

Avionics Summary

1. Basic avionics

Avionics is a rather broad subject. The word avionics is short for **aviation electronics**. But it actually encompasses much more than electronics. What exactly does avionics entail? How does it work? And how did it come to be? That is what this summary is all about. We start by examining the history of avionics, and some of the first systems that were developed.

1.1 History of avionics

The development of Avionics started as early as the 18th century. In 1783, the Montgolfier brothers used a **barometer** to measure altitude. The famous Wright brothers used avionics as well: an **anemometer** to measure airspeed. Soon after that, aircraft were equipped with **magnetic compasses** to measure heading, **angle of attack vanes** to measure attitude, **fuel-quantity gauges** to measure fuel levels, and so on. Yet in the early days, navigation was still done visually.

At the end of the 1920s, avionics had progressed so much, that the first blind flight and landing was performed: navigation was done solely based on **gyroscopes** and **radio navigation aids**. Over the 1930s, **radio navigation** and **landing aids** were further developed, and implemented on aircraft.

In the 1940s, the second world war started. This resulted in the development of **radar** for aircraft detection. Also, communication became more important. **VHF** and **UHF** communication was developed. But with all these systems, the pilot really had a hard time. The two decades after WW2 therefore mainly resulted in a reduction of the pilot workload. Systems like **autopilots**, **automated warning systems** and **integrated flight instruments** were developed. This was also the time where the **basic T** was implemented in aircraft.

The oil crisis of the 1970s changed the way of flying. Efficiency became an important topic. **Digital computers** were developed, aiding pilots in flying and navigating their airplane as efficiently as possible. Thanks to the development of **multi-function displays**, information could also be displayed in a much more flexible way. This resulted in so-called **glass cockpits**: cockpits with a lot of displays.

Today, glass cockpits are quite common. But we also have systems like **GPS navigation** and **digital communication links**. And many more advanced systems are yet to come. What will the future bring us? Only time will tell.

1.2 Pressure based systems

We can use pressure to measure things like altitude, vertical airspeed and airspeed itself. How do we do this?

1.2.1 Measuring altitude and vertical speed

Let's examine the **international standard atmosphere** (ISA). At **mean sea level** (MSL), we have $p_0 = 101325Pa$, $\rho_0 = 1.225kg/m^3$ and $T_0 = 288.15K$, $R = 287.05m^2/s^2K$. Up to a height of $h = 11km$ (the **troposphere**), the temperature changes by $\lambda = -6.5K/km$.

To calculate with this atmosphere, we define the **geopotential height** $H = \frac{g}{g_0}h$, with $g_0 = 9.81$. This

implies that

$$\frac{p}{p_0} = \left(\frac{T}{T_0}\right)^{-\frac{g_0}{\lambda R}} = \left(1 + \frac{\lambda H}{T_0}\right)^{-\frac{g_0}{\lambda R}} \quad \text{or} \quad H = \frac{T_0}{\lambda} \left(\left(\frac{p}{p_0}\right)^{-\frac{\lambda R}{g_0}} - 1 \right). \quad (1.2.1)$$

So, based on the pressure, we can find the **pressure altitude**. The pressure can simply be measured using an **aneroid**: a static air pressure meter.

It should be noted that the pressure altitude is not equal to the real attitude. So what's the use of using it? Well, the pressure altitude is mainly used for vertical separation of aircraft. During normal flight, all aircraft have set their altimeter to the ISA MSL pressure. (This is called **QNE**.) They may not know their exact altitude. But if they agree to fly on different pressure altitudes, collisions will at least be prevented.

But the pressure altitude can also be used during landing. Let's suppose that an aircraft comes close to an airport. The pilot then simply inserts the local pressure of the airport, converted to MSL, into his altimeter. (This is called **QNH**. The height at which the pilot does this – the **transition level** – differs per airport.) Now the pressure altitude approximately equals the actual altitude. So the pilot can use his altimeter for landing the airplane.

When we can measure height, we can also measure vertical speed. This time, we use a **variometer**, which measures pressure rates dp/dt . Based on this rate, we can find the vertical speed, using

$$\frac{dH}{dt} = -\frac{RT}{g_0 p} \frac{dp}{dt}. \quad (1.2.2)$$

1.2.2 Measuring airspeed

To measure the airspeed, we can use a **pitot tube**. This tube measures the difference between the **total pressure** p_t and the **static pressure** p . Bernoulli's law states that

$$p_t = p + q = p + \frac{1}{2}\rho V_t^2 \quad \Rightarrow \quad V_t = \sqrt{\frac{2(p_t - p)}{\rho}}. \quad (1.2.3)$$

In this equation, q is the **dynamic pressure** and V_t is the **true airspeed** (TAS). However, we usually do not know the density ρ when we're flying. So, instead, we use ρ_0 as density. Of course, this does not give us the true airspeed V_t anymore. Instead, it gives us the **equivalent airspeed** (EAS) $V_e = V_t \sqrt{\rho/\rho_0}$.

There's just one problem. The above method only works for incompressible flows. (So for roughly $M < 0.3$.) For higher velocities, we can use the isentropic relations to find that

$$p_t = p \left(1 + \left(\frac{\gamma - 1}{2} \right) M^2 \right)^{\frac{\gamma}{\gamma - 1}} \quad \text{or} \quad M^2 = \frac{2}{\gamma - 1} \left(\left(\frac{p_t}{p} \right)^{\frac{\gamma - 1}{\gamma}} - 1 \right). \quad (1.2.4)$$

In this equation, the **Mach number** M is defined such that $V_t = Ma$. a is the **speed of sound** and satisfies $a^2 = \gamma RT$.

Again, there's a problem. To find a , we need the **static air temperature** (SAT) T . One way to find it, is to measure the **total air temperature** (TAT) T_t . Then, we use the relation

$$T_t = T \left(1 + \frac{\gamma - 1}{2} \eta M^2 \right). \quad (1.2.5)$$

In this equation, η is the **recovery factor**. (Theoretically, it's 1, but in real life, it's closer to 0.9.) However, in airplanes, usually another method is used. Instead of using the real speed of sound a , we use $a_0 = \sqrt{\gamma RT_0} = 340.3 \text{ m/s}$. This gives us the **calibrated airspeed** (CAS) $V_c = Ma_0$. Again, V_c is not equal to V_t . But, for incompressible flows, we at least do have $V_c = V_e$.

2. Determining your orientation

The first step to knowing where you are, is knowing how you're oriented. How can we measure our orientation?

2.1 Indicating your orientation

Before we can determine our orientation, we need to be able to indicate it. For this, we first need a **reference frame**. There are several reference frames around, which we can use. We will examine two, both having their origin positioned at the center of gravity (CG) of the aircraft. In the **geodetical frame of reference** \mathcal{F}^g , the X_g axis points to the North, the Y_g axis points to the East and the Z_g axis points downward. In the **body frame of reference** \mathcal{F}^b , the X_b axis is the longitudinal axis of the aircraft, Y_b points to the right of the aircraft and Z_b points downward.

To go from one reference frame to another, we use **Euler angles**. These angles indicate rotations. For example, to go from \mathcal{F}^g to \mathcal{F}^b , we first rotate the coordinate system by the **heading angle** ψ about the Z axis. We then rotate it by the **pitch angle** θ about the Y axis. Finally, we rotate it by the **roll angle** ϕ about the X axis. These three angles indicate the orientation of the aircraft. The system measuring the orientation of the aircraft is called the **attitude and heading reference system** (AHRS).

2.2 Measuring orientation using gyroscopes

One way to measure orientation, is by using gyroscopes. Gyroscopes are discs that have been given a very big angular velocity about one axis. This makes them inert to rotations about the other two axes. (**Inertia** is the resistance to a change in momentum.)

2.2.1 Two degree of freedom gyroscopes

We can distinguish two important types of gyroscopes: **two degree of freedom gyroscopes** and **one degree of freedom gyroscopes**. First, let's examine the two degree of freedom gyroscopes, also known as **free gyros**. They can measure orientation with respect to two axes.

Within the category of free gyros, we can again distinguish two types. A **vertical gyro unit** (VGU) is set up such that its spin axis is vertical. It gives the **gyro horizon**. (You might be wondering, can't you use gravity for that? But when performing a turn, the apparent gravity doesn't point downward. That's why gyros come in handy here: they don't have this problem.) On the other hand, a **directional gyro unit** (DGU) point in a certain direction (e.g. North). DGUs are used to determine the heading of the aircraft.

