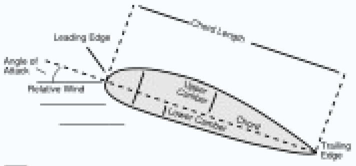
Airfoil

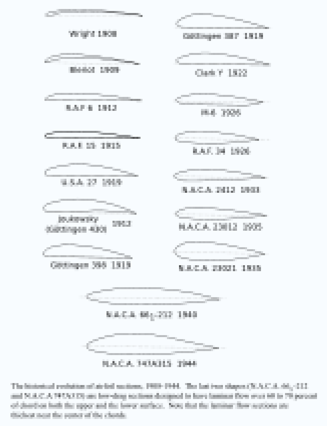


Various components of the airfoil.

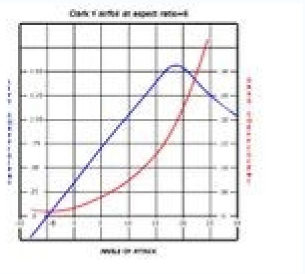
An **airfoil** (in American English, or **aerofoil** in British English) is the shape of a wing or blade (of a propeller, rotor or turbine) or sail as seen in cross-section.

An airfoil shaped body moved through a fluid produces a force perpendicular to the fluid called lift. Subsonic flight airfoils have a characteristic shape with a rounded leading edge, followed by a sharp trailing edge, often with asymmetric camber. Airfoils designed with water as the working fluid are also called **hydrofoils**.

Introduction



The historical evolution of airfoil sections, 1908 - 1944, NASA



Lift and Drag curves for a typical airfoil

A fixed-wing aircraft's wings, horizontal, and vertical stabilizers are built with airfoil-shaped cross sections, as are helicopter rotor blades. Airfoils are also found in propellers, fans, compressors and turbines. Sails are also airfoils, and the underwater surfaces of sailboats, such as the centerboard, and keel are similar in cross-section and operate on the same principles as airfoils. Swimming and flying creatures and even many plants and sessile organisms employ airfoils; common examples being bird wings, the bodies of fishes, and the shape of sand dollars. An airfoil shaped wing can create downforce on an automobile or other motor vehicle, improving traction.

Any object with an angle of attack in a moving fluid, such as a flat plate, a building, or the deck of a bridge, will generate an aerodynamic force perpendicular to the flow called lift. Airfoils are more efficient lifting shapes, generating lift with lower drag and maintaining lift at higher angles of attack. A lift and drag curve obtained in wind tunnel testing is shown on the right.

Airfoil design is a major facet of aerodynamics. Various airfoils serve different flight regimes. Asymmetric airfoils can generate lift at zero angle of attack, while a symmetric airfoil may better suit frequent inverted flight as in an aerobatic airplane. Supersonic airfoils are much more angular in shape and can have a very sharp leading edge. A supercritical airfoil, with its low camber, reduces transonic drag divergence. Movable high-lift devices, flaps and slats are fitted to airfoils on many aircraft.

Schemes have been devised to describe airfoils — an example is the NACA system. Various ad-hoc naming systems are also used. An example of a general purpose airfoil that finds wide application, and predates the NACA system, is the Clark-Y. Today, airfoils are designed for specific functions using inverse design programs such as PROFIL and XFOIL. Modern aircraft wings may have different airfoil sections along the wing span, each one optimized for the conditions in each section of the wing.

An airfoil designed for winglets (PSU 90-125WL)

Airfoil terminology

The various terms related to airfoils are defined below:[1]

- The *mean camber line* is a line drawn half way between the upper and lower surfaces.
- The *chord line* is a straight line connecting the leading and trailing edges of the airfoil, at the ends of the mean camber line.
- The *chord* is the length of the chord line and is the characteristic dimension of the airfoil section
- The *maximum thickness* and the location of maximum thickness are expressed as a percentage of the chord

An airfoil section is nicely displayed at the tip of this Denney Kitfox aircraft (G-FOXC), built in 1991.

Thin Airfoil Theory

A simple mathematical theory of 2-D thin airfoils was devised by Ludwig Prandtl and others in the 1920s.

The airfoil, centre-line equation y(x), is considered to produce a distribution of vorticity $\gamma(s)$ along the chord line s. By the Kutta condition, the vorticity is zero at the trailing edge. Since the airfoil is thin, x can be used instead of s, and all angles can be approximated as small.

From the Biot-Savart law, this vorticity produces a flow field $w(s)$ where

$$w(x) = \frac{1}{(2\pi)} \int \frac{\gamma(x')}{(x - x')} dx'$$

Since there is no flow normal to the curved surface of the airfoil, w(x) balances that from the component of main flow V which locally normal to the plate - the main flow is locally inclined to the plate by an angle $\alpha - dy / dx$. That is

$$V.(\alpha - dy/dx) = w(x) = \frac{1}{(2\pi)} \int \frac{\gamma(x')}{(x - x')} dx'$$

This integral equation can by solved for $\gamma(x)$, after replacing x by

$x = c(1 - cos(\theta)) / 2$,

as a Fourier series in $A_n sin(n\theta)$ with a modified lead term $A_0(1 + cos(\theta)) / sin(\theta)$

That is $\frac{\gamma(\theta)}{(2V)} = A_0 \frac{(1 + cos(\theta))}{sin(\theta)} + \sum A_n.sin(n\theta))$

(These terms are known as the Glauert integral).

The coefficients are given by $A_0 = \alpha - \frac{1}{\pi} \int ((dy/dx).d\theta$

and $A_n = \frac{2}{\pi} \int cos(n\theta)(dy/dx).d\theta$

By the Kutta-Joukowski theorem, the total lift force F is proportional to

$$\rho V \int \gamma(x).dx$$

and its moment M about the leading edge to

The calculated Lift coefficient depends only on the first two terms of the Fourier series, as

$$C_L = 2\pi(A_0 + A_1/2)$$

The moment M depends only on $A_0, A_1 and A_2$, as

$C_M = - 0.5\pi(A_0 + A_1 - A_2 / 2)$

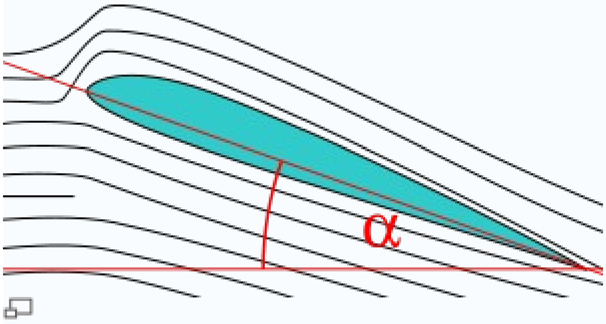
From this it follows that the center of lift is aft of the 'quarter-chord' point 0.25 c, by

$$\Delta x / c = 0.25\pi((A_1 - A_2) / C_L)$$

The aerodynamic center is at the quarter-chord point. The AC is where the pitching moment M' does not *vary* with angle of attack ie

$$\frac{\partial(C_{M'})}{\partial(C_L)} = 0$$

Angle of attack



In this diagram, the black lines represent the flow of the wind. The wing is shown end on. The angle α is the angle of attack.

Angle of attack (**AOA**, α, Greek letter alpha) is a term used in aerodynamics to describe the angle between the airfoil's chord line and the direction of airflow wind, effectively the direction in which the aircraft is currently moving. It can be described as the angle between where the wing is *pointing* and where it is *going*.

The amount of lift generated by a wing is directly related to the angle of attack, with greater angles generating more lift (and more drag). This remains true up to the stall point, where lift starts to decrease again because of flow separation.

Planes flying at high angles of attack can suddenly enter a stall if, for example, a strong wind gust changes the direction of the relative wind. Also, to maintain a given amount of lift, the angle of attack must be increased as speed through the air decreases. This is why stalling is an effect that occurs more frequently at low speeds.

Nonetheless, a wing (or any other airfoil) can stall at any speed. Planes that already have a high angle of attack, for example because they are pulling g or a heavy payload, will stall at speed well above the normal stall speed, since only a small increase in the angle of attack will take the wing above the critical angle.

The critical angle is typically around 15° for most airfoils. Using a variety of additional aerodynamic surfaces — known as high-lift devices — like leading edge extensions(leading edge wing root extensions), fighter aircraft have increased the potential flyable alpha from about 20° to over 45°, and in some designs, 90° or more. That is, the plane remains flyable when the wing's chord is perpendicular to the direction of motion.

Some aircraft are equipped with a built-in flight computer that automatically prevents the plane from lifting its nose any further when the maximum angle of attack is reached, in spite of pilot input. This is called the *angle of attack limiter* or *alpha limiter*. The pilot may disengage the alpha limiter at any time, thus allowing the plane to perform tighter turns (but with considerably higher risk of going into a stall). A famous military example of this is Pugachev's Cobra, a maneuver which has only been performed by the MiG-29, the Su-27/Su-33 family and some prototype Western aircraft, although some consider the F-15/F-16 to be capable if really pushed. Modern airliners which limit the angle of attack by means of computers include the Airbus 320, 330, 340 and 380 series. Currently, the highest angle of attack recorded for a duration of more than 10 seconds is 89.8, performed in the Russian Su-35 (Flanker-E)/Su-37 (Flanker-F) family. The F-35 is believed to be able to perform in even higher angles of attack for prolonged periods of time.

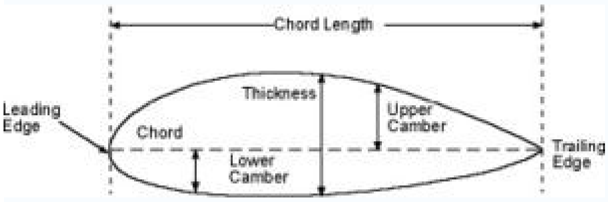
Sailing

In sailing, the **angle of attack** is the angle between a mid-sail and the direction of the wind. The physical principles involved are the same as for aircraft. See points of sail.

Aspect ratio

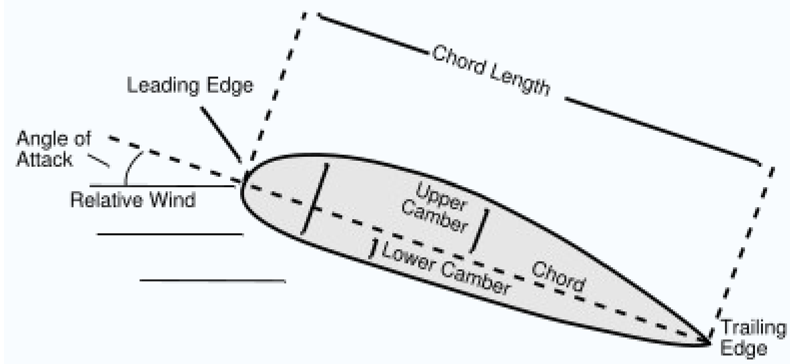
The **aspect ratio** of a two-dimensional shape is the ratio of its longer dimension to its shorter dimension. In aviation, the aspect ratio of aircraft tapered wings is found by dividing the square of the wing span by the wing area.

Camber (aerodynamics)

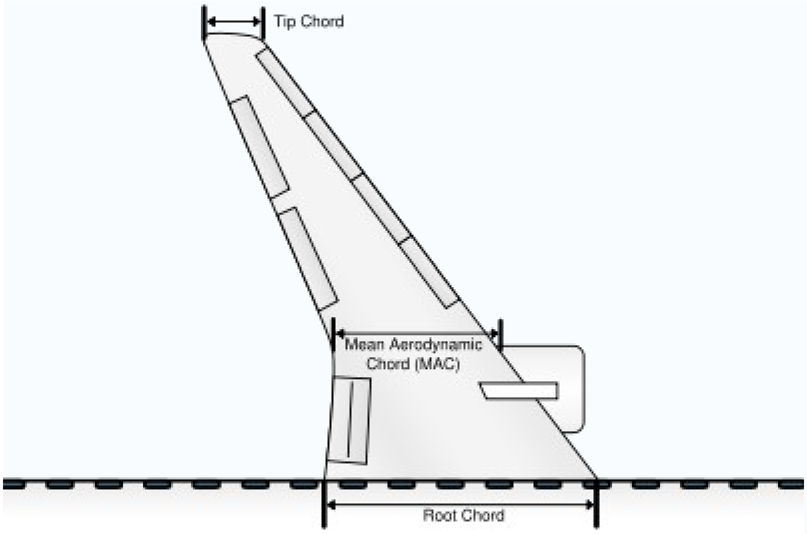


The **camber** in aerospace engineering is the asymmetry between the top and the bottom curves of an airfoil. Cambered airfoils generate lift at positive, zero, or even small negative angle of attack, whereas a symmetric airfoil only has lift at positive angles of attack.

Chord (aircraft)



Cross section of an airfoil showing chord



The various chords on a wing of an aircraft.

In reference to aircraft, **chord** refers to the distance between the leading edge and trailing edge of a wing, measured in the direction of the normal airflow. These front and back points are referred to as the leading edge and trailing edge.

Most wings change their chord over their width (or *span*). To give a characteristic figure which can be compared among various wing shapes, the *mean aerodynamic chord*, or *MAC*, is used. The MAC is somewhat more complex to calculate, because most wings vary in area over the span, growing narrower towards the outer tips. This means that more lift is generated on the wider inner portions, and the MAC moves the point to measure the chord to take this into account. (If a wing was rectangular, rather than tapering or swept, then the chord would simply be the width of the wing in the direction of airflow.)

Standard mean chord (SMC) is defined as wing area divided by wing span:

$$\mathrm{SMC} = \frac{S}{b},$$

where S is the wing area and b is the span of the wing. Thus, the SMC is the chord of a rectangular wing with the same area and span as those of the given wing. This is a purely geometric figure and is rarely used in aerodynamics.

Mean aerodynamic chord (MAC) is defined as

$$\mathrm{MAC} = \frac{2}{S}\int_0^{\frac{b}{2}} c^2(y)dy,$$

where y is the coordinate along the wing span and $c(y)$ is the chord at the coordinate y. Other terms are as for SMC.

Physically, MAC is the chord of a rectangular wing, which has the same area, full aerodynamic force and position of the center of pressure at a given angle of attack as the given wing has. Simply stated, MAC is the width of an equivalent rectangular wing in given conditions. Therefore, not only the measure but also the position of MAC is often important. In particular, the position of center of mass (CoM) of an aircraft is usually measured relative to the MAC, as the percentage of the distance from the leading edge of MAC to CoM with respect to MAC itself.

The ratio of the width (or *span*) of a wing to its chord is known as the aspect ratio an important indicator of the lift-induced drag the wing will create. In general planes with higher aspect ratios - wide skinny wings - will have less drag. This is why gliders have long wings.

Leading edge slot

A Zenair CH 701 STOL showing its fixed, full span leading edge slots in flight

Leading edge slots are fixed aerodynamic devices used on airplanes. Similar, but retractable, leading edge devices are called slats.

A leading edge slot is a fixed (non-moving) opening behind the wing's leading edge. The slot does not operate at low angles of attack, like those found in cruise flight. At low angles of attack the airflow just passes over and under the slot.

At progressively higher angles of attack air starts to move through the slot from the higher pressure air below the wing to the lower pressure air on top of the wing. The mixture of the air coming over the leading edge and through the slot has greater momentum and thus sticks to the upper surface of the wing to a higher angle of attack than if the slot were not there.

Leading edge slots are generally of two types: those that are full-span and those that are partial-span.

Full span slots are generally found on Short Take-off and Landing STOL aircraft, like the Zenair CH 701 STOL. Their primary purpose is to lower the stall speed of the aircraft, allowing slower landing speeds and short landing rolls.

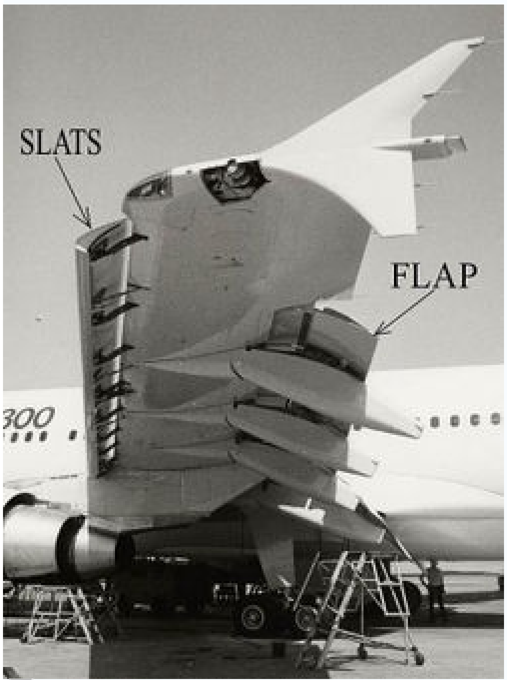
Partial-span slots are usually found only on the outboard portion of the leading edge of the wing where they ensure that that part of the wing will remain unstalled at higher angles of attack than the inboard portions of the wing. This ensures the wing root stalls first and contributes to docile stall behaviour and maintaining aileron control throughout the stall. Using slots in this manner produces a similar result to employing washout on a wing, but through a different means. An example of an aircraft with partial span slots is the Stinson 108.

Slots naturally exact a penalty on the aircraft they are used on. This is because at cruise airspeed they create some drag compared to a non-slotted wing and so reduce cruising speed.

One way to reduce the cruise drag of slots is to make them retractable. These are known as leading edge slats. Slats work in the same way as slots but slats retract at higher speeds when they are not needed. Slats, in turn, are heavier and more complex than slots.

Slats

Slats are small aerodynamic surfaces on the leading edge of the wings of fixed-wing aircraft which, when deployed, allow the wing to operate at a higher angle of attack. Lift is a product of angle of attack and speed, so by deploying slats an aircraft can fly slower or take off and land in a shorter distance. They are usually used while landing or performing manoeuvres which take the aircraft close to the stall, but are usually retracted in normal flight to minimise drag.



The position of the leading edge slats on an airliner (Airbus A-300). In this picture, the slats are extended.


The Fieseler Fi 156 *Storch* had permanently extended slots on its leading edges (fixed slats).

Types include:

- Automatic - the slat lies flush with the wing leading edge until reduced aerodynamic forces allow it to extend by way of springs when needed. This type is typically used on light aircraft.
- Fixed - the slat is permanently extended. This is rarely used, except on specialist low-speed aircraft (these are referred to as slots).
- Powered - the slat extension can be controlled by the pilot. This is commonly used on airliners.

The chord of the slat is typically only a few percent of the wing chord. They may extend over the outer third of the wing or may cover the entire leading edge. Slats work by increasing the camber of the wing, and also by opening a small gap (the slot) between the slat and the wing leading edge, allowing a small amount of high-pressure air from the lower surface to reach the upper surface, where it helps postpone the stall.

The slat has a counterpart found in the wings of some birds, the alula – a feather or group of feathers which the bird can extend under control of its "thumb".

History

Slats were first developed by Handley-Page in 1919; licensing the design became one of their major sources of income in the 1920s. The original designs were in the form of a fixed slot in the front of the wing, a design that was found on a number of STOL aircraft.

During World War II German aircraft commonly fitted a more advanced version that pushed back flush against the wing by air pressure to reduce drag, popping out on springs when the airflow decreased during slower flight. However, the most famous slats of that time belonged to the German Fieseler Fi 156 *Storch*. These were similar in design to retractaible slats, but were fixed non-retractable slots. The slotted wing allowed this aircraft to take off into a light wind in less than 45 m (150 ft), and land in 18 m (60 ft).

In the post-war era slats have generally been hydraulically or electrically operated, allowing for more complex and efficient designs.

Today slats are generally one of several high-lift devices used on airliners, complex flap systems running along the trailing edge of the wing as well.

Lift (force)

The **lift force**, **lifting force** or simply **lift** is a mechanical force generated by a solid object moving through a fluid.[1] Lift is the sum of all the fluid dynamic forces on a body perpendicular to the direction of the external flow approaching that body. The mathematical equations describing lift have been well established since the Wright Brothers experimentally determined a reasonably precise value for the "Smeaton coefficient" more than 100 years ago.[2] But the practical explanation of what those equations mean is still controversial, with persistent misinformation and pervasive misunderstanding.[3]

Sometimes the term **dynamic lift** or **dynamic lifting force** is used for the perpendicular force resulting from motion of the body in the fluid, as in an aerodyne, in contrast to the static lifting force resulting from buoyancy, as in an aerostat. Lift is commonly associated with the wing of a aircraft. However there are many other examples of lift such as propellers on both aircraft and boats, rotors on helicopters, sails and keels on sailboats, hydrofoils, wings on auto racing cars, and wind turbines. While the common meaning of the term "lift" suggests an upward action, the lift force is not necessarily directed up with respect to gravity.

Physical explanation

There are several ways to explain lift which are equivalent — they are different expressions of the same underlying physical principles:

Reaction due to deflection

Lift is created as the fluid flow is deflected by an airfoil or other body. The force created by this acceleration of the fluid creates an equal and opposite force according to Newton's third law of motion. Air deflected downward by an aircraft wing, or helicopter rotor, generating lift is known as downwash.

It is important to note that the acceleration of air flowing over an aircraft wing does not just involve the air molecules "bouncing off" the lower surface. Rather, air molecules closely follow both the top and bottom surfaces, and the airflow is deflected downward when the wing is producing lift. The acceleration of the air during the creation of lift can also been described as a "turning" of the airflow.

Many shapes, such as a flat plate set at an angle to the flow, will produce lift. This can be demonstrated simply by holding a sheet of paper at an angle in front of you as you move forward. However, lift generation by most shapes will be very inefficient and create a great deal of drag. One of the primary goals of airfoil design is to devise a shape that produces the most lift while producing the least Form drag.

It is possible to measure lift using the reaction model. The force acting on the wing is the negative of the time-rate-of-change of the momentum of the air. In a wind tunnel, the speed and direction of the air can be measured (using, for example, a Pitot tube or Laser Doppler velocimetry) and the lift calculated. Alternately, the force on the wind tunnel itself can be measured as the equal and opposite forces to those acting on the test body.

Bernoulli's principle

The force on the wing can also be examined in terms of the pressure differences above and below the wing, which can be related to velocity changes by Bernoulli's principle.

The total force (Lift + Drag) is the integral of pressure over the contour of the wing.

$$\mathbf{L} + \mathbf{D} = \oint_{\partial\Omega} p\mathbf{n}\, d\partial\Omega$$

where:

- **L** is the Lift,
- **D** is the Drag,
- $\partial\Omega$ is the frontier of the domain,
- p is the value of the pressure,
- **n** is the normal to the profile.

Since it is a two-dimensional vector equation, and since lift is perpendicular to drag, this equation suffices to predict both lift and drag. The drag component is Lift-induced drag rather than Form drag. This equation is always exactly true, by the definition of force and pressure.

One method for calculating the pressure is Bernoulli's equation, which is the mathematical expression of Bernoulli's principle. This method ignores the effects of viscosity, which can be important in the boundary layer and to predict drag, though it has only a small effect on lift calculations.

Bernoulli's principle states that in fluid flow, an increase in velocity occurs simultaneously with decrease in pressure. It is named for the Dutch-Swiss mathematician and scientist Daniel Bernoulli, though it was previously understood by Leonhard Euler and others. In a fluid flow with no viscosity, and therefore one in which a pressure difference is the only accelerating force, it is equivalent to Newton's laws of motion.

Bernoulli's principle also describes the Venturi effect that is used in carburetors and elsewhere. In a carburetor, air is passed through a Venturi tube in order to decrease its pressure. This happens because the air velocity has to increase as it flows through the constriction.

In order to solve for the velocity of inviscid flow around a wing, the Kutta condition must be applied to simulate the effects of inertia and viscosity. The Kutta condition allows for the correct choice among an infinite number of flow solutions that otherwise obey the laws of conservation of mass and conservation of momentum.

Some lay versions of this explanation use false information due to lack of understanding the Kutta condition, such as the incorrect assumption that the two parcels of air which separate at the leading edge of a wing must meet again at the trailing edge. There is no reason that a parcel of air on one side of the wing must rejoin a neighboring parcel with which it was originally synchronized on the other side. In fact, the requirement for circulation (see below) in order to generate non-zero lift specifies that parcels must never meet.

Circulation

A third way to calculate lift is to determine the mathematical quantity called circulation; (this concept is sometimes applied approximately to wings of large aspect ratio as "lifting-line theory"). Again, it is mathematically equivalent to the two explanations above. It is often used by practising aerodynamicists as a convenient quantity, but is not often useful for a layperson's understanding. (That said, the vortex system set up round a wing is both real and observable, and is one of the reasons that a light aircraft cannot take off immediately after a jumbo jet.)

The circulation is the line integral of the velocity of the air, in a closed loop around the boundary of an airfoil. It can be understood as the total amount of "spinning" (or vorticity) of air around the airfoil. When the circulation is known, the section lift can be calculated using the following equation:

$$l = \rho V \times \Gamma$$

where ρ is the air density, *V* is the free-stream airspeed, and Γ is the circulation. This is sometimes known as the **Kutta-Joukowski Theorem.**

A similar equation applies to the sideways force generated around a spinning object, the Magnus effect, though here the necessary circulation is induced by the mechanical rotation, rather than aerofoil action.

The Helmholtz theorem states that circulation is conserved; put simply this is conservation of the air's angular momentum. When an aircraft is at rest, there is no circulation. As the flow speed increases (that is, the aircraft accelerates in the air-body-fixed frame), a vortex, called the starting vortex, forms at the trailing edge of the airfoil, due to viscous effects in the boundary layer. Eventually the vortex detaches from the airfoil and gets swept away from it rearward. The circulation in the starting vortex is equal in magnitude and opposite in direction to the circulation around the airfoil. Theoretically, the starting vortex remains connected to the vortex bound in the airfoil, through the wing-tip vortices, forming a closed circuit. In reality, the starting vortex is dissipated by a number of effects, as are the wing-tip vortices far behind the aircraft. However, the net circulation in "the world" is still zero as the circulation from the vortices is transferred to the surroundings as they dissipate.

Common misconceptions

Equal transit-time

One misconception encountered in a number of explanations of lift is the "equal transit time" fallacy. This fallacy states that the parcels of air which are divided by an airfoil must rejoin again; because of the greater curvature (and hence longer path) of the upper surface of an airfoil, the air going over the top must go faster in order to "catch up" with the air flowing around the bottom.

Although it is true that the air moving over the top of the wing is moving faster (when the effective angle of attack is positive) there is no requirement for equal transit time. In fact if the air above and below an airfoil has equal transit time, there is no circulation, and therefore no lift. Only if the air flowing above has a *shorter* transit time than the air flowing below, is upward lift produced, with a downward deflection of the air behind the wing and a vortex at each wing tip. Wind tunnel smoke streamline pictures reveal this.[4][5]

A further flaw in this explanation is that it requires an airfoil to have a curvature in order to create lift. In fact, a thin, flat plate inclined to a flow of fluid will also generate lift.[6][7]

It is unclear why this explanation has gained such currency, except by repetition by authors of populist (rather than rigorously scientific) books and perhaps the fact that the explanation is easiest to grasp intuitively without mathematics. At least one common flight training book depicts the equal transit fallacy, adding to the confusion.[8]

Albert Einstein, in attempting to design a practical aircraft based on this principle, came up with an airfoil section that featured a large hump on its upper surface, on the basis that an even longer path must aid lift if the principle is true. Its performance was terrible.[9]

Coanda effect

A common misconception about aerodynamic lift is that the Coandă effect plays no part.

There are two techniques for increasing the lift on an airfoil. One is to decrease the pressure on the side of the airfoil normal to the direction of the desired lift and the other is to increase the pressure on the other side. (The latter is the primary cause of the lift of a paper airplane.) In order to generate lift one must create a pressure differential between the top and bottom of the airfoil.

The Coandă effect is the name given to the tendency of an airflow, under some conditions, to deflect toward a surface that curves away from the flow direction. This effect is caused by the decreased pressure on the curved surface where it curves away from the flow.

Jef Raskin and a few others have observed that the Coandă effect accounts for part of the lift generated by an airfoil. The decrease in pressure above the airfoil is caused by the interaction of the flow, at the microscopic level, with the curved surface. The effect is caused by a decrease of the pressure on the top of the wing as air particles are blown away from the surface (fewer particles, less pressure due to thermal molecular motion). This contributes to the pressure field under the integral sign in the lift equation.

For large angles of attack and/or high flow rates the Coandă effect results in vortices which may impinge normally on the surface thus increasing the pressure there. Under these circumstances the wing will lose lift and ultimately stall. This aspect of the Coandă effect has been used successfully in the design of the wings in Formula One race cars to pressurize the back of the car and partially offset drag.

For supersonic airplanes to be able to maintain lift at the low speeds necessary for safe landings on aircraft carriers, the stall-producing vortices must be dissipated. This is effected by blowing the boundary layer and other lift augmentation devices.

Some airliners exploit the Coandă effect by deploying slats at the leading edge of the wing. On takeoff when maximum lift is needed at low air speed, the slat moves away from the leading edge leaving a slot which allows some of the high pressure air from the bottom of the wing to blow up over the top of the wing, thus creating a lifting Coandă effect by disrupting vortices that would form there on takeoff. Another use of the Coandă effect to produce lift is the use of Fowler flaps, the aerodynamic surfaces that are deployed from the wing's trailing edge on takeoff and landing. These flaps in effect extend the curved surface of the wing. This extension utilizes the Coandă effect to decrease the pressure on the top of the wing and also "dams" the air as it passes under the wing thus increasing the pressure there. The latter is done at the expense of an increased drag but at the low speeds of takeoff and landing, the increased lift is much more beneficial than the increased drag is detrimental.

The Coandă effect provides one aspect of the lift generated on subsonic airfoils.

Venturi nozzle

Many web sites claim that an airfoil can be analyzed as a Venturi nozzle. The mass flow rate through a Venturi nozzle is constant, so the air must flow faster over the top of the wing. Therefore, there is a lower pressure over the top of the wing, producing lift. However, a Venturi nozzle requires that air is squeezed between surfaces. While this situation does exist with "infinite wing" experiments in wind tunnels, in an aircraft the top of a wing is only one surface. The air is not confined above the wing, therefore a wing is not a Venturi nozzle and it is incorrect to analyze it as such.

Coefficient of lift

The coefficient of lift is a dimensionless number. When the coefficient of lift is known, for instance from tables of airfoil data, lift can be calculated using the *Lift Equation*:

$$C_L = \frac{L}{\frac{1}{2}\rho V^2 A}$$

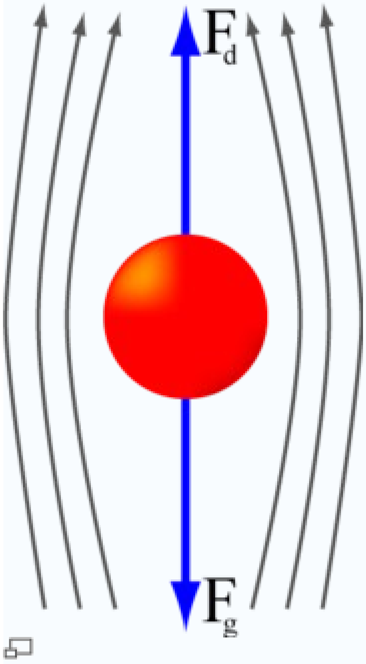
where:

- C_L is the *coefficient of lift*
- ρ is the density of air (1.225 kg/m^3 at sea level)*
- V is the freestream velocity, that is the airspeed far from the lifting surface
- A is the surface area of the lifting surface
- L is the lift force produced

This equation can be used in any consistent system. For instance, if the density is measured in kilograms per cubic metre, the velocity is measured in metres per second, and the area is measured in square metres, the lift will be calculated in newtons. Or, if the density is in slugs per cubic foot, the velocity is in feet per second, and the area is in square feet, the resulting lift will be in pounds force.

- Note that at altitudes other than sea level, the density can be found using the barometric formula

Drag (physics)



An object falling through a gas or liquid experiences a force in direction opposite to its motion. Terminal velocity is achieved when the drag force is equal to force of gravity pulling it down.

In fluid dynamics, **drag** is the force that resists the movement of a solid object through a fluid (a liquid or gas). Drag is made up of friction forces, which act in a direction parallel to the object's surface (primarily along its sides, as friction forces at the front and back cancel themselves out), plus pressure forces, which act in a direction perpendicular to the object's surface. For a solid object moving through a fluid or gas, the drag is the sum of all the aerodynamic or hydrodynamic forces in the direction of the external fluid flow. (Forces perpendicular to this direction are considered lift). It therefore acts to oppose the motion of the object, and in a powered vehicle it is overcome by thrust.

In astrodynamics, depending on the situation, **atmospheric drag** can be regarded as inefficiency requiring expense of additional energy during launch of the space object or as a bonus simplifying return from orbit.

Details

Types of drag are generally divided into three categories: parasitic drag, lift-induced drag and wave drag. Parasitic drag includes form drag, skin friction and interference drag. Lift-induced drag is only relevant when wings or a lifting body are present, and is therefore usually discussed only in the aviation perspective of drag. Wave drag occurs when a solid object is moving through a fluid at or near the speed of sound in that fluid. The overall drag of an object is characterized by a dimensionless number called the drag coefficient, and is calculated using the drag equation. Assuming a constant drag coefficient, drag will vary as the square of velocity. Thus, the resultant power needed to overcome this drag will vary as the cube of velocity. The standard equation for drag is one half the coefficient of drag multiplied by the fluid density, the cross sectional area of your specified green item, and the square of the velocity

Wind resistance is a layman's term used to describe drag. Its use is often vague, and is usually used in a relative sense (e.g. A badminton shuttlecock has more *wind resistance* than a squash ball).

Drag at low velocity; Stokes's Drag

The equation for **viscous resistance** is appropriate for small objects or particles moving through a fluid at relatively slow speeds. In this case, the force of drag is approximately proportional to velocity, but opposite in direction. [1] The equation for viscous resistance is:

$$\mathbf{F}_d = -b\mathbf{v}$$

where:

b is a constant that depends on the properties of the fluid and the dimensions of the object, and
$\mathbf{v}$ is the velocity of the object.

When an object falls from rest, its velocity will be

$$v(t) = \frac{mg}{b}(1 - e^{-bt/m})$$

which asymptotically approaches the terminal velocity $v_t = mg / b$. For a certain b, heavier objects fall faster.

For the special case of small spherical objects moving slowly through a viscous fluid (and thus at small Reynolds number), George Gabriel Stokes derived an expression for the drag coefficient,

$$b = 6\pi\eta r$$

where:

r is the Stokes radius of the particle, and
η is the fluid viscosity.

For example, consider a small sphere with radius r = 1 micrometre moving through water at a velocity $\mathbf{v}$ of 10 µm/s. Using 10^{-3} as the dynamic viscosity of water in SI units, we find a drag force of 0.2 pN. This is about the drag force that a bacterium experiences as it swims through water.

