illustration.



Independent Mode and Single Mode configurations (by courtesy of Airbus).

In normal operation, FMC-A provides data to FMS-1, FMC-B provides data to FMC-2 and FMC-C is the standby computer.

Of the two active computers, one FMC is the ‘master’ and the other is the ‘slave’, depending on which autopilot is active and the selected position of the FMS Source Select Switch.

The two active FMCs independently calculate data, and exchange, compare and synchronise these data. The standby computer does not perform any calculations, but is regularly updated by the master FMC.

Independent Mode

In the Independent Mode, FMS-1 and FMS-2 are both operative, but there is no data exchange between them because they disagree on one or more items such as aircraft position, gross weight, etc.

Single Mode

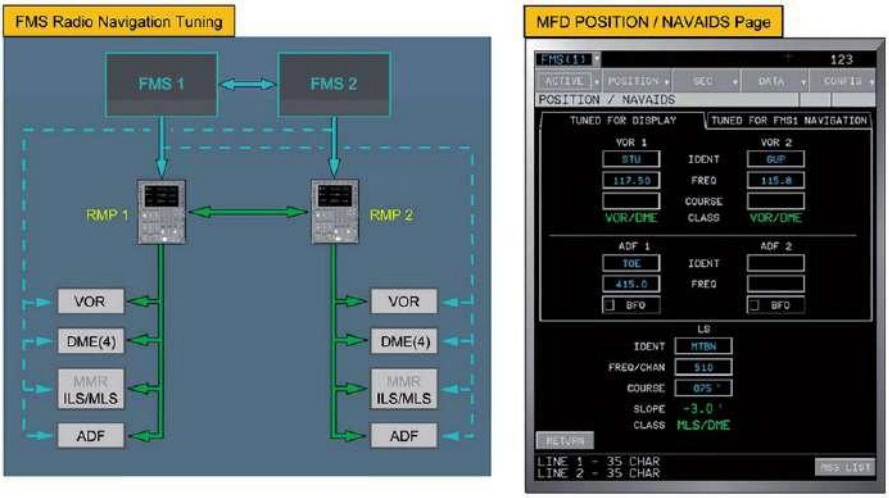
The loss of two FMC’s causes the loss of either FMS-1 or FMS-2.

The data from the operative FMS is displayed to the flight crew by operating the Source Select Switch.

Radio Navigation Tuning

The FMS automatically tunes the radio navigation aids, (NAVAIDs), used for the radio position computation, the NAVAIDS for display on the Navigation Displays and the landing system NAVAIDS.

In ‘Dual’ and ‘Independent’ modes each FMS tunes its onside NAVAIDS.



NAVAIDS tuning and POSITION/NAVAIDS Page displays on MFD (by courtesy of Airbus).

These comprise, in the case of the A380, one VOR, four DMEs, one ILS (MLS/GLS optional), one ADF (optional).

In ‘Single’ FMS mode, or in the case of a communications failure between an FMS and its onside Radio Management Panel (RMP), the available FMS will tune the NAVAIDS on both sides.

The tuning of the onside NAVAIDS passes through the onside RMP, to synchronise the NAVAIDS tuning between the FMS and the RMP.

Manual tuning always has priority over automatic tuning.