So how do they work? Well, let's apply a torque T to a disc. If the disc is not spinning, then the angular position ψ of the disc will satisfy

$$T = J\ddot{\psi} \quad \Rightarrow \quad \psi = \frac{1}{2} \frac{T}{J} t^2. \quad (2.2.1)$$

In this equation, J is the moment of inertia of the disc. The above equation implies that the angular position of the disc grows quadratically. However, if the disc is spinning with an angular velocity ω_R , things are different. If we apply the torque T about one of the axes about which the disc is not spinning, we have

$$T = H\dot{\theta} \quad \Rightarrow \quad \theta = \frac{T}{H} t = \frac{T}{J\omega_R} t. \quad (2.2.2)$$

Here, H is the angular momentum of the disc. This time, the angular position of the disc evolves only linearly. And if ω_R is very big, this happens very slowly as well. Thus gyroscopes are very **rigid**: they resist any rotation about an axis other than their rotational axis. We use this to measure our own orientation.

There is one important thing to remember. Let's suppose that the gyroscope is rotating about the Z axis. If we apply a torque T about the X axis, then the gyroscope will rotate about the Y axis. Similarly, if we apply the torque about the Y axis, the gyroscope will rotate about the X axis. This phenomenon (together with the above equation) is known as **precession**.

2.2.2 One degree of freedom gyroscopes

One degree of freedom gyroscopes are also known as **rate gyros**: they measure angular rates. Free gyros have two gimbals: one for each axis about which they're free to rotate. With a rate gyro, one of these gimbals is constrained by some sort of spring. A damping mechanism is attached as well.

Let's again suppose that the gyro is rotating about the Z axis. When the aircraft turns, about (for example) the Y axis, the gyro is forced to rotate along. However, this will cause the gyro to, in fact, rotate about the X axis instead. The spring will now cause a moment about the Y axis, based on the rotation of the gyroscope along the X axis. In other words, the bigger the rotation of the gyro about the X axis, the bigger the force of the spring, and the faster the aircraft is rotating. The force on the spring is thus an indication of how fast the aircraft rotates.

There is also a **rate integrating gyro** (RIG). This is a device that integrates the angular rate. It therefore gives an angular position as output. RIGs do have one downside. Because they integrate an angular velocity, errors are integrated as well. These errors accumulate over time. RIGs therefore lose precision after a couple of hours.

2.2.3 Drift and transport wander

Gyroscopes have several limitations. One of these is **drift**: when the spin axis of the gyroscope starts to rotate. Part of this is caused by **real drift**, due to imperfections in the gyroscope. On the other hand, there is also **apparent drift**. When the Earth rotates, the gyroscope won't rotate along. So, if we are on the equator, then after 6 hours, we will have an apparent drift of 90° . But, if we are on one of the poles, we won't have any apparent drift. Next to drift, there is also **transport wander**. This occurs when we carry a gyroscope over the Earth. (It is similar to apparent drift.)

For both apparent drift and transport wander, we can derive an equation. Let's suppose that we have a Northward velocity $V_N = V \cos \psi$ and an Eastward velocity $V_E = V \sin \psi$. Now, the error rates ω_{X_g} , ω_{Y_g} and ω_{Z_g} about the X_g , Y_g and Z_g axes, respectively, are

$$\omega_{X_g} = \omega_e \cos \Phi + \frac{V_E}{R_e + h}, \quad \omega_{Y_g} = -\frac{V_N}{R_e + h} \quad \text{and} \quad \omega_{Z_g} = -\omega_e \sin \Phi + \frac{V_E}{R_e + h} \tan \Phi. \quad (2.2.3)$$

In this equation, ω_e is the rotational velocity of Earth. The terms with ω_e indicate the apparent drift, while the terms with V_E and V_N indicate the transport wander.

2.2.4 Optical gyroscopes

Next to mechanical gyroscopes (with spinning disks), there are also **optical gyroscopes**. Light is used to measure rotational rates. This results in a higher accuracy, and (since there are no moving parts) less maintenance and more reliability.

We'll briefly explain how it works. In a so-called **ring laser gyro** (RLG), two laser beams are sent along a ring, one clockwise and one counterclockwise. They are being reflected by mirrors on the sides.

Normally, both light beams require the same amount of time $T = 2\pi R/c$ to complete their journey. But when the system rotates with an angular velocity $\dot{\theta}$, one light beam needs to travel an extra distance $R\dot{\theta}T$, whereas the other one needs to travel the same distance less. So, the difference in distance travelled is

$$\Delta L = 2R\dot{\theta}T = \frac{4\pi R^2\dot{\theta}}{c}. \quad (2.2.4)$$

The consequence of this is a shift in the frequency of the two signals. In fact, the difference in frequency Δf is given by

$$\frac{\Delta f}{f} = \frac{\Delta L}{L} = \frac{2R\dot{\theta}}{c} \quad \Rightarrow \quad \Delta f = \frac{4\pi R^2 f}{cL} \dot{\theta} = \frac{4\pi R^2}{\lambda L} \dot{\theta} = K\dot{\theta}. \quad (2.2.5)$$

The frequency difference Δf is thus proportional to the rotational rate $\dot{\theta}$. Also, the constant K is known as the **gyro scale factor** and λ is the wave length of the light signal.

Optical gyroscopes have one small downside. This is the **lock-in phenomenon**: at rotational rates $\dot{\theta}$ below the **lock-in rate**, the two laser beams show coupling effects. (This lock-in rate is in the order of $0.01^\circ/s$ to $0.1^\circ/s$.) This results in a frequency difference $\Delta f = 0$ when $\dot{\theta}$ is near zero. (We thus have a **dead zone**.) Of course this is undesirable. To solve it, we simply twist the optical gyroscope back and forth with a frequency of $\omega \approx 100Hz$. This is called **dithering**. Due to this, $\dot{\theta}$ is nowhere near zero. The lock-in effect will thus not play a role anymore, making the gyroscope accurate.

2.3 Using Earth's magnetic field

2.3.1 Earth's magnetic field, and its imperfections

Let's examine Earth's magnetic field. This field can be represented by a bar magnet. The North pole of this magnet (being the Magnetic North MN) lies at Earth's South pole. Similarly, the South pole of the magnet (the Magnetic South MS) lies at Earth's North pole.

Navigating by means of magnetic compasses is not very precise. This is because the magnetic poles aren't positioned at the exact geographical poles of our planet. In fact, they're slightly off. For this reason, the **magnetic variation** (or **declination**) is defined as the (horizontal) angle between the local magnetic meridian and the geographic meridian. The magnetic variation depends on where you are on our planet. It also changes over time. However, maps are available where the magnetic variation is given quite accurately. So, we can more or less compensate for the magnetic variation.

A magnetic compass aligns itself with magnetic field lines. However, these field lines aren't always horizontal. The **magnetic dip** (or **inclination**) is the (vertical) angle between the magnetic field lines and the local horizontal. So, at the magnetic equator, the inclination is zero, while at the magnetic poles, it is 90° . There are tricks to get rid of the effects of inclination. However, when manoeuvring the aircraft, these tricks will again cause errors. They are therefore not very reliable.

Finally, the **deviation** is defined as the angle between the true magnetic direction, and the direction indicated by the compass. There can be several causes for these deviations. First of all, the compass may be lagging during turns, due to friction. Also, magnetic fields caused by other aircraft systems can cause compass deviations.

2.3.2 Combining a magnetic compass with a gyroscope

We have seen that a compass doesn't work very well when the aircraft is turning, due to lagging. But during a long straight flight, they work fine. On the other hand, gyroscopes work well in turns. But during long straight flights, they will have drift. If we combine a magnetic compass with a DGU, we can get rid of these disadvantages.

So how does this work? During turns, our compass is inaccurate, so we use our gyro. On the other hand, during straight flight, we simply use the magnetic compass to keep our DGU accurate. This sounds like a great idea, but in practice, this means that you have to reset your DGU roughly every 15 minutes. This is quite cumbersome. It would be better if we could adjust the gyroscope electronically. But normal magnetic compasses don't give an electronic signal. However, magnetometers do. Let's examine these.

2.3.3 The magnetometer

A **magnetometer** uses electric signals to detect a magnetic field. Its working is quite complex, but we'll examine it nevertheless.

The main component of the magnetometer is a set of two easily magnetized spoke legs. The Earth's magnetic field runs through these spoke legs, thus magnetizing them. The amount of magnetization depends on the flux $B_1 = H \cos \theta$. Here, θ is the orientation of the coil and H is the magnetic field strength.

Now, let's put a **primary coil** around the two spoke legs. If we put an alternating current on this coil, the spoke legs will get an additional magnetization, with opposite polarity. This magnetization will result in another flux B_2 . But, whereas B_1 was a **static flux**, B_2 is an **alternating flux**.