Drag at high velocity

The **Drag equation** calculates the force experienced by an object moving through a fluid at relatively large velocity. The equation is attributed to Lord Rayleigh, who originally used L^2 in place of A (L being some length). The force on a moving object due to a fluid is:

$$\mathbf{F}_d = -\frac{1}{2}\rho v^2 A C_d \hat{\mathbf{v}}$$ see derivation

where

$\mathbf{F}_d$ is the force of drag,
ρ is the density of the fluid (*Note that for the Earth's atmosphere, the density can be found using the barometric formula. It is 1.293 kg/m*^{3} *at 0°C and 1 atmosphere.*),
$\mathbf{v}$ is the speed of the object relative to the fluid,
A is the reference area,
C_d is the drag coefficient (a dimensionless constant, e.g. 0.25 to 0.45 for a car), and

$\hat{\mathbf{v}}$ is the unit vector indicating the direction of the velocity (the negative sign indicating the drag is opposite to that of velocity).

The reference area A is related to, but not exactly equal to, the area of the projection of the object on a plane perpendicular to the direction of motion (i.e., cross sectional area). Sometimes different reference areas are given for the same object in which case a drag coefficient corresponding to each of these different areas must be given. The reference for a wing would be the plane area rather than the frontal area.

Power

The power required to overcome the aerodynamic drag is given by:

$$P_d = \mathbf{F}_d \cdot \mathbf{v} = -\frac{1}{2}\rho v^3 A C_d$$

Note that the power needed to push an object through a fluid increases as the cube of the velocity. A car cruising on a highway at 50 mph (80 km/h) may require only 10 horsepower (7.5 kW) to overcome air drag, but that same car at 100 mph (160 km/h) requires 80 hp (60 kW). With a doubling of speed the drag (force) quadruples per the formula. Exerting four times the force over a fixed distance produces four times as much work. At twice the speed the work (resulting in displacement over a fixed distance) is done twice as fast. Since power is the rate of doing work, four times a work in half the time requires eight times the power.

It should be emphasized here that the drag equation is an approximation, and does not necessarily give a close approximation in every instance. Thus one should be careful when making assumptions using these equations.

Velocity of falling object

The velocity as a function of time for an object falling through a non-dense medium is roughly given by a function involving a hyperbolic tangent:

$$v(t) = \sqrt{\frac{2mg}{\rho A C_d}} \tanh\left(t\sqrt{\frac{g\rho C_d A}{2m}}\right)$$

In other words, velocity asymptotically approaches a maximum value called the Terminal velocity:

$$v_t = \sqrt{\frac{2mg}{\rho A C_d}}$$

With all else (gravitational acceleration, density, cross-sectional area, drag constant, etc.) being equal, heavier objects fall faster.

For a potato-shaped object of average diameter d and of density ρ_{obj} terminal velocity is about

$$v_t = \sqrt{gd\frac{\rho_{obj}}{\rho}}$$

For objects of water-like density (raindrops, hail, live objects - animals, birds, insects, etc.) falling in air near the surface of the Earth at sea level, terminal velocity is roughly equal to

$$v_t = 90\sqrt{d},$$

For example, for human body (d~0.6 m) v_t ~70 m/s, for a small animal like a cat (d~0.2 m) v_t ~40 m/s, for a small bird (d~0.05 m) v_t ~20 m/s, for an insect (d~0.01 m) v_t ~9 m/s, for a fog droplet (d~0.0001 m) v_t ~0.9 m/s, for a pollen or bacteria (d~0.00001 m) v_t ~0.3 m/s and so on. Actual terminal velocity for very small objects (pollen, etc) is even smaller due to the viscosity of air.

As can be seen, terminal velocity for grown-up human and large animals is deadly on impact, but for small animals and birds can be survivable, and for insects - not deadly at all. It is known that sometimes small babies survive a fall that would be fatal for a grown-up - which can be explained by their smaller terminal velocity.

Stall (flight)

In aerodynamics, a **stall** is a sudden reduction in the lift forces generated by an airfoil. This most usually occurs when the critical angle of attack for the airfoil is exceeded.

Because stalls are most commonly discussed in connection with aviation, this article discusses stalls mainly as they relate to aircraft. In layman's terms, a stall in an aircraft is an event that causes the aircraft to drop suddenly (see the overview below).

Note that an aerodynamic stall does *not* mean that an aircraft's engines have stopped or that the aircraft has stopped moving.

Overview

Aircraft are supported in the air by an aerodynamic force called lift, which is generated by the wings of the aircraft as air is forced past the wings by the forward movement of the aircraft. The wings of the aircraft generate lift when they are pointed slightly upward with respect to the direction of the air flowing towards them. If the pilot tilts the aircraft upward, the wings form a larger angle with the airflow, and lift increases. This angle is called the *angle of attack*, or *AOA*. The heavier the aircraft and/or the slower the aircraft is flown the greater must be the angle of attack to generate the lift force necessary to maintain altitude.

Although raising the nose of the aircraft increases angle of attack and thus increases lift, this cannot be done without limit. Up to a certain angle of attack, called the *critical angle of attack,* pointing the wings upward continues to produce more lift. However, beyond the critical angle of attack, the airflow behind the wing separates from the wing and becomes turbulent, and the aerodynamic effects that produce the lifting force largely disappear, and the wing *stalls*—that is, it ceases to provide enough lift to support the aircraft. At the same time, the turbulence greatly increases drag, which slows the aircraft down as it moves through the air, and this also reduces lift. As a result of these changes, the aircraft begins to sink rapidly towards the ground.

Recovering from a stall is simple. Since the stall is caused by an excessive angle of attack, simply pointing the nose of the aircraft downward will stop the stall, by reducing the angle between the wings and the flow of air. Some aircraft have a natural tendency to pitch downward (sometimes dramatically) when the wings stall; others must be directed downward by the pilot. As soon as the angle of attack drops below the critical angle, the aerodynamic stall of the wings will cease i.e. the wings will produce lift and far less drag. However, the aircraft may still be flying too slowly to generate enough lift to prevent the aircraft from continuing to descend: Recovery from the stall includes regaining this necessary speed.

Typically a stall is caused by the pilot attempting to fly the aircraft too slowly, or to pull up too quickly from a dive, or to turn too steeply. Each of these causes the nose to be lifted until the wing's critical angle of attack is exceeded. Increasing engine power counteracts the increased drag caused by the stall and also increases air speed, and this helps in recovery from a stall. The critical action is recovering from a stall is, however, reduction in the angle of attack i.e. lowering the nose.

Altitude (height above the ground) is lost by the aircraft during the stall itself but considerably more height can be lost during the recovery i.e. while regaining enough speed to generate enough lift to maintain altitude. If the aircraft is already at a high altitude this is not a problem. If the aircraft is very close to the ground, however, a stall may cause the aircraft to lose so much altitude that it hits the ground before recovery from the stall is possible. For this reason, pilots are especially careful to avoid stalls during take-off and landing procedures, when the aircraft is often very close to the ground.

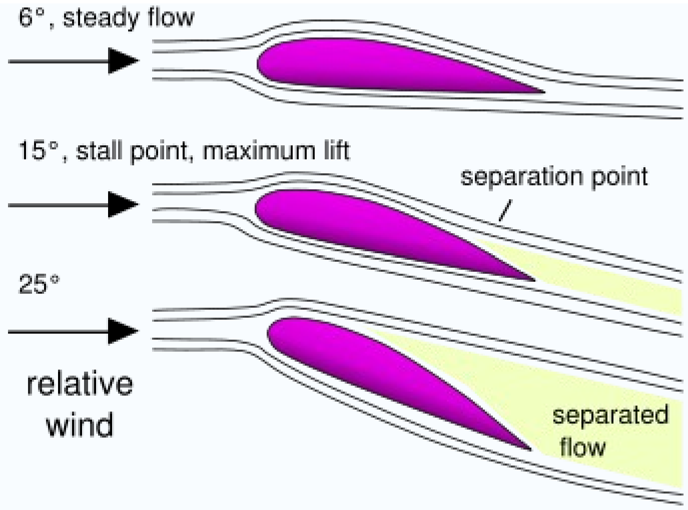
Stalls in aircraft usually do not occur without warning. Sensors in the aircraft alert the pilot when the aircraft is about to stall, and experienced pilots can often sense an approaching stall in the changing behavior of the aircraft. Since the conditions that produce stalls are very well understood, pilots can easily avoid stalls, and many pilots never experience stalls outside of their pilot training. Standard pilot training includes training in the proper ways to avoid, recognize, and recover from stalls.

Stalls can be alarming for non-pilots, because the aircraft may drop very suddenly and pitch forward in a frightening way. However, recovery is simple, and stalls are not a cause for concern unless they occur in close proximity to the ground. Commercial airliners never experience stalls in normal flight, and commercial pilots are especially careful to avoid stalls in order to avoid making passengers uncomfortable.

A few types of aircraft with a T-shaped tail or rear-mounted engines can enter a *deep stall* or *superstall*. This is a type of stall that produces turbulence behind the wings that can interfere with the operation of engines or the tail of the aircraft. Recovery from a deep stall can be impossible, resulting in a crash. Some aircraft with such characteristics are fitted with special control devices to prevent the aircraft from ever approaching a position that can cause a deep stall. An example of such a device is a *stick pusher*, which forces the nose of the aircraft down whenever it approaches a stall, regardless of any actions taken by the pilot.

The remainder of this article describes stalls in more technical terms.

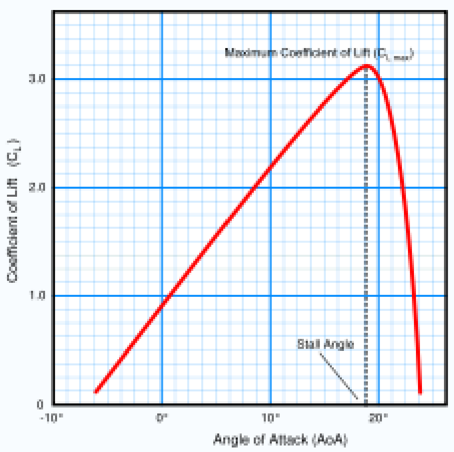
Formal definition



A **stall** is a condition in aerodynamics and aviation where as the angle between the wing's chord line and the relative wind (the angle of attack) increases beyond a certain point the lift, rather than increasing, reduces. The angle at which this occurs is called the *critical angle of attack*. This angle is typically 12 to 15 degrees for many subsonic airfoils. The critical angle of attack is the angle of attack on the lift coefficient versus angle-of-attack curve at which the maximum lift coefficient occurs, and it usually represents the boundary between the wing's linear and nonlinear airflow regimes. Flow separation begins to occur at this point, decreasing lift, increasing drag, and changing the wing's centre of lift. A fixed-wing aircraft during a stall may experience buffeting, a change in attitude (nose up or nose down). Most aircraft are designed to have a gradual stall with characteristics that will warn the pilot and give the pilot time to react. For example an aircraft that does not buffet before the stall may have a stick shaker installed to simulate the feel of a buffet by vibrating the stick fore and aft. The critical angle of attack in steady straight and level flight can only be attained at low airspeed. Attempts to increase the angle of attack at higher airspeeds can cause a high speed stall or may merely cause the aircraft to climb.

Because air no longer flows smoothly over the wings during a stall aileron control of roll becomes less effective, whereas the tendency for the ailerons to generate adverse yaw increases. Any yaw will increase the lift from the advancing wing and may cause the aircraft to increase rather than reduce the roll.

Depending on the aircraft's design, a stall can expose extremely adverse properties of balance and control. The ease with which a particular craft will recover from a stall depends on the dynamics of the aircraft itself and the skill of the pilot. If the stall persists a high rate of descent will occur and a spin may also develop.



Typical behavior of most airfoils.

Graph

The graph shows that the greatest amount of lift is produced as the critical angle of attack is reached (which in early 20th century aviation was called the "burble point"). This angle is 17.5 degrees in this case but changes from airfoil to airfoil. The graph shows that as the angle of attack is exceeded beyond the critical angle, the lift produced by the wing decreases significantly. The airfoil is now stalled.

Note that this graph shows the stall angle, yet in practice most pilots discuss stalling in terms of airspeed. This is because in general terms one can relate the angle of attack to airspeed - a lower speed requires a greater angle of attack to produce the necessary lift and vice versa. Thus as speed falls, AoA can increase, until the critical angle is reached. The airspeed at which this angle is reached is the (1g, unaccelerated) stalling speed of the aircraft in that particular configuration. Deploying flaps/slats decreases the stall speed to allow the aircraft to land at a lower speed.

The stall speed will be higher if the aircraft is experiencing more than one-g of longitudinal acceleration. The stall speeds found in many aircraft manuals only apply to unaccelerated flight.

Aerodynamic description of a stall

Stalling an aeroplane

If attempting the stall for flight training purposes, be sure to carry out correct checks before hand such as the HASELL check. This ensures that the engine is in the right condition and the area around the aircraft is safe and acceptable.

An aeroplane can be made to stall in any pitch attitude or bank angle or at any airspeed but is commonly practised by reducing the speed to the unaccelerated stall speed, at a safe altitude. Unaccelerated (1g) stall speed varies on different aeroplanes and is represented by colour codes on the air speed indicator. As the plane flies at this speed the angle of attack must be increased to prevent any loss of altitude or gain in airspeed (which corresponds to the stall angle described above). The pilot will notice the flight controls have become less responsive and may also notice some buffeting, an aerodynamic vibration caused by the airflow starting to detach from the wing surface.

In most light aircraft, as the stall is reached the aircraft will start to descend (because the wing is no longer producing enough lift to support the aeroplane's weight) and the nose will pitch down. Recovery from this stalled state usually involves the pilot decreasing the angle of attack and increasing the air speed, until smooth air flow over the wing is resumed. Normal flight can be resumed once recovery from the stall is complete. The manoeuvre is normally quite safe and if correctly handled leads to only a small loss in altitude. It is taught and practised in order to help pilots recognize, avoid, and recover from stalling the aeroplane.

The most common stall-spin scenarios occur on takeoff (departure stall) and during landing (base to final turn) because of insufficient airspeed during these manoeuvres. Stalls also occur during a go-around manoeuvre if the pilot does not properly respond to the out-of-trim situation resulting from the transition from low power setting to high power setting at low speed. Stall speed is increased when the upper wing surfaces are contaminated with ice or frost creating a rougher surface.

A special form of asymmetric stall in which the aircraft also rotates about its yaw axis is called a spin. A spin will occur if an aircraft is stalled and there is an asymmetric yawing moment applied to it. This yawing moment can be aerodynamic (sideslip angle, rudder, adverse yaw from the ailerons), thrust related (p-factor, one engine inoperative on a multi-engine non-centreline thrust aircraft), or from any number of possible sources of yaw.

Since most aircraft have an engine, some confusion exists between an aerodynamic versus engine **stall**. Many people seem to believe that an aircraft will drop out of the sky as soon as the engine stops in flight. In reality, the pilot can simply lower its nose to generate enough airspeed to maintain lift over the wings and so prevent a stall. The aircraft will then descend at a steady airspeed. The pilot then has time to find a suitable landing area or to restart the engine.

Put differently, all powered aircraft (even the biggest ones) become gliders when they lose all thrust. There have been cases of airliners running out of fuel at high altitude that landed successfully at airports a hundred kilometres away. However the distance which an aircraft can glide is directly related to the airspeed, but most of all the density altitude which the aircraft is at. The Gimli Glider is a celebrated example.

Stalls can occur at higher speeds if the wings already have a high angle of attack. Attempting to increase the angle of attack at 1g by moving the control column back simply causes the aircraft to rise. However the aircraft may experience higher g, for example when it is pulling out of a dive. In this case, the wings will already be generating more lift to provide the necessary upwards acceleration and so there will be higher angle of attack. Increasing the g still further, by pulling back on the control column, can cause the stalling angle to be exceeded even at a high speed. High speed stalls produce the same buffeting characteristics as 1g stalls and can also initiate a spin if there is also any yawing.

Symptoms of an approaching stall

One symptom of an approaching stall is slow and sloppy controls. As the speed of the aeroplane decreases approaching the stall, there is less air moving over the wing and therefore less will be deflected by the control surfaces (ailerons, rudder and elevator) at this slower speed. Some buffeting may also be felt from the turbulent flow above the wings as the stall is reached. However during a turn this buffeting will not be felt and immediate action must be taken to recover from the stall. The stall warning will sound, if fitted, in most aircraft 5 to 10 knots above the stall speed.

Stalling characteristics

Different aircraft types have different stalling characteristics. A benign stall is one where the nose drops gently and the wings remain level throughout. Slightly more demanding is a stall where one wing stalls slightly before the other, causing that wing to drop sharply, with the possibility of entering a spin. A dangerous stall is one where the nose rises, pushing the wing deeper into the stalled state and potentially leading to an unrecoverable *deep stall*. This can occur in some T-tailed aircraft where the turbulent airflow from the stalled wing can blanket the control surfaces at the tail.

"Stall speed"

Stalls depend more on angle of attack rather than airspeed. However, since, for every weight of every aircraft, there is an airspeed at which the wing's angle of attack will exceed the critical angle of attack, airspeed in a given configuration is often used as an indirect indicator of approaching stall conditions.

There are multiple V speeds which are used to indicate when a stall will occur:

- V_S: the stalling speed or the minimum steady flight speed at which the airplane is controllable. Usually synonymous with V_{S1}.
- V_{S0}: the stalling speed or the minimum steady flight speed in the landing configuration.
- V_{S1}: the stalling speed or the minimum steady flight speed obtained in a specific configuration (usually a "clean" configuration of flaps, landing gear and other sources of drag).
- V_{SR}: reference stall speed.
- V_{SR0}: reference stall speed in the landing configuration.
- V_{SR1}: reference stall speed in a specific configuration.
- V_{SW}: speed at which onset of natural or artificial stall warning occurs.

On an airspeed indicator, V_{S0} is indicated by the bottom of the white arc, while V_S is indicated by the bottom of the green arc.

Deep stall

A *deep stall* (also called a *superstall*) is a dangerous type of stall that affects certain aircraft designs, notably those with a T-tail configuration. In these designs, the turbulent wake of a stalled main wing "blanks" the horizontal stabilizer, rendering the elevators ineffective and preventing the aircraft from recovering from the stall.

Although effects similar to deep stall had long been known to occur on many aircraft designs, the name first came into widespread use after a deep stall caused the prototype BAC 1-11 to crash, killing its crew. This led to changes to the aircraft, including the installation of a stick shaker (see below) in order to clearly warn the pilot of the problem before it occurred. Stick shakers are now a part of all commercial airliners. Nevertheless, the problem continues to periodically haunt new designs; in the 1980s a prototype of the latest model of the Canadair Challenger business jet entered deep stall during testing, killing one of the test pilots who was unable to jump from the plane in time. Also, paragliders are sometimes known to enter a deep stall condition.

Deep stall is possible with some sailplanes, as their most common designs are T-tail configurations. The IS-29 glider is one of the gliders that are vulnerable to deep stalls when the CG and the overall weight are between certain limits.

In the early 1980s, a Schweizer SGS 1-36 sailplane was modified for NASA's controlled deep-stall flight program.[1]

Stall warning and safety devices

Aeroplanes can be equipped with a variety of devices to prevent or postpone a stall or to make it less (or in some cases more) severe, or to make recovery easier.

- An **aerodynamic twist** can be introduced to the wing with the leading edge near the wing tip twisted downward. This is called **washout** and causes the wing root to stall before the wing tip. This makes the stall gentle and progressive. Since the stall is delayed at the wing tips, where the ailerons are, roll control is maintained when the stall begins.
- A **stall strip** is a small sharp-edged device which, when attached to the leading edge of a wing, encourages the stall to start there in preference to any other location on the wing. If attached close to the wing root it makes the stall gentle and progressive; if attached near the wing tip it encourages the aircraft to drop a wing when stalling.

- **Vortex generators**, tiny strips of metal or plastic placed on top of the wing near the leading edge that protrude past the boundary layer into the free stream. As the name implies they energize the boundary layer by mixing free stream airflow with boundary layer flow thereby creating vortices, this increases the inertia of the boundary layer. By increasing the inertia of the boundary layer airflow separation and the resulting stall may be delayed.
- An **anti-stall strake** is a wing extension at the root leading edge which generates a vortex on the wing upper surface to postpone the stall.
- A **stick-pusher** is a mechanical device which prevents the pilot from stalling an aeroplane by pushing the controls forwards as the stall is approached.
- A **stick shaker** is a similar device which shakes the pilot's controls to warn of the onset of stall.
- A **stall warning** is an electronic or mechanical device which sounds an audible warning as the stall speed is approached. The majority of aircraft contain some form of this device that warns the pilot of an impending stall. The simplest such device is a 'stall warning horn', which consists of either a pressure sensor or a movable metal tab that actuates a switch, and produces an audible warning in response.
- An **angle of attack limiter** or an "alpha" limiter is a flight computer that automatically prevents pilot input from causing the plane to rise over the stall angle. Some alpha limiters can be disabled by the pilot.

If a forward canard is used for pitch control rather than an aft tail, the canard is designed to stall at a slightly greater angle of attack than the wing (i.e. the canard stalls first). When the canard stalls, the nose drops, lowering the angle of attack thus preventing the wing from stalling. Thus the wing virtually never stalls.

If an aft tail is used, the wing is designed to stall before the tail. In this case, the wing can be flown at higher lift coefficient (closer to stall) to produce more overall lift.

Many aircraft have an angle of attack indicator among the pilot's instruments which lets the pilot know precisely how close to the stall point the aircraft is.

Spoilers

In most circumstances, a stall is an undesirable event. Spoilers, however, are devices that are intentionally deployed to create a carefully controlled stall over part of an aircraft's wing, in order to reduce the lift it generates, and allow it to descend without gaining speed. Spoilers are also deployed asymmetrically (i.e. on one wing only) to enhance roll control.

Bernoulli's principle

Bernoulli's Principle states that in an ideal fluid (low speed air is a good approximation), with no work being performed on the fluid, an increase in velocity occurs simultaneously with decrease in pressure or a change in the fluid's gravitational potential energy.

This principle is a simplification of Bernoulli's equation, which states that the sum of all forms of energy in a fluid flowing along an enclosed path (a streamline) is the same at any two points in that path. It is named after the Dutch/Swiss mathematician/scientist Daniel Bernoulli, though it was previously understood by Leonhard Euler and others. In fluid flow with no viscosity, and, therefore, one in which a pressure difference is the only accelerating force, it is equivalent to Newton's laws of motion. It is important to note that the only cause of the change in fluid velocity is the difference in balanced pressure on either side of it. It may be misunderstood to be that a change in velocity simply causes a change in pressure; the Bernoulli principle does not make any such statement.

Incompressible flow

The original form, for incompressible flow in a uniform gravitational field, is:

$$\frac{v^2}{2} + gh + \frac{p}{\rho} = \text{constant}$$

v = fluid velocity along the streamline
g = acceleration due to gravity
h = height of the fluid
p = pressure along the streamline
ρ = density of the fluid

These assumptions must be met for the equation to apply:

- Inviscid flow − viscosity (internal friction) = 0
- Steady flow
- Incompressible flow − ρ = constant along a streamline. Density may vary from streamline to streamline, however.
- Generally, the equation applies along a streamline. For constant-density potential flow, it applies throughout the entire flow field.

The decrease in pressure, and the corresponding increase in velocity as predicted by the equation, is often called Bernoulli's principle. The equation is named for Daniel Bernoulli although it was first presented in the above form by Leonhard Euler.

Compressible flow

A second, more general form of Bernoulli's equation may be written for compressible fluids, in which case, following a streamline:

$$\frac{v^2}{2} + \phi + w = \text{constant}$$

ϕ= gravitational potential energy per unit mass, $\phi = gh$ in the case of a uniform gravitational field
w= fluid enthalpy per unit mass, which is also often written as h (which conflicts with the use of h in this article for "height"). Note that $w = \epsilon + \frac{p}{\rho}$ where ϵ is the fluid thermodynamic energy per unit mass, also known as the specific internal energy or "sie".

The constant on the right hand side is often called the Bernoulli constant and denoted *b*. For steady inviscid adiabatic flow with no additional sources or sinks of energy, *b* is constant along any given streamline. More generally, when *b* may vary along streamlines, it still proves a useful parameter, related to the "head" of the fluid (see below).

When shock waves are present, in a reference frame moving with a shock, many of the parameters in the Bernoulli equation suffer abrupt changes in passing through the shock. The

Bernoulli parameter itself, however, remains unaffected. An exception to this rule is radiative shocks, which violate the assumptions leading to the Bernoulli equation, namely the lack of additional sinks or sources of energy.

Derivation of Bernoulli equation for incompressible fluids

Incompressible fluids

The Bernoulli equation for incompressible fluids can be derived by integrating the Euler equations, or applying the law of conservation of energy in two sections along a streamline, ignoring viscosity, compressibility, and thermal effects.

The simplest derivation is to first ignore gravity and consider constrictions and expansions in pipes that are otherwise straight, as seen in Venturi effect. Let the x axis be directed down the axis of the pipe.

The equation of motion for a parcel of fluid on the axis of the pipe is

$$\rho\frac{dv}{dt} = -\frac{dp}{dx}$$

In steady flow, $v = v(x)$ so

$$\frac{dv}{dt} = \frac{dv}{dx}\frac{dx}{dt} = \frac{dv}{dx}v = \frac{d}{dx}\frac{v^2}{2}$$

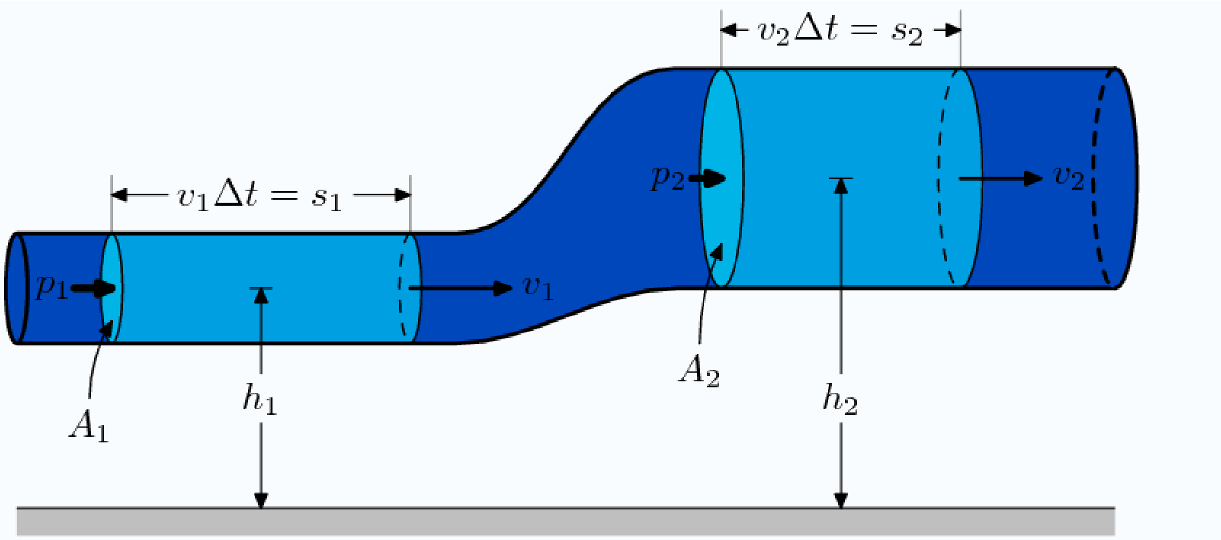
With ρ constant, the equation of motion can be written as

$$\frac{d}{dx}\left(\rho\frac{v^2}{2} + p\right) = 0$$

or

$$\frac{v^2}{2} + \frac{p}{\rho} = C$$

where C is a constant, sometimes referred to as the Bernoulli constant. We deduce that where the speed is large, pressure is low. In the above derivation, no external work-energy principle is invoked. Rather, the work-energy principle inherently derived by a simple manipulation of the momentum equation. The derivation that follows includes gravity and applies to a curved trajectory, but a work-energy principle must be assumed.



A streamtube of fluid moving to the right. Indicated are pressure, height, velocity, distance (s), and cross-sectional area.

Applying conservation of energy in form of the work-kinetic energy theorem we find that:

> the change in KE of the system equals the net work done on the system;
> $W = \Delta KE.$

Therefore,

> the work done by the forces in the fluid + decrease in potential energy = increase in kinetic energy.

The work done by the forces is

$$F_1 s_1 - F_2 s_2 = p_1 A_1 v_1 \Delta t - p_2 A_2 v_2 \Delta t.$$

The decrease of potential energy is

$$mgh_1 - mgh_2 = \rho g A_1 v_1 \Delta t h_1 - \rho g A_2 v_2 \Delta t h_2$$

The increase in kinetic energy is

$$\frac{1}{2} m v_2^2 - \frac{1}{2} m v_1^2 = \frac{1}{2}\rho A_2 v_2 \Delta t v_2^2 - \frac{1}{2}\rho A_1 v_1 \Delta t v_1^2.$$

Putting these together,

$$p_1 A_1 v_1 \Delta t - p_2 A_2 v_2 \Delta t + \rho g A_1 v_1 \Delta t h_1 - \rho g A_2 v_2 \Delta t h_2 = \frac{1}{2}\rho A_2 v_2 \Delta t v_2^2 - \frac{1}{2}\rho A_1 v_1 \Delta t v_1^2$$

or

$$\frac{\rho A_1 v_1 \Delta t v_1^2}{2} + \rho g A_1 v_1 \Delta t h_1 + p_1 A_1 v_1 \Delta t = \frac{\rho A_2 v_2 \Delta t v_2^2}{2} + \rho g A_2 v_2 \Delta t h_2 + p_2 A_2 v_2 \Delta t.$$

After dividing by Δt, ρ and $A_1 v_1$ (= rate of fluid flow = $A_2 v_2$ as the fluid is incompressible):

$$\frac{v_1^2}{2} + gh_1 + \frac{p_1}{\rho} = \frac{v_2^2}{2} + gh_2 + \frac{p_2}{\rho}$$

or, as stated in the first paragraph:

$$\frac{v^2}{2} + gh + \frac{p}{\rho} = C$$

Further division by g implies

$$\frac{v^2}{2g} + h + \frac{p}{\rho g} = C$$

A free falling mass from a height h (in vacuum), will reach a velocity

$$v = \sqrt{2gh},\text{ or } h = \frac{v^2}{2g}.$$

The term $\frac{v^2}{2g}$ is called the *velocity head*.

The hydrostatic pressure or *static head* is defined as

$$p = \rho gh,\text{ or } h = \frac{p}{\rho g}.$$

The term $\frac{p}{\rho g}$ is also called the *pressure head*.

A way to see how this relates to conservation of energy directly is to multiply by density and by unit volume (which is allowed since both are constant) yielding:

$$v^2\rho + P = constant \text{ and}$$
$$mV^2 + P \cdot volume = constant$$

Compressible fluids

The derivation for compressible fluids is similar. Again, the derivation depends upon (1) conservation of mass, and (2) conservation of energy. Conservation of mass implies that in the above figure, in the interval of time Δt, the amount of mass passing through the boundary defined by the area A_1 is equal to the amount of mass passing outwards through the boundary defined by the area A_2:

$$0 = \Delta M_1 - \Delta M_2 = \rho_1 A_1 v_1 \, \Delta t - \rho_2 A_2 v_2 \, \Delta t.$$

Conservation of energy is applied in a similar manner: It is assumed that the change in energy of the volume of the streamtube bounded by A_1 and A_2 is due entirely to energy entering or leaving through one or the other of these two boundaries. Clearly, in a more complicated situation such as a fluid flow coupled with radiation, such conditions are not met. Nevertheless, assuming this to be the case and assuming the flow is steady so that the net change in the energy is zero,

$$0 = \Delta E_1 - \Delta E_2$$

where ΔE_1 and ΔE_2 are the energy entering through A_1 and leaving through A_2, respectively.

The energy entering through A_1 is the sum of the kinetic energy entering, the energy entering in the form of potential gravitational energy of the fluid, the fluid thermodynamic energy entering, and the energy entering in the form of mechanical $p\,dV$ work:

$$\Delta E_1 = \left[\frac{1}{2}\rho_1 v_1^2 + \phi_1\rho_1 + \epsilon_1\rho_1 + p_1\right] A_1 v_1\,\Delta t$$

A similar expression for ΔE_2 may easily be constructed. So now setting $0 = \Delta E_1 - \Delta E_2$:

$$0 = \left[\frac{1}{2}\rho_1 v_1^2 + \phi_1\rho_1 + \epsilon_1\rho_1 + p_1\right] A_1 v_1\,\Delta t - \left[\frac{1}{2}\rho_2 v_2^2 + \phi_2\rho_2 + \epsilon_2\rho_2 + p_2\right] A_2 v_2\,\Delta t$$

which can be rewritten as:

$$0 = \left[\frac{1}{2}v_1^2 + \phi_1 + \epsilon_1 + \frac{p_1}{\rho_1}\right] \rho_1 A_1 v_1\,\Delta t - \left[\frac{1}{2}v_2^2 + \phi_2 + \epsilon_2 + \frac{p_2}{\rho_2}\right] \rho_2 A_2 v_2\,\Delta t$$

Now, using the previously-obtained result from conservation of mass, this may be simplified to obtain

$$\frac{1}{2}v^2 + \phi + \epsilon + \frac{p}{\rho} = \text{constant} \equiv b$$

which is the Bernoulli equation for compressible flow.

Newtonian fluid

A **Newtonian fluid** (named for Isaac Newton) is a fluid that flows like water—its stress / rate of strain curve is linear and passes through the origin. The constant of proportionality is known as the viscosity.

A simple equation to describe Newtonian fluid behaviour is

$$\tau = \mu\frac{du}{dx}$$

where

τ is the shear stress exerted by the fluid ("drag") [Pa]
μ is the fluid viscosity - a constant of proportionality [Pa·s]

$\frac{du}{dx}$ is the velocity gradient perpendicular to the direction of shear [s^{-1}]

In common terms, this means the fluid continues to flow, regardless of the forces acting on it. For example, water is Newtonian, because it continues to exemplify fluid properties no matter how fast it is stirred or mixed. Contrast this with a non-Newtonian fluid, in which stirring can leave a "hole" behind (that gradually fills up over time - this behaviour is seen in materials such as pudding, oobleck, or, to a less rigorous extent, sand), or cause the fluid to become thinner, the drop in viscosity causing it to flow more (this is seen in non-drip paints, which brush on easily but become more viscous when on walls).

For a Newtonian fluid, the viscosity, by definition, depends only on temperature and pressure (and also the chemical composition of the fluid if the fluid is not a pure substance), not on the forces acting upon it.