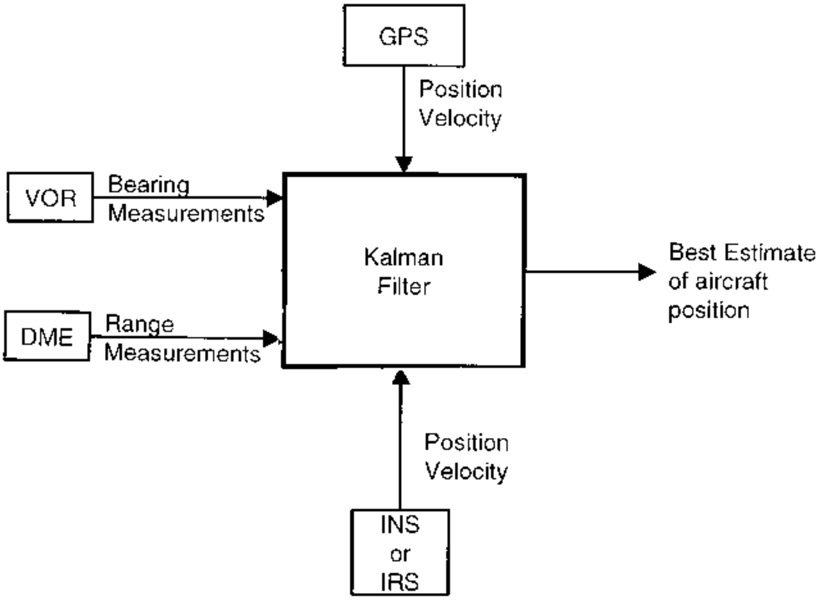
Navigation

The FMS combines the data from the navigational sources, comprising the inertial systems, GPS and the radio navigation systems, in a Kalman filter to derive the best estimate of the aircraft position.

The accuracy of this estimate is also evaluated.

Each FMS computes the aircraft position and the position accuracy.

The FMS computed position is an optimum combination of the inertial position and the GPS or radio position, depending on which equipment provides the most accurate data.



Kalman filtering of navigational data sources.

This results in four navigation modes in decreasing order of priority:

- Inertial (IRS) – GPS.
- Inertial (IRS) – DME/DME.
- Inertial (IRS) – VOR/DME.
- Inertial (IRS) only.

The FMS aircraft position always uses the inertial position. This computation is not possible if the inertial position is not valid, and in this case all the FMS navigation and flight planning functions are no longer available.

The FMS continually computes the Estimated Position Uncertainty (EPU), and the EPU is used, together with the Required Navigation Performance (RNP,) to define the aircraft navigation accuracy.

The FMS continuously compares the actual EPU with the current RNP, and defines the navigation class as:

- HIGH, if the EPU is less than, or equal to the RNP.
- LOW, if the EPU is greater than the RNP.

The navigation class has to satisfy the Airworthiness Authorities Accuracy Requirements (AAAR).

The FMS computes ground speed, track, wind direction and velocity. (It should be noted that the air data system provides the height information for vertical navigation.) As stated earlier, the FMS provides both lateral and vertical guidance signals to the autopilot to control the aircraft flight path.

In the lateral case, the FMS computes the aircraft position relative to the flight plan and the lateral guidance signals to capture and track the flight path specified by the flight plan.

Three-dimensional vertical guidance is provided to control the vertical flight profile including the time dimension as will be explained in more detail later.

This is of particular benefit during the descent and approach.

Flight Planning

As explained earlier, a major function of an FMS is to help the flight crew with flight planning and it contains a database of:

- *Radio NAVAIDS* – VOR, DME, VORTAC, TACAN, NDB, comprising identification, latitude/longitude, altitude, frequency, magnetic variation, class, airline figure of merit.
- *Waypoints* – usually beacons.
- *Airways* – identifier, sequence number, waypoints, magnetic course.
- *Airports* – identifier, latitude, longitude, elevation, alternative airport.
- *Runways* – length, heading, elevation, latitude, longitude.
- *Airport procedures* – ICAO code, type, SID, STAR, ILS, profile descent.
- *Company routes* – original airport, destination airport, route number, type, cruise altitudes, cost index.

The navigation data base is updated every 28 days, according to the ICAO AiRAC cycle, and is held in non-volatile memory.