Let's take the two spoke legs, and put another coil (the **pick-off coil**) around them. Since the spokes are magnetized, an electric current will run through this pick-off coil. Due to the opposite polarity, the alternating flux of these two spokes will cancel out at the pick-off coil. We remain with a rather useless constant signal. This method thus doesn't work yet.

We therefore need to apply another trick. And this trick is based on **saturation**: we can't magnetize the two spoke legs indefinitely. After a certain amount of magnetic force H , we simply can't magnetize the spoke legs any further. (The flux density B will remain constant, even if we increase H .) We have reached the **saturation point**. We now make sure that, by applying our alternating current, we exactly reach this saturation point.

Let's ask ourselves, what happens if we also add the magnetic flux B_1 due to Earth's magnetic field to this? This will then imply that the magnetization of the spoke legs will lose its sinusoid shape. On one side of the graph, the flux has reached its maximum. The tips of the sinusoid are thus 'cut off'. The resulting voltage at the pick-off coil will then contain several jumps. The amplitude of these **voltage jumps** is an indication of the angle with respect to the magnetic field lines.

Now we can get an indication of the angle with respect to the magnetic field lines. But this still leaves some ambiguity. For example, no distinction can be made between an angle of 0° and 180° . For this reason, we usually use 3 spokes, being at 120° angles with respect to each other. By combining the signals from the three spokes, the exact direction with respect to Earth's magnetic field can be found.

By using a magnetometer, the downsides of magnetic compasses are mostly gone. Only during steep turns can there still be problems. For this reason, we often combine a magnetometer with a directional gyro. This will result in a **Magnetic Heading Reference System** (MHRS) called a **GYROSYN compass**.

3. Navigation basics and inertial navigation

Navigation is determining your own position and velocity. **Guidance** is the process of reaching a certain destination. Naturally these two topics are closely linked: you can't reach a destination if you don't even know where you are. In this chapter, we'll deal with the basics of navigation and also discuss inertial navigation. In the next chapter, we will discuss more navigation methods, and also deal with guidance.

3.1 Navigation basics

3.1.1 The reference ellipsoid

Before we can determine our position, we need to know with respect to what we want to know our position. One option is to approximate the Earth as a sphere with radius $R_e = 6378km$. There's just one downside. The gravity isn't always perpendicular to the surface for such a sphere. Instead, the **apparent gravity field \mathbf{g}** satisfies

$$\mathbf{g} = \mathbf{G} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}), \quad (3.1.1)$$

where \mathbf{G} is the Newtonian gravity field, $\boldsymbol{\Omega}$ is Earth's rotation vector and \mathbf{R} your position with respect to Earth's center of mass. To solve this problem, we define the so-called **World Geodetic System Ellipsoid**. This ellipsoid represents the Earth surface. It is defined such that the average mean-square deviation between the direction of the apparent gravity and the normal to the ellipsoid is minimized over Earth's surface.

This ellipsoid has several important parameters. First of all, there are the **equatorial radius R_E** and the **polar radius R_N** . Second, there is the **eccentricity $\epsilon = \sqrt{\frac{R_E^2 - R_N^2}{R_E^2}}$** . We need these parameters to define the **meridian radius of curvature R_M** and the **prime radius of curvature R_P** . They are given by

$$R_M = \frac{R_E(1 - \epsilon^2)}{(1 - \epsilon^2 \sin^2 \Phi_T)^{3/2}}, \quad \text{and} \quad R_P = \frac{R_E}{(1 - \epsilon^2 \sin^2 \Phi_T)^{1/2}}. \quad (3.1.2)$$

The definition of the **geodetic latitude Φ_T** will follow in the upcoming paragraph.

3.1.2 Indicating your position

To indicate our position, we still need a reference frame. We'll examine two of them, both having their origin positioned at the center of the Earth. In the **inertial frame of reference \mathcal{F}^i** , the Z_i axis points North, X_i points to a convenient star and Y_i is perpendicular to the other two axes. This reference frame does not rotate as the Earth rotates. On the other hand, the **Earth frame of reference \mathcal{F}^e** does rotate along with Earth. Again, the Z_e axis points North, but now the X_e axis points through the Greenwich Meridian. The Y_e axis is again perpendicular to the other two axes.

To indicate the position of an aircraft, we can use **geocentric spherical coordinates**; relative to a spherical Earth. The aircraft then has a **longitude λ** (with East being positive) and a **(geocentric) latitude Φ** (with North being positive). Also, the distance from the center of the Earth is $R_e + h$, with $R_e = 6378km$ the radius of the Earth and h the height of the airplane. In this way, the position of the aircraft can be expressed by the numbers (λ, Φ, h) .

We can also use coordinates with respect to an ellipsoidal Earth. This gives us the **geocentric spherical coordinates (λ, Φ_T, h)** . h is the height above the reference ellipsoid. Φ_T is the **geodetic latitude**: the angle between the line normal to the ellipsoid surface and the equatorial plane. Finally, λ is still the longitude. (It remains unchanged.)

Let's suppose that we're flying with a velocity V_N to the North and V_E to the East. How will λ and Φ_T change? They will do this according to

$$\dot{\Phi}_T = \frac{V_N}{R_M + h} \quad \text{and} \quad \dot{\lambda} = \frac{V_E}{(R_P + h) \cos \Phi_T}. \quad (3.1.3)$$

3.1.3 History of navigation

Navigation became very important at the time when several European countries sent out ships across the world. The navigators on these ships could easily determine their latitude. All they needed to measure was the angle of the sun above the horizon at noon. However, finding the longitude was a lot harder. In fact, the problem was so hard that, in 1714 in London, an enormous prize of 20,000 pounds was promised for a solution.

John Harrison solved the problem. His solution was as impressive as it was simple: you just use a clock. The only problem was that, back in the early 18th century, clocks weren't accurate enough. But Harrison's clock, using springs, only had a deviation of about 0.7 seconds per day. With this clock, navigation was possible around the globe.

So how does Harrison's method work? First, you measure the local time, using the sun. Then, using your clock, you measure a certain reference time (like e.g. the London time). Then, by using the time difference, you can find the relative longitude, with respect to London.

3.2 Categorizing navigation

3.2.1 Navigation system categories

Navigation systems must satisfy several **performance requirements**. They must have...

- **Accuracy** – The information that they give must be close to the actual value.
- **Integrity** – If the system can not give sufficiently accurate information, it must notify the user of this in time.
- **Availability** – It may not occur that the system is unexpectedly unavailable.
- **Continuity of service** – If the system stops working after, for example, 2 years, it's not really useful.

There are several ways to categorize navigation systems.

- A **primary means** navigation system is a system that meets the accuracy and integrity requirements, but does not necessarily meet the other two requirements all the time.
- A **sole means** navigation system is a system that, for a given phase of the flight, satisfies all four requirements. However, in real life, no single sole means navigation system exists. Currently, we always need combinations of primary means systems.
- A **supplemental means** navigation system is a system that must be used in conjunction with a sole means navigation system. An example is the GPS system: it is not integer enough. When the GPS system gives a wrong signal, it often needs 10 to 15 minutes before it can notify the user of this.

Another way to characterize navigation systems is as follows.

- **Positioning systems** measure the state vector of an aircraft without regard to the past. (Examples are using radio navigation or GPS.) The error of positioning systems is usually constant in time.
- **Dead reckoning systems** derive the state vector from a continuous series of measurements relative to an initial point. (The most important example here is the **inertial navigation system** (INS), using gyroscopes and accelerometers.) For such systems, the error accumulates over time.

3.2.2 Categorizing navigation errors

Let's suppose that we want to fly to some destination. If we fly directly towards our destination in a straight line, then it is called **direct steering** or **area navigation**. However, if we fly along preplanned airways, we are performing **airway steering**.

Of course, when we fly along airways, we don't exactly stay on the flight path. There will be deviations. There are several possible causes for these deviations.

- **Navigation sensor errors** – Errors when measuring your position.
- **Computer errors** – Errors when processing measured data.
- **Data entry errors** – For example, when the wrong position of a beacon is entered in the computer.
- **Display errors** – When humans read a display incorrectly.
- **Flight-technical errors** (FTE) – Due to reasons outside the navigation system.

The first four points of this list together form the **navigation system error** (NSE). If we add this up to the FTE, we get the **total system error** (TSE).

The NSE is a very important parameter. It is used to calculate the minimum distance between two airways or, similarly, between two runways. It can also be used to determine the risk of airborne collisions.