If the fluid is incompressible and viscosity is constant across the fluid, the equation governing the shear stress, in the Cartesian coordinate system, is

$$\tau_{ij} = \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$$

with comoving stress tensor $\mathbb{P}$(also written as σ)

$$\mathbb{P}_{ij} = -p\delta_{ij} + \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$$

where, by the convention of tensor notation,

τ_{ij} is the shear stress on the i^{th} face of a fluid element in the j^{th} direction
u_i is the velocity in the i^{th} direction
x_j is the j^{th} direction coordinate

If a fluid does not obey this relation, it is termed a non-Newtonian fluid, of which there are several types

Non-Newtonian fluid

A **non-Newtonian fluid** is a fluid in which the viscosity changes with the applied strain rate. As a result, non-Newtonian fluids may not have a well-defined viscosity.

Although the concept of viscosity is commonly used to characterize a material, it can be inadequate to describe the mechanical behavior of a substance, particularly non-Newtonian fluids. They are best studied through several other rheological properties which relate the relations between the stress and strain tensors under many different flow conditions, such as oscillatory shear, or extensional flow which are measured using different devices or rheometers. The rheological properties are better studied using tensor-valued constitutive equations, which are common in the field of continuum mechanics.

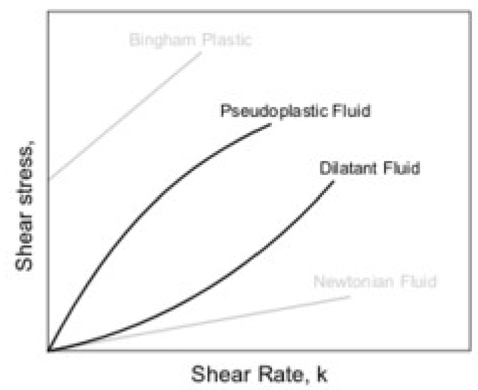
Common examples

An inexpensive, non-toxic sample of a non-Newtonian fluid sometimes known as oobleck can be made very easily by adding corn starch (cornflour) to a cup of water. Add the starch in small portions and stir it in slowly. When the suspension nears the critical concentration - becoming like single cream (light cream) in consistency - the so called "shear thickening" property of this non-Newtonian fluid becomes apparent. The application of force - for example by stabbing the surface with a finger, or rapidly inverting the container holding it - leads to the fluid behaving like a solid rather than a liquid. More gentle treatment, such as slowly inserting a spoon, will leave it in its liquid state. Trying to jerk the spoon back out again, however, will trigger the return of the temporary solid state. A person moving quickly and/or applying sufficient force with his feet can literally walk across such a liquid.

Shear thickening fluids of this sort are being researched for bullet resistant body armor, useful for their ability to absorb the energy of a high velocity projectile impact but remain soft and flexible while worn.

A familiar example of the opposite, a shear-thinning fluid, is paint: one wants the paint to flow readily off the brush when it is being applied to the surface being painted, but not to drip excessively.

Classification types



Principal types of non-Newtonian fluid include:

Type of fluid	Behaviour	Characteristics	Examples

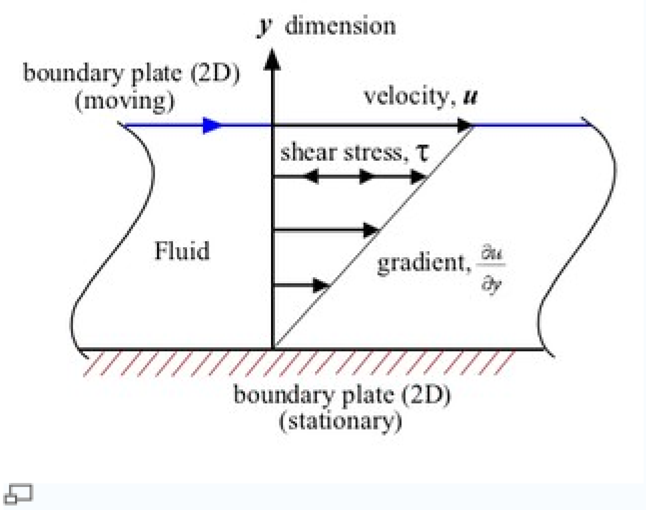
Plastic solids	Perfectly plastic	Strain does not result in opposing stress	Ductile metals past the yield point
	Bingham plastic	Linear relationship between shear stress and rate of strain once threshold shear stress exceeded	Mud, some colloids
	Yield pseudo-plastic	Pseudo-plastic above some threshold shear stress	
	Yield dilatant	Dilatant above some threshold shear stress	
Power-law fluids	Pseudoplastic or "shear thinning"	Apparent viscosity reducing with rate of shear	Some colloids, clay, milk, gelatin, blood and liquid cement
	Dilatant or "shear thickening"	Apparent viscosity increasing with rate of shear	Concentrated solution of sugar in water, suspensions of rice starch or corn starch
Viscoelastic - having both viscous and elastic properties	Maxwell material	"Series" linear combination of elastic and viscous effects	metals, composite materials
	Oldroyd-B fluid	Linear combination of Maxwell and Newtonian behaviour	Bitumen, dough, nylon, and Silly Putty
	Kelvin material	"Parallel" linear combination of elastic and viscous effects	
	Anelastic	Material returns to a well-defined "rest shape"	

Time-dependent viscosity	Rheopectic	Apparent viscosity increases with duration of stress	Some lubricants
	Thixotropic	Apparent viscosity decreases with duration of stress	Non-drip paints and tomato ketchup and most honey varieties.
Generalized Newtonian fluids		Stress depends on normal and shear strain rates and also the pressure applied on it	Blood, Custard

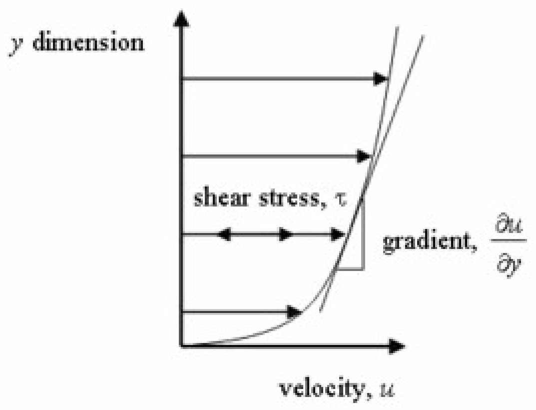
Viscosity

Viscosity is a measure of the resistance of a fluid to deform under shear stress. It is commonly perceived as "thickness", or resistance to flow. Viscosity describes a fluid's internal resistance to flow and may be thought of as a measure of fluid friction. Thus, water is "thin", having a lower viscosity, while vegetable oil is "thick" having a higher viscosity. All real fluids (except superfluids) have some resistance to shear stress, but a fluid which has no resistance to shear stress is known as an **ideal fluid** or **inviscid fluid** (Symon 1971).

Newton's theory



Laminar shear of fluid between two plates. Friction between the fluid and the moving boundaries causes the fluid to shear. The force required for this action is a measure of the fluid's viscosity. This type of flow is known as a Couette flow.



Laminar shear, the non-linear gradient, is a result of the geometry the fluid is flowing through (e.g. a pipe).

In general, in any flow, layers move at different velocities and the fluid's viscosity arises from the shear stress between the layers that ultimately opposes any applied force.

Isaac Newton postulated that, for straight, parallel and uniform flow, the shear stress, τ, between layers is proportional to the velocity gradient, ∂*u*/∂*y*, in the direction perpendicular to the layers, in other words, the relative motion of the layers.

$$\tau = \eta \frac{\partial u}{\partial y}.$$

Here, the constant η is known as the *coefficient of viscosity*, the *viscosity*, or the *dynamic viscosity*. Many fluids, such as water and most gases, satisfy Newton's criterion and are known as Newtonian fluids. Non-Newtonian fluids exhibit a more complicated relationship between shear stress and velocity gradient than simple linearity.

The relationship between the shear stress and the velocity gradient can also be obtained by considering two plates closely spaced apart at a distance *y*, and separated by a homogeneous substance. Assuming that the plates are very large, with a large area *A*, such that edge effects may be ignored, and that the lower plate is fixed, let a force *F* be applied to the upper plate. If this force causes the substance between the plates to undergo shear flow (as opposed to just shearing elastically until the shear stress in the substance balances the applied force), the substance is called a fluid. The applied force is proportional to the area and velocity of the plate and inversely proportional to the distance between the plates. Combining these three relations results in the equation $F = \eta(Au/y)$, where η is the proportionality factor called the *absolute viscosity* (with units Pa·s = kg/(m·s) or slugs/(ft·s)). The absolute viscosity is also known as the *dynamic viscosity*, and is often shortened to simply *viscosity*. The equation can be expressed in terms of shear stress; $\tau = F/A = \eta(u/y)$. The rate of shear deformation is u / y and can be also written as a shear velocity, *du/dy*. Hence, through this method, the relation between the shear stress and the velocity gradient can be obtained.

In many situations, we are concerned with the ratio of the viscous force to the inertial force, the latter characterised by the fluid density ρ. This ratio is characterised by the *kinematic viscosity*, defined as follows:

$$\nu = \frac{\eta}{\rho}.$$

James Clerk Maxwell called viscosity *fugitive elasticity* because of the analogy that elastic deformation opposes shear stress in solids, while in viscous fluids, shear stress is opposed by *rate* of deformation.

Measuring viscosity

Viscosity is measured with various types of viscometer, typically at 20 °C (standard state). For some fluids, it is a constant over a wide range of shear rates. The fluids without a constant viscosity are called Non-Newtonian fluids.

In paint industries, viscosity is commonly measured with a Zahn cup, in which the efflux time is determined and given to customers. The efflux time can also be converted to kinematic viscosities (cSt) through the conversion equations.

Also used in paint, a Stormer viscometer uses load-based rotation in order to determine viscosity. It uses units, Krebs units (KU), unique to this viscometer.

Units

Viscosity (dynamic/absolute viscosity): η or μ

The IUPAC symbol for viscosity is the Greek symbol eta (η), and dynamic viscosity is also commonly referred to using the Greek symbol mu (μ). The SI physical unit of dynamic viscosity is the pascal-second (Pa·s), which is identical to 1 $kg \cdot m^{-1} \cdot s^{-1}$. If a fluid with a viscosity of one Pa·s is placed between two plates, and one plate is pushed sideways with a shear stress of one pascal, it moves a distance equal to the thickness of the layer between the plates in one second. The name *poiseuille* (Pl) was proposed for this unit (after Jean Louis Marie Poiseuille who formulated Poiseuille's law of viscous flow), but not accepted internationally. Care must be taken in not confusing the poiseuille with the *poise* named after the same person!

The cgs physical unit for dynamic viscosity is the *poise*[1] (P; IPA: [pwaz])) named after Jean Louis Marie Poiseuille. It is more commonly expressed, particularly in ASTM standards, as *centipoise* (cP). The centipoise is commonly used because water has a viscosity of 1.0020 cP (at 20 °C; the closeness to one is a convenient coincidence).

$$1\ P = 1\ g \cdot cm^{-1} \cdot s^{-1}$$

The relation between Poise and Pascal-second is:

$$10\ P = 1\ kg \cdot m^{-1} \cdot s^{-1} = 1\ Pa \cdot s$$
$$1\ cP = 0.001\ Pa \cdot s = 1\ mPa \cdot s$$

Kinematic viscosity: ν

Kinematic viscosity (Greek symbol: ν) has SI units ($m^2 \cdot s^{-1}$). The cgs physical unit for kinematic viscosity is the *stokes* (abbreviated S or St), named after George Gabriel Stokes. It is sometimes expressed in terms of *centistokes* (cS or cSt). In U.S. usage, *stoke* is sometimes used as the singular form.

$$1\ stokes = 100\ centistokes = 1\ cm^2 \cdot s^{-1} = 0.0001\ m^2 \cdot s^{-1}.$$
$$1\ centistokes = 1\ mm^2/s$$

Dynamic versus kinematic viscosity

Conversion between kinematic and dynamic viscosity, is given by $\nu\rho = \eta$. Note that the parameters must be given in SI units not in P, cP or St.

For example, if $\nu = 1$ St (=0.0001 $m^2 \cdot s^{-1}$) and $\rho = 1000$ $kg \cdot m^{-3}$ then $\eta = \nu\rho = 0.1$ $kg \cdot m^{-1} \cdot s^{-1} = 0.1$ Pa·s [1].

For a plot of kinematic viscosity of air as a function of absolute temperature, see James Ierardi's Fire Protection Engineering Site

Molecular origins

The viscosity of a system is determined by how molecules constituting the system interact. There are no simple but correct expressions for the viscosity of a fluid. The simplest exact expressions are the Green-Kubo relations for the linear shear viscosity or the Transient Time Correlation Function expressions derived by Evans and Morriss in 1985. Although these expressions are each exact in order to calculate the viscosity of a dense fluid, using these relations requires the use of molecular dynamics computer simulation.

Gases

Viscosity in gases arises principally from the molecular diffusion that transports momentum between layers of flow. The kinetic theory of gases allows accurate prediction of the behaviour of gaseous viscosity, in particular that, within the regime where the theory is applicable:

- Viscosity is independent of pressure; and
- Viscosity increases as temperature increases.

Liquids

In liquids, the additional forces between molecules become important. This leads to an additional contribution to the shear stress though the exact mechanics of this are still controversial.[citation needed] Thus, in liquids:

- Viscosity is independent of pressure (except at very high pressure); and
- Viscosity tends to fall as temperature increases (for example, water viscosity goes from 1.79 cP to 0.28 cP in the temperature range from 0 °C to 100 °C); see temperature dependence of liquid viscosity for more details.

The dynamic viscosities of liquids are typically several orders of magnitude higher than dynamic viscosities of gases.

Viscosity of materials

The viscosity of air and water are by far the two most important materials for aviation aerodynamics and shipping fluid dynamics. Temperature plays the main role in determining viscosity.

Viscosity of air

The viscosity of air depends mostly on the temperature. At 15.0 °C, the viscosity of air is 1.78×10^{-5} kg/(m·s). You can get the viscosity of air as a function of altitude from the eXtreme High Altitude Calculator

Viscosity of water

The viscosity of water is 8.90 × 10^{-4} Pa·s or 8.90 × 10^{-3} dyn·s/cm²at about 25 °C. as a function of temperature: μ=A × $10^{B/(T-C)}$
Where A=2.414 × 10-5 N*s/m2 ; B = 247.8 Kelvin ; C = 140 Kelvin

Viscosity of various materials

Example of the viscosity of milk and water. Liquids with higher viscosities will not make such a splash.

The Sutherland's formula can be used to derive the dynamic viscosity as a function of the temperature:

$$\eta = \eta_0 \frac{T_0 + C}{T + C} \left(\frac{T}{T_0}\right)^{3/2}$$

where:

- η = viscosity in (Pa·s) at input temperature T
- η_0 = reference viscosity in (Pa·s) at reference temperature T_0
- T = input temperature in kelvin
- T_0 = reference temperature in kelvin
- C = Sutherland's constant

Valid for temperatures between 0 < T < 555 K with an error due to pressure less than 10% below 3.45 MPa

Sutherland's constant and reference temperature for some gases

Gas	***C*** **[K]**	**T_0** **[K]**	**η_0** **[10^{-6} Pa s]**
air	120	291.15	18.27

nitrogen	111	300.55	17.81
oxygen	127	292.25	20.18
carbon dioxide	240	293.15	14.8
carbon monoxide	118	288.15	17.2
hydrogen	72	293.85	8.76
ammonia	370	293.15	9.82
sulphur dioxide	416	293.65	12.54

Some dynamic viscosities of Newtonian fluids are listed below:

Gases (at 0 °C):

	viscosity [Pa·s]
hydrogen	8.4×10^{-6}
air	17.4×10^{-6}
xenon	21.2×10^{-6}

Liquids (at 25 °C):

	viscosity [Pa·s]
acetone	[a] 0.306×10^{-3}
methanol	[a] 0.544×10^{-3}
benzene	[a] 0.604×10^{-3}
ethanol	[a] 1.074×10^{-3}
mercury	[a] 1.526×10^{-3}
nitrobenzene	[a] 1.863×10^{-3}
propanol	[a] 1.945×10^{-3}
sulfuric acid	[a] 24.2×10^{-3}
olive oil	81×10^{-3}
glycerol	[a] 934×10^{-3}
castor oil	985×10^{-3}
HFO-380	2022×10^{-3}
pitch	2.3×10^{8}

[a] Data from CRC Handbook of Chemistry and Physics, 73rd edition, 1992-1993.

Fluids with variable compositions, such as honey, can have a wide range of viscosities.

A more complete table can be found here

Can solids have a viscosity?

If on the basis that all solids flow to a small extent in response to shear stress then yes, substances know as Amorphous solids, such as glass, may be considered to have viscosity. This has led some to the view that solids are simply liquids with a very high viscosity, typically greater than 10^{12} Pa•s. This position is often adopted by supporters of the widely held misconception that glass flow can be observed in old buildings. This distortion is more likely the result of glass making process rather than the viscosity of glass.

However, others argue that solids are, in general, elastic for small stresses while fluids are not. Even if solids flow at higher stresses, they are characterized by their low-stress behavior. Viscosity may be an appropriate characteristic for solids in a plastic regime. The situation becomes somewhat confused as the term *viscosity* is sometimes used for solid materials, for example Maxwell materials, to describe the relationship between stress and the rate of change of strain, rather than rate of shear.

These distinctions may be largely resolved by considering the constitutive equations of the material in question, which take into account both its viscous and elastic behaviors. Materials for which both their viscosity and their elasticity are important in a particular range of deformation and deformation rate are called *viscoelastic*. In geology, earth materials that exhibit viscous deformation at least three times greater than their elastic deformation are sometimes called rheids.

One example of solids flowing which has been observed since 1930 is the Pitch drop experiment.

Bulk viscosity

The trace of the stress tensor is often identified with the negative-one-third of the thermodynamic pressure,

$$T_a^a = -\frac{1}{3}p,$$

which only depends upon the equilibrium state potentials like temperature and density (equation of state). In general, the trace of the stress tensor is the sum of thermodynamic pressure contribution plus another contribution which is proportional to the divergence of the velocity field. This constant of proportionality is called the **bulk viscosity**.

Eddy viscosity

In the study of turbulence in fluids, a common practical strategy for calculation is to ignore the small-scale *vortices* (or *eddies*) in the motion and to calculate a large-scale motion with an *eddy viscosity* that characterizes the transport and dissipation of energy in the smaller-scale flow. Values of eddy viscosity used in modeling ocean circulation may be from 5×10^4 to 10^6 Pa·s depending upon the resolution of the numerical grid.

Fluidity

The reciprocal of viscosity is *fluidity*, usually symbolized by $\varphi = 1 / \eta$ or $F = 1 / \eta$, depending on the convention used, measured in *reciprocal poise* ($cm\cdot s\cdot g^{-1}$), sometimes called the *rhe*. *Fluidity* is seldom used in engineering practice.

The concept of fluidity can be used to determine the viscosity of an ideal solution. For two components *a* and *b*, the fluidity when *a* and *b* are mixed is

$$F \approx \chi_a F_a + \chi_b F_b$$

which is only slightly simpler than the equivalent equation in terms of viscosity:

$$\eta \approx \frac{1}{\chi_a/\eta_a + \chi_b/\eta_b}$$

where χ_a and χ_b is the mole fraction of component *a* and *b* respectively, and η_a and η_b are the components pure viscosities.

The linear viscous stress tensor

(See *Hooke's law* and *strain tensor* for an analogous development for linearly elastic materials.)

Viscous forces in a fluid are a function of the rate at which the fluid velocity is changing over distance. The velocity at any point $\mathbf{r}$ is specified by the velocity field $\mathbf{v}(\mathbf{r})$. The velocity at a small distance $d\mathbf{r}$ from point $\mathbf{r}$ may be written as a Taylor series:

$$\mathbf{v}(\mathbf{r} + d\mathbf{r}) = \mathbf{v}(\mathbf{r}) + \frac{d\mathbf{v}}{d\mathbf{r}} d\mathbf{r} + \ldots$$

where $\frac{d\mathbf{v}}{d\mathbf{r}}$ is shorthand for the dyadic product of the del operator and the velocity:

$$\frac{d\mathbf{v}}{d\mathbf{r}} = \begin{bmatrix} \frac{\partial v_x}{\partial x} & \frac{\partial v_x}{\partial y} & \frac{\partial v_x}{\partial z} \\ \frac{\partial v_y}{\partial x} & \frac{\partial v_y}{\partial y} & \frac{\partial v_y}{\partial z} \\ \frac{\partial v_z}{\partial x} & \frac{\partial v_z}{\partial y} & \frac{\partial v_z}{\partial z} \end{bmatrix}$$

This is just the Jacobian of the velocity field. Viscous forces are the result of relative motion between elements of the fluid, and so are expressible as a function of the velocity field. In other words, the forces at $\mathbf{r}$ are a function of $\mathbf{v}(\mathbf{r})$ and all derivatives of $\mathbf{v}(\mathbf{r})$ at that point. In the case of linear viscosity, the viscous force will be a function of the Jacobian tensor alone. For almost all practical situations, the linear approximation is sufficient.

If we represent *x*, *y*, and *z* by indices 1, 2, and 3 respectively, the *i,j* component of the Jacobian may be written as $\partial_i v_j$ where ∂_i is shorthand for $\partial/\partial x_i$. Note that when the first and higher derivative terms are zero, the velocity of all fluid elements is parallel, and there are no viscous forces.

Any matrix may be written as the sum of an antisymmetric matrix and a symmetric matrix, and this decomposition is independent of coordinate system, and so has physical significance. The velocity field may be approximated as:

$$v_i(\mathbf{r} + d\mathbf{r}) = v_i(\mathbf{r}) + \frac{1}{2}\left(\partial_i v_j - \partial_j v_i\right) dr_i + \frac{1}{2}\left(\partial_i v_j + \partial_j v_i\right) dr_i$$

where Einstein notation is now being used in which repeated indices in a product are implicitly summed. The second term on the left is the asymmetric part of the first derivative term, and it represents a rigid rotation of the fluid about $\mathbf{r}$ with angular velocity ω where:

$$\omega = \nabla \times \mathbf{v} = \frac{1}{2}\begin{bmatrix} \partial_2 v_3 - \partial_3 v_2 \\ \partial_3 v_1 - \partial_1 v_3 \\ \partial_1 v_2 - \partial_2 v_1 \end{bmatrix}$$

For such a rigid rotation, there is no change in the relative positions of the fluid elements, and so there is no viscous force associated with this term. The remaining symmetric term is responsible for the viscous forces in the fluid. Assuming the fluid is isotropic (i.e. its properties are the same in all directions), then the most general way that the symmetric term (the rate-of-strain tensor) can be broken down in a coordinate-independent (and therefore physically real) way is as the sum of a constant tensor (the rate-of-expansion tensor) and a traceless symmetric tensor (the rate-of-shear tensor):

$$\frac{1}{2}\left(\partial_i v_j + \partial_j v_i\right) = \frac{1}{3}\partial_k v_k \delta_{ij} + \left(\frac{1}{2}\left(\partial_i v_j + \partial_j v_i\right) - \frac{1}{3}\partial_k v_k \delta_{ij}\right)$$

where δ_{ij} is the unit tensor. The most general linear relationship between the stress tensor σ and the rate-of-strain tensor is then a linear combination of these two tensors (Landau & Lifshitz 1997):

$$\sigma_{visc;ij} = \zeta \partial_k v_k \delta_{ij} + \eta\left(\partial_i v_j + \partial_j v_i - \frac{2}{3}\partial_k v_k \delta_{ij}\right)$$

where ζ is the coefficient of bulk viscosity (or "second viscosity") and η is the coefficient of (shear) viscosity.

The forces in the fluid are due to the velocities of the individual molecules. The velocity of a molecule may be thought of as the sum of the fluid velocity and the thermal velocity. The viscous stress tensor described above gives the force due to the fluid velocity only. The force on an area element in the fluid due to the thermal velocities of the molecules is just the hydrostatic pressure. This pressure term ($p\delta_{ij}$) must be added to the viscous stress tensor to obtain the total stress tensor for the fluid.

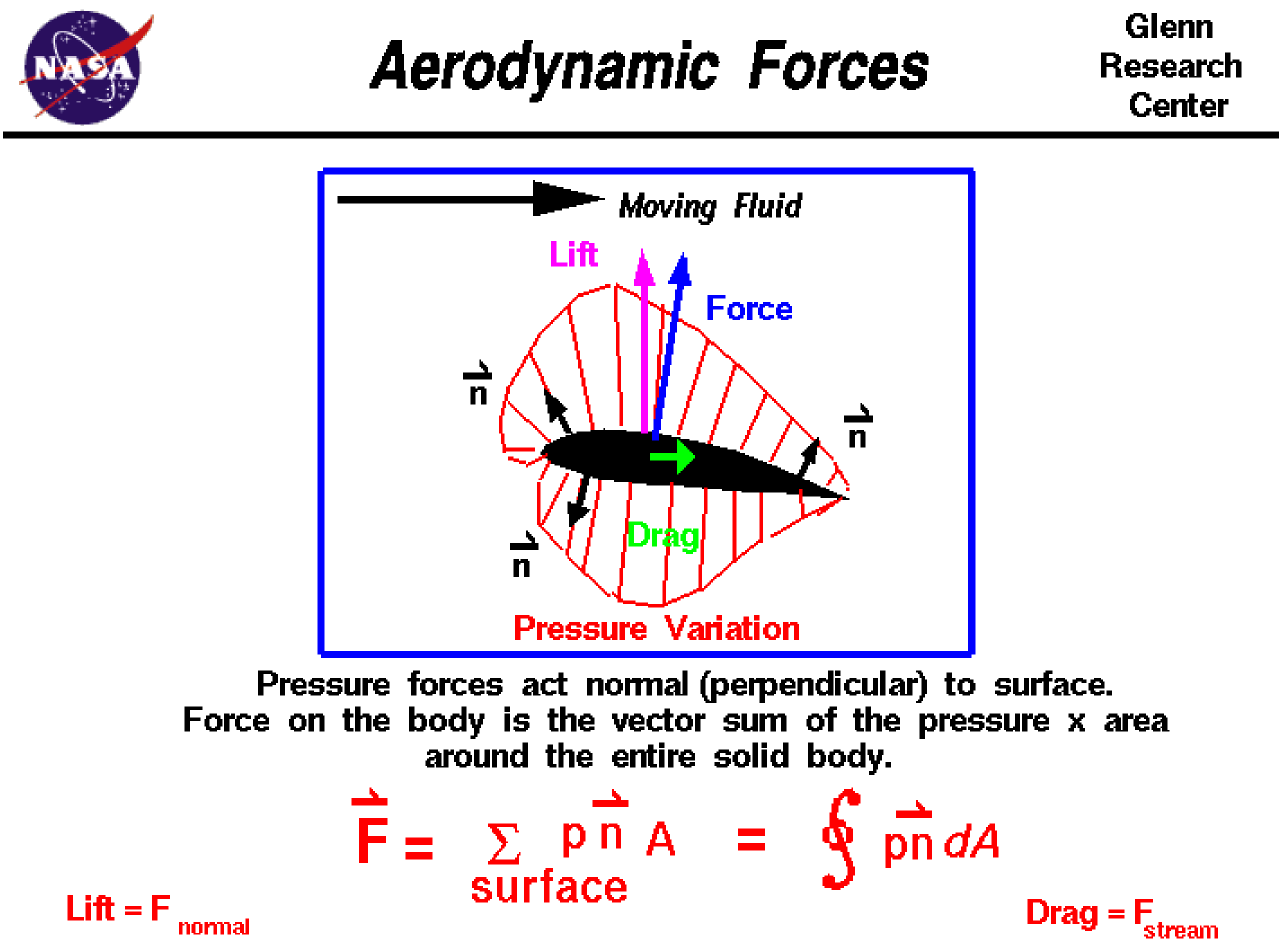
$$\sigma_{ij} = p\delta_{ij} + \sigma_{visc;ij}$$

The infinitesimal force dF_i on an infinitesimal area dA_i is then given by the usual relationship:

$$dF_i = \sigma_{ij} dA_j$$

Etymology

The word "viscosity" derives from the Latin word "viscum" for mistletoe. A viscous glue was made from mistletoe berries and used for lime-twigs to catch birds.



When two solid objects interact in a mechanical process, forces are transmitted, or applied, at the point of contact. But when a solid object interacts with a fluid, things are more difficult to describe because the fluid can change its shape. For a solid body immersed in a fluid, the "point of contact" is every point on the surface of the body. The fluid can flow around the body and maintain physical contact at all points. The transmission, or application, of mechanical forces between a solid body and a fluid occurs at every point on the surface of the body. And the transmission occurs through the fluid pressure.

Variation in Pressure

The magnitude of the force acting over a small section of an object immersed in a fluid equals the pressure **p** times the area **A** of the section. A quick units check shows that:

p * A = (force/area) * area = force

As discussed on the fluid pressure slide, **pressure** is a scalar quantity related to the momentum of the molecules of a fluid. Since a force is a vector quantity, having both magnitude and direction, we must determine the direction of the force. Pressure acts **perpendicular (or normal)** to the solid surface of an object. So the direction of the force on the small section of the object is along the normal to the surface. We denote this direction by the letter **n**.

The normal direction changes from the front of the airfoil to the rear and from the top to the bottom. To obtain the net mechanical force over the entire solid object, we must sum the contributions from all the small sections. Mathematically, the summation is indicated by the

Greek letter **sigma** (Σ) The net aerodynamic force **F** is equal to the sum of the product of the pressure **p** times the area **A** in the normal direction.

$F = \Sigma p * A * n$

In the limit of infinitely small sections, this gives the integral of the pressure times the area around the closed surface. Using the symbol S dA for integration, we have:

F = S (p * n) dA

where the integral is taken all around the body. On the figure, that is why the integral sign has a circle through it.

If the pressure on a closed surface is a constant, there is no net force produced because the summation of the directions of the normal adds up to zero. For every small section there is another small section whose normal points in exactly the opposite direction.

F = S (p * n) dA = p * S n dA = 0

For a fluid in motion, the velocity has different values at different locations around the body. The local pressure is related to the local velocity, so the pressure also varies around the closed surface and a net force is produced. **Summing** the pressure perpendicular to the surface times the area around the body produces a **net force**.

F = S (p * n) dA

Definitions of Lift and Drag

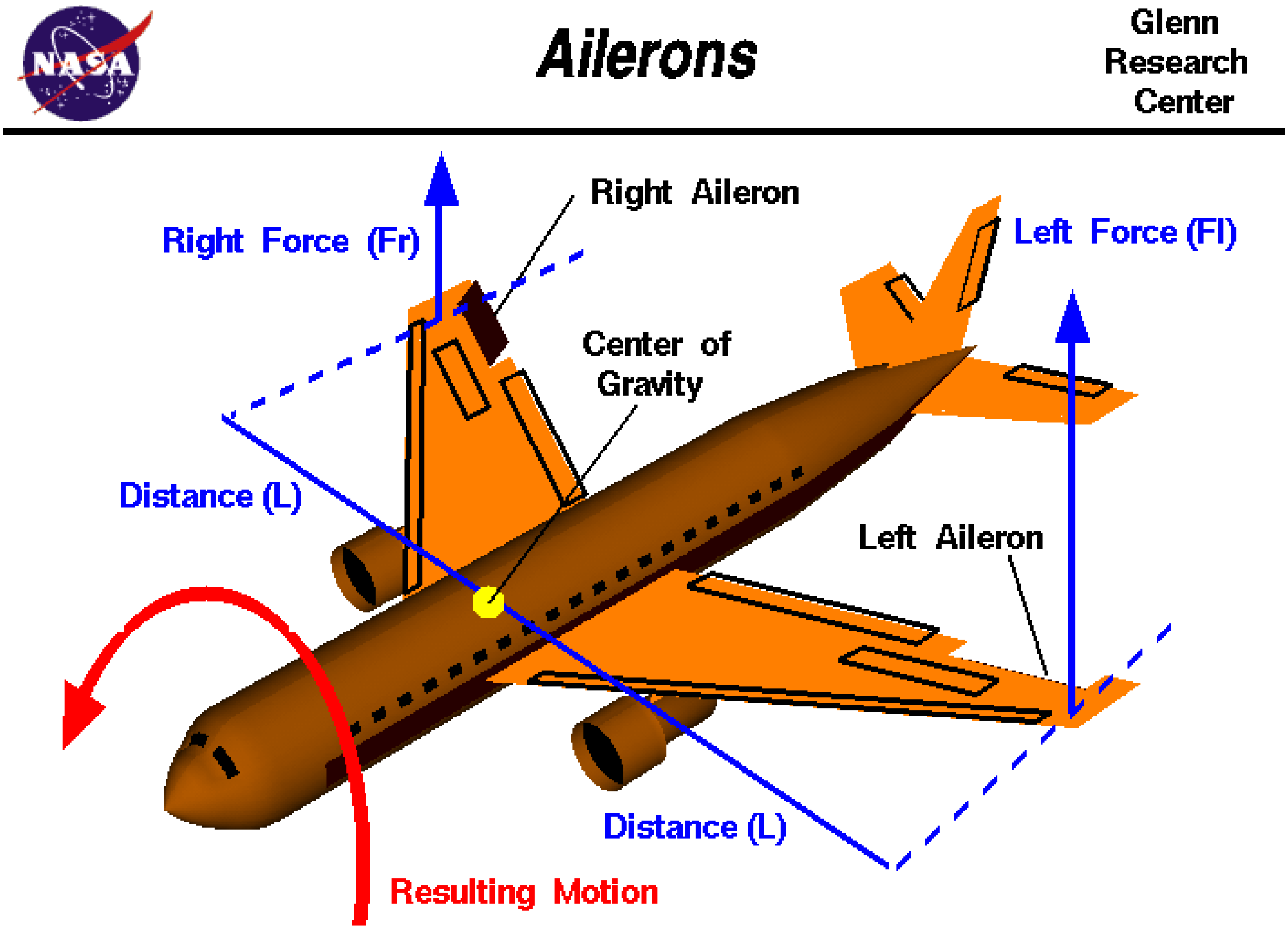
Since the fluid is in motion, we can define a flow direction along the motion. The component of the net force perpendicular (or normal) to the flow direction is called the lift; the component of the net force along the flow direction is called the drag. These are definitions. In reality, there is a single, net, integrated force caused by the pressure variations along a body. This **aerodynamic force** acts through the average location of the pressure variation which is called the center of pressure.