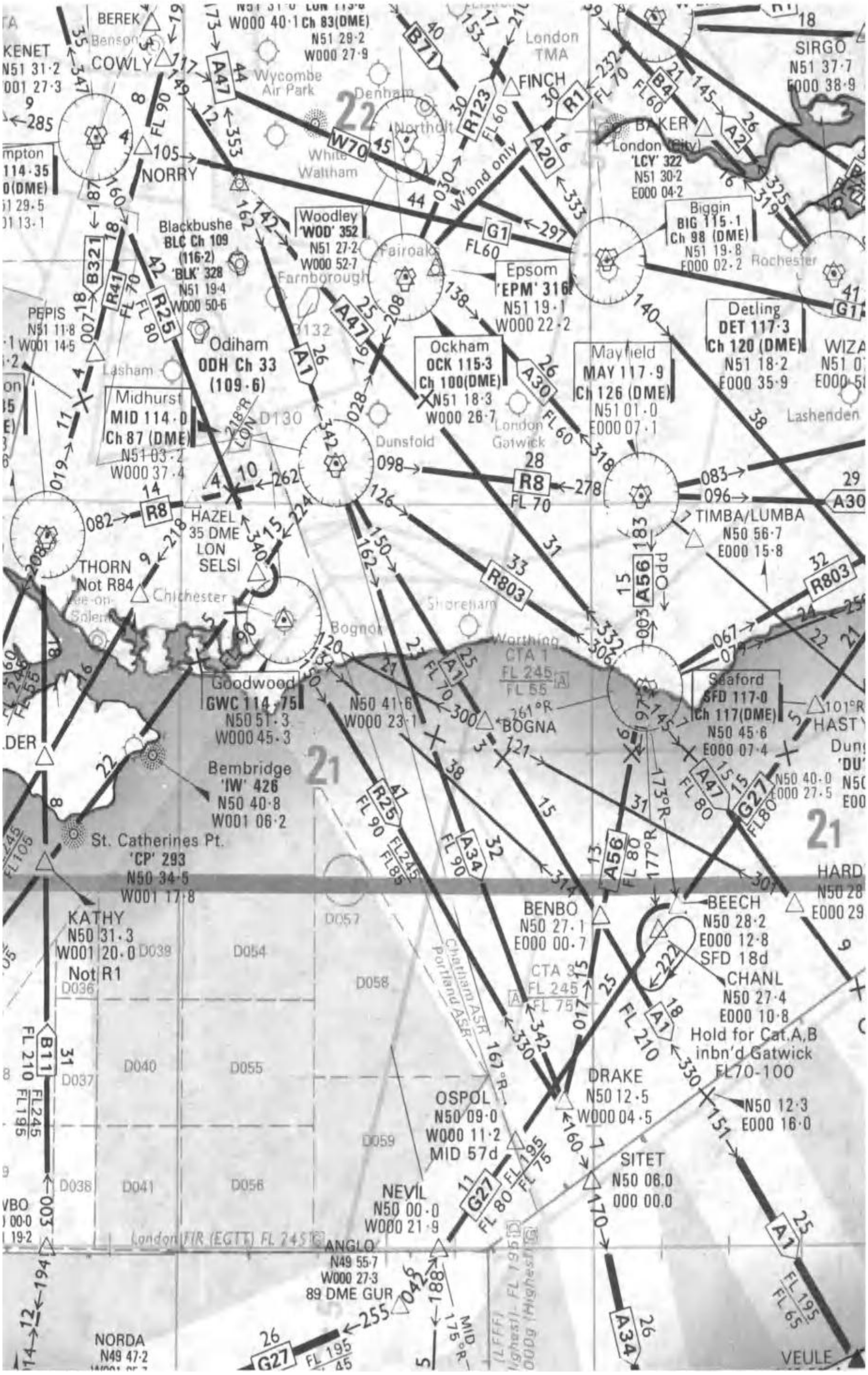
It is clearly essential to maintain the recency and quality of the data base and the operator is responsible for the detail contents of the data base which is to ARINC 424 format.

The flight crew can change the flight plan at any time; a change to the lateral plan is called a ‘lateral revision’ and a change to the vertical plan a ‘vertical revision’.

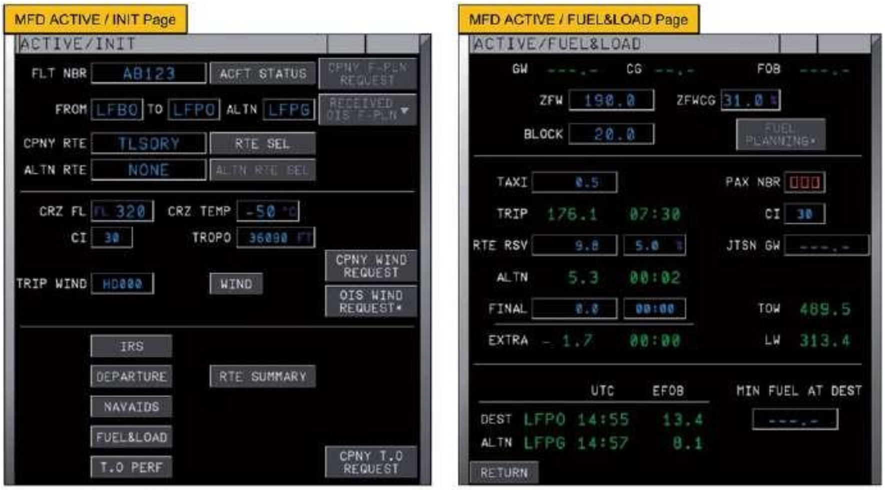
The FMS can simultaneously memorise four flight plans:

- One *active* flight plan for lateral and vertical long term guidance and for radio navigation auto-tuning.
- Three *secondary* flight plans with drafts to compare predictions, to anticipate a diversion or to store company, ATC and Onboard Information System flight plans.