3.3 Inertial navigation

3.3.1 Inertial navigation system basics

An **inertial navigation system** (INS) use accelerometers to measure accelerations. These accelerations are then integrated along the axes to find the change in position. Advantages of such a system is that it is continuously available, self-contained, autonomous, and accurate. However, it is also quite expensive and, since it is a dead reckoning system, its performance degrades over time.

An **accelerometer** is not much more than a box containing a mass and a spring. If the box is accelerated, the spring needs to exert a force on the mass to tug it along. (The mass has inertia.) The deflection of the mass from the equilibrium position gives an indication of the acceleration. However, what accelerometers measure is not exactly the acceleration. Instead, it is the **specific force** \mathbf{f} . To find the acceleration \mathbf{a} , we also need to take into account gravity. Thus, $\mathbf{f} = \mathbf{a} - \mathbf{g}$. To know in which direction \mathbf{g} points, we have to know the orientation of the aircraft. For this, we can use gyroscopes.

One of the downsides of inertial navigation is the performance degradation over time. Let's examine this more closely. If an accelerometer has an **accelerometer bias** B then, after a time t , the distance will be off by $\frac{1}{2}Bt^2$. But the accelerometers aren't the only problems. We also have gyros. If we have a constant **gyro drift rate** W , then we will get a **platform tilt error** $\Delta\theta = Wt$. This will result in an acceleration error of $g\Delta\theta = gWt$. The distance error thus becomes $\frac{1}{6}gWt^3$.

3.3.2 Types of inertial systems

There are two important types of inertial systems. First, there are **stable (gimballed) platform** systems. Such a system have a stable horizontal platform. It consists of two loops. First, there is the **fast loop**. In this loop, gyroscopes measure the orientation of the platform. If it is not exactly horizontal, servo motors are used to turn the platform back to a horizontal position. Second, there is the **slow loop**. This loop contains accelerometers. These accelerometers are used to find the position of the aircraft. Of course this position is what we wanted to know. But the position data is also used to precess the gyroscopes, such that they follow the Earth's curvature. (If we don't, then apparent drift and transport wander will start to play a role.)

The second type of system is the **strapdown** system. Now, the INS is mounted on the airplane. Gyroscopes measure the orientation of the INS. Then, computers use this data, together with the output of the accelerometers, to determine the position of the aircraft. Strapdown systems thus do not have a real stable platform. But in the computer, there is some kind of virtual 'stable platform', known as the **analytic platform**. And, just as in the stable platform system, the strapdown system also consists of two loops. But again, these loops are only present in the computer.

Since the 1980s, when computers have become widely used, strapdown systems have become the dominant type of inertial systems. They have reduced mechanical complexity, improved reliability, lower power consumption, lower volume/weight and lower cost.

3.3.3 Schuler tuning

The downside of inertial navigation systems, is that they require tuning: you need to set the initial orientation. This is not very difficult, when the system is standing still. But now let's suppose that we accidentally do give the platform a horizontal acceleration a . This will result in a platform, tilted by an angle

$$\theta = \arctan\left(\frac{a}{g}\right). \quad (3.3.1)$$

The effect of this tilt angle is somewhat surprising. Because of the (unknown) tilt angle, the system senses a component of the gravity for which it does not compensate. It is thus interpreted as an acceleration. It now seems as if the aircraft is (for example) accelerating: the system will say the aircraft flies faster than it really does. Because of this, the slow loop of the platform will compensate more for the effects of the curved Earth. The tilt angle of the platform thus decreases. After a while, this will result in a tilt of the platform in the other way. This will then cause the aircraft to decelerate. And the whole process happens again, but now in the opposite way. This repetitive process is called **Schuler tuning**.

We can derive equations for the effects of errors. A platform tilt angle $\Delta\theta$ (or similarly, a constant accelerometer bias $B = g\Delta\theta$) will give a distance error of

$$D_e(t) = \frac{B}{\omega_s^2} (1 - \cos \omega_s t), \quad (3.3.2)$$

where $\omega_s = \sqrt{\frac{g}{R_e}}$ is the **undamped natural frequency** of the Schuler loop. Similarly, the distance error due to a constant gyro drift rate W is

$$D_e(t) = WR_e \left(t - \frac{1}{\omega_s} \sin(\omega_s t) \right). \quad (3.3.3)$$

So, both errors result in an oscillation. This **Schuler oscillation** has a fixed period $T_s = 2\pi\sqrt{\frac{R_e}{g}} = 84.4$ minutes, irrespective of where you are on Earth. And surprisingly, the Schuler oscillation doesn't only occur for the real platform of a stable platform system. It also occurs for the analytic platform of a strapdown system.

4. Radio based navigation techniques

When navigating, we can also use a certain type of electromagnetic radiation, known as radio signals. In this chapter, we'll examine how this works. First we examine some basic principles. Then we examine certain types of beacons. And after that, we look at how we can use space based signals for navigation.

4.1 Radio navigation basics

4.1.1 Information carrying waves

Radio navigation makes use of **electro-magnetic waves**. These waves travel at the speed of light $c = 2.998 \cdot 10^8 m/s$. Their wavelength λ and frequency f satisfy $c = \lambda f$. To send a signal, we usually take a **wire/antenna** with length $L = \lambda/2$ and put an alternating current on it with the corresponding frequency.

But by just emitting a wave, we don't send any information. We need to be coding information on the radio waves. This principle is called **modulation**. When applying modulation, we modify a (high-frequency sinusoidal) **carrier wave**. There are three important types of modulation.

- **On-off modulation** – When the data signal is 1, we emit the signal. And when the data signal is 0, we emit nothing. The signal thus consists of pulses.
- **Amplitude modulation (AM)** – The amplitude of the carrier wave is proportional to the data signal. We thus have

$$V(t) = V_M(t) \cos(\omega_c t + \phi), \quad \text{where} \quad V_M(t) = V_0(1 + m \cos \omega_s t). \quad (4.1.1)$$

In this equation, $V_M(t)$ is the **data signal**, ω_s is the **modulating signal frequency**, ω_c is the **carrier frequency** and m is the **depth in modulation**, satisfying $0 \leq m \leq 1$.

- **Frequency modulation (FM)** – The frequency of the carrier wave is proportional to the data signal. We now have

$$V(t) = V_0 \cos(\omega_c t + \Delta\omega \sin(\omega_s t) t). \quad (4.1.2)$$

Radio signals can be categorized, based on their carrier frequencies. We have **very low frequency (VLF)**, **low frequency (LF)**, **medium frequency (MF)**, **high frequency (HF)**, **very high frequency (VHF)**, **ultra high frequency (UHF)**, **super high frequency (SHF)** and **extremely high frequency (EHF)**. The frequency for VLF ranges from 3 to 30kHz. The frequency of LF ranges from 30 to 300kHz. This continues to increase with a factor 10 every step until we reach EHF, ranging from 30 to 300GHz.

4.1.2 Propagation of radio waves

Normally, electro-magnetic waves simply travel in a straight line. Therefore, they only allow so-called **line-of-sight navigation**: you can only navigate using beacons you can actually see. It is now important to know over what distance we can see beacons. This **range** R can be found, using the equation

$$R = 1.2\sqrt{h_T} + 1.2\sqrt{h_R}. \quad (4.1.3)$$

In this equation, R is the range in nautical miles, h_T is the transmitter height in feet and h_R is the receiver height in feet. (These units are important. The above equation won't work for other units.) For flying aircraft, the range is often roughly 200 nautical miles.

The downside of line-of-sight navigation is its limited range. However, there are waves that don't travel in a straight line. For frequencies up to $3MHz$ (HF), radio waves tend to follow the curvature of the Earth. These **ground waves** thus enable long-range radio-navigation systems, like the **Long Range Navigation** (LORAN) system. The downside of ground waves is that they quickly lose power with increasing distances.

For frequencies up to $30MHz$ (VHF), there are also **sky waves**. This time, the signals are 'reflected' by the ionosphere of the Earth. Since sky waves need to travel through the atmosphere first, they won't work on short distances. Only on relatively long distances are sky waves useful.

4.1.3 Multipath effects

Let's suppose we send out a low-frequency signal. This signal can now reach its destination along various paths: as a ground wave and as a sky wave. This means that **multipath effects** may occur. The two signals have travelled a different distance, and thus may have a phase difference. This can cause the signal to be **amplified**, or it may be **fading** away.

But this isn't the only situation where multipath effects occur. Let's suppose we have a transmitter at a height h_T . When sending out a signal, this signal can reach a nearby receiver directly. However, it can also reach the receiver after being reflected by the ground. The result of this is that a sort of **lobing** pattern occurs around the transmitter. In some lobes, the signal is amplified by the multipath effects. But in other lobes, the signal fades away.