Velocity Distribution

For an ideal fluid with no boundary layers, the surface of an object is a streamline. If the velocity is low, and no energy is added to the flow, we can use Bernoulli's equation along a streamline to determine the pressure distribution for a known velocity distribution. If boundary layers are present, things are a little more confusing, since the external flow responds to the edge of the boundary layer and the pressure on the surface is imposed from the edge of the boundary layer. If the boundary layer separates from the surface, it gets even more confusing. How do we determine the velocity distribution around a body? **Specifying** the velocity is the source of error in two of the more popular incorrect theories of lift. To correctly determine the velocity distribution, we have to solve equations expressing a conservation of mass, momentum, and energy for the fluid passing the object. In some cases, we can solve simplified versions of the equations to determine the velocity.

Summary

To summarize, for any object immersed in a fluid, the mechanical forces are transmitted at every point on the surface of the body. The forces are transmitted through the pressure, which acts perpendicular to the surface. The net force can be found by integrating (or summing) the pressure times the area around the entire surface. For a moving flow, the pressure will vary from point to point because the velocity varies from point to point. For some simple flow problems, we can determine the pressure distribution (and the net force) if we know the velocity distribution by using Bernoulli's equation.



Ailerons can be used to generate a rolling motion for an aircraft. **Ailerons** are small hinged sections on the outboard portion of a wing. Ailerons usually work in opposition: as the right aileron is deflected upward, the left is deflected downward, and vice versa. This slide shows what happens when the pilot deflects the right aileron upwards and the left aileron downwards.

The ailerons are used to bank the aircraft; to cause one wing tip to move up and the other wing tip to move down. The banking creates an unbalanced side force component of the large wing lift force which causes the aircraft's flight path to curve. (Airplanes turn because of banking created by the ailerons, not because of a rudder input.

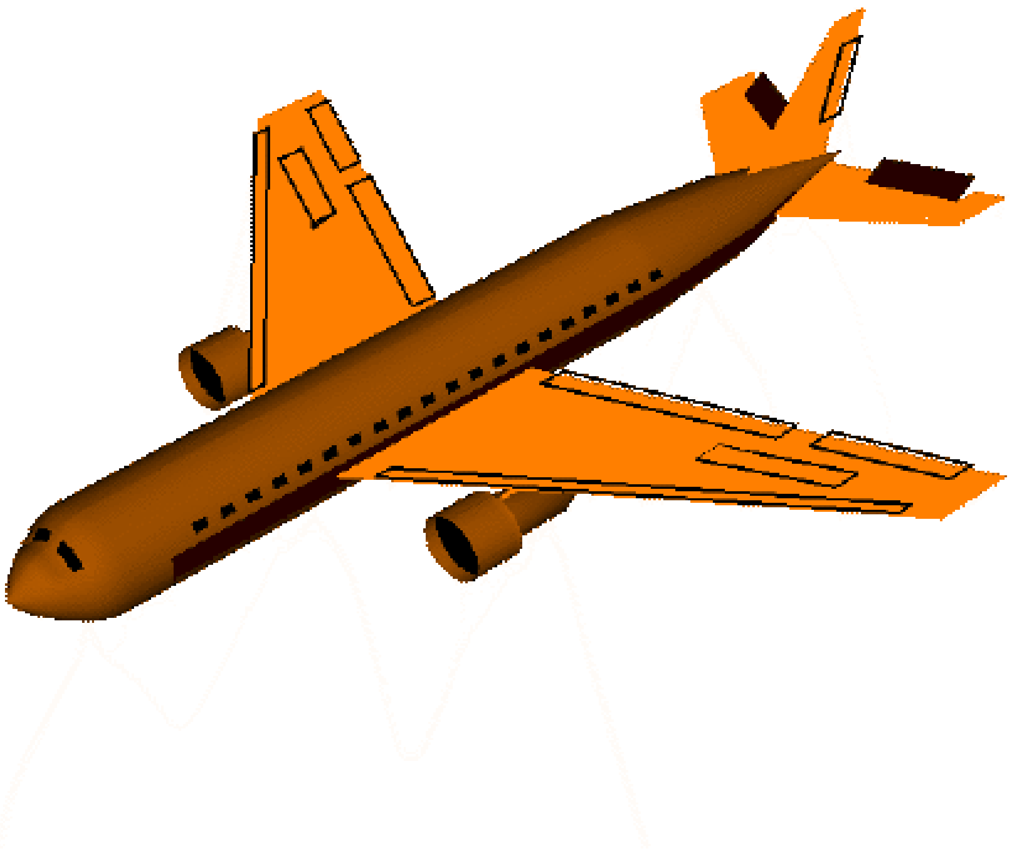
The ailerons work by changing the effective shape of the airfoil of the outer portion of the wing. As described on the shape effects slide, changing the angle of deflection at the rear of an airfoil will change the amount of lift generated by the foil. With greater downward deflection, the lift will increase in the upward direction. Notice on this slide that the aileron on the left wing, as viewed from the rear of the aircraft, is deflected down. The aileron on the right wing is deflected up. Therefore, the lift on the left wing is increased, while the lift on the right wing is decreased. For both wings, the lift force (Fr or Fl) of the wing section through the aileron is applied at the aerodynamic center of the section which is some distance (L) from the aircraft center of gravity. This creates a torque

T = F * L

about the center of gravity. If the forces (and distances) are equal there is no net torque on the aircraft. But if the forces are unequal, there is a net torque and the aircraft rotates about its center of gravity. For the conditions shown in the figure, the resulting motion will roll the aircraft to the right (clockwise) as viewed from the rear. If the pilot reverses the aileron deflections (right aileron down, left aileron up) the aircraft will roll in the opposite direction. We have chosen to name the left wing and right wing based on a view from the back of the aircraft towards the nose, because that is the direction in which the pilot is looking.

Aircraft Pitch Motion

Glenn Research Center

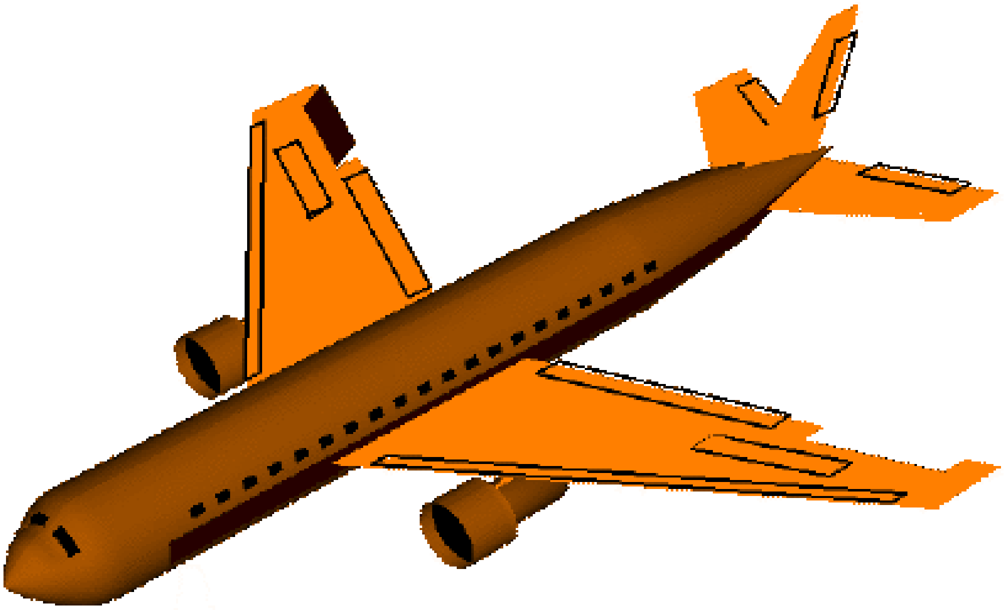


In flight, any aircraft will rotate about its center of gravity, a point which is the average location of the mass of the aircraft. We can define a three dimensional coordinate system through the center of gravity with each axis of this coordinate system perpendicular to the other two axes. We can then define the orientation of the aircraft by the amount of rotation of the parts of the aircraft along these **principal axes**. The **pitch axis** is perpendicular to the aircraft centerline and lies in the plane of the wings. A **pitch motion** is an up or down movement of the nose of the aircraft as shown in the animation.

The pitching motion is being caused by the deflection of the elevator of this aircraft. The elevator is a hinged section at the rear of the horizontal stabilizer. There is usually an elevator on each side of the vertical stabilizer. The elevators work in pairs; when the right elevator goes up, the left elevator also goes up.

As described on the shape effects slide, changing the angle of deflection at the rear of an airfoil changes the amount of lift generated by the foil. With greater downward deflection, lift increases in the upward direction. With greater upward deflection, lift increases in the downward direction. The lift generated by the elevator acts through the center of pressure of the elevator and horizontal stabilizer and is located at some distance from the center of gravity of the aircraft. The change in lift created by deflecting the elevator generates a torque about the center of gravity which causes the airplane to rotate. The pilot can use this ability to make the airplane loop. Or, since many aircraft loop naturally, the deflection can be used to trim or balance the aircraft, thus preventing a loop.

On many aircraft, the horizontal stabilizer and elevator create a symmetric airfoil like the one shown on the left of the shape effects slide. This produces no lift when the elevator is aligned with the stabilizer and allows the combination to produce either positive or negative lift, depending on the deflection of the elevator. On many fighter planes, in order to meet their high maneuvering requirements, the stabilizer and elevator are combined into one large moving surface called a stabilator. The change in force is created by changing the inclination of the entire surface, not by changing its effective shape.

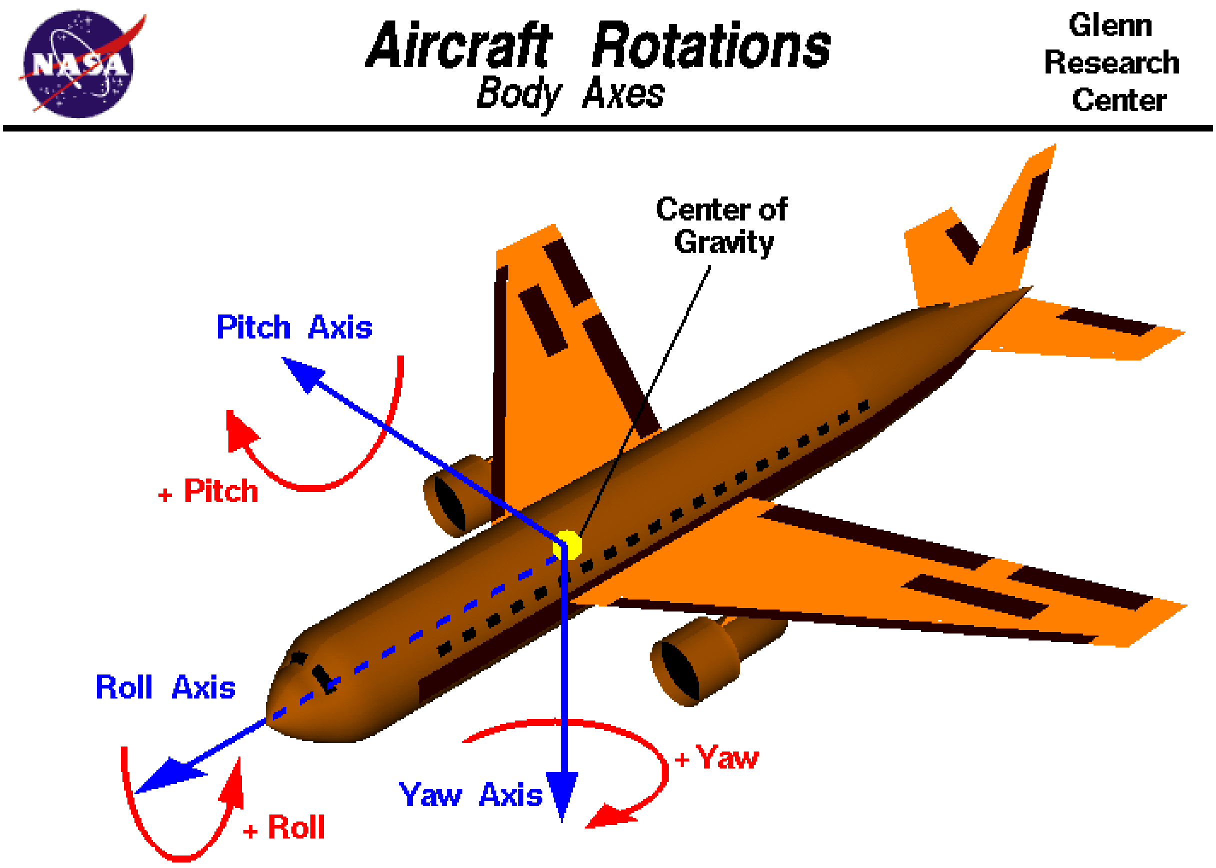


In flight, any aircraft will rotate about its center of gravity, a point which is the average location of the mass of the aircraft. We can define a three dimensional coordinate system through the center of gravity with each axis of this coordinate system perpendicular to the other two axes. We can then define the orientation of the aircraft by the amount of rotation of the parts of the aircraft along these **principal axes**. The **roll axis** lies along the aircraft centerline. A **roll motion** is an up and down movement of the wings of the aircraft as shown in the animation.

The rolling motion is being caused by the deflection of the ailerons of this aircraft. The aileron is a hinged section at the rear of each wing. The ailerons work in opposition; when the right aileron goes up, the left aileron goes down.

As described on the shape effects slide, changing the angle of deflection at the rear of an airfoil will change the amount of lift generated by the foil. With greater downward deflection, the lift will increase in the upward direction; with greater upward deflection, the lift will decrease in the upward direction. Since the ailerons work in pairs, the lift on one wing increases as the lift on the opposite wing decreases. Because the forces are not equal, there is a net twist, or torque about the center of gravity and the aircraft rotates about the roll axis. The pilot can use this ability to bank the aircraft which causes the airplane to turn.

On this page we have demonstrated an aircraft roll induced by movement of the ailerons, but there are other ways to produce a rolling motion on an aircraft. The Wright brothers used a method called wing warping. Their wings were wired together in such a way that the outer panels of each wing could be twisted relative to the inner panel. The twisting changed the local angle of attack of sections of the wing which changed the lift being generated by that section. Unequal forces on the wings caused the aircraft to roll. Many modern airliners use a spoiler to roll the aircraft. A spoiler is a plate that is raised between the leading and trailing edges of the wing. The spoiler effectively changes the shape of the airfoil, disrupts the flow over the wing, and causes a section of the wing to decrease its lift. This produces an unbalanced force with the other wing, which causes the roll. Airliners use spoilers because spoilers can react more quickly than ailerons and require less force to activate, but they always decrease the total amount of lift for the aircraft. It's an interesting trade! You can tell whether an airliner is using spoilers or ailerons by noticing where the moving part is located. At the trailing edge, it's an aileron; between the leading and trailing edges, it's a spoiler. (Now you can dazzle the person sitting next to you on the plane!)



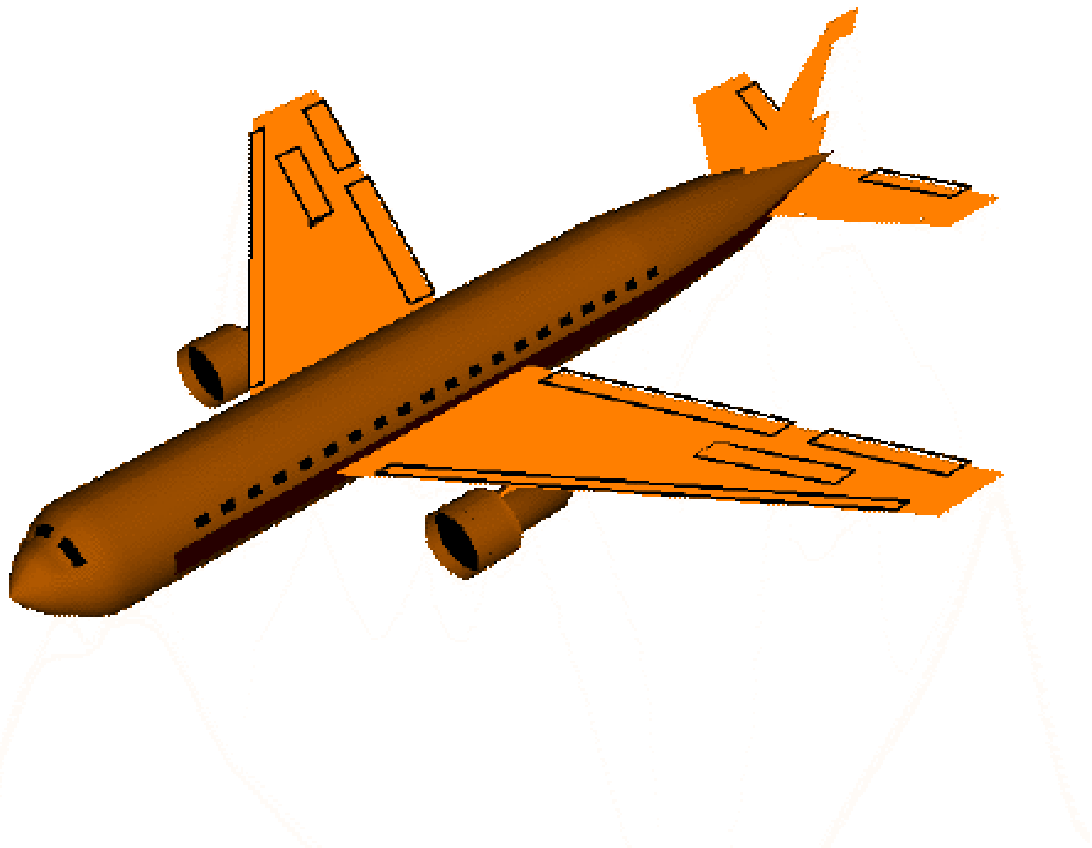
Since we live in a three dimensional world, it is necessary to control the **attitude** or orientation of a flying aircraft in all three dimensions. In flight, any aircraft will rotate about its center of gravity, a point which is the average location of the mass of the aircraft. We can define a three dimensional coordinate system through the center of gravity with each axis of this coordinate system perpendicular to the other two axes. We can then define the orientation of the aircraft by the amount of rotation of the parts of the aircraft along these **principal axes**.

The **yaw axis** is defined to be perpendicular to the plane of the wings with its origin at the center of gravity and directed towards the bottom of the aircraft. A yaw motion is a movement of the nose of the aircraft from side to side. The **pitch axis** is perpendicular to the yaw axis and is parallel to the plane of the wings with its origin at the center of gravity and directed towards the right wing tip. A pitch motion is an up or down movement of the nose of the aircraft. The **roll axis** is perpendicular to the other two axes with its origin at the center of gravity, and is directed towards the nose of the aircraft. A rolling motion is an up and down movement of the wing tips of the aircraft.

In flight, the control surfaces of an aircraft produce aerodynamic forces. These forces are applied at the center of pressure of the control surfaces which are some distance from the aircraft cg and produce torques (or moments) about the principal axes. The torques cause the aircraft to rotate. The elevators produce a pitching moment, the rudder produces a yawing moment, and the ailerons produce a rolling moment. The ability to vary the amount of the force and the moment allows the pilot to maneuver or to trim the aircraft. The first aircraft to demonstrate active control about all three axes was the Wright brothers' 1902 glider.

Aircraft Yaw Motion

Glenn Research Center

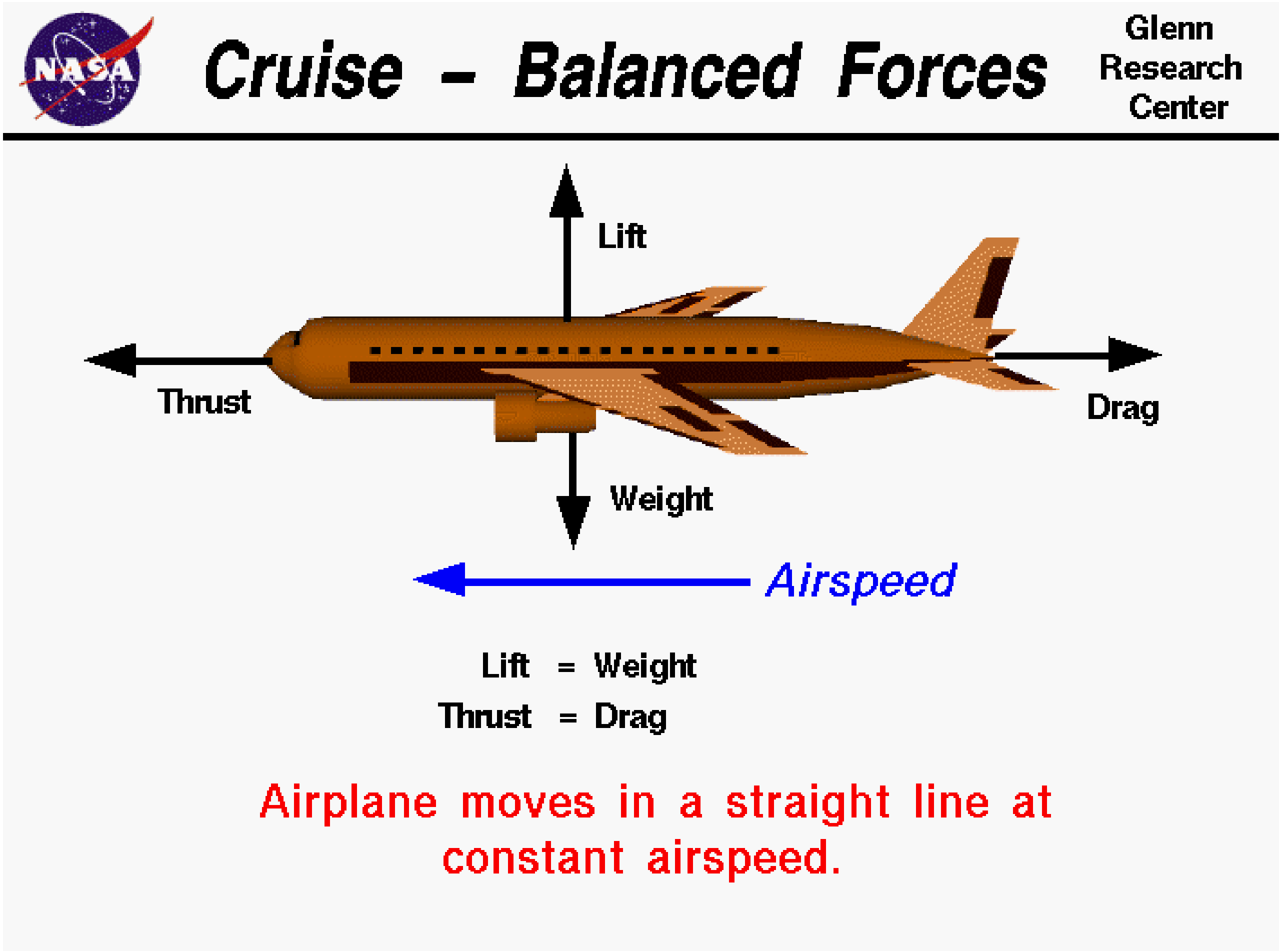


In flight, any aircraft will rotate about its center of gravity, a point which is the average location of the mass of the aircraft. We can define a three dimensional coordinate system through the center of gravity with each axis of this coordinate system perpendicular to the other two axes. We can then define the orientation of the aircraft by the amount of rotation of the parts of the aircraft along these **principal axes**. The **yaw axis** is perpendicular to the wings and lies in the plane of the aircraft centerline. A **yaw motion** is a side to side movement of the nose of the aircraft as shown in the animation.

The yawing motion is being caused by the deflection of the rudder of this aircraft. The rudder is a hinged section at the rear of the vertical stabilizer.

As described on the shape effects slide, changing the angle of deflection at the rear of an airfoil changes the amount of lift generated by the foil. For the vertical stabilizer and rudder, the orientation of the airfoil causes a side force to be generated. With greater deflection of the rudder to the left, the side force increases to the right. With greater deflection to the right, the side force increases to the left. The lift generated by the rudder acts through the center of pressure of the rudder and vertical stabilizer and is located at some distance from the center of gravity of the aircraft. The change in side force created by deflecting the rudder generates a torque about the center of gravity which causes the airplane to rotate. The pilot uses this ability to keep the nose of the aircraft pointed in the direction of travel.

On all aircraft, the vertical stabilizer and rudder create a symmetric airfoil. This produces no side force when the rudder is aligned with the stabilizer and allows the combination to produce either positive or negative side force, depending on the deflection of the rudder. Some fighter planes have two vertical stabilizers and rudders because of the need to control the plane with multiple, very powerful engines

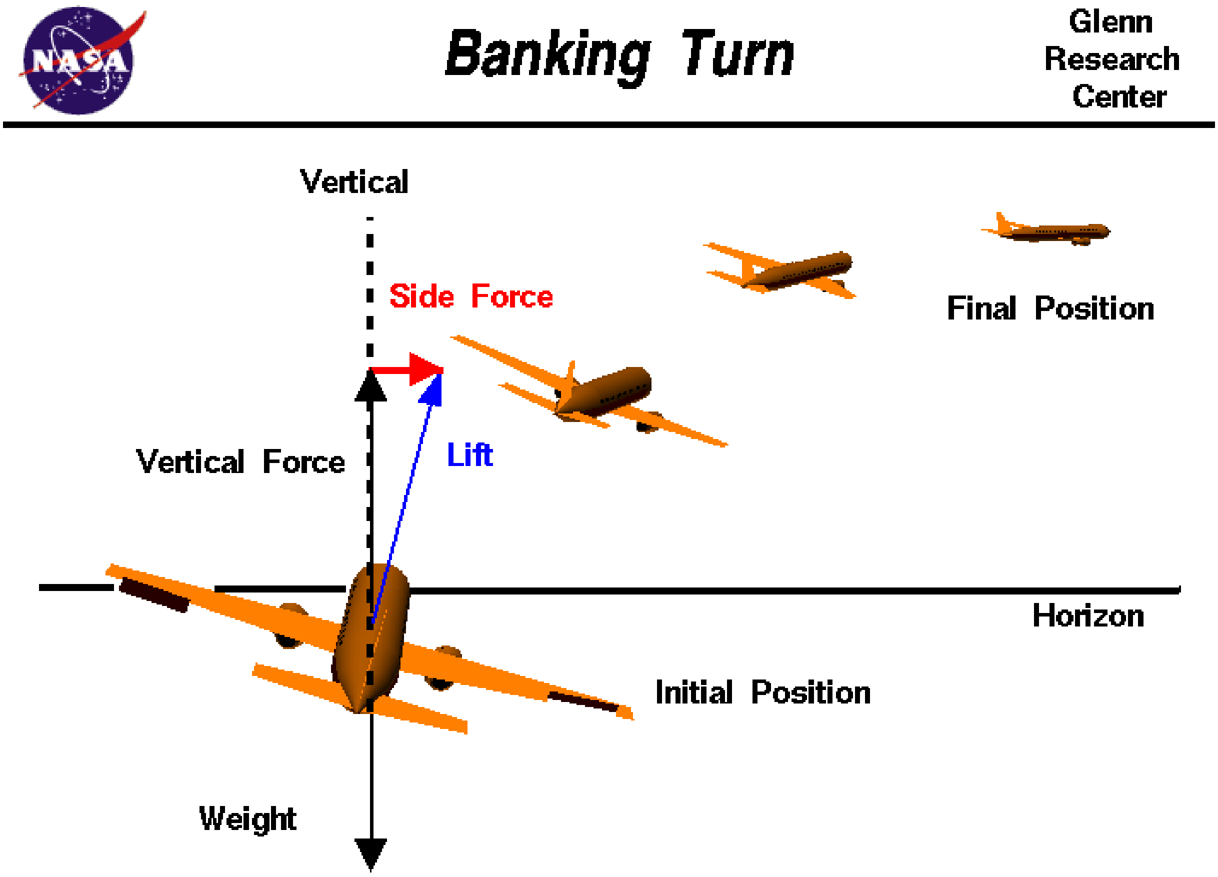


There are four forces that act on an aircraft in flight: lift, weight, thrust, and drag. From Newton's first law of motion we know that an object at rest will stay at rest, and an object in motion (constant velocity) will stay in motion unless acted on by an external force. If there is no net external force, the object will maintain a constant velocity.

In an ideal situation, the forces acting on an aircraft in flight can produce no net external force. In this situation the lift is equal to the weight, and the thrust is equal to the drag. The closest example of this condition is a **cruising** airliner. While the weight decreases due to fuel burned, the change is very small relative to the total aircraft weight. The aircraft maintains a constant airspeed called the **cruise velocity**.

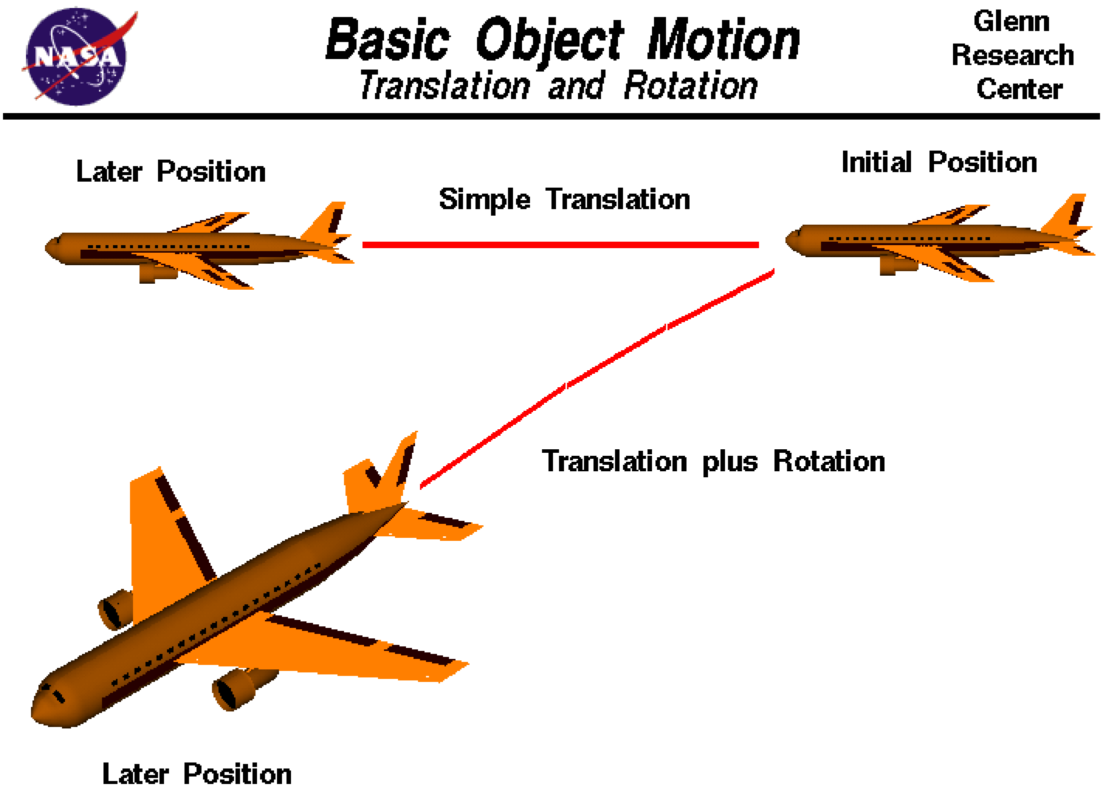
If we take into account the relative velocity of the wind, we can determine the ground speed of a cruising aircraft. The ground speed is equal to the airspeed plus the wind speed using vector addition. The motion of the aircraft is a pure translation. With a constant ground speed it is relatively easy to determine the aircraft range, the distance the airplane can fly with a given load of fuel.

If the pilot changes the throttle setting, or increases the wing angle of attack, the forces become unbalanced. The aircraft will move in the direction of the greater force, and we can compute acceleration of the aircraft from Newton's second law of motion.



A fundamental aircraft motion is a **banking turn**. This maneuver is used to change the aircraft heading. The turn is initiated by using the ailerons or spoilers to roll, or bank, the aircraft to one side. On the figure, the airliner is banked to the right by lowering the left aileron and raising the right aileron. The lift of the wings of the aircraft is a vector quantity which is always directed perpendicular to the flight path and perpendicular to the wings generating the lift. As the aircraft is rolled, the lift vector is tilted in the direction of the roll. We can break the lift vector into two components. One component is vertical and opposed to the weight which is always directed towards the center of the earth. The other component is an unopposed side force which is in the direction of the roll, and perpendicular to the flight path.

As long as the aircraft is banked, the side force is a constant, unopposed force on the aircraft. The resulting motion of the center of gravity of the aircraft is a circular arc. When the wings are brought level by an opposing motion of the ailerons, the side force is eliminated and the aircraft continues to fly in a straight line along a new heading. Notice that the rudder is not used to turn the aircraft. The aircraft is turned through the action of the side component of the lift force. The rudder is used during the turn to **coordinate** the turn, i.e. to keep the nose of the aircraft pointed along the flight path. If the rudder is not used, one can encounter an **adverse yaw** in which the drag on the outer wing pulls the aircraft nose away from the flight path.



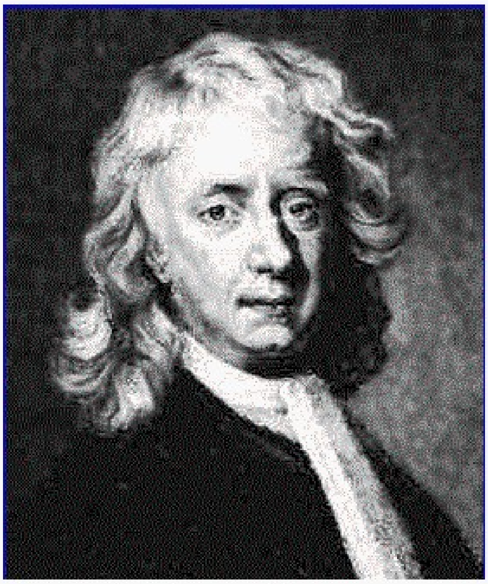
We live in a world that is defined by three spatial dimensions and one time dimension. Objects move within this domain in two ways. An object translates, or changes **location**, from one point to another. And an object rotates, or changes its **attitude**. In general, the motion of any object involves both translation and rotation. The translations are in direct response to external forces. The rotations are in direct response to external torques or moments (twisting forces).

The motion of an aircraft is particularly complex because the rotations and translations are coupled together; a rotation affects the magnitude and direction of the forces which affect translations. To understand and describe the motion of an aircraft, we usually try to break down the complex problem into a series of easier problems. We can, for instance, assume that the aircraft translates from one point to another as if all the mass of the aircraft were collected into a single point called the center of gravity. We can describe the motion of the center of gravity by using Newton's laws of motion. There are four forces acting on the aircraft; the lift, drag, thrust, and weight. Depending on the relative magnitudes and directions of these forces, the aircraft will climb (increase in altitude), dive (decrease in altitude), or bank (roll to one side). The magnitude of the aerodynamic forces depends on the attitude of the aircraft during the translations. The attitude depends on the rotations about the center of gravity when the aircraft is trimmed.