The lateral flight plan includes the departure, cruise and arrival and is composed of waypoints that are linked with flight plan legs and transitions between legs.



Section of radio navigation chart (by courtesy of British Airways AERAD).



‘Active Initialisation’ Page and ‘Active Fuel’ Page displays on MFD (by courtesy of Airbus).

The ease and visibility of the data entry process can be appreciated.
A flight plan can be created in three ways:

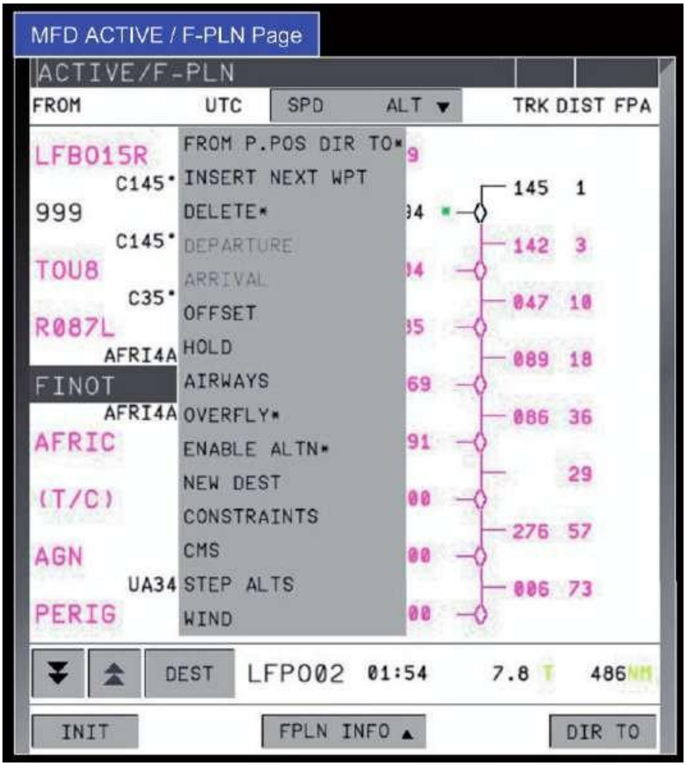
1. By inserting an origin/destination pair and then manually selecting the departure, waypoints, airways and arrival.
2. By inserting a company route stored in the database.
3. By sending a company request to the ground for an active Flight Plan (F-PLN) uplink.

The flight crew can perform the following lateral revisions:

- Delete and insert waypoints.
- Departure procedures: Takeoff runway, Standard Instrument Departure (SID) and transition.
- Arrival procedures: Runway, type of approach, Standard Terminal Arrival Route (STAR), via, transition.
- Airways segments.

The flight crew can also perform the following vertical revisions:

- Time constraints.
- Speed constraints.
- Constant Mach segments.
- Altitude constraints.
- Step altitudes.
- Wind.



‘Active Flight Plan’ Page display on MFD (by courtesy of Airbus).

Performance Prediction and Flight Path Optimisation

The FMS is able to optimise specific aspects of the flight plan from a knowledge of the aircraft type, weight, engines and performance characteristics, information on the wind and air temperature and the aircraft state – airspeed, Mach number, height, etc.

The FMS continually monitors the aircraft envelope and ensures that the speed envelope restrictions are not breached.

It also computes the optimum speeds for the various phases of the flight profile. This is carried out taking into account factors such as:

- Aircraft weight – computed from a knowledge of the take-off weight and the fuel consumed (measured by the engine flow meters). It should be noted that fuel can account for over 50% of the aircraft weight at take off.
- CG position – computed from known aircraft loading and fuel consumed.
- Flight level and flight plan constraints.
- Wind and temperature models.
- Company route cost index.