Of course, it's inconvenient if we can't receive a signal at certain places. How can we solve the lobing problem? One solution is to raise the height h_T of the transmitter. This will result in a lobing pattern with a lot of thin lobes. So, we will never be without a signal for long. Another possibility is to apply **counterpoising**. This is a fancy word for placing a plate under the transmitter, to make it seem that the transmitter is placed on the ground. This will result in one very big lobe in which the signal can be received.

4.1.4 Navigation system types

Now let's try to use radio signals to navigate our airplane. How can we find our position? One way is to just look in which direction there are beacons. This will give us the **radial** θ of the corresponding beacons. (Beacons that give you the radial are called **θ -systems**. An example is the **VHF omnidirectional range** (VOR) beacon.) To find our position in 2D in this way, we need at least two such beacons.

Another type of beacons are **ρ -systems**. These beacons give you the distance ρ with respect to that beacon. An example is the **Distance Measuring Equipment** (DME) beacon. This time, two beacons can't unambiguously say where you are. You thus need at least three beacons to find your position.

We can also combine a θ -system with a ρ -system. This **ρ - θ -system** then tells us the radial and the distance with respect to the beacon. Examples now include the co-located VOR/DME beacon or the **TACAN beacon**. Now, only one beacon is sufficient to tell us our position.

Finally, there are **hyperbolic systems**, like the LORAN system. With hyperbolic systems, you measure the time difference between two signals arriving from two beacons. This data then results in a hyperbolic line on which your position must lie. You now need at least two beacons (and sometimes even three) to find your position.

4.1.5 Geometric dilution of precision

All the systems treated above are called **line of position** systems. A beacon alone can't tell you your position, but only a 'line' on which your position must be. In reality, the system can't even tell you that.

And this is because accuracy needs to be taken into account.

Let's suppose that we're trying to find our position using two θ -systems. These systems give us the radials θ_1 and θ_2 . However, these radials aren't exactly accurate. They can be off by (for example) 0.5° on either side. We now don't get a point anymore where our position must be. Instead, we get a region. This is displayed in figure 4.1.

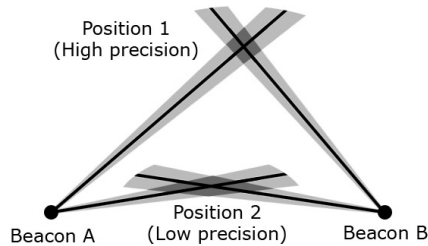


Figure 4.1: Geometric dilution of precision (GDOP), shown in action.

This region, however, has a special property. Its size depends on our position. If we are placed inconveniently with respect to the beacons, then the region is quite big. This 'dependence' of the precision, based on our position, is called **geometric dilution of precision** (GDOP). The GDOP is smallest when the aircraft is positioned on two perpendicular lines of position.

4.2 Navigation systems

It's time examine some navigation systems, and take a look at how they work.

4.2.1 Non-directional beacons and automatic direction finders

A **non-directional beacon** (NDB) is a beacon on the ground, sending a signal in all directions. To use it, we need to equip our aircraft with an **automatic direction finder** (ADF). This ADF has two types of antennas.

- The **loop antenna** has the shape of a loop with two arms AB and CD . When the loop is aligned with the direction of the signal, the arms AB and CD will get the same excitation: nothing happens. This is the **null position** of the loop antenna. On the other hand, when the loop is placed facing the signal, then AB gets a (for example) positive current, while CD gets a negative. Thus, a voltage is created in the loop. This voltage has maximal value V_{max} when the loop exactly faces the direction of the signal.
- The **sense antenna** is a lot simpler. It is omni-directional. It always measures the maximum voltage V_{max} .

The loop antenna alone can't detect the direction: it has a 180° ambiguity. But, by adding up the loop antenna voltage V_1 and the sense antenna voltage V_2 , we get a **cardioid** pattern. Now, the ADF can find the direction of the beacon. The ADF thus yields a bearing to an NDB ground station.

4.2.2 VHF omnidirectional range

A **VHF omnidirectional range** (VOR) beacon gives three types of informations: a voice (for example, to carry weather information), the identity of the VOR beacon and the radial on which the aircraft is positioned, relative to the beacon. This means that a lot of data is put on one wave. In fact, we first take a sub-carrier wave of $30Hz$. This wave is frequency-modulated with the data signal. Then, we use this wave to amplitude-modulate the actual carrier wave of $9960Hz$.

Think this is complex? The working of the VOR itself isn't very easy either. To explain the working principle of the VOR, we examine a 'lighthouse' with two 'lights'. Light 1 is always on, but the direction in which it shines varies. In fact, the light beam rotates at 30 RPM. Light 2 is only on when light 1 shines North. At this moment, light 2 shines in all directions. At every other time, light 2 is off. The result of this is a repeating pattern with a period of $T = 1/30s$. To find the radial, we can now use the equation

$$\frac{t_{\text{light 1 on}} - t_{\text{light 2 on}}}{T} \cdot 360^\circ. \quad (4.2.1)$$

In reality, we don't have two lights. Instead, we use a special **limacon** antenna. This antenna sends signals in a limacon pattern: some sides get a strong signal, while other sides get a weak signal. And the limacon antenna is also rotating at 30 RPS. This results in another modulation: a $30Hz$ amplitude modulation. The signal that is sent thus contains two $30Hz$ signals. By measuring the phase difference between them, we can find our radial θ with respect to the beacon.

VORs have been around for quite a while. In fact, they were used when defining airways. When following an airway, you usually keep on flying from one VOR to the next, until you have reached your destination.

4.2.3 Distance measuring equipment

A **distance measuring equipment** (DME) system is based on two-way communication. First, an airborne **DME interrogator** sends out a UHF pulse. (This is done about 150 times each second.) The ground-based **DME transponder** receives the signal, waits exactly $50\mu s$, and then sends a signal back. The difference between the frequency of the incoming signal and the frequency of the replied signal is always $63MHz$. The airborne equipment now computes the **slant range** d , being the line-of-sight distance. (So, if you're flying directly above a DME transponder, the slant range is not zero, but h .) This slant range is given by

$$d = \frac{1}{2}c(\Delta T - 50\mu s). \quad (4.2.2)$$

One downside of the DME system is its relatively low accuracy. The accuracy is only $\frac{1}{4}NM + 0.0125R$. Another downside is that it requires active communication. A DME beacon can only send out a limited amount of signals. This means that at most 50 to 100 aircraft can make use of a beacon simultaneously.

But, with 50 aircraft using a beacon, how do we know which signal is meant for us? A first suggestion to solve this problem might be to just look at multiple intervals. Let's suppose that, every time we send the signal, we get a response a fixed amount of time Δt later. Then this will probably be the reply to our signal! Sadly, if all aircraft send their signals at regular intervals, this trick doesn't work. But the solution to this problem is **jitter**. Instead of sending out our signal at constant intervals, we apply very small variations in our interval length. (Of up to $10\mu s$ only.) Now, the trick that we just described does work: there's only one signal with a constant time difference Δt .

4.2.4 Long range navigation system

The **long range navigation** (LORAN) system is, surprisingly, used for long range navigation. It's especially useful for oceanic regions or other remote places where no other beacons are present. LORAN makes use of ground waves for navigation. It thus uses LF signals.

The LORAN system consists of transmitting stations (beacons) put together in groups known as **chains**. Every group has one **master station** and at least two (but often more) **secondary (slave) stations**. This is because at least three beacons are needed for a position fix. All beacons in a LORAN chain send out pulses in all directions. This is done at a regular pattern. (This pattern can even be used to identify the LORAN chain.)

A receiver, listening to the LORAN stations, measures the time differences between receiving the pulses. The receiver also knows the time difference between the sending of the pulses. Based on this, he can calculate the difference in distance with respect to the beacons. If he does this for multiple pairs of beacons, he can derive his own position.

The accuracy of the LORAN system varies with position (GDOP). The accuracy is best when the receiver is on the **baseline**: the line between two beacons. Furthermore, the accuracy decreases as the distance from the beacons increases. Up to $350NM$, the accuracy is roughly $130m$. Up to $1000NM$, it is about $550m$. Starting from $1500NM$, multipath effects due to sky waves start to play a role. The accuracy then drops to over $10NM$.

4.3 Satellite radio navigation

Radio signals don't always need to come from ground-based beacons. We might as well use satellites. But how does this work?