Bernoulli and Newton

Glenn Research Center

Daniel Bernoulli



Sir Isaac Newton

Lift is the force that holds an aircraft in the air. How is lift generated? There are many explanations for the generation of lift found in encyclopedias, in basic physics textbooks, and on Web sites. Unfortunately, many of the explanations are misleading and incorrect. Theories on the generation of lift have become a source of great controversy and a topic for heated arguments for many years.

The proponents of the arguments usually fall into two camps: (1) those who support the "Bernoulli" position that lift is generated by a pressure difference across the wing, and (2) those who support the "Newton" position that lift is the reaction force on a body caused by deflecting a flow of gas. ***Notice that we place the names in quotation marks because neither Newton nor Bernoulli ever attempted to explain the aerodynamic lift of an object***. The names of these scientists are just labels for two camps.

Looking at the lives of Bernoulli and Newton we find more similarities than differences. On the figure at the top of this page we show portraits of Daniel Bernoulli, on the left, and Sir Isaac Newton, on the right. Newton worked in many areas of mathematics and physics. He developed the theories of gravitation in 1666, when he was only 23 years old. Some twenty years later, in 1686, he presented his three laws of motion in the *Principia Mathematica Philosophiae Naturalis* . He and Gottfried Leibnitz are also credited with the development of the mathematics of Calculus. Bernoulli also worked in many areas of mathematics and physics and had a degree in medicine. In 1724, at age 24, he had published a mathematical work in which he investigated a problem begun by Newton concerning the flow of water from a container and several other problems involving differential equations. In 1738, his work *Hydrodynamica* was published. In this work, he applied the conservation of energy to fluid mechanics problems.

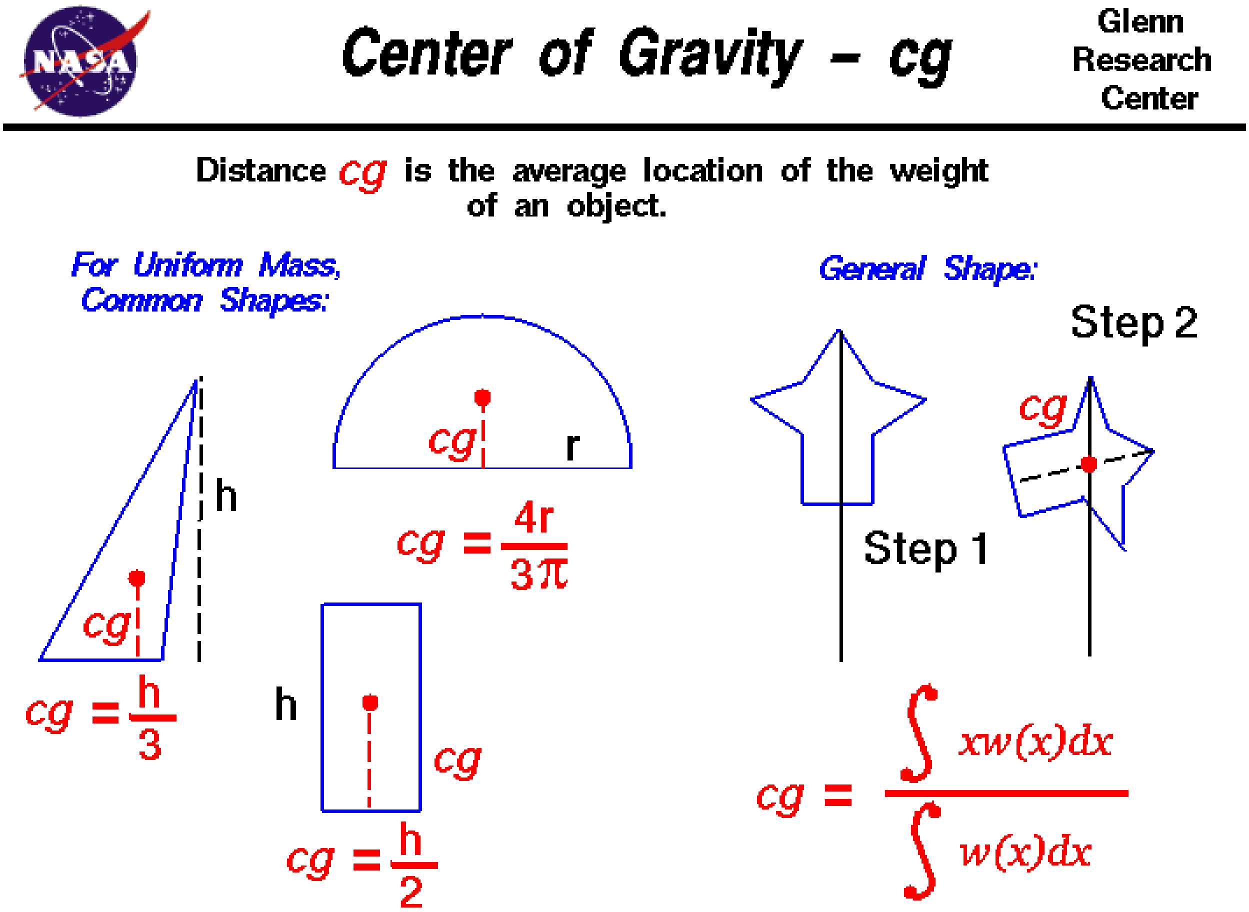
Which camp is correct? How is lift generated?

When a gas flows over an object, or when an object moves through a gas, the molecules of the gas are free to move about the object; they are not closely bound to one another as in a solid. Because the molecules move, there is a velocity associated with the gas. Within the gas, the velocity can have very different values at different places near the object. Bernoulli's equation, which was named for Daniel Bernoulli, relates the pressure in a gas to the local velocity; so as the velocity changes around the object, the pressure changes as well. Adding up (integrating) the pressure variation times the area around the entire body determines the aerodynamic force on the body. The lift is the component of the aerodynamic force which is perpendicular to the original flow direction of the gas. The drag is the component of the aerodynamic force which is parallel to the original flow direction of the gas. Now adding up the velocity variation around the object instead of the pressure variation also determines the aerodynamic force. The integrated velocity variation around the object produces a net turning of the gas flow. From Newton's third law of motion, a turning action of the flow will result in a re-action (aerodynamic force) on the object. *So **both** "Bernoulli" and "Newton" are correct*. Integrating the effects of either the pressure or the velocity determines the aerodynamic force on an object. We can use equations developed by each of them to determine the magnitude and direction of the aerodynamic force.

What is the argument?

Arguments arise because people mis-apply Bernoulli and Newton's equations and because they over-simplify the description of the problem of aerodynamic lift. The most popular incorrect theory of lift arises from a mis-application of Bernoulli's equation. The theory is known as the "equal transit time" or "longer path" theory which states that wings are designed with the upper surface longer than the lower surface, to generate higher velocities on the upper surface because the molecules of gas on the upper surface have to reach the trailing edge at the same time as the molecules on the lower surface. The theory then invokes Bernoulli's equation to explain lower pressure on the upper surface and higher pressure on the lower surface resulting in a lift force. The error in this theory involves the specification of the velocity on the upper surface. In reality, the velocity on the upper surface of a lifting wing is much higher than the velocity which produces an equal transit time. If we know the correct velocity distribution, we can use Bernoulli's equation to get the pressure, then use the pressure to determine the force. But the equal transit velocity is not the correct velocity. Another incorrect theory uses a Venturi flow to try to determine the velocity. But this also gives the wrong answer since a wing section isn't really half a Venturi nozzle. There is also an incorrect theory which uses Newton's third law applied to the bottom surface of a wing. This theory equates aerodynamic lift to a stone skipping across the water. It neglects the physical reality that both the lower and upper surface of a wing contribute to the turning of a flow of gas.

The real details of how an object generates lift are very complex and do not lend themselves to simplification. For a gas, we have to simultaneously conserve the mass, momentum, and energy in the flow. Newton's laws of motion are statements concerning the conservation of momentum. Bernoulli's equation is derived by considering conservation of energy. So both of these equations are satisfied in the generation of lift; both are correct. The conservation of mass introduces a lot of complexity into the analysis and understanding of aerodynamic problems. For example, from the conservation of mass, a change in the velocity of a gas in one direction results in a change in the velocity of the gas in a direction perpendicular to the original change. This is very different from the motion of solids, on which we base most of our experiences in physics. The simultaneous conservation of mass, momentum, and energy of a fluid (while neglecting the effects of air viscosity) are called the Euler Equations after Leonard Euler. Euler was a student of Johann Bernoulli, Daniel's father, and for a time had worked with Daniel Bernoulli in St. Petersburg. If we include the effects of viscosity, we have the Navier-Stokes Equations which are named after two independent researchers in France and in England. To truly understand the details of the generation of lift, one has to have a good working knowledge of the Euler Equations.



The **center of gravity** is a geometric property of any object. The center of gravity is the average location of the weight of an object. We can completely describe the motion of any object through space in terms of the **translation** of the center of gravity of the object from one place to another, and the **rotation** of the object about its center of gravity if it is free to rotate. If the object is confined to rotate about some other point, like a hinge, we can still describe its motion. In flight, both airplanes and rockets rotate about their centers of gravity. A kite, on the other hand, rotates about the bridle point. But the trim of a kite still depends on the location of the center of gravity relative to the bridle point, because for every object the weight always acts through the center of gravity.

Determining the center of gravity is very important for any flying object. How do engineers determine the location of the center of gravity for an aircraft which they are designing?

In general, determining the center of gravity (cg) is a complicated procedure because the mass (and weight) may not be uniformly distributed throughout the object. The general case requires the use of calculus which we will discuss at the bottom of this page. If the mass is uniformly distributed, the problem is greatly simplified. If the object has a line (or plane) of **symmetry**, the cg lies on the line of symmetry. For a solid block of uniform material, the center of gravity is simply at the average location of the ^M physical dimensions. (For a rectangular block, 50 X 20 X 10, the center of gravity is at the point (25,10, 5)). For a triangle of height h, the cg is at h/3, and for a semi-circle of radius r, the cg is at (4*r/(3*pi)) where pi is ratio of the circumference of the circle to the diameter. There are tables of the location of the center of gravity for many simple shapes in math and science books. The tables were generated by using the equation from calculus shown on the slide.

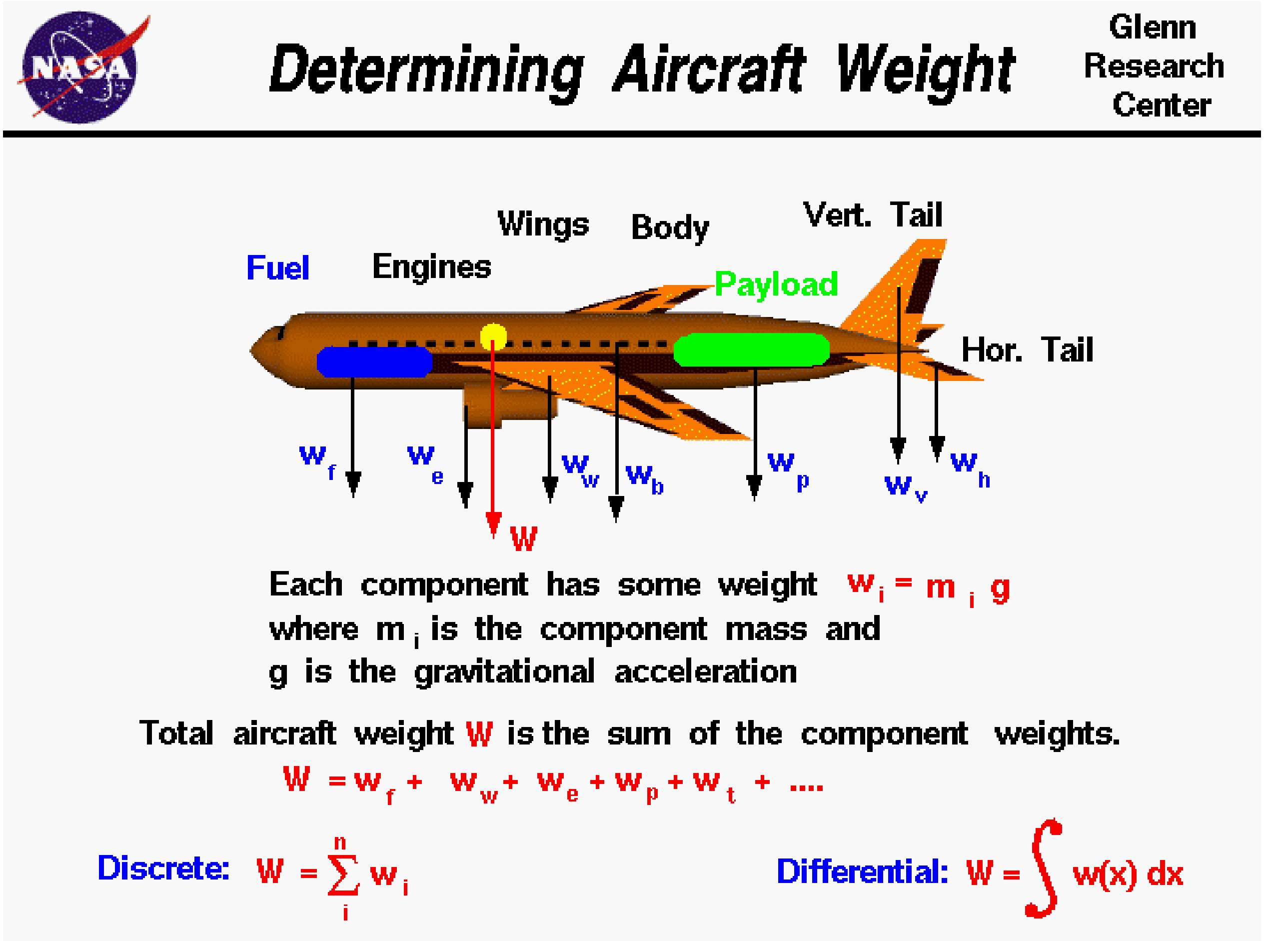
For a general shaped object, there is a simple mechanical way to determine the center of gravity:

1. If we just balance the object using a string or an edge, the point at which the object is balanced is the center of gravity. (Just like balancing a pencil on your finger!)
2. Another, more complicated way, is a two step method shown on the slide. In Step 1, you hang the object from any point and you drop a weighted string from the same point. Draw a line on the object along the string. For Step 2, repeat the procedure from another point on the object You now have two lines drawn on the object which intersect. The center of gravity is the point where the lines intersect. This procedure works well for irregularly shaped objects that are hard to balance.

If the mass of the object is not uniformly distributed, we can characterize the mass distribution by a function w(x) which indicates that the weight is some function of distance x from a reference line. If we can determine the form of the function, there are methods to perform a calculus integration of the equation. We will use the symbols "S[ ]dx" to denote the integration of a continuous function. Then the center of gravity can be determined from:

cg = (S[x * w(x)]dx) / (S[w(x)]dx)

If we don't know the actual functional form, we can numerically integrate the equation using a spreadsheet by dividing the distance into a number of small distance segments and determining the average value of the **mass** (or weight) over that small segment. Taking the sum of the average value times the distance times the distance segment divided by the sum of the average value times the distance segment will produce the center of gravity.



Weight is the force generated by the gravitational attraction of the earth on the airplane. Each part of the aircraft has a unique weight and mass, and for some problems it is important to know the distribution. But for total aircraft maneuvering, we only need to be concerned with the total weight and the location of the center of gravity. The center of gravity is the average location of the mass of any object.

How do engineers determine the weight of an airplane which they are designing?

An airplane is a combination of many parts; the wings, engines, fuselage, and tail, plus the payload and the fuel. Each part has a weight associated with it which the engineer can estimate, or calculate, using Newton's weight equation:

$w = m * g$

where **w** is the weight, **m** is the mass, and **g** is the gravitational constant which is 32.2 ft/square sec in English units and 9.8 meters/square sec in metric units. The mass of an individual component can be calculated if we know the size of the component and its chemical composition. Every material (iron, plastic, aluminum, gasoline, etc.) has a unique density. Density **r** is defined to be the mass divided by the volume **v**:

$r = m / v$

If we can calculate the volume

$m = r * v$

The total weight **W** of the aircraft is simply the sum of the weight of all of the individual components.

W = w(fuselage) + w(wing) + w(engines) + w(payload) + w(fuel) + ...

To generalize, if we have a total of "n" discrete components, the weight of the aircraft is the sum of the individual **i** component weights with the index i going from 1 to n. The greek mathematical symbol sigma is used by mathematicians to denote this addition. (Sigma is a zig-zag symbol with the index designation being placed below the bottom bar, the total number of additions placed over the top bar, and the variable to be summed placed to the right of the sigma with each component designated by the index.)

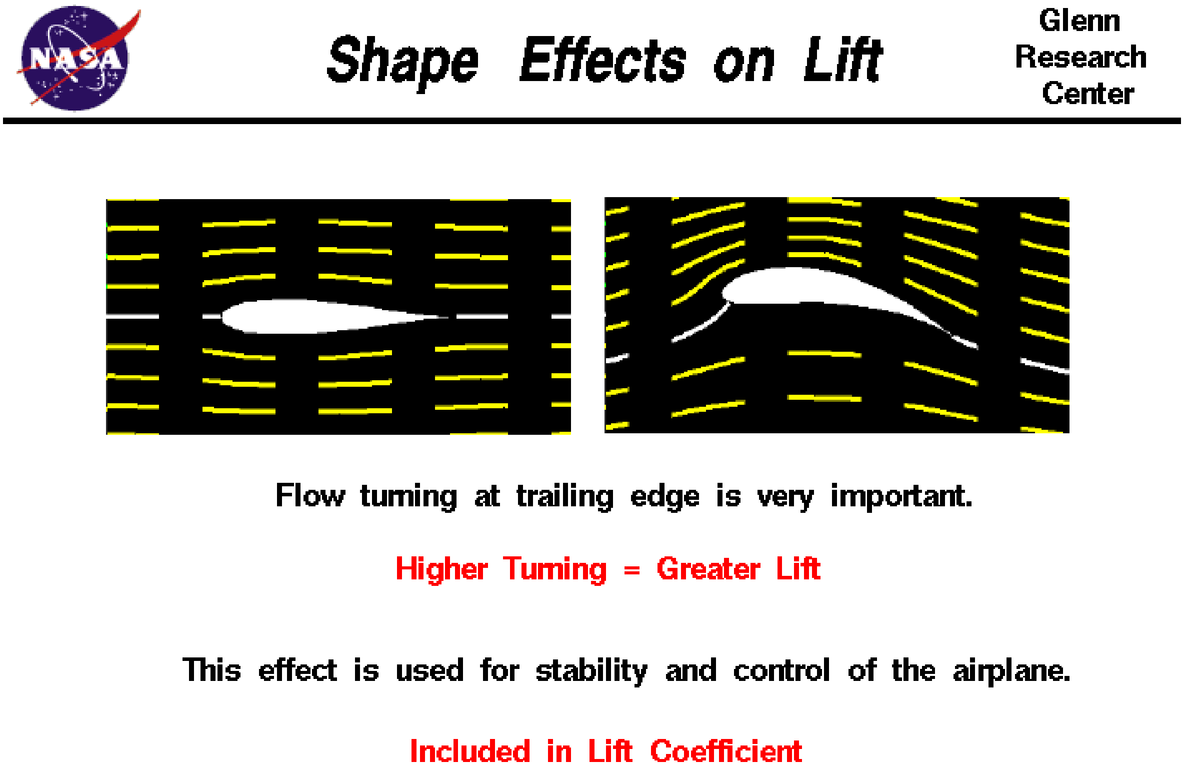
W = SUM(i=1 to i=n) [wi]

This equation says that the weight of the airplane is equal to the sum of the weight of "n" discrete parts.

What if the parts are not discrete? What if we had a continuous change of mass from front to back? The continuous change can be computed using integral calculus. The sigma designation is changed to the integral "S" shaped symbol to denote a continuous variation.

W = S w(x) dx

The discrete weight is replaced with **w(x)** which indicates that the weight is some function of distance **x**. If we are given the form of the function, there are methods to solve the integration. If we don't know the actual functional form, we can still numerically integrate the equation using a spread sheet by dividing the distance up into a number of small distance segments and determining the average value of the weight over that small segment, then summing up the value.

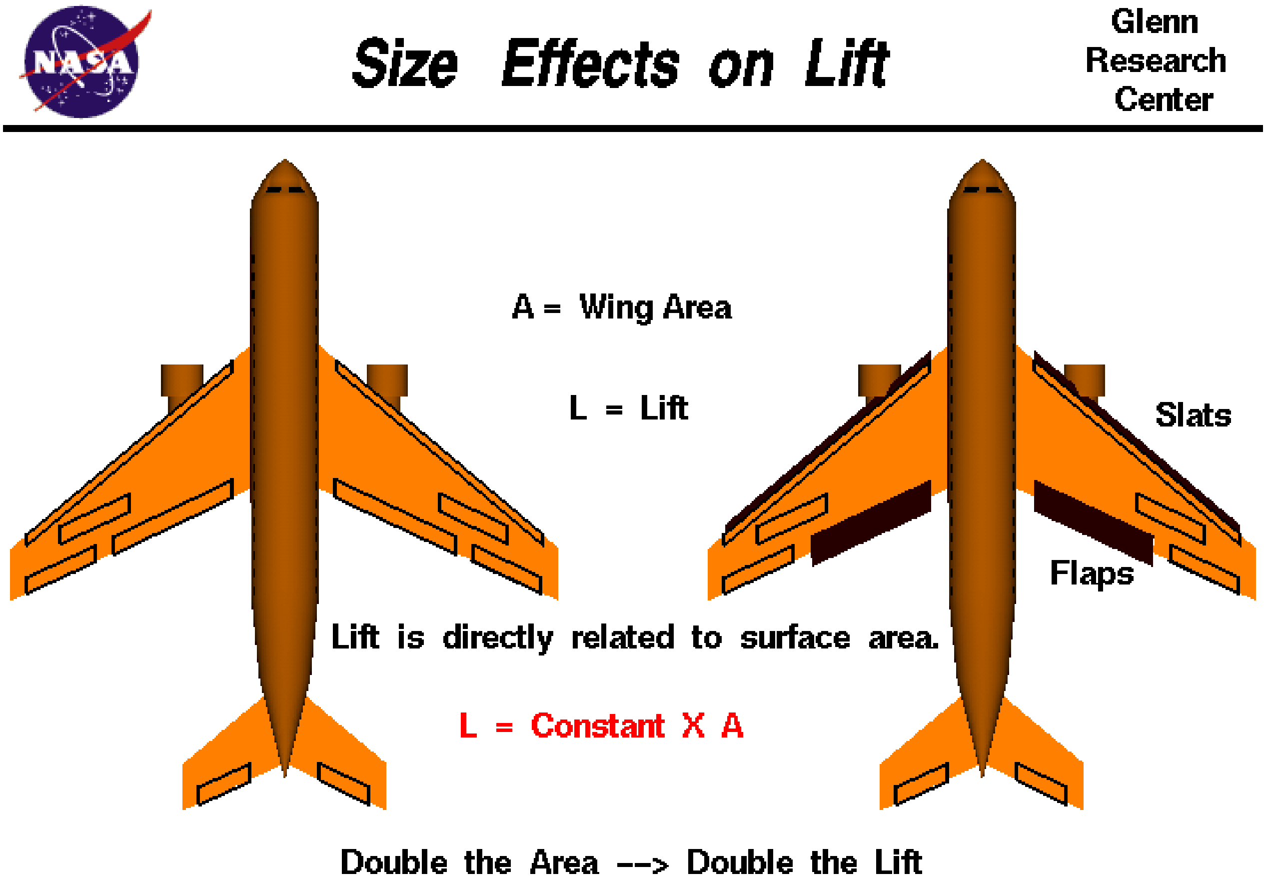


The amount of lift generated by an object depends on how much the flow is turned, which depends on the shape of the object. In general, the lift is a very complex function of the shape. Aerodynamicists model the shape effect by a lift coefficient which is normally determined through wind tunnel testing. For some simple shapes, we can develop mathematical equations to determine the lift coefficient. The simplest model, the two dimensional Kutta-Joukowski airfoil, is studied by undergraduate students. The FoilSim computer program provides the results of this analysis in a form readily usable by students. A result of the analysis shows that **the greater the flow turning, the greater the lift** generated by an airfoil.

This slide shows the flow fields for two different airfoils. The airfoil on the left is a symmetric airfoil; the shapes above and below the white centerline are the same. The airfoil on the right is curved near the trailing edge. The yellow lines on each figure show the streamlines of flow from left to right. The left figure shows no net turning of the flow and produces no lift; the right figure shows a large amount of turning and generates a large amount of lift. The front portions of both airfoils are nearly identical. The aft portion of the right airfoil creates the higher turning.

The example shown above explains why the aft portion of wings have hinged sections to control and maneuver an aircraft. Deflecting the aft section down produces a geometry similar to the figure on the right producing more lift. Similarly, if the aft section is deflected up, it creates less lift (or even negative lift). The ability to vary the amount of lift over a portion of the wing gives the pilot the ability to maneuver an aircraft. The following slides show the deflection of the control surfaces and the resulting motion of the aircraft:

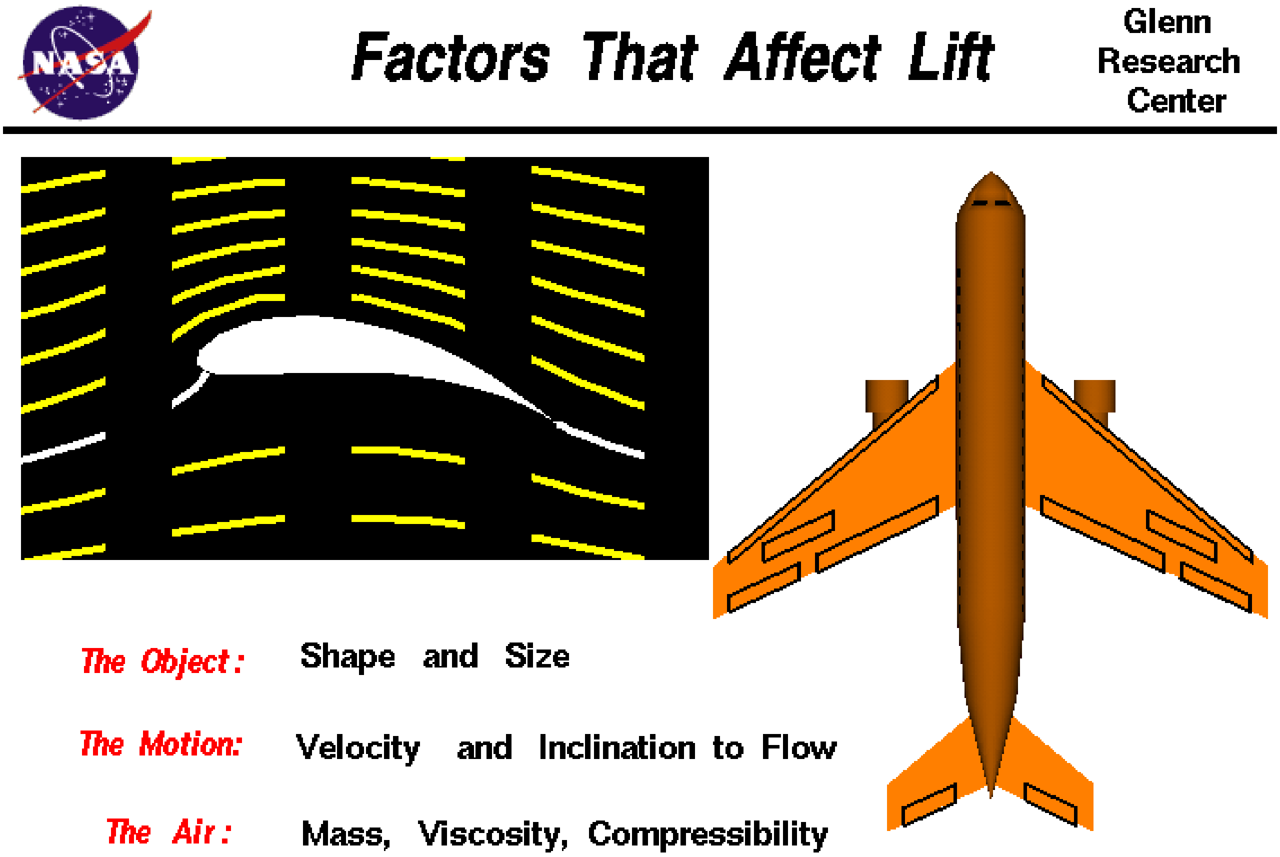
- Elevator controls pitching motion.
- Rudder controls yawing motion.
- Ailerons control rolling motion.



The amount of lift generated by an object depends on the size of the object. Lift is an **aerodynamic force** and therefore depends on the pressure variation of the air around the body as it moves through the air. The total aerodynamic force is equal to the pressure times the surface area around the body. Lift is the component of this force perpendicular to the flight direction. Like the other aerodynamic force, drag, the lift is directly proportional to the area of the object. **Doubling the area doubles the lift.**

There are several different areas from which to choose when developing the **reference area** used in the lift equation. Since most of the lift is generated by the wings, and lift is the force perpendicular to the flight direction, the logical choice is the wing planform area. The planform area is the area of the wing as viewed from above the wing, looking along the "lift" direction. It is a flat plane, and is NOT the total surface area (top and bottom) of the entire wing, although it is almost half that number for most wings. We could, in theory, use the total surface area as the reference area. The total surface area is proportional to the wing planform area. Since the lift coefficient is determined experimentally, by measuring the lift and measuring the area and performing the necessary math to produce the coefficient, we are free to use **any area** which can be easily measured. If we choose the total surface area, the computed coefficient has a different value than if we choose the wing planform area, but the lift is the same, and the coefficients are related by the ratio of the areas.

This slide shows the projected surface area for two different aircraft. The airplane on the left is shown in a cruise condition while the airplane on the right is shown in a takeoff or landing condition. Takeoff and landing are times of relatively low velocity, so to keep the lift high (to avoid the ground!) designers try to increase the wing area. This is done by sliding the flaps backwards along metal tracks and shifting the slats forward to increase the wing area. The next time you fly in an airliner, watch the wings during takeoff and landing to see the change in wing area.

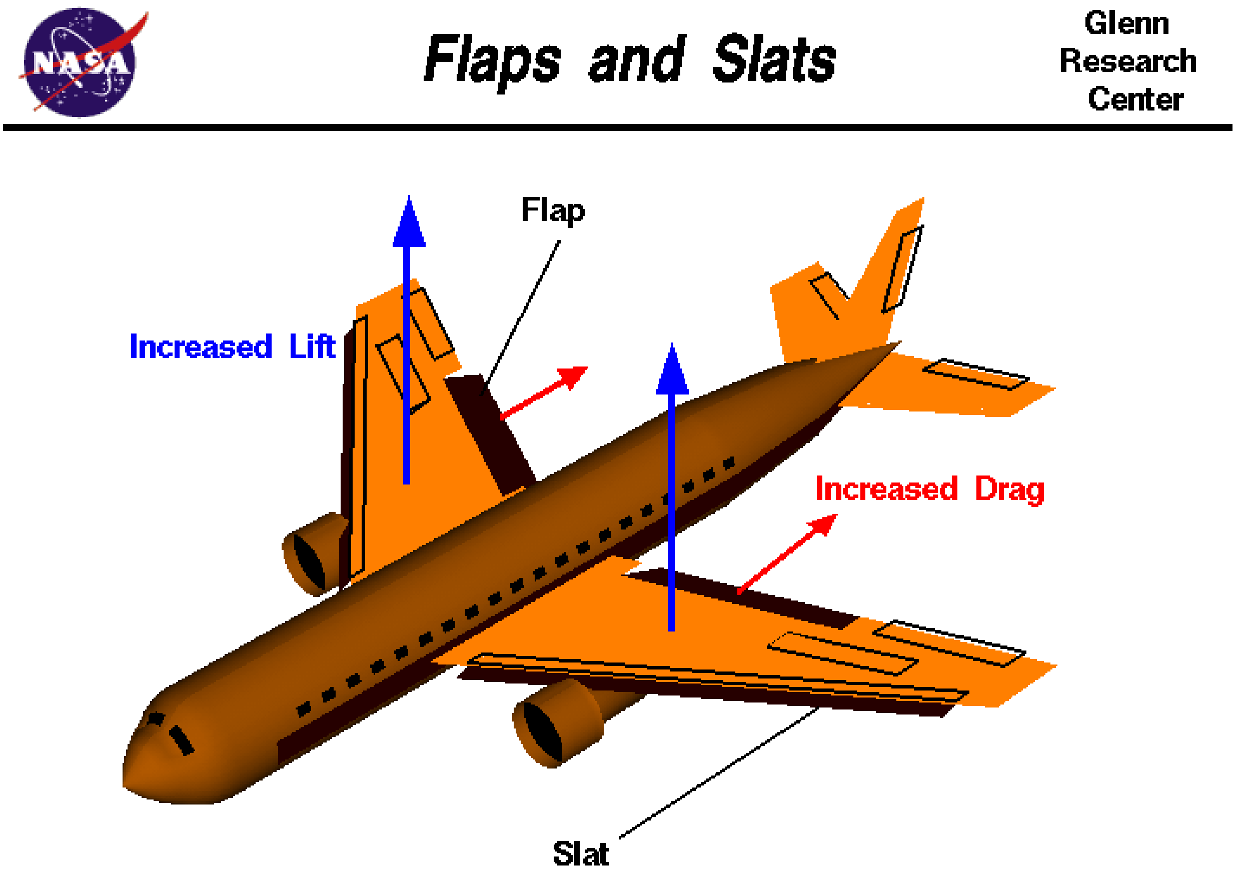


All that is necessary to create lift is to turn a flow of air. An aerodynamic, curved airfoil will turn a flow. But so will a simple flat plate, if it is inclined to the flow. The fuselage of an airplane will also generate lift if it is inclined to the flow. For that matter, an automobile body also turns the flow through which it moves, generating a lift force. Lift is a big problem for NASCAR racing machines and race cars now include spoilers on the roof to kill lift in a spin. Any physical body moving through a fluid can create lift if it produces a net turning of the flow.