The recommended cruise altitude and the maximum altitude are also computed from the above information.

The flight crew enter the following data to enable the performance computations and flight plan predictions to be made.

- Zero Fuel Weight (ZFW) and Zero Fuel Centre of Gravity (ZFCG).
- Block fuel.
- Airline Cost Index (CI).
- Flight conditions (Cruise Flight Level (CRZ FL), temperature, wind).

The FMS computes the following predictions from the flight plan and the flight crew data entries:

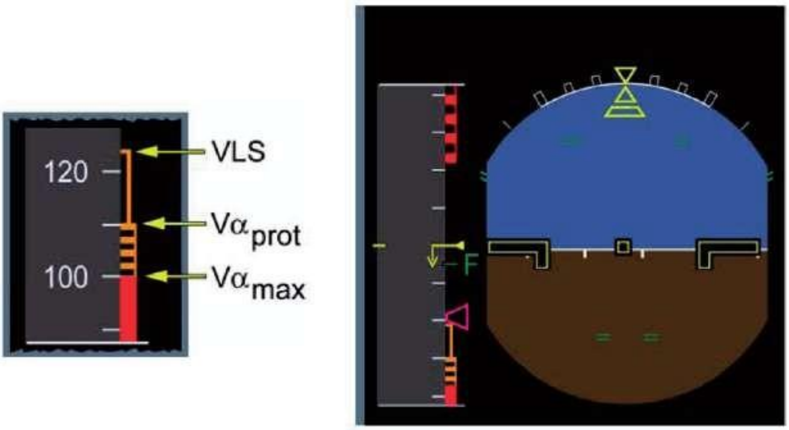
- Wind and temperature.
- Speed changes.
- Pseudo waypoint computation: T/C, T/D, LVL OFF.
- For each waypoint or pseudo waypoint:
  - Distance
  - Estimated Time of Arrival (ETA)
  - Speed
  - Altitude
  - Estimated Fuel on Board (EFOB)
  - Wind for each waypoint or pseudo waypoint
- For primary and alternate destination
  - ETA
  - Distance to destination
  - EFOB at destination

These predictions are continually updated depending on:

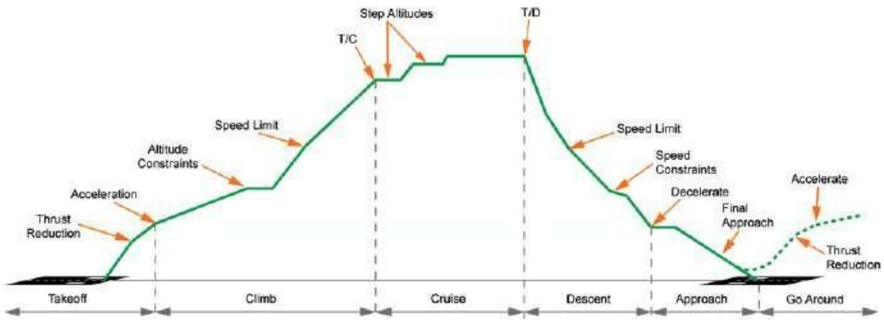
- Revisions to the lateral and vertical flight plans.
- Current winds and temperature.
- Actual position versus lateral and vertical flight plans.
- Current guidance modes.

The predictions and the lateral flight plan combine to form a vertical profile that has six flight phases.

‘Active Performance’ Page display on MFD (by courtesy of Airbus).



Speed scale showing safe limiting values on PFD (by courtesy of Airbus).



Flight plan – vertical definition (by courtesy of Airbus).

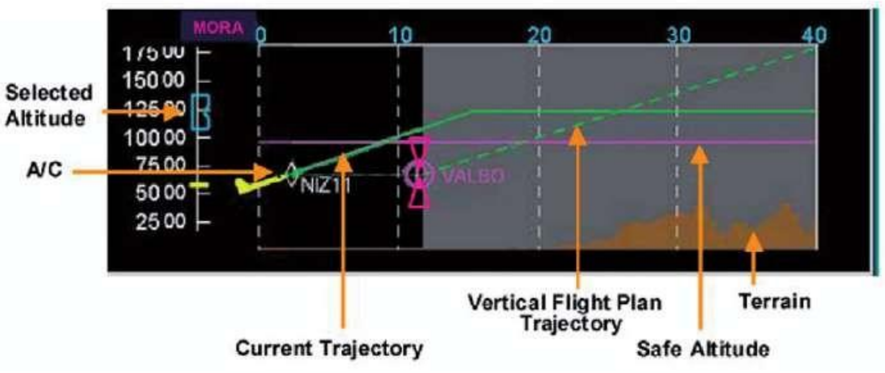
Control of the Vertical Flight Path Profile

The FMS selects the speeds, altitudes and engine power settings during climbs, cruises and descents taking into account the flight plan, the prevailing conditions and the optimisation of the operation of the aircraft.