4.3.1 Satellite navigation equations

Satellite navigation is all about timing. The signal is sent by the satellite and received by the receiver. The **pseudo range** PR_i to satellite i is now given by

$$PR_i = c\tau = R_i + c\Delta t_u + c\Delta t_{s_i} + \epsilon_{PR_i}, \quad \text{with} \quad R_i = \sqrt{(x_{s_i} - x_u)^2 + (y_{s_i} - y_u)^2 + (z_{s_i} - z_u)^2}. \quad (4.3.1)$$

τ is the **time delay**, Δt_u is the **user clock error**, Δt_{s_i} is the **satellite clock error** and ϵ_{PR_i} is the sum of **other measurement errors**. (Think of atmospheric effects, multipath effects, etcetera.) Now let's assume that Δt_{s_i} and ϵ_{PR_i} are either known (due to models) or ignored; we can eliminate them. Also, the position of the satellites is known. We then remain with four unknowns: x_u , y_u , z_u and Δt_u . We thus also need at least four satellites to find these unknowns. Note that the four equations that we can find in this way are nonlinear. Solving it manually isn't easy. Instead, we usually use some sort of least-squares algorithm to solve it.

GPS systems have an accuracy of about 5 to 10 meters. The main cause for the inaccuracies are atmospheric effects. However, there is a nice way to increase the accuracy of GPS, called **differential GPS**. How does this work? First, we need a reference station. This station has an accurately known position. It compares this position to the position measured by the GPS system. Of course, there will be some very small differences. These **differential pseudorange corrections** are then sent to the user. The user can then also apply these corrections to make his own measurements more accurate. In this way, all **common errors** are cancelled. How well this method works depends mainly on the distance between the user and the reference station.

4.3.2 The global positioning system

It is time to examine a satellite system. Let's just start with the most well-known example: the **global positioning system** (GPS). This system was intended to have 24 satellites, although it often has a bit more. These satellites fly in 6 different orbital planes, each having an inclination of 55° and a height of approximately $20.000km$. This ensures that there are always 5 to 10 satellites in sight.

With the GPS system, every satellite has a **pseudo random noise** (PRN) code $C(t)$. This so-called **coarse/acquisition code** (C/A code), having a frequency of $1.023MHz$, consists of a series of -1 s and 1 s. The code is then multiplied by the $1575.42MHz$ carrier signal to find the **spread spectrum signal** $S(t)$. The reason for this is that energy is spread out over different frequencies. So, the signal can't be jammed anymore by any enemies we might have. Also, since every satellite has its own PRN code, we can use the PRN to identify the satellite.

Next to the PRN, we also send a $50Hz$ **data message** $D(t)$ along with the signal. This data signal contains information like the position/velocity of the satellite, its clock error, and so on. The full signal that is sent is thus

$$S(t) = C(t)D(t) \cos(2\pi ft). \quad (4.3.2)$$

However, the receiver receives the signal

$$S_{rec}(t) = S(t - \tau) = C(t - \tau)D(t - \tau) \cos(2\pi(f t - \tau)), \quad (4.3.3)$$

where τ still is the time delay. The most important problem is that we don't know τ . However, we know the PRN codes of the satellites. So, we just try to match these codes with the signal as well as possible. The best match will give us the time delay τ . And once we know the time delay, we can also find the message $D(t)$ that was sent along with the signal.

5. How avionics can assist pilots

Avionics isn't just about finding out where you are and how you're oriented. It's also about helping the pilots with their job. The question that we'll examine in this chapter is, how can avionics help the pilots?

5.1 Landing guidance systems

Avionics can help pilots land an aircraft. The system involved is the **instrument landing system** (ILS). We'll examine how it works. But first, let's examine landing categories.

5.1.1 Landing categories

The approach and landing are the most dangerous part of the flight. When the pilot has sufficient visibility, he is allowed to perform the landing under **visual flight rules** (VFR). However, when visibility is reduced (e.g. due to bad weather), the pilot has to use **instrument flight rules** (IFR): he has to land, using only his instruments.

The **International Civil Aviation Organization** (ICAO) has defined three **categories of visibility**. These categories are based on the **decision height** (DH) and the **runway visual range** (RVR). The DH is the height above the runway at which the landing must be aborted, if the runway is not in sight. The RVR is the visibility at the runway surface. The three categories are as follows.

- **Cat I** – DH \geq 200 ft and RVR \geq 2600 ft.
- **Cat II** – DH \geq 100 ft and RVR \geq 1200 ft.
- **Cat III** – DH $<$ 100 ft and RVR $<$ 1200 ft.

Category III is further divided into regions a, b and c. But we won't go into detail on that.

5.1.2 The ILS localizer antenna

When landing, the aircraft should follow the glide slope. The **ILS localizer** is a subsystem of the ILS. It makes sure that the ILS system knows whether the aircraft is left or right from the glide slope. So how does it work?

The main part of the ILS localizer subsystem is the ILS localizer antenna array. The central antenna of this array sends out a **base signal**. This base signal is the sum of a $90Hz$ AM carrier wave and a $150Hz$ AM carrier wave. No phase difference is present. (The signal thus consists of a **carrier and side bands** (CSB).) Next to this, there are also the **right antennas**. They send out a sum of a $90Hz$ AM carrier wave and a $150Hz$ AM carrier wave as well. However, this time a phase difference is introduced. The $90Hz$ signal has a phase lag of 90° , while the $150Hz$ signal has a phase lag of 270° . (This signal has **side bands only** (SBO).) For the **left antennas**, this is exactly opposite. The $90Hz$ signal has a phase lag of 270° and the $150Hz$ signal has a phase lag of 90° . (It is SBO as well.)

The result of this will be a lobe pattern. If we're exactly on the vertical plane through the runway, then we will measure the $90Hz$ AM signal with equal strength as the $150Hz$ signal. But now let's suppose that we're off to the left of the runway (while approaching). In this case, the $90Hz$ signal from the antennas will be amplified, whereas the $150Hz$ signal will be more or less faded. The $90Hz$ signal is thus dominant. Similarly, on the right side, the $150Hz$ signal will be dominant. In this way, we will know whether we're off to the left or to the right of the runway.

5.1.3 The ILS glide slope antenna

The **ILS glide slope** subsystem measures whether the aircraft is above or below the glide slope. It works in a similar way as the ILS localizer. But there are a few differences. Of course, the ILS glide slope uses a different carrier wave frequency. (In this way, aircraft can distinguish the ILS localizer signal from the ILS glide slope signal.) Also, instead of using a lot of antennas, we now use multipath effects (via the ground).

The ILS glide slope antenna array consists of only two antennas. Antenna A is $10\lambda \approx 9m$ above the ground. It sends out a CSB signal: the $90Hz$ and the $150Hz$ AM signals don't have a phase difference. On the other hand, antenna B is $5\lambda \approx 4.5m$ above the ground. It sends out an SBO signal: the $90Hz$ part has a phase lag of 180° . (The $150Hz$ signal does not have a phase lag.)

Because of multipath effects, a lobing pattern will be present. Antenna A will give maxima at 1.5° , 4.5° , 7.5° and so on. Antenna B gives maxima at 3° , 9° , 15° and so on. The result will be as follows. From 0 to 3 degrees above the horizontal, $150Hz$ dominates. From 3 to 9 degrees, $90Hz$ dominates. From 9 to 15 degrees, $150Hz$ dominates again. And this continues in steps of 6 degrees. So, if an airplane detects a $90Hz$ dominant signal, it will be above the 3° glide slope. But, if it detects a $150Hz$ dominant signal, it is probably below it (or way above it).

The ILS system works quite well. However, it does require a lot flat terrain. Also, obstacles (like buildings and even aircraft and other vehicles) can disturb the signal.

5.1.4 The microwave landing system

The **microwave landing system** (MLS) gets rid of some of the disadvantages of the ILS system. It works for angles from -40° to 40° horizontally (with respect to the runway) and from 0° to 20° vertically. Its range is $20NM$ up to a maximum height of $20.000ft$. And it uses a **time-reference scanning beam** (TRSB).

How does the TRSB work? We'll examine the horizontal positioning first. TRSB is based on a rotating antenna. Most of the time, this antenna is pointing at the 40° direction. But in regular intervals, it sweeps to the -40° line and back again. An airplane in the corresponding region is thus 'hit' twice by the signal. Based on the time between these two hits, the **azimuth angle** θ can be found.

Finding the (vertical) **elevation angle** ϕ works exactly the same. But, instead of an **azimuth antenna**, we now use an **elevation antenna**. The MLS also has a DME beacon to keep track of the distance of the aircraft.