There are many factors that affect the turning of the flow, which creates lift. We can group these factors into(a) those associated with the object, (b) those associated with the motion of the object through the air, and (c) those associated with the air itself:

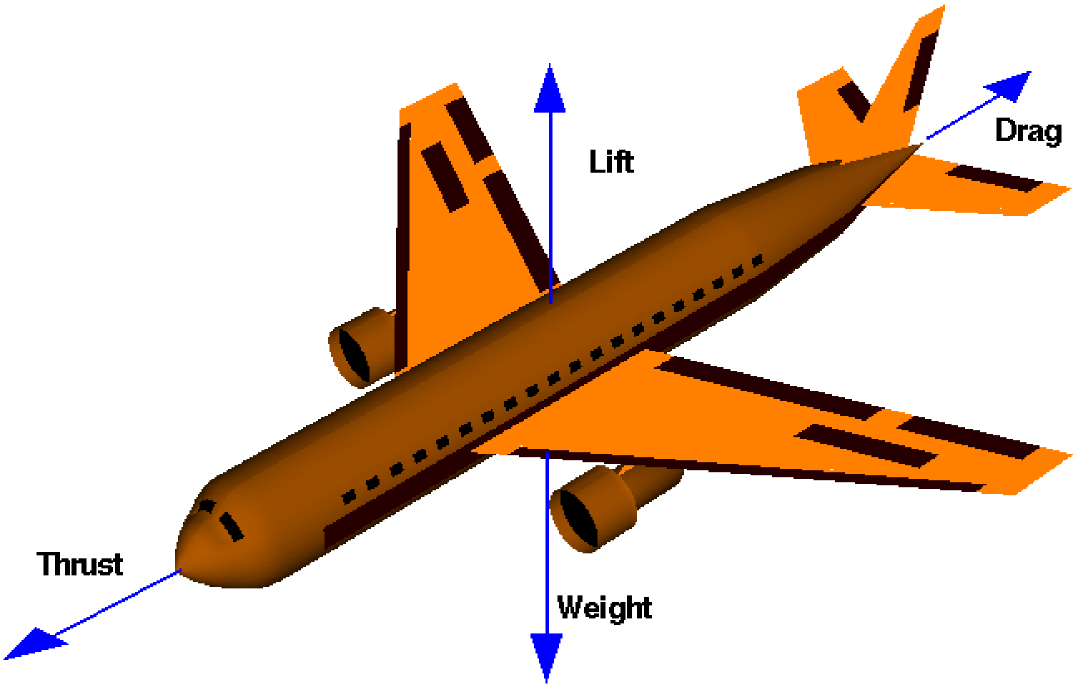
1. **Object:** At the top of the figure, aircraft wing geometry has a large effect on the amount of lift generated. The airfoil shape and wing size will both affect the amount of lift. The ratio of the wing span to the wing area also affects the amount of lift generated by a wing.
2. **Motion:** To generate lift, we have to move the object through the air. The lift then depends on the velocity of the air and how the object is inclined to the flow.
3. **Air:** Lift depends on the mass of the flow. The lift also depends in a complex way on two other properties of the air: its viscosity and its compressibility.

We can gather all of this information on the factors that affect lift into a single mathematical equation called the Lift Equation. With the lift equation we can predict how much lift force will be generated by a given body moving at a given speed.



The amount of lift generated by a wing depends on the shape of the airfoil, the wing area, and the aircraft velocity.

During takeoff and landing the airplane's velocity is relatively low. To keep the lift high (to avoid objects on the ground!), airplane designers try to increase the wing area and change the airfoil shape by putting some moving parts on the wings' leading and trailing edges. The part on the leading edge is called a **slat**, while the part on the trailing edge is called a **flap**. The flaps and slats move along metal tracks built into the wings. Moving the flaps **aft** (toward the tail) and the slats forward increases the wing area. Pivoting the leading edge of the slat and the trailing edge of the flap downward increases the effective camber of the airfoil, which increases the lift. In addition, the large aft-projected area of the flap increases the drag of the aircraft. This helps the airplane slow down for landing.



A force may be thought of as a push or pull in a specific direction. A force is a vector quantity so a force has both a magnitude and a direction. When describing forces, we have to specify both the magnitude and the direction. This slide shows the forces that act on an airplane in flight.

Weight

Weight is a force that is always directed toward the center of the earth. The magnitude of the weight depends on the mass of all the airplane parts, plus the amount of fuel, plus any payload on board (people, baggage, freight, etc.). The weight is distributed throughout the airplane. But we can often think of it as collected and acting through a single point called the center of gravity. In flight, the airplane rotates about the center of gravity.

Flying encompasses two major problems; overcoming the weight of an object by some opposing force, and controlling the object in flight. Both of these problems are related to the object's weight and the location of the center of gravity. During a flight, an airplane's weight constantly changes as the aircraft consumes fuel. The distribution of the weight and the center of gravity also changes. So the pilot must constantly adjust the controls to keep the airplane balanced, or trimmed.

Lift

To overcome the weight force, airplanes generate an opposing force called lift. Lift is generated by the motion of the airplane through the air and is an aerodynamic force. "**Aero**" stands for the air, and "**dynamic**" denotes motion. Lift is directed **perpendicular** to the flight direction. The magnitude of the lift depends on several factors including the shape, size, and velocity of the aircraft. As with weight, each part of the aircraft contributes to the aircraft lift force. Most of the lift is generated by the wings. Aircraft lift acts through a single point called the center of pressure. The center of pressure is defined just like the center of gravity, but using the pressure distribution around the body instead of the weight distribution.

The distribution of lift around the aircraft is important for solving the control problem. Aerodynamic surfaces are used to control the aircraft in roll, pitch, and yaw.

Drag

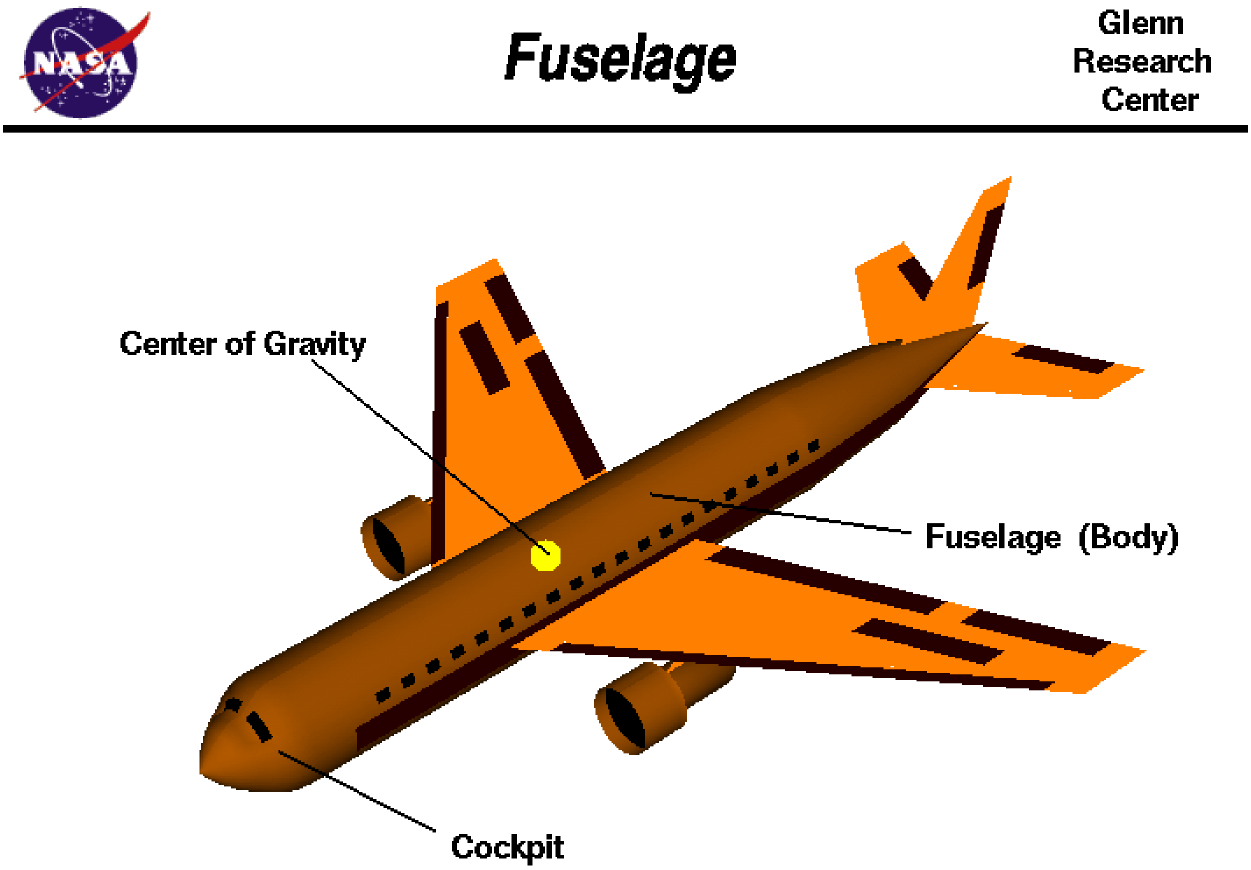
As the airplane moves through the air, there is another aerodynamic force present. The air resists the motion of the aircraft and the resistance force is called drag. Drag is directed **along and opposed** to the flight direction. Like lift, there are many factors that affect the magnitude of the drag force including the shape of the aircraft, the "stickiness" of the air, and the velocity of the aircraft. Like lift, we collect all of the individual components' drags and combine them into a single aircraft drag magnitude. And like lift, drag acts through the aircraft center of pressure.

Thrust

To overcome drag, airplanes use a propulsion system to generate a force called thrust. The direction of the thrust force depends on how the engines are attached to the aircraft. In the figure shown above, two turbine engines are located under the wings, parallel to the body, with thrust acting along the body centerline. On some aircraft, such as the Harrier, the thrust direction can be varied to help the airplane take off in a very short distance. The magnitude of the thrust depends on many factors associated with the propulsion system including the type of engine, the number of engines, and the throttle setting.

For jet engines, it is often confusing to remember that aircraft thrust is a reaction to the hot gas rushing out of the nozzle. The hot gas goes out the back, but the thrust pushes towards the front. Action <--> reaction is explained by Newton's Third Law of Motion.

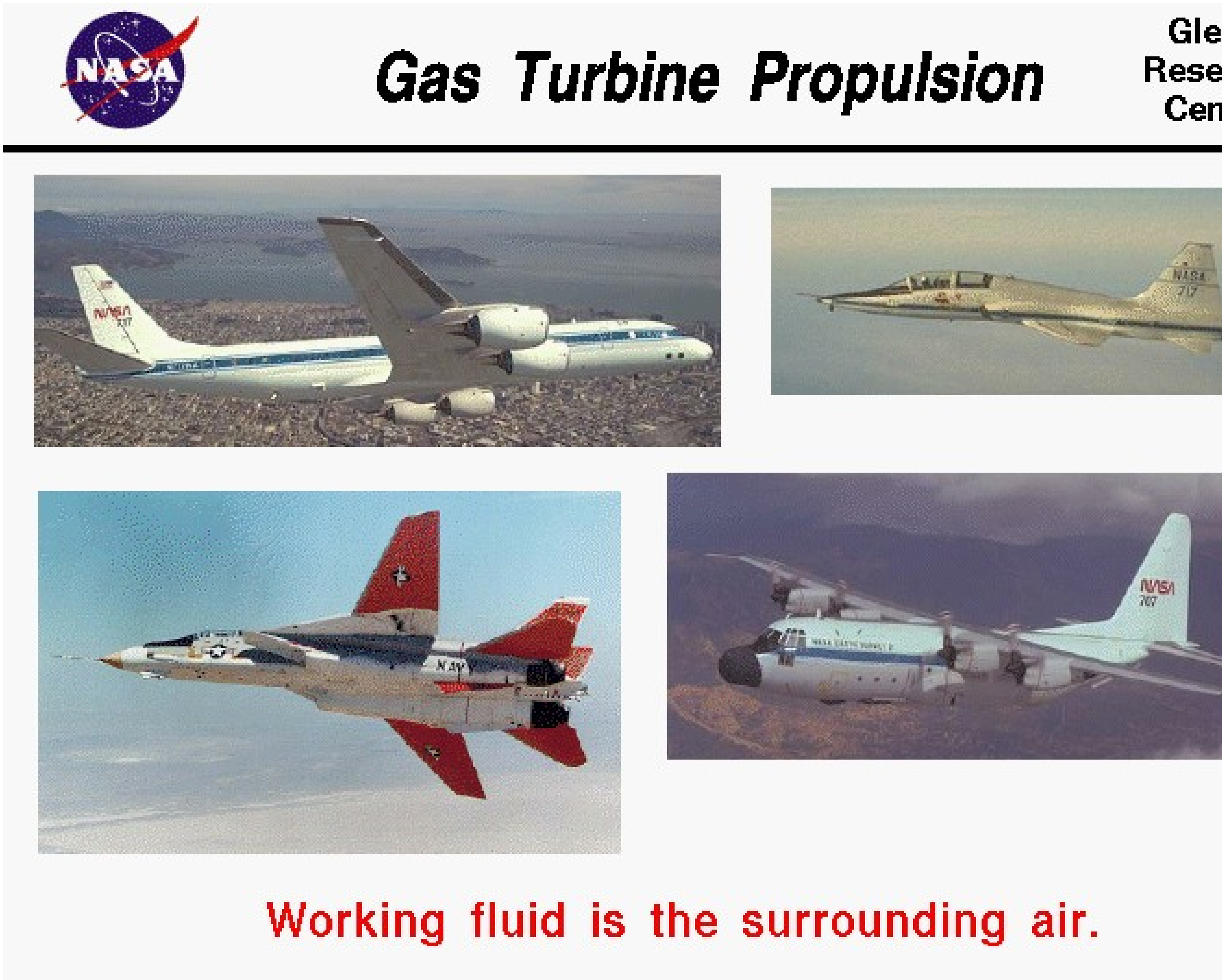
The motion of the airplane through the air depends on the relative strength and direction of the forces shown above. If the forces are balanced, the aircraft cruises at constant velocity. If the forces are unbalanced, the aircraft accelerates in the direction of the largest force.



Airplanes are transportation devices which are designed to move people and cargo from one place to another. Airplanes come in many different shapes and sizes depending on the mission of the aircraft. The airplane shown on this slide is a turbine-powered airliner which has been chosen as a representative aircraft.

The **fuselage**, or body of the airplane, is a long hollow tube which holds all the pieces of an airplane together. The fuselage is hollow to reduce weight. As with most other parts of the airplane, the shape of the fuselage is normally determined by the mission of the aircraft. A supersonic fighter plane has a very slender, streamlined fuselage to reduce the drag associated with high speed flight. An airliner has a wider fuselage to carry the maximum number of passengers. On an airliner, the pilots sit in a **cockpit** at the front of the fuselage. Passengers and cargo are carried in the rear of the fuselage and the fuel is usually stored in the wings. For a fighter plane, the cockpit is normally on top of the fuselage, weapons are carried on the wings, and the engines and fuel are placed at the rear of the fuselage.

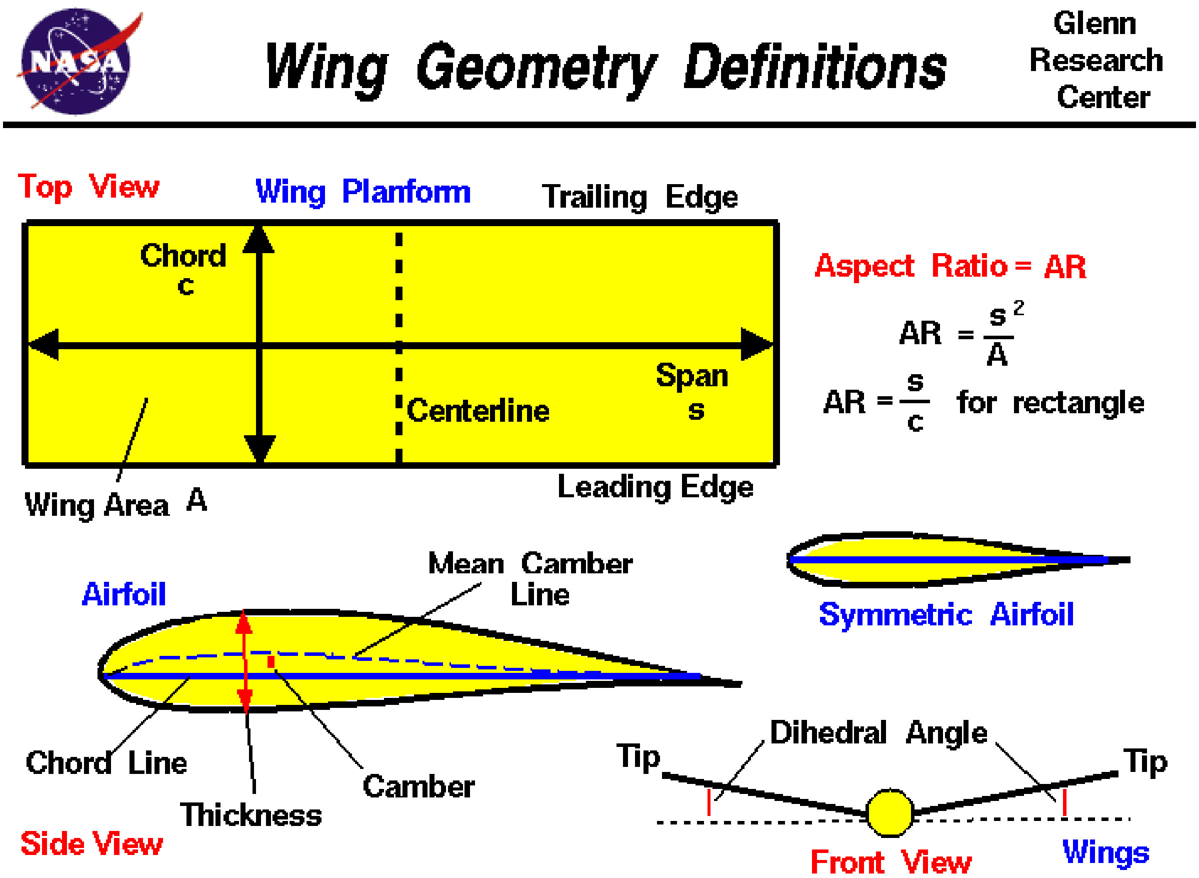
The weight of an aircraft is distributed all along the aircraft. The fuselage, along with the passengers and cargo, contribute a significant portion of the weight of an aircraft. The center of gravity of the aircraft is the average location of the weight and it is usually located inside the fuselage. In flight, the aircraft rotates around the center of gravity because of torques generated by the elevator, rudder, and ailerons. The fuselage must be designed with enough strength to withstand these torques.



Thrust is the force which moves any aircraft through the air. Thrust is generated by the propulsion system of the aircraft. Different propulsion systems develop thrust in different ways, but all thrust is generated through some application of Newton's third law of motion. For every action there is an equal and opposite reaction. In any propulsion system, a **working fluid** is accelerated by the system and the reaction to this acceleration produces a force on the system. A general derivation of the thrust equation shows that the amount of thrust generated depends on the mass flow through the engine and the exit velocity of the gas.

During World War II, a new type of airplane engine was developed independently in Germany and in England. This engine was called a **gas turbine** engine. We sometimes call this engine a **jet engine**. Early gas turbine engines worked much like a rocket engine creating a hot exhaust gas which was passed through a nozzle to produce thrust. But unlike the rocket engine which must carry its oxygen for combustion, the turbine engine gets its oxygen from the surrounding air. A turbine engine does not work in outer space because there is no surrounding air. For a gas turbine engine, the accelerated gas, or **working fluid**, is the jet exhaust. Most of the mass of the jet exhaust comes from the surrounding atmosphere. Most modern, high speed passenger and military aircraft are powered by gas turbine engines. Because gas turbine engines are so important for modern life, we will be providing a lot of information about turbine engines and their operation.

Turbine engines come in a wide variety of shapes and sizes because of the many different aircraft missions. All gas turbine engines have some parts in common, however. On the slide we see pictures of four different aircraft equipped with gas turbine engines. Each aircraft has a unique mission and therefore a unique propulsion requirement. At the upper left is a DC-8 airliner. Its mission is to carry large loads of passengers or cargo for a long distance at high speed. It spends most of its life in high speed cruise. At the lower left is an F-14 fighter plane. Its mission is to shoot down other aircraft in air-to-air combat. It spends most of its life in cruise, but needs high acceleration when in combat. At the lower right is a C-130 cargo aircraft. Like the DC-8, it carries cargo a long distance, but it does not have the high speed requirement of the DC-8. At the upper right is a T-38 trainer. It is used to teach pilots how to fly jet aircraft and does not have the acceleration requirements of the F-14. The DC-8 is powered by four high-bypass turbofan engines, the F-14 by two afterburning low-bypass turbofans, the C-130 by four turboprop engines, and the T-38 by two turbojet engines.



This slide gives technical definitions of a wing's geometry, which is one of the chief **factors** affecting airplane lift and drag. The terminology is used throughout the airplane industry and is also found in the FoilSim interactive airfoil simulation program developed here at NASA Glenn. Actual aircraft wings are complex three-dimensional objects, but we will start with some simple definitions. The figure shows the wing viewed from three directions; the upper left shows the view from the top looking down on the wing, the lower right shows the view from the front looking at the wing leading edge, and the lower left shows a side view from the left looking in towards the centerline. The side view shows an airfoil shape with the leading edge to the left.

Top View

The top view shows a simple wing geometry, like that found on a light general aviation aircraft. The front of the wing (at the bottom) is called the **leading edge**; the back of the wing (at the top) is called the **trailing edge**. The distance from the leading edge to the trailing edge is called the **chord**, denoted by the symbol **c**. The ends of the wing are called the **wing tips**, and the distance from one wing tip to the other is called the **span**, given the symbol **s**. The shape of the wing, when viewed from above looking down onto the wing, is called a **planform**. In this figure, the planform is a rectangle. For a rectangular wing, the chord length at every location along the span is the same. For most other planforms, the chord length varies along the span. The **wing area, A,** is the projected area of the planform and is bounded by the leading and trailing edges and the wing tips. **Note:** *The wing area is NOT the total surface area of the wing. The* ***total surface area*** *includes both upper and lower surfaces. The wing area is a projected area and is almost half of the total surface area.*

Aspect ratio is a measure of how long and slender a wing is from tip to tip. The **Aspect Ratio** of a wing is defined to be the square of the span divided by the wing area and is given the symbol **AR**. For a rectangular wing, this reduces to the ratio of the span to the chord length as shown at the upper right of the figure.

$AR = s^2 / A = s^2 / (s * c) = s / c$

High aspect ratio wings have long spans (like high performance gliders), while low aspect ratio wings have either short spans (like the F-16 fighter) or thick chords (like the Space Shuttle). There is a component of the drag of an aircraft called induced drag which depends inversely on the aspect ratio. A higher aspect ratio wing has a lower drag and a slightly higher lift than a lower aspect ratio wing. Because the glide angle of a glider depends on the ratio of the lift to the drag, a glider is usually designed with a very high aspect ratio. The Space Shuttle has a low aspect ratio because of high speed effects, and therefore is a very poor glider. The F-14 and F-111 have the best of both worlds. They can change the aspect ratio in flight by pivoting the wings--large span for low speed, small span for high speed.

Front View

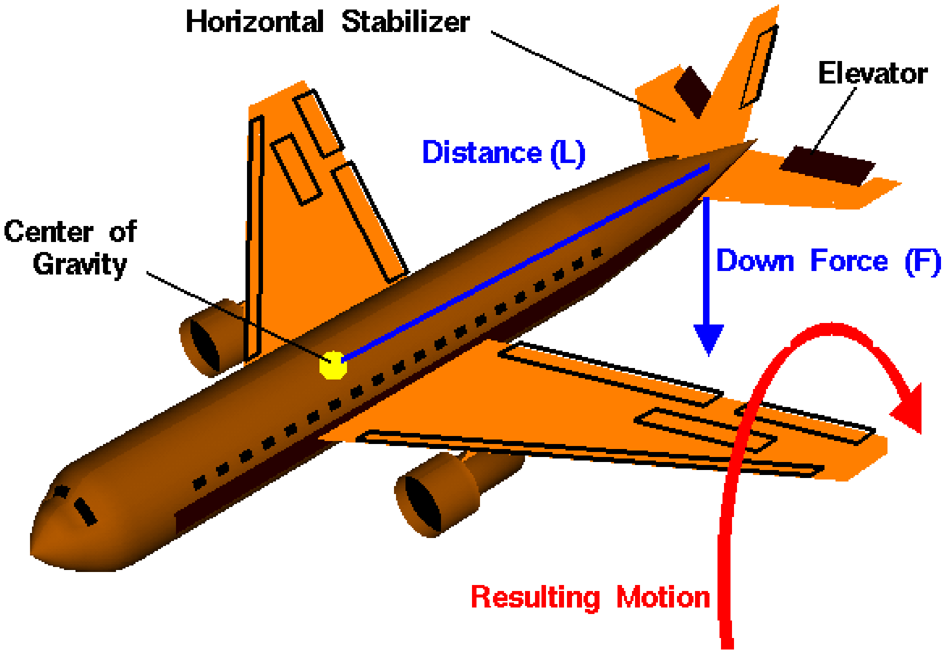
The front view of this wing shows that the left and right wing do not lie in the same plane but meet at an angle. The angle that the wing makes with the local horizontal is called the **dihedral angle**. Dihedral is added to the wings for roll stability; a wing with some dihedral will naturally return to its original position if it encounters a slight roll displacement. You may have noticed that most large airliner wings are designed with diherdral. The wing tips are farther off the ground than the wing root. Highly maneuverable fighter planes, on the other hand do not have dihedral. In fact, some fighter aircraft have the wing tips lower than the roots giving the aircraft a high roll rate. A negative dihedral angle is called **anhedral** . **Historical Note:** *The Wright brothers designed their 1903 flyer with a slight anhedral to enhance the aircraft roll performance.*

Side View

A cut through the wing perpendicular to the leading and trailing edges will show the cross-section of the wing. This side view is called an **airfoil**, and it has some geometry definitions of its own as shown at the lower left. The straight line drawn from the leading to trailing edges of the airfoil is called the **chord line**. The chord line cuts the airfoil into an upper surface and a lower surface. If we plot the points that lie halfway between the upper and lower surfaces, we obtain a curve called the **mean camber line**. For a **symmetric airfoil** (upper surface the same shape as the lower surface) the mean camber line will fall on top of the chord line. But in most cases, these are two separate lines. The maximum distance between the two lines is called the **camber**, which is a measure of the curvature of the airfoil (high camber means high curvature). The maximum distance between the upper and lower surfaces is called the **thickness**. Often you will see these values divided by the chord length to produce a non-dimensional or "percent" type of number. Airfoils can come with all kinds of combinations of camber and thickness distributions. **NACA** (the precursor of NASA) established a method of designating classes of airfoils and then wind tunnel tested the airfoils to provide lift coefficients and drag coefficients for designers.

Horizontal Stabilizer – Elevator

Glenn Research Center



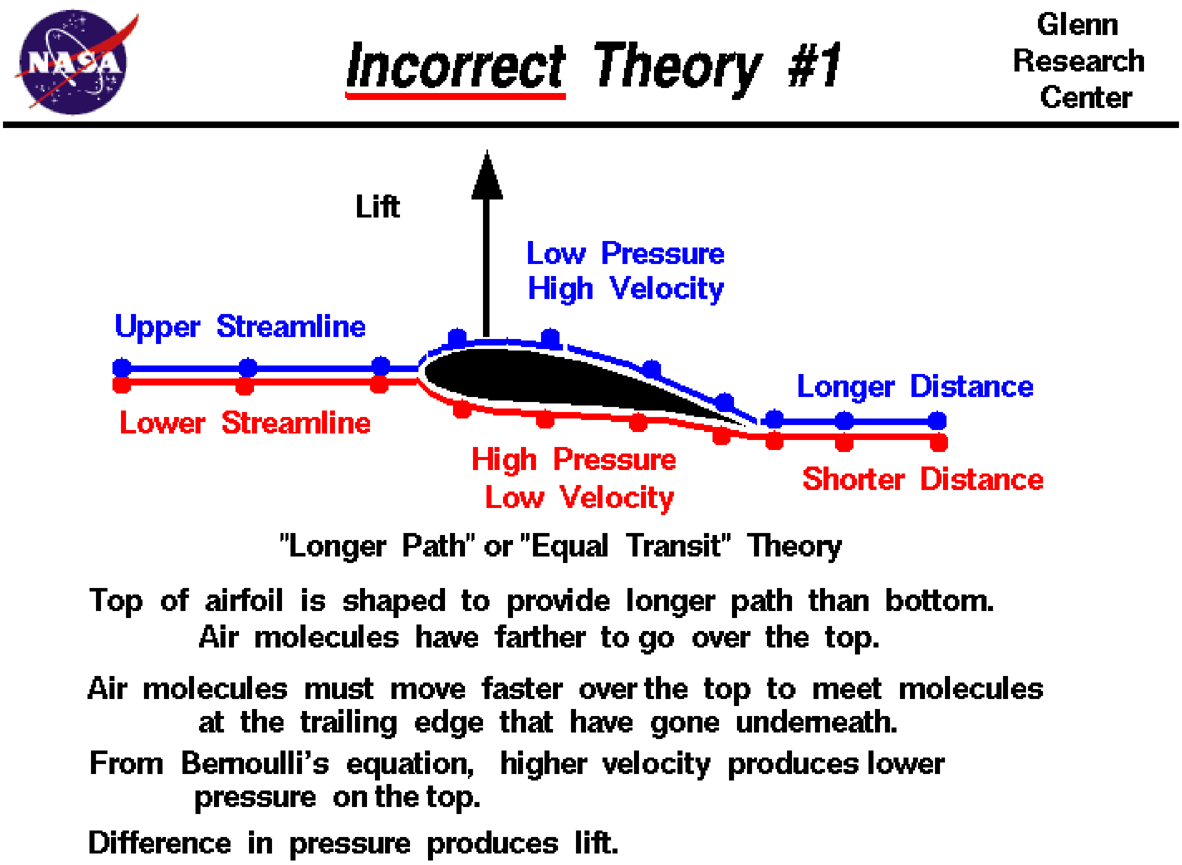
At the rear of the fuselage of most aircraft one finds a **horizontal stabilizer** and an **elevator**. The stabilizer is a fixed wing section whose job is to provide stability for the aircraft, to keep it flying straight. The horizontal stabilizer prevents up-and-down, or pitching, motion of the aircraft nose. The elevator is the small moving section at the rear of the stabilizer that is attached to the fixed sections by hinges. Because the elevator moves, it varies the amount of force generated by the tail surface and is used to generate and control the pitching motion of the aircraft. There is an elevator attached to each side of the fuselage. The elevators work in pairs; when the right elevator goes up, the left elevator also goes up. This slide shows what happens when the pilot deflects the elevator.

The elevator is used to control the position of the nose of the aircraft and the angle of attack of the wing. Changing the inclination of the wing to the local flight path changes the amount of lift which the wing generates. This, in turn, causes the aircraft to climb or dive. During take off the elevators are used to bring the nose of the aircraft up to begin the climb out. During a banked turn, elevator inputs can increase the lift and cause a tighter turn. That is why elevator performance is so important for fighter aircraft.

The elevators work by changing the effective shape of the airfoil of the horizontal stabilizer. As described on the shape effects slide, changing the angle of deflection at the rear of an airfoil changes the amount of lift generated by the foil. With greater downward deflection of the trailing edge, lift increases. With greater upward deflection of the trailing edge, lift decreases and can even become negative as shown on this slide. The lift force (F) is applied at center of pressure of the horizontal stabilzer which is some distance (L) from the aircraft center of gravity. This creates a torque

$T = F * L$

on the aircraft and the aircraft rotates about its center of gravity. The pilot can use this ability to make the airplane loop. Or, since many aircraft loop naturally, the deflection can be used to trim or balance the aircraft, thus preventing a loop. If the pilot reverses the elevator deflection to down, the aircraft pitches in the opposite direction.

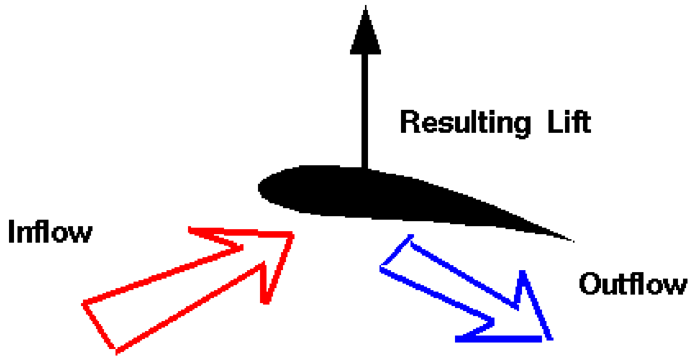


There are many theories of how lift is generated. Unfortunately, many of the theories found in encyclopedias, on web sites, and even in some textbooks are **incorrect**, causing unnecessary confusion for students.

The theory described on this slide is one of the most widely circulated, incorrect explanations. The theory can be labeled the "Longer Path" theory, or the "Equal Transit Time" theory. The theory states that airfoils are shaped with the upper surface longer than the bottom. The air molecules (the little colored balls on the figure) have farther to travel over the top of the airfoil than along the bottom. In order to meet up at the trailing edge, the molecules going over the top of the wing must travel faster than the molecules moving under the wing. Because the upper flow is faster, then, from Bernoulli's equation, the pressure is lower. The difference in pressure across the airfoil produces the lift.