The tasks which can be carried out and the facilities provided by the FMS during the various phases of the flight are briefly summarised below:

- TAKE OFF – The critical speeds V1, VR, and V2 are inserted by the crew and displayed on the primary flight displays.
- CLIMB – The FMS uses the manually input speed, the ATC constraint speed or the economical speed. It determines the start of the climb during take-off and predicts the end of the climb and the optimum cruising flight level.
- CRUISE – Five flight levels can be defined manually in the FMS. Two flight levels can be stored for every route in the navigation data base. During the cruise, ATC or the crew may change the cruise altitude and the FMS can perform a 'step' climb at economical speed or a 'step' descent at 1000 ft/min at an economical speed. These events are also displayed symbolically on the navigation display.
- DESCENT – The FMS uses the manually input speed, the ATC constraint speed or the economic speed. The altitude and speed during the descent are computed as a function of the distance to the destination and a geometric profile is formed. The flight path is then computed backwards to satisfy the constraints.
- APPROACH – The FMS can be coupled to the autopilot or alternatively provide guidance information to the pilot for manual control of the aircraft. Speed is critical during this phase and the approach speed is computed with respect to V_{REF} and the landing configuration (flaps, slats, etc.) and the wind at the destination.

  The approach mode is entered at the end of the descent and the approach ends either with landing or go around. Lateral guidance is provided by the FMS from



Aircraft trajectory in vertical plane displayed on lower section of Navigation Display (by courtesy of Airbus).

the computed aircraft position and vertical guidance from barometric altitude when an RNAV approach has been selected. The FMS also provides speed control.
At the end of an RNAV approach the crew takes control to carry out the landing using visual references.

When an ILS approach has been selected, the FMS tunes the ILS frequency and selects the runway heading as required for the runway selected by the crew.

The approach and landing guidance is carried out by the autopilot using the ILS localiser references for horizontal guidance and the ILS glide slope references for the vertical guidance until the glide extension and flare phases, unless the crew elect to carry out an automatic go around or elect to take over control.

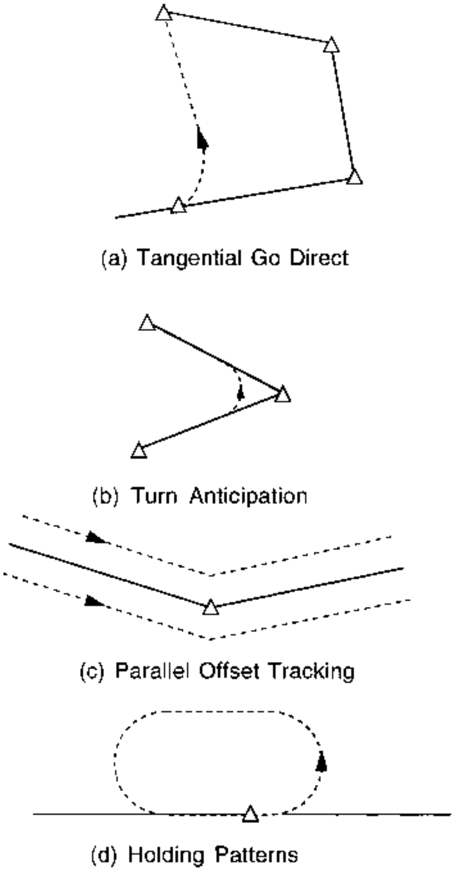
- GO AROUND – This is always assumed.

The FMS manages the climb to the accelerating altitude or a selected altitude and provides track guidance from the outbound track defined in the flight plan.

The lower section of the Navigation Display is known as the 'Vertical Display Zone' and is used to display the vertical flight profile. (Refer to Figure 2.37 in Chapter 2, which shows a typical Navigation Display.)