5.2 Aircraft instruments

It's nice to know where you are. But how do you tell this to the pilots in the right way? That's what we will examine now.

5.2.1 An overview of the cockpit

To inform the pilot of the status of the aircraft, the cockpit contains several displays. Among them are the **heads up display** (HUD), the **primary flight display** (PFD), the **navigation display** (ND), the **multi-function display** (MFD), the **control display unit** (CDU), the **engine indicating & crew alerting system** (EICAS) and the **mode control panel** (MCP). An overview of these displays can be seen in figure 5.1.

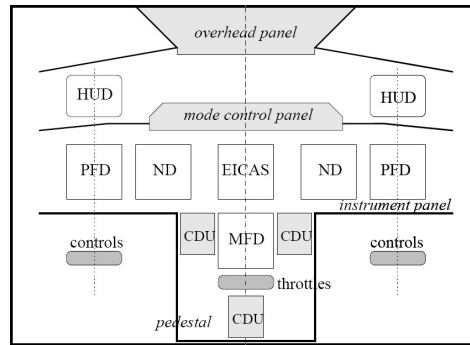


Figure 5.1: The cockpit human machine interface.

5.2.2 Creating efficient displays

When displaying data, we want to make things as easy as possible for the pilots. It should cost them as little time as possible to acquire information. How can we reduce the time pilots need to absorb data?

- A first step is to put displays that are related to each other close together. This will result in an **additive display**. An example is the well-known **basic T**.
- But we can do more than switching displays around. We can also take several pieces of data and integrate them in a single display. This gives us an **accumulative display**. An example now is the **attitude direction indicator** (ADI). It contains an artificial horizon, a roll angle index, a bank angle scale, a deviation from ILS, a rate-of-turn indicator, a slip indicator, and much more.
- Finally, we can process available data in a rather advanced way. This would result in an **integrative display**. This time, an example is the **flight director** (FD).

You might be wondering, what is a flight director? A **flight director** (FD) is a system that tells the pilot how he should fly the aircraft. For example, the pilot wants to land at a certain landing strip. He inserts this into the FD. The FD now computes the pitch angle, roll angle, etc. which the pilot should maintain. The flight director should not be confused with an autopilot. The two systems are slightly similar. However, an autopilot controls the airplane on its own. A flight director only tells the pilot how he should control the aircraft. Then, all the pilot has to do, is follow the needle.

5.2.3 Electronic displays and other future trends

When flight displays were first developed, only mechanical systems could be made. And, since it's impossible to put a huge amount of mechanical systems in a small display, these displays were quite limited. But as time progressed, electronic displays became available. This resulted in the so-called **glass cockpit**. However, initially all that aircraft designers did was take the old mechanical displays, and turn them into a digital display.

Slowly, people got the idea to do more with electronic displays. This resulted in the **electronic flight instrument system** (EFIS). It consists of two parts. First, there is the **electronic attitude and direction indicator** (EADI), also known as the **primary flight display** (PFD). It mainly tells you how you're oriented. Second, there is the **electronic horizontal situation indicator** (EHSI), also known as the **navigation display** (ND). This display mainly informs you about your position.

Still, the EFIS displays the same information as the old mechanical systems were displaying. In the future, this might very well change. One development is the **tunnel in the sky** display: a tunnel is displayed in the sky, which the pilot should follow. Another development is the **head up display** (HUD). With this system, information is superimposed on the pilot's normal view, through the windshield of the aircraft.

6. Controlling air traffic

Air traffic needs to be controlled. The first step is to know where all aircraft are. CNS is important here. Next to this, flights also need to be managed properly, both by air crew and ground crew. This chapter discusses several aspects of these subjects.

6.1 CNS

6.1.1 What is CNS?

CNS stands for **communication, navigation and surveillance**. It consists of three parts. These are, surprisingly,

- **Communication** – Ensuring that the telecommunication necessary for the safety, regularity and efficiency of air navigation are continuously available. An important communication system is the **aircraft communications addressing and reporting system** (ACARS). Currently, aircraft communication systems are low-bandwidth, expensive and local.
- **Navigation** – When airborne, determining your own position and velocity. This usually done by following beacons. But another possibility is **area navigation**, also known as **random navigation** (RNAV). Now, the aircraft still uses beacons for navigation, but flies in a straight line to its destination. This is much more efficient.
- **Surveillance** – Determining the position and velocity of an airborne vehicle. To do this, we often use **radar: radio detection and ranging**. In fact, radar is the primary means of **air traffic control** (ATC) for surveillance. The only places where radar can not be applied is in oceanic regions and other remote areas. On these places we use **procedural voice reporting**: roughly every 30 minutes, the pilots must report their position to air traffic control.

Two types of radar systems can be distinguished: **primary surveillance radar** (PR) and **secondary surveillance radar** (SSR). We will examine PR first. After that, we go into depth on SSR.

6.1.2 Primary surveillance radar

Let's look at what a PR system does. First, it sends pulses for $4\mu s$ in a certain direction. (This is the **pulse width**.) It then starts to listen for $2496\mu s$. In that time, the pulses will be scattered back by the aircraft that are in the neighbourhood. From the signals that come back, the distance and heading of the aircraft, relative to the radar, can be derived.

Note that there is a cycle of $2500\mu s$. Thus, 400 **pulses per second** (PPS) are sent. And, since light can travel only 400 nautical miles $2500\mu s$, the **maximum range** of the radar is $200NM$. The **minimum range** depends on the pulse width. Another interesting thing is the accuracy of the radar system. Let's suppose that the **antenna beam width** is 1.8° . A radar can only measure whether a signal is reflected back. It can not measure where in that 1.8° the aircraft actually is. So, if the aircraft is $200NM$ away, then the size of the aircraft will appear to be 1 nautical mile.

6.1.3 Secondary surveillance radar

SSR systems works slightly different than PR systems. When using SSR, there has to be a **transponder** on the aircraft. When this transponder receives a radar signal, it automatically sends a signal back. The

exact message, which is sent by the aircraft, depends on the SSR mode. There are three important SSR modes.

- In **SSR mode A**, the transponder replies with the **aircraft identification code** (ACID), which is defined by ATC. This message consists of only 12 bits. Thus, $2^{12} = 4096$ possible ACID codes are possible. (SSR Mode A is used, when the interval between the P_1 and P_3 pulses is $8\mu s$.)
- In **SSR mode C**, the transponder replies with the aircraft **flight level** (FL) (being the pressure altitude in feet, divided by 100). So, when using SSR, you can find out which aircraft have which heading, distance and altitude. (SSR Mode C is used, when the interval between the P_1 and P_3 pulses is $21\mu s$.)
- In **SSR mode S**, a unique 24-bit **Mode S address** is assigned to each aircraft. Thus $2^{24} \approx 17$ million IDs are possible. Using the mode S address, aircraft can be unambiguously identified worldwide. (This mode is relatively new, and might replace mode A one day.)

An SSR system sends three pulses: P_1 , P_2 and P_3 . P_1 and P_3 can be seen as the most important signals. The interval between them determines the interrogation mode. However, when transmitting P_1 and P_3 , a lobing pattern will occur: there will be one main lobe and several side lobes. When the aircraft comes close to the SSR beacon, it may start reading the side-lobes. This so-called **side-lobe interrogation** is undesired. To prevent it, we use a signal P_2 . This signal is omni-directional. It is stronger than the side-lobes, but weaker than the main lobe. Thus, an aircraft will only reply to an SSR if the P_1 and P_3 signals are somewhat stronger than the P_2 signal.

Sadly, there are a couple of problems with SSR. First of all, aircraft transponders might become **over-interrogated**, when multiple SSRs are around. Also, problems may arise when two aircraft have the same distance from an SSR system. In this case, the signals from the two aircraft may become merged, such that they lose their meaning. This is called **garbling**.

6.2 The flight management system

The **flight management system** (FMS) is the center of the avionics system. It helps the crew in the planning and execution of the control, monitoring and management of the flight. How did it come into existence?

In the 1970s, the oil crisis forced airlines to fly as efficiently as possible. This led to the development of **aircraft operating manuals** (AOM). This manual contained data on fuel consumption, as a function of cruise altitude, trip distance, and more. The crew then needed to find the optimal flight. Not only was this a lot of work for the crew. But the AOM also didn't take into account other data, like fuel prices, crew salary, etcetera. Thus, the flight management system was developed.

The FMS consists of three important subsystems. The **flight management computer** (FMC) performs the important calculations, the **flight data storage unit** (FDSU) contains almost all data that is necessary and available, and the **command/display unit** (CDU) is the interface of the FMS with the crew. (Although the FMS can display data on the PFD and the ND as well.) The FMS also has three main tasks.