Incorrect Theory #2

Glenn Research Center

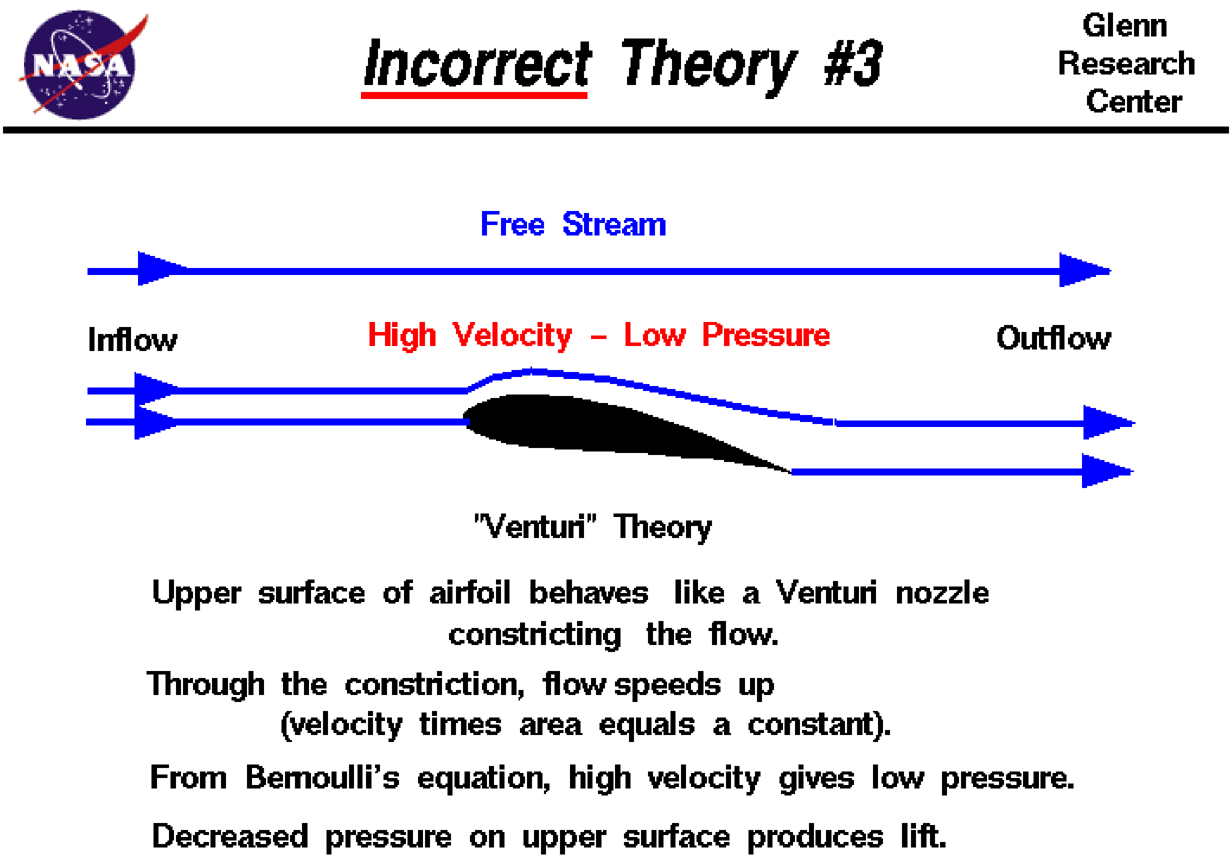


"Skipping Stone" Theory

Lift is the result of simple action <--> reaction as air molecules strike bottom of the airfoil imparting momentum to the foil.

There are many theories of how lift is generated. Unfortunately, many of the theories found in encyclopedias, on web sites, and even in some textbooks are **incorrect**, causing unnecessary confusion for students.

The theory described on this slide is often seen on web sites and in popular literature. The theory is based on the idea that lift is the reaction force to air molecules striking the bottom of the airfoil as it moves through the air. Because this is similar to the way in which a flat rock thrown at a shallow angle skips across a body of water, it is called the "Skipping Stone" theory of lift. *It is sometimes called a Newtonian theory of lift, since it involves Newton's third law, but to avoid confusion with the* ***correct*** *Newtonian theory of flow turning, we shall call it the "Skipping Stone" theory.*



There are many theories of how lift is generated. Unfortunately, many of the theories found in encyclopedias, on web sites, and even in some textbooks are **incorrect**, causing unnecessary confusion for students.

The theory described on this slide is often seen on web sites and in popular literature. The theory is based on the idea that the airfoil upper surface is shaped to act as a nozzle which accelerates the flow. Such a nozzle configuration is called a **Venturi nozzle** and it can be analyzed classically. Considering the conservation of mass, the mass flowing past any point in the nozzle is a constant; the mass flow rate of a Venturi nozzle is a constant. The mass flow rate **m dot** is equal to the density **r** times the velocity **V** times the flow area **A**:

m dot = r * V * A = constant

For a constant density, decreasing the area increases the velocity.

Turning to the **incorrect** airfoil theory, the top of the airfoil is curved, which constricts the flow. Since the area is decreased, the velocity over the top of the foil is increased. Then from Bernoulli's equation, higher velocity produces a lower pressure on the upper surface. The low pressure over the upper surface of the airfoil produces the lift.

Lift from Flow Turning

Glenn Research Center

Lift is a force.

Force = mass X acceleration

$F = m\,a$

Force = mass X change in velocity with time

$$F = \frac{m\,(V_1 - V_0)}{(t_1 - t_0)}$$

Velocity has both magnitude (speed) and direction.

Changing either the speed or direction of a flow generates a force.

Lift is a force generated by turning a moving fluid.

Lift can be generated by a wide variety of objects, including airplane wings, rotating cylinders, spinning balls, and flat plates. Lift is the force that holds an aircraft in the air. Lift can be generated by any part of the airplane, but most of the lift on a normal airliner is generated by the wings. How is lift generated?

Force = Mass x Acceleration

Lift is a force. From Newton's second law of motion, a force **F** is produced when a mass **m** is accelerated **a**:

F = m * a

An **acceleration** is a change in velocity **V** with a change in time **t**.

F = m * (V1 - V0) / (t1 - t0)

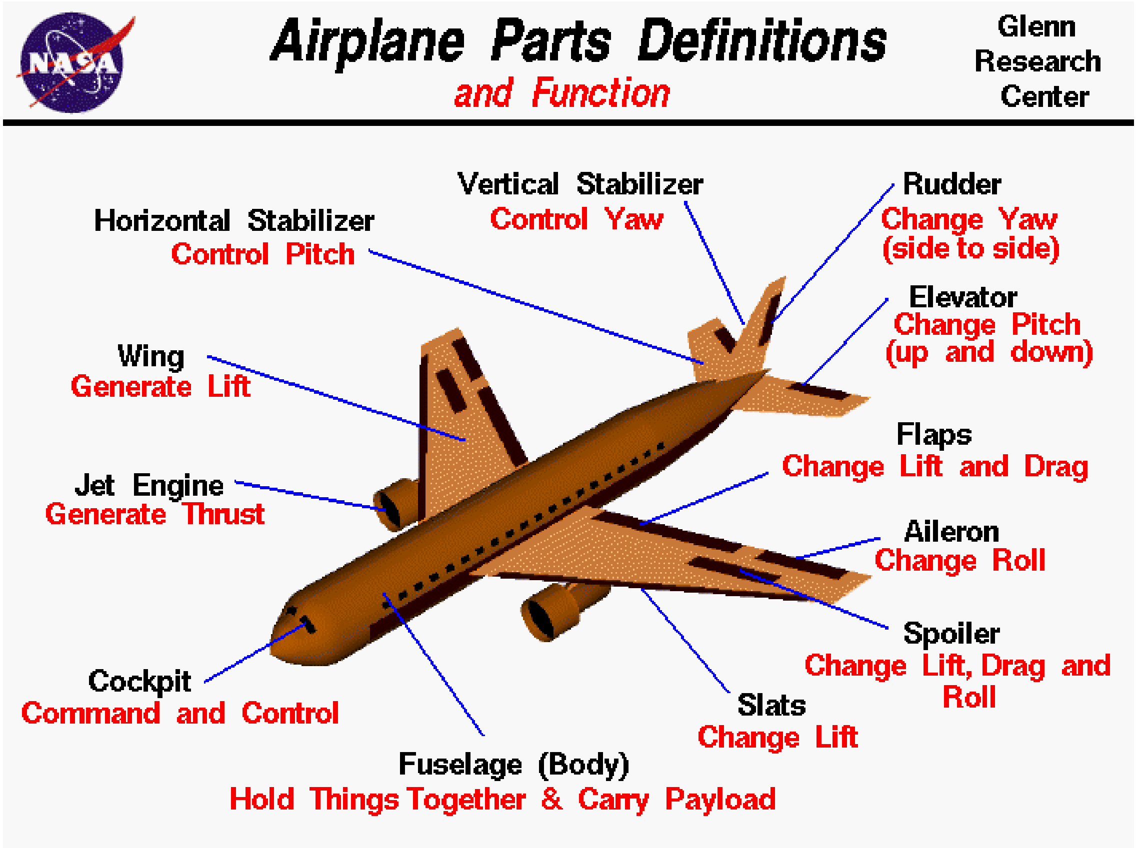
We have written this relationship as a difference equation, but it is recognized that the relation is actually a differential from calculus.

F = m * dV/dt

The important fact is that a force causes a change in velocity; and, likewise, a change in velocity generates a force. The equation works both ways. A velocity has both a **magnitude** called the speed and a direction associated with it. Scientists and mathematicians call this a vector quantity. So, to change either the speed or the direction of a flow, you must impose a force. And if either the speed or the direction of a flow is changed, a force is generated.

Lift Generated in a Moving Fluid

For a body immersed in a moving fluid, the fluid remains in contact with the surface of the body. If the body is shaped, moved, or inclined in such a way as to produce a net deflection or turning of the flow, the local velocity is changed in magnitude, direction, or both. Changing the velocity creates a net force on the body. It is very important to note that the turning of the fluid occurs because the molecules of the fluid stay in contact with the solid body since the molecules are free to move. Any part of the solid body can deflect a flow. Parts facing the oncoming flow are said to be **windward**, and parts facing away from the flow are said to be **leeward**. Both windward and leeward parts deflect a flow. Ignoring the leeward deflection leads to a popular incorrect theory of lift.



This page shows the parts of an airplane and their functions. Airplanes are transportation devices which are designed to move people and cargo from one place to another. Airplanes come in many different shapes and sizes depending on the mission of the aircraft. The airplane shown on this slide is a turbine-powered airliner which has been chosen as a representative aircraft.

For any airplane to fly, you must lift the weight of the airplane itself, the fuel, the passengers, and the cargo. The wings generate most of the lift to hold the plane in the air. To generate lift, the airplane must be pushed through the air. The jet engines, which are located beneath the wings, provide the thrust to push the airplane forward through the air. The air resists the motion in the form of aerodynamic drag. Some airplanes use propellers for the propulsion system instead of jets.

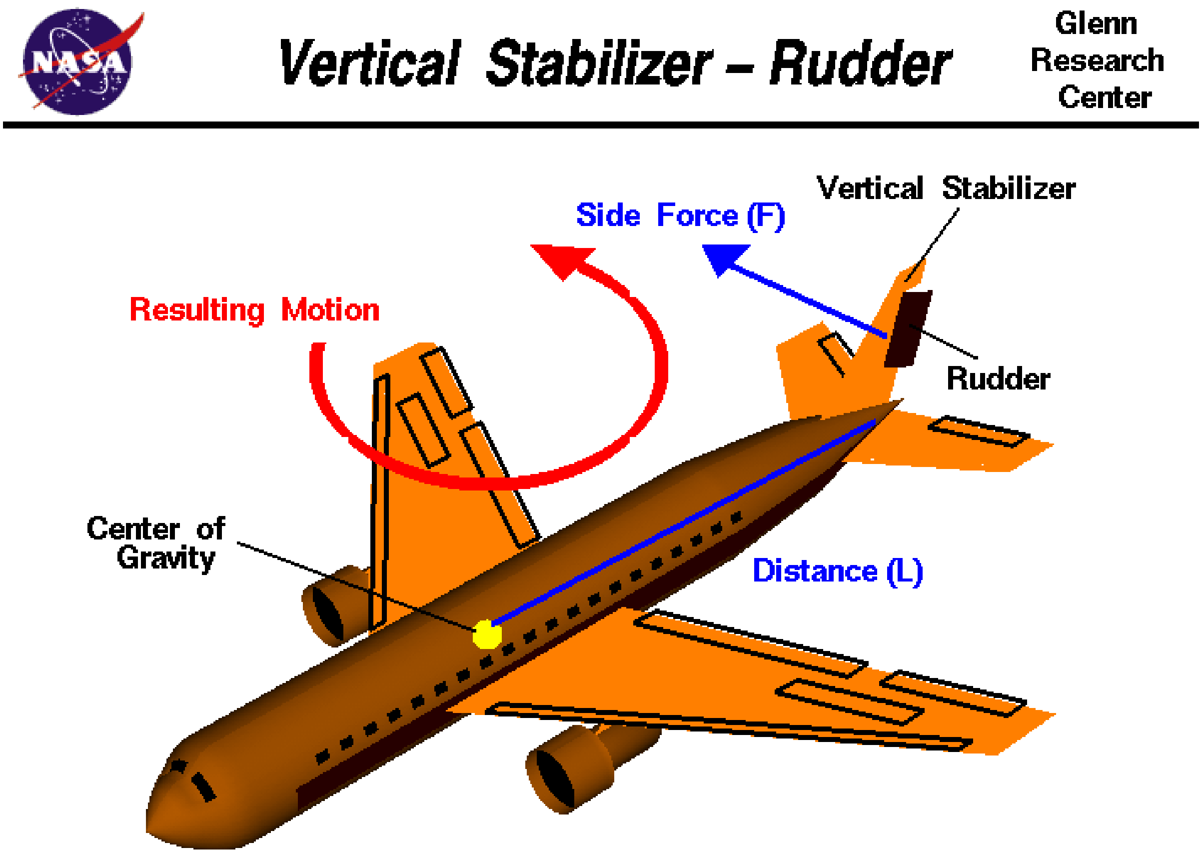
To control and maneuver the aircraft, smaller wings are located at the tail of the plane. The tail usually has a fixed horizontal piece (called the horizontal stabilizer) and a fixed vertical piece (called the vertical stabilizer). The stabilizers' job is to provide stability for the aircraft, to keep it flying straight. The **vertical stabilizer** keeps the nose of the plane from swinging from side to side, while the **horizontal stabilizer** prevents an up-and-down motion of the nose. (On the Wright brother's first aircraft, the horizontal stabilizer was placed in front of the wings. Such a configuration is called a **canard** after the French word for "duck").

At the rear of the wings and stabilizers are small moving sections that are attached to the fixed sections by hinges. In the figure, these moving sections are colored brown. Changing the rear portion of a wing will change the amount of force that the wing produces. The ability to change forces gives us a means of controlling and maneuvering the airplane. The hinged part of the vertical stabilizer is called the rudder; it is used to deflect the tail to the left and right as viewed from the front of the fuselage. The hinged part of the horizontal stabilizer is called the elevator; it is used to deflect the tail up and down. The outboard hinged part of the wing is called the aileron; it is used to roll the wings from side to side. Most airliners can also be rolled from side to side by using the spoilers. Spoilers are small plates that are used to disrupt the flow over the wing and to change the amount of force by decreasing the lift when the spoiler is deployed.

The wings have additional hinged, rear sections near the body that are called flaps. Flaps are deployed downward on takeoff and landing to increase the amount of force produced by the wing. On some aircraft, the front part of the wing will also deflect. **Slats** are used at takeoff and landing to produce additional force. The spoilers are also used during landing to slow the plane down and to counteract the flaps when the aircraft is on the ground. The next time you fly on an airplane, notice how the wing shape changes during takeoff and landing.

The fuselage or body of the airplane, holds all the pieces together. The pilots sit in the **cockpit** at the front of the fuselage. Passengers and cargo are carried in the rear of the fuselage. Some aircraft carry fuel in the fuselage; others carry the fuel in the wings.

As mentioned above, the aircraft configuration in the figure was chosen only as an example. Individual aircraft may be configured quite differently from this airliner. The Wright Brothers 1903 Flyer had pusher propellers and the elevators at the front of the aircraft. Fighter aircraft often have the jet engines buried inside the fuselage instead of in pods hung beneath the wings. Many fighter aircraft also combine the horizontal stabilizer and elevator into a single stabilator surface. There are many possible aircraft configurations, but any configuration must provide for the four forces needed for flight.



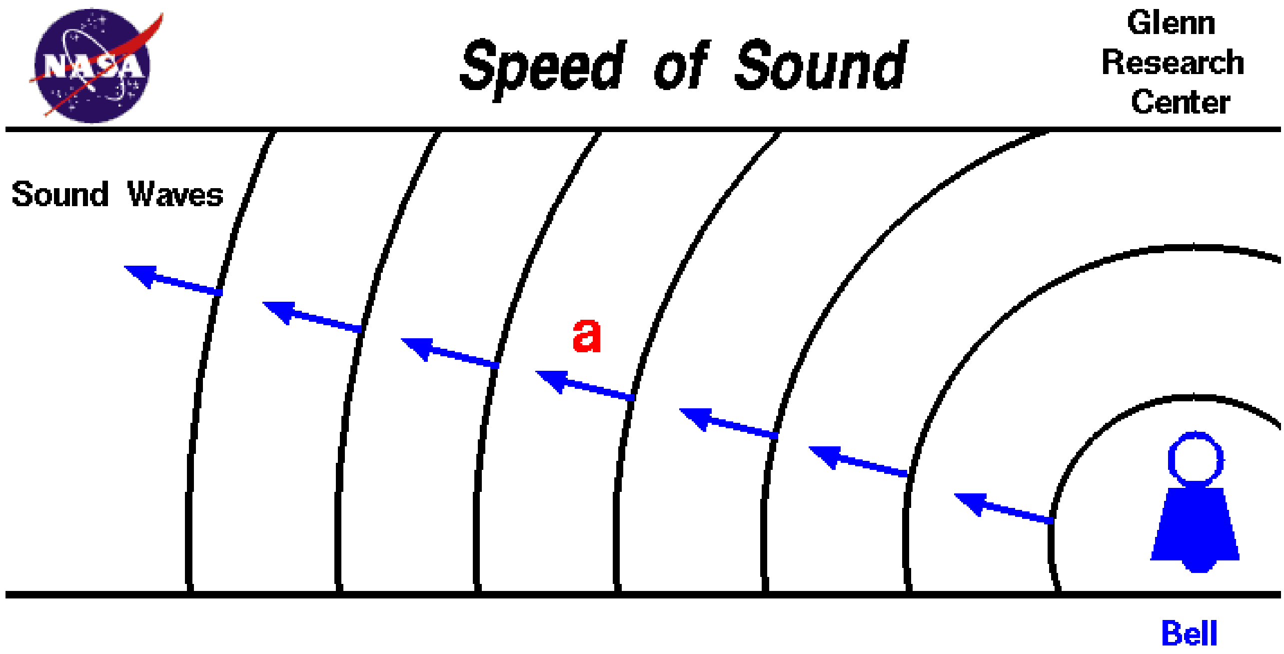
At the rear of the fuselage of most aircraft one finds a **vertical stabilizer** and a **rudder**. The stabilizer is a fixed wing section whose job is to provide stability for the aircraft, to keep it flying straight. The vertical stabilizer prevents side-to-side, or yawing, motion of the aircraft nose. The rudder is the small moving section at the rear of the stabilizer that is attached to the fixed sections by hinges. Because the rudder moves, it varies the amount of force generated by the tail surface and is used to generate and control the yawing motion of the aircraft. This slide shows what happens when the pilot deflects the **rudder**, a hinged section at the rear of the vertical stabilizer.

The rudder is used to control the position of the nose of the aircraft. Interestingly, it is NOT used to turn the aircraft in flight. Aircraft turns are caused by banking the aircraft to one side using either ailerons or spoilers. The banking creates an unbalanced side force component of the large wing lift force which causes the aircraft's flight path to curve. The rudder input insures that the aircraft is properly aligned to the curved flight path during the maneuver. Otherwise, the aircraft would encounter additional drag or even a possible **adverse yaw** condition in which, due to increased drag from the control surfaces, the nose would move farther off the flight path.

The rudder works by changing the effective shape of the airfoil of the vertical stabilizer. As described on the shape effects slide, changing the angle of deflection at the rear of an airfoil will change the amount of lift generated by the foil. With increased deflection, the lift will increase in the opposite direction. The rudder and vertical stabilizer are mounted so that they will produce forces from side to side, not up and down. The side force (F) is applied through the center of pressure of the vertical stabilizer which is some distance (L) from the aircraft center of gravity. This creates a torque

$$T = F * L$$

on the aircraft and the aircraft rotates about its center of gravity. With greater rudder deflection to the left as viewed from the back of the aircraft, the force increases to the right. If the pilot reverses the rudder deflection to the right, the aircraft will yaw in the opposite direction. We have chosen to base the deflections on a view from the back of the aircraft towards the nose, because that is the direction in which the pilot is looking



Speed of sound (a) depends on the type of medium and the temperature of the medium.

$$a = \text{sqrt}(\gamma R T)$$

γ = ratio of specific heats (1.4 for air at STP)

R = gas constant (286 $m^2/s^2/K$ for air)

T = absolute temperature (273.15 + °C)

Air is a gas, and a very important property of any gas is the **speed of sound** through the gas. Why are we interested in the speed of sound? The **speed of "sound"** is actually the speed of transmission of a small disturbance through a medium. **Sound** itself is a sensation created in the human brain in response to sensory inputs from the inner ear. (We won't comment on the old "tree falling in a forest" discussion!)

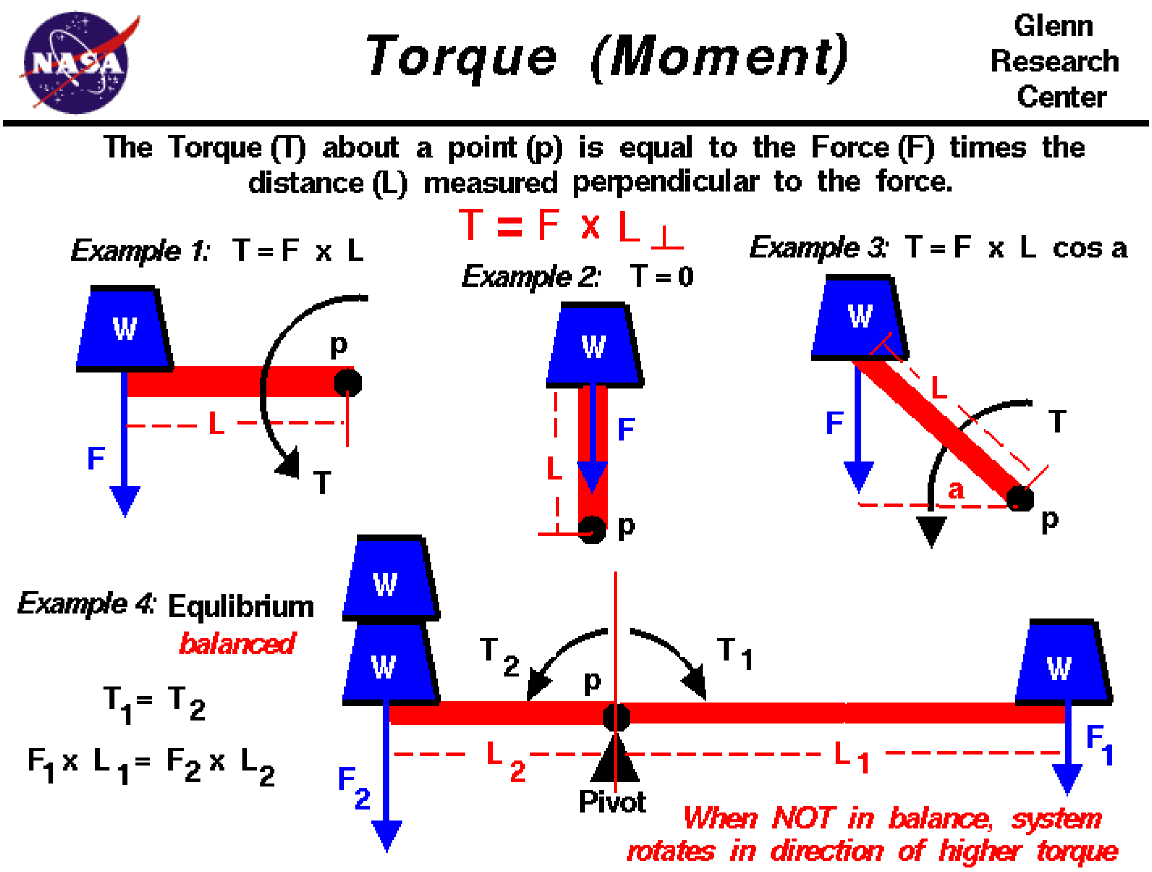
Disturbances are transmitted through a gas as a result of collisions between the randomly moving molecules in the gas. The transmission of a small disturbance through a gas is an isentropic process. The conditions in the gas are the same before and after the disturbance passes through. Because the speed of transmission depends on molecular collisions, the speed of sound depends on the state of the gas. The speed of sound is a constant within a given gas and the value of the constant depends on the type of gas (air, pure oxygen, carbon dioxide, etc.) and the temperature of the gas. An analysis based on conservation of mass and momentum shows that the speed of sound **a** is equal to the square root of the ratio of specific heats **g** times the gas constant **R** times the temperature **T**.

a = sqrt [g * R * T]

*Notice that the temperature must be specified on an absolute scale (Kelvin or Rankine). The dependence on the type of gas is included in the gas constant **R**. which equals the universal gas constant divided by the molecular weight of the gas, and the ratio of specific heats.*

The speed of sound in air depends on the type of gas and the temperature of the gas. On Earth, the atmosphere is composed of mostly diatomic nitrogen and oxygen, and the temperature depends on the altitude in a rather complex way. Scientists and engineers have created a mathematical model of the atmosphere to help them account for the changing effects of temperature with altitude. Mars also has an atmosphere composed of mostly carbon dioxide.

There is a similar mathematical model of the Martian atmosphere. We have created an atmospheric calculator to let you study the variation of sound speed with planet and altitude.



A force may be thought of as a push or pull in a specific direction. When a force is applied to an object, the resulting motion of the object depends on where the force is applied and how the object is confined. If the object is unconfined and the force is applied through the center of gravity, the object moves in pure translation, as described by Newton's laws of motion. If the object is confined (or pinned) at some location called a **pivot**, the object rotates about the pivot, but does not translate. The force is transmitted through the pivot and the details of the rotation depend on the distance from the applied force to the pivot. If the object is unconfined and the force is applied at some distance from the center of gravity, the object both translates and rotates about the center of gravity. The details of the rotation depend on the distance from the applied force to the center of gravity. The motion of flying objects is described by this third type of motion; a combination of translation and rotation.

A force **F** is a vector quantity, which means that it has both a magnitude and a direction associated with it. The direction of the force is important because the resulting motion of the object is in the same direction as the force. The product of the force and the **perpendicular distance** to the center of gravity for an unconfined object, or to the pivot for a confined object, is^M called the **torque** or the **moment**. A torque is also a vector quantity and produces a rotation in the same way that a force produces a translation. Namely, an object at rest, or rotating at a constant angular velocity, will continue to do so until it is subject to an external torque. A torque produces a change in angular velocity which is called an angular acceleration.

The distance **L** used to determine the torque **T** is the distance from the pivot **p** to the force, but measured perpendicular to the direction of the force. On the figure, we show four examples of torques to illustrate the basic principles governing torques. In each example a blue weight **W** is acting on a red bar, which is called an arm.

In Example 1, the force (weight) is applied perpendicular to the arm. In this case, the perpendicular distance is the length of the bar and the torque is equal to the product of the length and the force.

T = F * L

In Example 2, the same force is applied to the arm, but the force now acts right through the pivot. In this case, the distance from the pivot perpendicular to the force is zero. So, in this case, the torque is also zero. Think of a hinged door. If you push on the edge of the door, towards the hinge, the door doesn't move because the torque is zero.

Example 3 is the general case in which the force is applied at some angle **a** to the arm. The perpendicular distance is given by trigonometry as the length of the arm (L) times the cosine (cos) of the angle. The torque is then given by:

T = F * L * cos(a)

Examples 1 and 2 can be derived from this general formula, since the cosine of 0 degrees is 1.0 (Example 1), and the cosine of 90 degrees is 0.0 (Example 2).

In Example 4, the pivot has been moved from the end of the bar to a location near the middle of the bar. Weights are added to both sides of the pivot. To the right a single weight **W** produces a force **F1** acting at a distance **L1** from the pivot. This creates a torque **T1** equal to the product of the force and the distance.

T1 = F1 * L1

To the left of the pivot two weights **W** produce a force **F2** at a distance **L2**. This produces a torque **T2** in a direction opposite from T1 because the distance is in the opposite direction.

T2 = F2 * L2

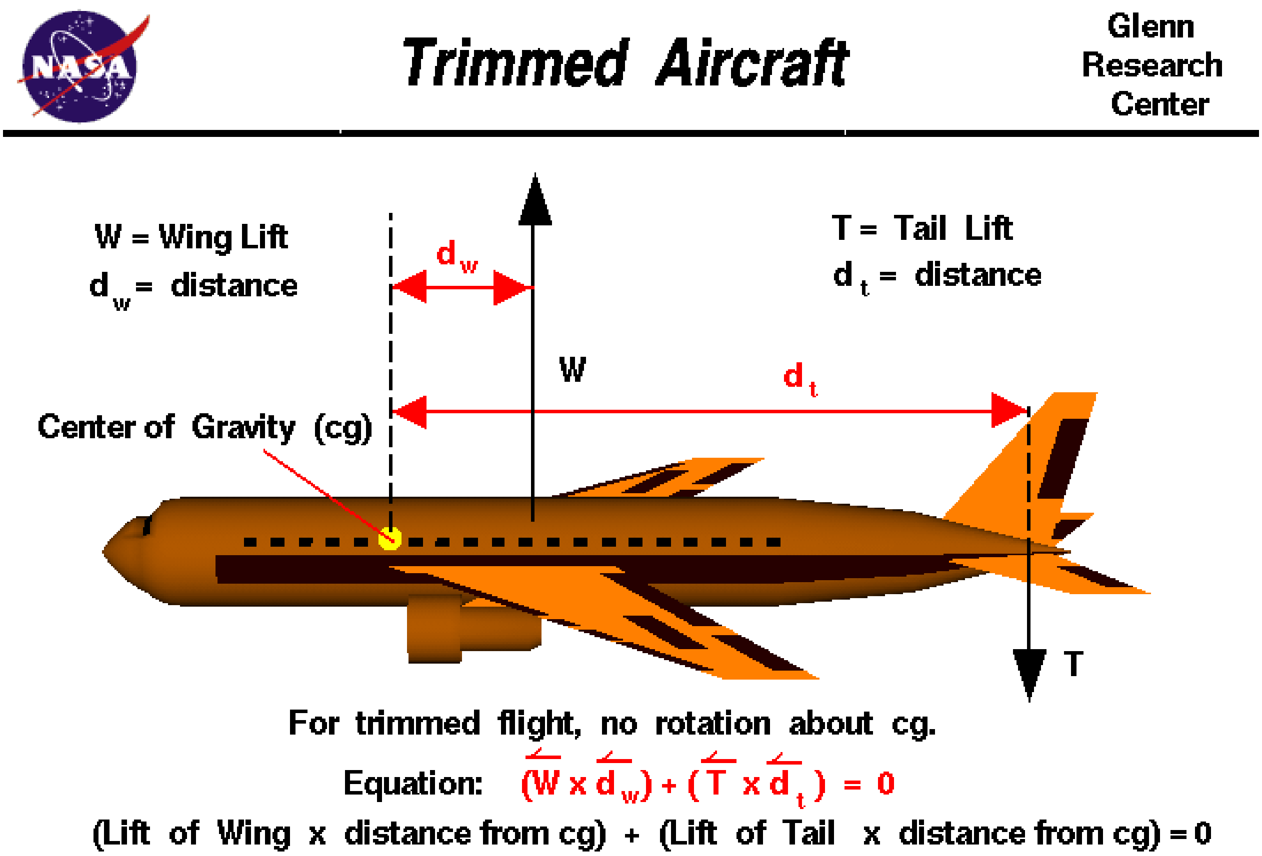
If the system were in **equilibrium**, or balanced, the torques would be equal and no net torque would act on the system.

T1 = T2 or T1 - T2 = 0

F1 * L1 = F2 * L2

If the system is not in equilibrium, or unbalanced, the bar rotates about the pivot in the direction of the higher torque. If F2 = 2 * F1, what is the relation between L1 and L2 to balance the system? If F2 = 2 * F1, and L1 = L2, in which direction would the system rotate?

Aeronautical engineers use the torque generated by aerodynamic surfaces to stabilize and control aircraft. On airplanes, the control surfaces produce aerodynamic forces. These forces are applied at some distance from the aircraft cg and therefore cause the aircraft to rotate. The elevators produce a pitching moment, the rudder produce a yawing moment, and the ailerons produce a rolling moment. The ability to vary the amount of the force and the moment allows the pilot to maneuver or to trim the aircraft. On model rockets, the fins are used to generate a torque about the rocket center of gravity to provide stability during powered flight. On kites, the aerodynamic and weight forces produce a torque about the bridle point. The distance from the bridle point and the magnitude of the forces has a strong effect on the performance of the kite.



As described on the forces slide, the aircraft lift is the sum of the lift of all of the parts of the airplane and acts through the aircraft center of pressure. Each part of the aircraft has its own lift component and its own center of pressure. The major part of the lift comes from the wings, but the horizontal stabilizer and elevator also produce lift which can be varied to maneuver the aircraft.

The average location of the weight of the aircraft is the center of gravity (cg). Any force acting at some distance from the cg produces a torque about the cg. Torque is defined to be the product of the force times the distance. A torque is a "twisting force" that produces rotations of an object. In flight, during maneuvers, an airplane rotates about its cg. But when the aircraft is not maneuvering, we want the rotation about the cg to be zero. When there is no rotation about the cg the aircraft is said to be **trimmed**.

On most aircraft, the center of gravity of the airplane is located near the center of pressure of the wing. If the center of presure of the wing is aft of the center of gravity, its lift produces a counter-clockwise rotation about the cg. The center of pressure for the elevator is aft of the center of gravity for the airliner shown in the figure. A positive lift force from the tail produces a counter-clockwise rotation about the cg. To trim the aircraft it is necessary to balance the torques produced by the wing and the tail. But since both rotations are counter-clockwise, it is impossible to balance the two rotations to produce no rotation. However, if the tail lift is negative it then produces a clockwise rotation about the cg which can balance the wing rotation.

Let us look carefully at the torques produced by the wing and the tail. The torque from the wing **TW** is equal to the lift of the wing **W** times the distance from the cg to the center of pressure of the wing **dw**.

TW = W * dw

The torque from the tail **TT** is equal to the lift of the tail **T** times the distance from the cg to the center of pressure of the tail **dt**. The lift of the wing and the lift of the tail are both forces and forces are vector quantities which have both a magnitude and a direction. We must include a minus sign on the lift of the tail because the direction of this force is negative.