Operational Modes

The FMS provides a number of very useful operational modes



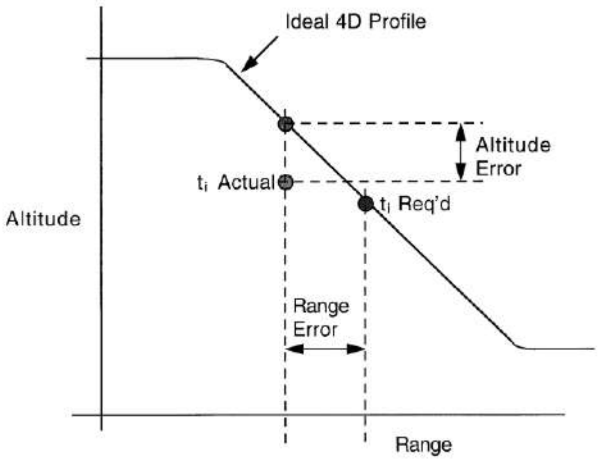
Operational modes.

- *Tangential go direct to mode* – This is and provides navigation from the current position to any waypoint in the flight plan or entered during the flight.

- *Turn anticipation* – This is shown and avoids overshooting waypoints. It reduces both the distance flown and off-track manoeuvring.

- *Parallel offset tracking* – This is illustrated. The lateral offset allows ATC to increase traffic flows in certain cases.

- *Holding pattern* – This is illustrated. The FMS produces a precision holding pattern based on published ICAO entry procedure to reduce the pilot work load.



4D descent control variables.

4D Flight Management

4D navigation has already been briefly referred to a number of times; in fact '4D flight management' is a better description of the process. 4D flight management covers the optimisation of the aircraft's flight along the most fuel conservative 3D path through climb, cruise and descent within the constraints of the air traffic control environment.

Most importantly, the arrival time of the aircraft is also controlled so that it fits into the air traffic flow without incurring or causing delays. This is achieved by the automatic closed-loop control exercised by the FMS through the autopilot and auto-throttle systems.

These control the aircraft's flight path so that its 3D position at any time corresponds closely with the optimum time referenced flight path generated by the FMS computer.

When instructed to resume the descent the FMS then re-engages and guides the aircraft down to the ATC specified position at the pre-established time.

The process of automatic 4D descent control is very briefly described below. An ideal 4D trajectory, which is defined by a table of time, range and altitude, is pre-computed as briefly explained earlier.

The air traffic control system also specifies the arrival time of the aircraft to the nearest integral minute, so the aircraft must have at least 30 seconds of arrival time flexibility to fit. Delays caused by traffic or weather also require some flexibility. Besides arrival time flexibility, there must also be some path flexibility. For example, air traffic control may require the aircraft to vector off course because of traffic.

The 4D guidance and control exercised by the FMS then automatically re-engages the descent and computes the new course to the position specified by the ground ATC and makes good the original or newly specified arrival time.

Occasionally aircraft are required to level-off in their descent, for example when the high altitude air traffic controller hands the aircraft off to the low altitude controller.

The first major step towards integrating avionic systems was taken in the mid-1950s with the establishment of the 'weapon system' concept.

These concepts were incorporated in the 1960s generation of aircraft, some of which are still in service. The concept requires a total system approach to the task of carrying out the mission effectively with a high probability of success.

The aircraft, weapons and the avionic systems required by the crew to carry out the mission effectively must thus be con- sidered as an integrated combination

As an example of the overall system approach, consider the requirements for a naval strike aircraft. The aircraft must be able to operate from an aircraft carrier in all weathers and be able to find the target and attack it with a suitable weapon (or weapons) with a high probability of success.

- Radar – target acquisition in all weather conditions.
- Doppler – accurate ((4 knots) velocity sensor for DR navigation. (Note: IN systems capable of accurate initial alignment at sea on a moving carrier were still under development in the early 1960s.)
- Attitude heading reference system (or master reference gyro system – UK terminology) – attitude and heading information for pilot's displays, navigation computer, weapon aiming computer, autopilot.
- Air data computer – height, calibrated airspeed, true airspeed, Mach number information for pilot's displays, weapon aiming, reversionary DR navigation, autopilot.
- Radio altimeter – very low level flight profile during attack phase and all weather operation.
- Navigation computer – essential for mission.
- Autopilot – essential for reduction of pilot work load.
- Weapon aiming computer – essential for mission.
- HUD – all the advantages of the HUD plus weapon aiming for low level attack; for example, 'toss' bombing.
- Stores management system – control and release of the weapons.
- Electronic warfare (EW) systems – radar warning receivers, radar jamming equipment. Essential for survivability in hostile environment.