- **Flight planning** – The pilots enter the initial position, the destination, and some other data into the FMS. Next to this, the FDSU also contains data on airports, airways, beacons, waypoints, and so on. Based on this data, the FMS plans the flight of the aircraft.
- **Navigation and guidance** – The FMS takes data from various sources (INS, GPS, etc.) to find the best possible estimate of the aircraft position and velocity. It also computes the ground speed, the wind direction and velocity and much more. In this way, the FMS can provide navigation and guidance for the aircraft.

- **Flight optimization and performance prediction** – The FMS has knowledge on the aircraft type, the aircraft weight, the engine types, the aircraft CG position, the wind properties, the flight plan constraints and so on. Based on this data, it can calculate an optimal flight plan. Next to this, the FMS can also predict the performance, like the fuel usage, the altitude at waypoints, the arrival times, and more.

6.3 Air traffic control

6.3.1 Air traffic services

The present **air navigation system** (ANS) provides **air traffic services** (ATS) for civil aviation. The purpose of ATC is to accommodate air traffic. It should enable aircraft operators to stick to their planning as well and as safely as possible. Air traffic services can be split up into three parts:

- **Flight information service** (FIS) – Collect and handle information to assist pilots. An example system is the **automatic terminal information service** (ATIS). It provides data like weather reports, the QNH, the transition altitude, and more.
- **Alerting service** (AL) – Initiate an early search and rescue activity for aircraft in distress.
- **Air traffic management** (ATM) – This again consists of three subparts.
 - **Air traffic control** (ATC) – Maintain a safe distance between aircraft and obstacles. (By the way, this is mostly only necessary for IFR. In VFR, pilots are often responsible for separation themselves. But (almost) all commercial flights are IFR flights.)
 - **Air space management** (ASM) – Maximize the use of airspace. (For example, based on dynamic time sharing.)
 - **Air traffic flow management** (ATFM) – Ensure an optimal flow of aircraft through busy regions.

6.3.2 Splitting up airspace

There are a lot of **flight information regions** (FIR) around. Such regions can generally be distinguished into two categories. In **uncontrolled airspace**, no ATS is provided. On the other hand, in **controlled airspace** there is ATS. But you do need clearance from the ATC to enter controlled airspace.

Near an airport, there are always several regions. First, there is the **control zone** (CZ), having a radius of only $r = 5NM$ and a height of $h = 3000ft$. (These values can differ a bit per airport though.) Around it is the much bigger **terminal control area** (TMA). It starts at a height of $1500ft$ (below that is lower airspace, where free flying is allowed) and ends at a height of $10500ft$. Around the TMA is the **control area** (CTA), which is up to roughly $19500ft$. Above this is the **upper control area** (UTA). In this area, control is often done over bigger regions (e.g. Eurocontrol). Airports also have a **waiting stack**, where airplanes wait before they can land. In this stack, airplanes fly a holding pattern on an assigned altitude.

Next to air regions, there are also **ATS routes**. These ‘highways-in-the-sky’ often go from beacon to beacon. Most airways are **dual airways**: they have two lanes to separate aircraft. Less busy airlines often have **separation by altitude**.

6.3.3 Controlling aircraft near airports

There are three important types of control centers on an airport.

- The **area control center** (ACC) controls incoming/outgoing airport traffic and en-route traffic that flies through the CTA. It uses a **long-range air-route surveillance radar** (LAR).
- The **aerodrome control** (TWR) controls air traffic in the CTR (so very near to the airport). Also, taxiing aircraft are controlled by the TWR. The TWR uses both **airport surface detection equipment** (ASDE) and visual inspection (binoculars) for surveillance.
- Finally, the **approach/departure control** (APP) forms the link between ACC and TWR. It controls the approach and departure of aircraft in the TMA. The APP uses a **terminal area surveillance radar** (TAR).

When aircraft leave from an airport, they often use a **standard instrument departure** (SID). These are procedures (defined by APP) that make use of beacons. They connect the aircraft departure with an ATS route. Similarly, arriving aircraft use **standard terminal arrival routes** (STARs). SIDs and STARs are defined for three main reasons: to separate incoming and outgoing traffic, to reduce noise around the aircraft and to reduce the necessary communication between the pilot and the controller.

To improve the control possibilities of aircraft, the aircraft crew must develop a **flight plan** before every flight and hand it to ATC authorities. This plan must consist of (among others) aircraft data, whether IFR or VFR is used, what navigation equipment is present, what TAS and cruise altitude is planned, what the origin and destination of the flight are, which departure time and route are planned, how much fuel and people are on board, and so on. The flight plan is then sent to all relevant ATS units through the **aeronautical fixed telecommunications network** (AFTN).

6.3.4 Airplane noise and safety

ATC has three priorities (in this order): safety, noise abatement and efficiency. To reduce the amount of noise, we can do several things. There are technical measures (more silent engines), political measures (reduce night flying) and operational measures (e.g. defining SIDs and STARs well, or using specific runways).

To ensure safety, **separation criteria** are used. For airways, a **lateral separation** of 1 to 6 nautical miles is used. For **longitudinal separation**, this is doubled. (So 2 to 12 nautical miles is used.) For **vertical separation**, often 1000 *ft* is used (or sometimes 2000 *ft* at higher altitudes).

Next to this, relatively big aircraft also must have an **airborne collision avoidance system** (ACAS). This system autonomously prevents mid-air collisions of aircraft. The most-used ACAS system is the **traffic alert and collision avoidance system II** (TCAS-II). This system interrogates SSR transponders of nearby aircraft and looks for possible conflicts. 40 seconds before the **closest point of approach** (CPA), it issues a **traffic advisory** (TA): where is the ‘dangerous’ aircraft and what is its heading? The pilots can then look for a solution. If they don’t, then 25 seconds before CPA, the system will give a **resolution advisory** (RA): it shows a vertical escape manoeuvre (e.g. ‘go up’ or ‘go down’) to avoid a collision.

6.4 The future air navigation system

6.4.1 Necessity of a new system

In 1983, the ICAO noticed several shortcomings of the ANS system. Routes and airways are often very indirect and have inconsistent procedures. Widely different aircraft use the same high density traffic area, causing great complexity in traffic flow and control. SIDs and STARs are fixed, which results in very little flexibility. And, in general, insufficient use is made of all the technological possibilities. Thus, improvement was necessary.

The ICAO therefore formed a committee to investigate **future air navigation systems** (FANS). This resulted in ideas for a new **FANS CNS/ATM** system. Ideas for communication, navigation and surveillance were all present. Let's have a look at these ideas.

6.4.2 Communication

The goal was that airplanes could communicate **around the world**, including remote and oceanic regions. Quite likely, digital data links would be used for this. For this, we could use VHF data links: ground stations are quite cheap and already widely available. Next to this, also SSR Mode S is a possible option; although this may not interfere with the primary task of SSR, being surveillance.

However, these two options both use line-of-sight communication. The third option is developing an **aeronautical mobile satellite service** (AMSS) system. This would ensure high-quality global communication, except at the poles. Also, ground stations are not required. But sadly, this system would be rather expensive. These three systems together could then form an **aeronautical telecommunications network** (ATN).

6.4.3 Navigation

For navigation, it would be nice if there is one system, providing adequate navigation **all over the world** for **all phases of flight**, for **all users in all meteorological conditions**, without having to follow beacons. A **global navigation satellite system** (GNSS) like GPS, combined with RNAV, can approach this ideal. However, currently, such systems are not always accurate and integer enough.

Luckily, there are methods to reduce these problems. With **receiver autonomous integrity monitoring** (RAIM), the receiver autonomously monitors the integrity of the signal. Also, **local/regional aera augmentation systems** (LAAS/RAAS) are ground-based systems that monitor the status of the GNSS system. The resulting information is then sent to the aircraft to increase the on-board accuracy and integrity.

6.4.4 Surveillance

For **surveillance**, we can use **automatic dependent surveillance** (ADS). This is an on-board avionics system that automatically transmits (via a digital data link) various aircraft data. In **ADS-Broadcast** (ADSB), aircraft periodically transmit data, like their position, velocity, altitude and other kinds of data.

When such a system is in place, we might actually achieve the possibility of **free flight**. In this case, aircraft can choose their own path and speed in real time. All ATC needs to do is make sure that the traffic flow in a region is not exceeded, and ensure separation. The latter is done based on a **protected zone** and a smaller **alert zone** around the aircraft. These zones must always stay clear of other aircraft.