- Identification system (identification friend or foe – ‘IFF’) – essential to avoid attack by friendly forces.
- Radio navigation aids – location of parent ship on return from mission.
- Communications radio suite – essential for communicating to parent ship, co-operating aircraft, etc.

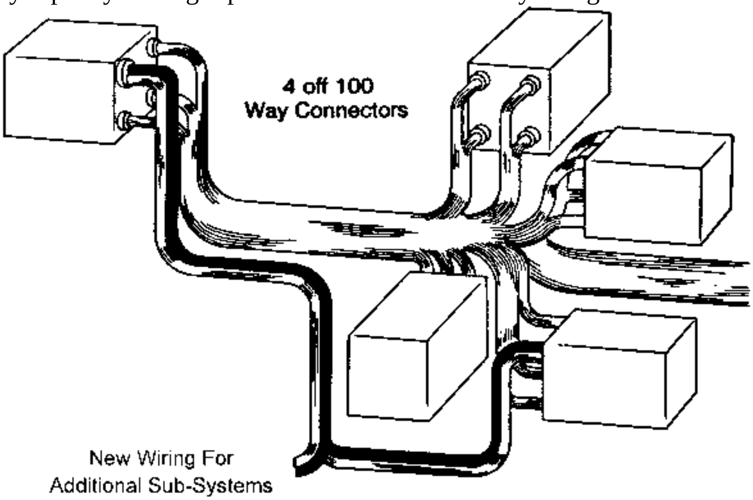
A significant degree of integration was required between the avionic sub-systems. For example, the weapon aiming system required the integration of the HUD, weapon aiming computer, AHRS, air data computer and the radar system.

.

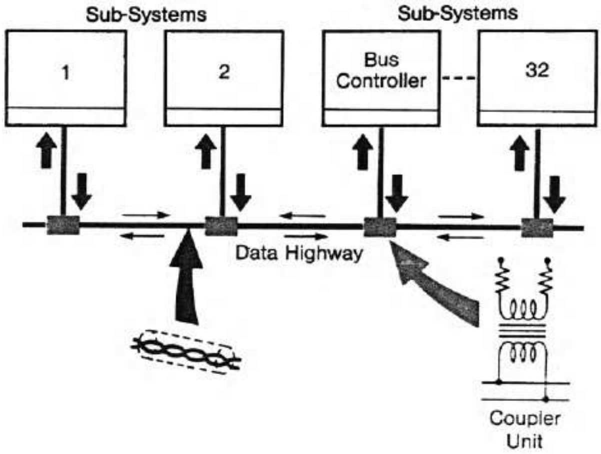
These first generation airborne digital computers were very expensive and it is interesting to note that a number of proposals were made during the 1960s to promote the concept of carrying out as many of the avionic sub-system computing tasks as possible with a powerful central digital computer. These proposals were not taken up for a number of reasons, such as:

1. Vulnerability – ‘all the eggs in one basket’, whereby a failure in the central digital computer affected all the sub-systems sharing its computing facilities.
2. Inflexibility – changes in an individual sub-system could involve changes in the main computer software with possible ramifications and ‘knock on’ effects on all the other sub-systems sharing the computing facility. This was particularly relevant with the computing speeds achievable at the time and the high cost and limited capacity of the non-volatile memory technology which depended on magnetic core stores. (Typical store sizes were 8K to 16K at that time.)
3. Cost and weight of redundant central computer configurations was unacceptable.

The availability today, however, of affordable, very powerful processors with large memory capacity and high speed data buses has radically changed the situation.



Interconnection of avionic sub-systems before introduction of multiplexed data buses.



Interconnections of avionic sub-systems by multiplexed data bus.

By the mid-1970s, it became possible to implement many more avionic sub-systems digitally, exploiting task orientated processors and the newly developed micro-processors and so eliminate the analogue computing elements and the analogue input/output components such as synchros and potentiometers, etc.