TT = -T * dt

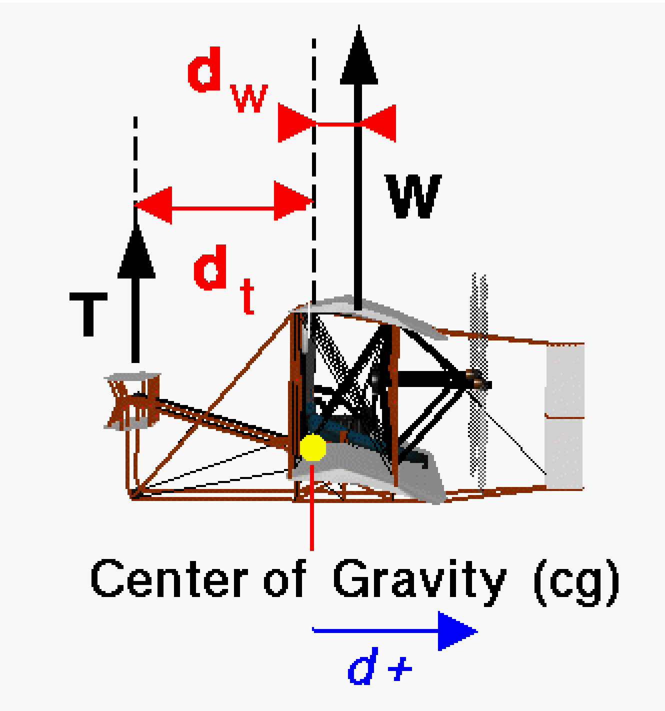
In trimmed flight, these two torques are equal:

TW = TT

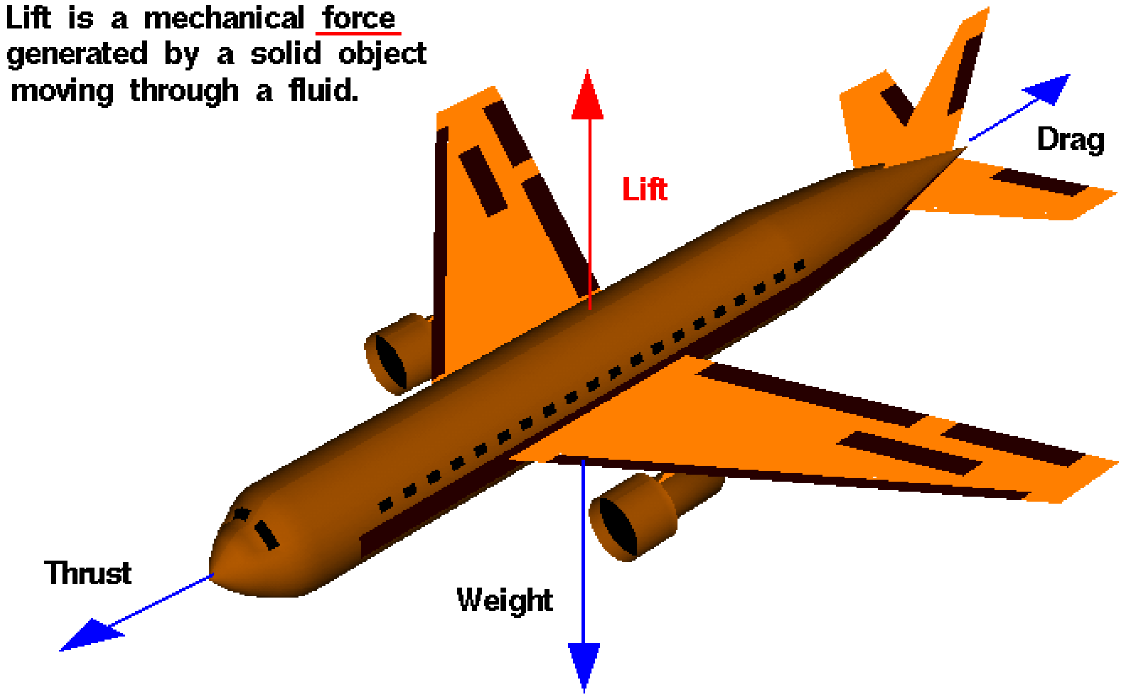
W * dw = -T * dt

W * dw + T * dt = 0

The torque equation, as written here, is a vector equation. All of the quantities are vector quantities having a magnitude and a direction. If the distances are both positive (same side of the center of gravity), then the direction of the tail force must be different than the direction of the wing force to produce no net torque or rotation. However, if the distance to the tail were negative, then the lift of the of the tail could be positive and there would be no net torque. A negative distance to the tail would imply that the tail is on the front of the aircraft, ahead of the center of gravity. A tail at the front of the aircraft is called a **canard** and was the configuration first used by the Wright brothers.



The total lift of the aircraft is the vector sum of the wing lift and the tail lift. For the airliner, the total lift is less than the wing lift; for the Wright brothers, the total lift is greater than the wing lift. The added lift was important for the Wright brothers because their aircraft had a very small engine and flew at low speeds (35mph). Since lift depends on the square of the velocity, it is hard to generate enough lift for flight at such low speeds.



Lift is the force that directly opposes the weight of an airplane and holds the airplane in the air. Lift is generated by every part of the airplane, but most of the lift on a normal airliner is generated by the wings. Lift is a mechanical aerodynamic force produced by the motion of the airplane through the air. Because lift is a force, it is a vector quantity, having both a magnitude and a direction associated with it. Lift acts through the center of pressure of the object and is directed **perpendicular** to the flow direction. There are several factors which affect the magnitude of lift.

HOW IS LIFT GENERATED?

There are many explanations for the generation of lift found in encyclopedias, in basic physics textbooks, and on Web sites. Unfortunately, many of the explanations are misleading and incorrect. Theories on the generation of lift have become a source of great controversy and a topic for heated arguments. To help you understand lift and its origins, a series of pages will describe the various theories and how some of the popular theories fail.

Lift occurs when a moving flow of gas is turned by a solid object. The flow is turned in one direction, and the lift is generated in the opposite direction, according to Newton's Third Law of action and reaction. Because air is a gas and the molecules are free to move about, any solid surface can deflect a flow. For an aircraft wing, both the upper and lower surfaces contribute to the flow turning. Neglecting the upper surface's part in turning the flow leads to an incorrect theory of lift.

NO FLUID, NO LIFT

Lift is a mechanical force. It is generated by the interaction and contact of a solid body with a fluid (liquid or gas). It is not generated by a **force field**, in the sense of a gravitational field,or an **electromagnetic field**, where one object can affect another object without being in physical contact. For lift to be generated, the solid body must be in contact with the fluid: no fluid, no lift. The Space Shuttle does not stay in space because of lift from its wings but because of orbital mechanics related to its speed. Space is nearly a vacuum. Without air, there is no lift generated by the wings.

NO MOTION, NO LIFT

Lift is generated by the difference in velocity between the solid object and the fluid. There must be motion between the object and the fluid: no motion, no lift. It makes no difference whether the object moves through a static fluid, or the fluid moves past a static solid object. Lift acts perpendicular to the motion. Drag acts in the direction opposed to the motion.

You can learn more about the factors that affect lift at this web site. There are many small interactive programs here to let you explore the generation of lift.

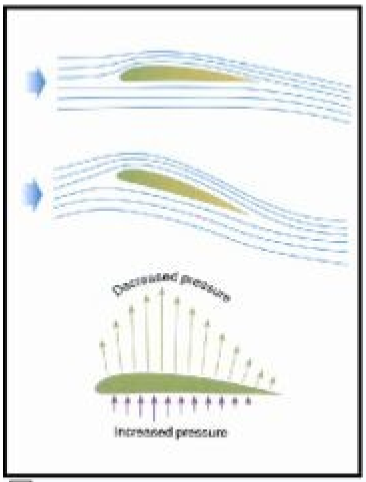
Adverse pressure gradient

An **adverse pressure gradient** occurs when the static pressure increases in the direction of the flow. This is important for boundary layers, since increasing the fluid pressure is akin to increasing the potential energy of the fluid, leading to a reduced kinetic energy and a deceleration of the fluid. Since the fluid in the inner part of the boundary layer is relatively slower, it is more greatly affected by the increasing pressure gradient. For a large enough pressure increase, this fluid may slow to zero velocity or even become reversed. When flow reversal occurs, the flow is said to be separated from the surface. This has very significant consequences in aerodynamics since flow separation significantly modifies the pressure distribution along the surface and hence the lift and drag characteristics.

Turbulent boundary layers tend to be able to sustain an adverse pressure gradient better than an equivalent laminar boundary layer. The more efficient mixing which occurs in a turbulent boundary layer transports kinetic energy from the edge of the boundary layer to the low momentum flow at the solid surface, often preventing the separation which would occur for a laminar boundary layer under the same conditions. This physical fact has led to a variety of schemes to actually produce turbulent boundary layers when boundary layer separation is dominant at high Reynolds numbers. The dimples on a golf ball, the fuzz on a tennis ball, or the seams on a baseball are good examples. Aeroplane wings are often engineered with vortex generators on the upper surface to produce a turbulent boundary layer.

Aerodynamics

Aerodynamics (shaping of objects that affect the flow of air, liquid or gas) is a branch of fluid dynamics concerned with the study of forces generated on a body in a flow. The solution of an aerodynamic problem normally involves calculating for various properties of the flow, such as velocity, pressure, density, and temperature, as a function of space and time. Understanding the flow pattern makes it possible to calculate or approximate the forces and moments acting on bodies in the flow. This mathematical analysis and empirical approximation form the scientific basis for heavier-than-air flight.



A profile of an aircraft wing.

Aerodynamic problems can be classified in a number of ways. The flow environment defines the first classification criterion. *External* aerodynamics is the study of flow around solid objects of various shapes. Evaluating the lift and drag on an airplane, the shock waves that form in front of the nose of a rocket or the flow of air over a hard drive head are examples of external aerodynamics. *Internal* aerodynamics is the study of flow through passages in solid objects. For instance, internal aerodynamics encompasses the study of the airflow through a jet engine or through an air conditioning pipe.

The ratio of the problem's characteristic flow speed to the speed of sound comprises a second classification of aerodynamic problems. A problem is called subsonic if all the speeds in the problem are less than the speed of sound, transonic if speeds both below and above the speed of sound are present (normally when the characteristic speed is approximately the speed of sound), supersonic when the characteristic flow speed is greater than the speed of sound, and hypersonic when the flow speed is much greater than the speed of sound. Aerodynamicists disagree over the precise definition of hypersonic flow; minimum Mach numbers for hypersonic flow range from 3 to 12. Most aerodynamicists use numbers between 5 and 8.

The influence of viscosity in the flow dictates a third classification. Some problems involve only negligible viscous effects on the solution, in which case viscosity can be considered to be nonexistent. The approximations to these problems are called inviscid flows. Flows for which viscosity cannot be neglected are called viscous flows.

Aerodynamics in other fields

Aerodynamics is important in a number of applications other than aerospace engineering. It is a significant factor in any type of vehicle design, including automobiles. It is important in the prediction of forces and moments in sailing. It is used in the design of small components such as hard drive heads. Structural engineers also use aerodynamics, and particularly aeroelasticity, to calculate wind loads in the design of large buildings and bridges. Urban aerodynamics seeks to help town planners and designers improve comfort in outdoor spaces, create urban microclimates and reduce the effects of urban pollution. The field of environmental aerodynamics studies the ways atmospheric circulation and flight mechanics affects ecosystems. The aerodynamics of internal passages is important in heating/ventilation, gas piping, and in automotive engines where detailed flow patterns strongly affect the performance of the engine. If a solid object was to push on the other solid object, since it can't change shape like fluid it will stop it in its tracks but with fluid, since it can change shape and can go all around the object which then puts friction of aerodynamic forces!

Continuity assumption

Bernoulli's principle: Gases are composed of molecules which collide with one another and solid objects. In j, and velocity are taken to be well-defined at infinitely small points, and are assumed to vary continuously from one point to another. The discrete, molecular nature of a gas is ignored.

The continuity assumption becomes less valid as a gas becomes more rarefi=d. In these cases, statistical mechanics is a more valid method of solving the problem than aerodynamics.

Conservation laws

Aerodynamic problems are solved using the conservation laws, or equations derived from the conservation laws. In aerodynamics, three conservation laws are used:

- Conservation of mass: Matter is not created or destroyed. If a certain mass of fluid enters a volume, it must either exit the volume or increase the mass inside the volume.
- Conservation of momentum: Also called Newton's second law of motion
- Conservation of energy: Although it can be converted from one form to another, the total energy in a given system remains constant.

Boundary layer

The concept of boundary layer is important in most aerodynamic problems. The viscosity and fluid friction in the air is usually important only in this thin layer. This principle makes aerodynamics much more tractable mathematically and also intuitively.

Low-speed aerodynamics

Low-speed aerodynamics is the study of inviscid, incompressible and irrotational aerodynamics where the differential equations used are a simplified version of the governing equations of fluid dynamics.[1]. It is a special case of Subsonic aerodynamics.

In solving a subsonic problem, one decision to be made by the aerodynamicist is whether or not to incorporate the effects of compressibility. Compressibility is a description of the amount of change of density in the problem. When the effects of compressibility on the solution are small, the aerodynamicist may choose to assume that density is constant. The problem is then an incompressible low-speed aerodynamics problem. When the density is allowed to vary, the problem is called a compressible problem. In air, compressibility effects can be ignored when the Mach number in the flow does not exceed 0.3. Above 0.3, the problem should be solved using compressible aerodynamics.

Subsonic aerodynamics

In a subsonic aerodynamic problem, all of the flow speeds are less than the speed of sound. This class of problems encompasses nearly all internal aerodynamic problems, as well as external aerodynamics for most unpowered and propeller driven aircraft, model aircraft, and automobiles. Notable exceptions are propellers and rotors whose tip speeds can become transonic or even supersonic.

Transonic aerodynamics

Transonic aerodynamic problems are defined as problems in which both supersonic and subsonic flow exist. Normally the term is reserved for problems in which the characteristic Mach number is very close to one.

Transonic flows are characterized by shock waves and expansion waves. A shock wave or expansion wave is a region of very large changes in the flow properties. In fact, the properties change so quickly they are nearly discontinuous across the waves.

Transonic problems are arguably the most difficult to solve. Flows behave very differently at subsonic and supersonic speeds, therefore a problem involving both types is more complex than one in which the flow is either purely subsonic or purely supersonic.

Supersonic aerodynamics

Supersonic aerodynamic problems are those involving flow speeds greater than the speed of sound. Calculating the lift on the Concorde during cruise can be an example of a supersonic aerodynamic problem.

Supersonic flow behaves very differently from subsonic flow. Fluids react to differences in pressure; pressure changes are how a fluid is "told" to respond to its environment. Therefore, since sound is in fact an infinitesmal pressure difference propagating through a fluid, the speed of sound in that fluid can be considered the fastest speed that "information" can travel in the flow. This difference most obviously manifests itself in the case of a fluid striking an object. In front of that object, the fluid builds up a stagnation pressure as impact with the object brings the moving fluid to rest. In Gas traveling at subsonic speed, this pressure disturbance can propagate upstream, changing the flow pattern ahead of the object and giving the impression that the fluid "knows" the object is there and is avoiding it. However, in a supersonic flow, the pressure disturbance cannot propagate upstream. Thus, when the fluid finally does strike the object, it is forced to change its properties -- temperature, density, pressure, and Mach number -- in an extremely violent and irreversible fashion called a shock wave. The presence of shock waves, along with the compressibility effects of high-velocity (see Reynolds number) fluids, is the central difference between supersonic and subsonic aerodynamics problems.

References

1. ^ Katz, Joseph (1991). *Low-speed aerodynamics: from wing theory to panel methods*, McGraw-Hill series in aeronautical and aerospace engineering. McGraw-Hill.

Blasius boundary layer

A **Blasius boundary layer**, in physics and fluid mechanics, describes the steady two-dimensional boundary layer that forms on a semi-infinite plate which is held parallel to a constant unidirectional flow *U*.

Within the boundary layer the usual balance between viscosity and convective inertia is struck, resulting in the scaling argument

$$\frac{U^2}{L} \approx \nu \frac{U}{\delta^2},$$

where δ is the boundary-layer thickness and ν is the kinematic viscosity.

However the semi-infinite plate has no natural length scale *L* and so the steady, two-dimensional boundary-layer equations

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$$
$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \nu\frac{\partial^2 u}{\partial y^2}$$

(note that the x-independence of U has been accounted for in the boundary-layer equations) admit a similarity solution. In the system of partial differential equations written above it is assumed that a fixed solid body wall is parallel to the x-direction whereas the y-direction is normal with respect to the fixed wall. u and v denote here the x- and y-components of the fluid velocity vector. Furthermore, from the scaling argument it is apparent that the boundary layer grows with the downstream coordinate x, e.g.

$$\delta(x) \approx \left(\frac{\nu x}{U}\right)^{1/2}.$$

This suggests adopting the similarity variable

$$\eta = \frac{y}{\delta(x)} = y\left(\frac{U}{\nu x}\right)^{1/2}$$

and writing

$$u = Uf'(\eta).$$

It proves convenient to work with the streamfunction, in which case

$$\psi = (\nu U x)^{1/2} f(\eta)$$

and on differentiating, to find the velocities, and substituting into the boundary-layer equation we obtain the Blasius equation

$$f''' + \frac{1}{2} f f'' = 0$$

subject to $f = f' = 0$ on $\eta = 0$ and $f' \to 1$ as $\eta \to \infty$. This non-linear ODE must be solved numerically, with the shooting method proving an effective choice. The shear stress on the plate

$$\sigma_{xy} = \frac{f''(0)\rho U^2 \sqrt{\nu}}{\sqrt{Ux}}.$$

can then be computed. The numerical solution gives $f''(0) \approx 0.332$.

Falkner-Skan boundary layer

A generalisation of the Blasius boundary layer that considers outer flows of the form $U = cx^m$ results in a boundary-layer equation of the form

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=c^2mx^{2m-1}+\nu\frac{\partial^2 u}{\partial y^2}.$$

Under these circumstances the appropriate similarity variable becomes

$$\eta=\frac{y}{\delta(x)}=\frac{\sqrt{c}y}{\sqrt{\nu}x^{(1-m)/2}},$$

and, as in the Blasius boundary layer, it is convenient to use a stream function

$\psi = U(x)\delta(x)f(\eta) = cx^m\delta(x)f(\eta)$

This results in the Falkner-Skan equation

$$f'''+\frac{1}{2}(m+1)ff''-mf'^2+m=0$$

(note that $m = 0$ produces the Blasius equation).

References

- Schlichting, H. (2004), *Boundary-Layer Theory*, Springer. ISBN 3-540-66270-7
- Pozrikidis, C. (1998), *Introduction to Theoretical and Computational Fluid Dynamics*, Oxford. ISBN 0-19-509320-8
- Blasius, H. (1908), *Grenzschichten in Flussigkeiten mit kleiner Reibung*, Z. Math. Phys. vol 56, pp. 1-37. http://naca.larc.nasa.gov/reports/1950/naca-tm-1256 (English translation)

Boundary layer

In physics and fluid mechanics, the **boundary layer** is that layer of fluid in the immediate vicinity of a bounding surface. In the atmosphere the boundary layer is the air layer near the ground affected by diurnal heat, moisture or momentum transfer to or from the surface. On an aircraft wing the boundary layer is the part of the flow close to the wing. The *Boundary layer effect* occurs at the field region in which all changes occur in the flow pattern. The boundary layer distorts surrounding nonviscous flow. It is a phenomenon of viscous forces. This effect is related to the Leidenfrost effect and the Reynolds number.

Laminar boundary layers come in various forms and can be loosely classified according to their structure and the circumstances under which they are created. The thin shear layer which develops on an oscillating body is an example of a Stokes layer, whilst the Blasius boundary layer refers to the well-known similarity solution for the steady boundary layer attached to a flat plate held in an oncoming unidirectional flow. When a fluid rotates, viscous forces may be balanced by Coriolis effects, rather than convective inertia, leading to the formation of an Eckman layer.

Contents

[hide]

Aerodynamics

The aerodynamic boundary layer was first defined by Ludwig Prandtl in a paper presented at the third Congress of Mathematicians in Heidelberg, Germany. It allows aerodynamicists to simplify the equations of fluid flow by dividing the flow field into two areas: one inside the boundary layer, where viscosity is dominant and the majority of the drag experienced by a body immersed in a fluid is created, and one outside the boundary layer where viscosity can be neglected without significant effects on the solution. This allows a closed-form solution for the flow in both areas, which is a significant simplification over the solution of the full Navier-Stokes equations. The majority of the heat transfer to and from a body also takes place within the boundary layer, again allowing the equations to be simplified in the flow field outside the boundary layer.

The thickness of the velocity boundary layer is normally defined as the distance from the solid body at which the flow velocity is 99% of the freestream velocity, that is, the velocity that is calculated at the surface of the body in an inviscid flow solution. The no-slip condition requires that the flow velocity at the surface of a solid object is zero and that the fluid temperature is equal to the temperature of the surface. The flow velocity will then increase rapidly within the boundary layer, governed by the boundary layer equations, below. The thermal boundary layer thickness is similarly the distance from the body at which the temperature is 99% of the temperature found from an inviscid solution. The ratio of the two thicknesses is governed by the Prandtl number. If the Prandtl number is 1, the two boundary layers are the same thickness. If the Prandtl number is greater than 1, the thermal boundary layer is thinner than the velocity boundary layer. If the Prandtl number is less than 1, which is the case for air at standard conditions, the thermal boundary layer is thicker than the velocity boundary layer.

In high-performance designs, such as sailplanes and commercial transport aircraft, much attention is paid to controlling the behavior of the boundary layer to minimize drag. Two effects must to be considered. First, the boundary layer adds to the effective thickness of the body, through the displacement thickness, hence increasing the pressure drag. Secondly, the shear forces at the surface of the wing create skin friction drag.

At high Reynolds numbers, typical of full-sized aircraft, it is desirable to have a laminar boundary layer. This results in a lower skin friction due to the characteristic velocity profile of laminar flow. However, the boundary layer inevitably thickens and becomes less stable as the flow develops along the body, and eventually becomes turbulent, the process known as boundary layer transition. One way of dealing with this problem is to suck the boundary layer away through a porous surface (see Boundary layer suction). This can result in a reduction in drag, but is usually impractical due to the mechanical complexity involved.

At lower Reynolds numbers, such as those seen with model aircraft, it is relatively easy to maintain laminar flow. This gives low skin-friction, which is desirable. However, the same velocity profile which gives the laminar boundary layer its low skin friction also causes it to be badly affected by adverse pressure gradients. As the pressure begins to recover over the rear part of the wing chord, a laminar boundary layer will tend to separate from the surface. Such separation causes a large increase in the pressure drag, since it greatly increases the effective size of the wing section. In these cases, it can be advantageous to deliberately trip the boundary layer into turbulence at a point prior to the location of laminar separation, using a turbulator. The fuller velocity profile of the turbulent boundary layer allows it to sustain the adverse pressure gradient without separating. Thus, although the skin friction is increased, overall the drag is decreased. This is the principle behind the dimpling on golf balls, as well as vortex generators on light aircraft. Special wing sections have also been designed which tailor the pressure recovery so that laminar separation is reduced or even eliminated. This represents an optimum compromise between the pressure drag from flow separation and skin friction the induced turbulence.

Boundary layer equations

The deduction of the **boundary layer equations** was perhaps one of the most important advances in aerodynamics. Using an order of magnitude analysis, the well-known governing Navier-Stokes equations of viscous fluid flow can be greatly simplified within the boundary layer. Notably, the characteristic of the partial differential equations (PDE) becomes parabolic, rather than the elliptical form of the full Navier-Stokes equations. This greatly simplifies the solution of the equations. By making the boundary layer approximation, the flow is divided into an inviscid portion (which is easy to solve by a number of methods) and the boundary layer, which is governed by an easier to solve PDE. The Navier-Stokes equations for a two-dimensional steady incompressible flow in cartesian coordinates are given by

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$$

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \nu(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2})$$

$$u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial y} + \nu(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2})$$

where u and v are the velocity components, ρ is the density, p is the pressure, and ν is the kinematic viscosity of the fluid at a point.

The approximation states that, for a sufficiently high Reynolds number the flow over a surface can be divided into an outer region of inviscid flow unaffected by viscosity (the majority of the flow), and a region close to the surface where viscosity is important (the boundary layer). Let *u* and *v* be streamwise and transverse (wall normal) velocities respectively inside the boundary layer. Using asymptotic analysis, it can be shown that the above equations of motion reduce within the boundary layer to become

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$$

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=-\frac{1}{\rho}\frac{\partial p}{\partial x}+\nu\frac{\partial^2 u}{\partial y^2}$$

and the remarkable result that

$$\frac{1}{\rho}\frac{\partial p}{\partial y}=0$$

The asymptotic analysis also shows that v, the wall normal velocity, is small compared with u the streamwise velocity, and that variations in properties in the streamwise direction are generally much lower than those in the wall normal direction.

Since the static pressure p is independent of y, then pressure at the edge of the boundary layer is the pressure throughout the boundary layer at a given streamwise position. The external pressure may be obtained through an application of Bernoulli's Equation. Let u_0 be the fluid velocity outside the boundary layer, where u and u_0 are both parallel. This gives upon substituting for p the following result

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=u_0\frac{\partial u_0}{\partial x}+\nu\frac{\partial^2 u}{\partial y^2}$$

with the boundary condition

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0$$

For a flow in which the static pressure p also does not change in the direction of the flow then

$$\frac{\partial p}{\partial x}=0$$

so u_0 remains constant.

Therefore, the equation of motion simplifies to become

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=\nu\frac{\partial^2 u}{\partial y^2}$$

These approximations are used in a variety of practical flow problems of scientific and engineering interest. The above analysis is for any instantaneous laminar or turbulent boundary layer, but are used mainly in laminar flow studies since the mean flow is also the instantaneous flow because there are no velocity fluctuations present.

Turbulent boundary layers

The treatment of turbulent boundary layers is far more difficult due to the time-dependent variation of the flow properties. One of the most widely used techniques in which turbulent flows are tackled is to apply Reynolds decomposition. Here the instantaneous flow properties are decomposed into a mean and fluctuating component. Applying this technique to the boundary layer equations give the full turbulent boundary layer equations not often given in literature, viz.

$$\frac{\partial \overline{u}}{\partial x} + \frac{\partial \overline{v}}{\partial y} = 0$$
$$\overline{u}\frac{\partial \overline{u}}{\partial x} + \overline{v}\frac{\partial \overline{u}}{\partial y} = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x} + \nu\left(\frac{\partial^2 \overline{u}}{\partial x^2} + \frac{\partial^2 \overline{u}}{\partial y^2}\right) - \frac{\partial}{\partial y}(\overline{u'v'}) - \frac{\partial}{\partial x}(\overline{u'^2})$$
$$\overline{u}\frac{\partial \overline{v}}{\partial x} + \overline{v}\frac{\partial \overline{v}}{\partial y} = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial y} + \nu\left(\frac{\partial^2 \overline{v}}{\partial x^2} + \frac{\partial^2 \overline{v}}{\partial y^2}\right) - \frac{\partial}{\partial x}(\overline{u'v'}) - \frac{\partial}{\partial y}(\overline{v'^2})$$

Using the same order-of-magnitude analysis as for the instantaneous equations, these turbulent boundary layer equations generally reduce to become in their classical form:

$$\frac{\partial \overline{u}}{\partial x} + \frac{\partial \overline{v}}{\partial y} = 0$$
$$\overline{u}\frac{\partial \overline{u}}{\partial x} + \overline{v}\frac{\partial \overline{u}}{\partial y} = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x} + \nu\frac{\partial^2 \overline{u}}{\partial y^2} - \frac{\partial}{\partial y}(\overline{u'v'})$$
$$\frac{\partial \overline{p}}{\partial y} = 0$$

The additional term $\overline{u'v'}$ in the turbulent boundary layer equations is known as the Reynolds shear stress and is unknown a priori. The solution of the turbulent boundary layer equations therefore necessitate the use of a turbulence model, which aims to express the Reynolds shear stress in terms of known flow variables or derivatives. The lack of accuracy and generality of such models is the single major obstacle which inhibits the successful prediction of turbulent flow properties in modern aerodynamics.

Boundary layer turbine

This effect was exploited in the Tesla turbine, patented by Nikola Tesla in 1913. It is referred to as a bladeless turbine because it uses the boundary layer effect and not a fluid impinging upon the blades as in a conventional turbine. Boundary layer turbines are also known as cohesion-type turbines, bladeless turbine, and Prandtl layer turbine (after Ludwig Prandtl).

Boundary layer suction

Boundary layer suction is technique in which an air pump is used to extract the boundary layer at the wing or the inlet of an aircraft. Improving the air flow can reduce drag. Improvements in fuel efficiency have been estimated as high as 30%.

The boundary layer

The air molecules at the surface of a wing are effectively stationary (see the no-slip condition). If the flow is smooth, known as laminar flow, the velocity of the air increases steadily as measurements are taken further away from the surface. However the smooth flow is often disturbed by the boundary layer breaking away from the surface and creating a low pressure region immediately behind the airfoil (see flow separation). This low pressure region results in increased overall drag. Attempts have been made over the years to delay the onset of this flow separation by careful design and smooth surfaces.

Use of suction

As flow separation results from the velocity deficit that is characteristic of boundary layers, suction attempts to remove the boundary layer from the surface before it can separate. The technology was first developed by Werner Pfenninger in the Second World War and has been researched almost continuously since. In the 1990s tests [1] were done by NASA with a F16XL.

Loek Boermans is researching the technology for use in gliders at the Technical University of Delft. However about 500 watts of power would be needed to drive the pumps and this would mean covering the glider with solar panels. This would increase the cost greatly. There are also structural problems to be overcome before the ultimate glider could be manufactured.

Boundary layer transition

The process of a laminar boundary layer becoming turbulent is known as **boundary layer transition**. This process is an extroadinarily complicated process which at present is not fully understood. However, as the result of many decades of intensive research, certain features have become gradually clear, and it is known that the process proceeds through a series of stages.

Transition stages

The initial stage of the natural transition process is known as the receptivity phase and consists of the transformation of external disturbances in the outer freestream flow over the boundary layer (such as freestream turbulence, surface roughness, acoustic noise etc.) into internal instability oscillations within the boundary layer. Upon entering the boundary layer, a wide spectrum of disturbances are present. Many of these disturbances decay however, and only a limited number become amplified with further downstream development.

The second stage of the process, is the exponential growth of the few unstable disturbances. Since this stage is linear, it can be well-described by linear stability theory by following the most unstable mode. It is generally accepted that for subsonic incompressible boundary layers, these initial instabilities which cause transition and ultimately lead to turbulent flow, take the form of Tollmien-Schlichting waves.

In the third stage, the amplitudes of the disturbances now become large enough to introduce non-linearity effects. In low Reynolds number flows, the initial amplitude of the disturbances are insufficient to cause immediate transition. These waves must first develop within the boundary layer over a finite distance to trigger non-linear effects characteristic of the transition process. Here the uniform spanwise mean flow begins to become modulated by the non-linear interaction of the disturbances. In this third phase, the mean boundary layer profile now begins to become distorted and the boundary layer thickness varies strongly in the streamwise direction.

Due to the distortion of the boundary layer, inflexional mean profiles develop and a fourth stage is reached, where the boundary layer becomes unstable to three-dimensional high-frequency disturbances. The frequencies observed in this phase are typically an order-of-magnitude greater than those observed in the initial stages, and this is generally referred to as secondary instability. Finally, an explosive growth of these high-frequency disturbances initiates the fifth and final phase, the breakdown into turbulence.

Numerous experiments in recent decades have revealed that the extent of the amplification region, and hence the location of the transition point on the body surface, is strongly dependent not only upon the amplitude and/or the spectrum of external disturbances but also on their physical nature. Some of the disturbances easily penetrate into the boundary layer and turn into Tollmien-Schlichting waves, whilst others do not. Consequently, the concept of boundary layer transition is a complex one and still lacks a complete theoretical exposition.

Displacement thickness

Displacement thickness is distance by which a surface would have to be moved parallel to itself towards the reference plane in an ideal fluid stream of velocity u_0 to give the same volumetric flow as occurs between the surface and the reference plane in a real fluid.

In practical aerodynamics, the displacement thickness essentially modifies the shape of a body immersed in a fluid. It is commonly used in aerodynamics to overcome the difficulty inherent in the fact that the fluid velocity in the boundary layer approaches asymptotically to the free stream value as distance from the wall increases at any given location. The mathematical definition of the displacement thickness for incompressible flow is given by

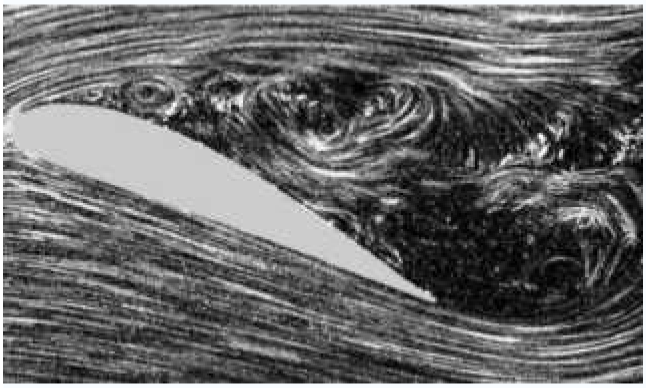
$$\delta^* = \int_0^\delta (1 - \frac{u}{u_0}) dy$$

and for compressible flow, by

$$\delta^* = \int_0^\delta (1 - \frac{\rho u}{\rho_0 u_0}) dy$$

where ρ_0 and u_0 refer to the density and velocity outside the boundary layer.

Flow separation


Airflow separating from a wing which is at a high angle of attack

All solid objects travelling through a fluid (or alternatively a stationary object exposed to a moving fluid) acquire a boundary layer of fluid around them where friction between the fluid molecules and the object's rough surface occurs. Boundary layers can be either laminar or turbulent. A calculation of the Reynolds number of the local flow conditions is necessary to determine which form the flow will take.

Flow separations occurs when the boundary layer encounters a sufficiently large adverse pressure gradient. The fluid flow becomes detached from the surface of the object, and instead takes the forms of eddies and vortices. In aerodynamics, flow separation can often result in increased drag, particularly pressure drag which is caused by the pressure differential between the front and rear surfaces of the object as it travels through the fluid. For this reason much effort and research has gone into the design of aero- or hydrodynamic surfaces which keep the local flow attached for as long as possible; examples of this include the dimples on a golf ball, turbulators on a glider, vortex generators on light aircraft and leading edge extensions on aircraft such as the F/A-18 Hornet for high angles of attack.

Laminar flow

Laminar flow (bottom) and turbulent flow (top) over a submarine hull.

Laminar flow, sometimes known as streamline flow, occurs when a fluid flows in parallel layers, with no disruption between the layers. In fluid dynamics, laminar flow is a flow regime characterized by high momentum diffusion, low momentum convection, and pressure and velocity independence from time. It is the opposite of turbulent flow. In nonscientific terms laminar flow is "smooth," while turbulent flow is "rough."

The (dimensionless) Reynolds number is an important parameter in the equations that describe whether flow conditions lead to laminar or turbulent flow. In laminar flow, the Reynolds number is less than 2100. Creeping motion or Stokes flow, an extreme case of laminar flow where viscous (friction) effects are much greater than inertial forces, occurs when the Reynolds number is much less than 1.

For example, consider the flow of air over an airplane wing. The boundary layer is a very thin sheet of air lying over the surface of the wing (and all other surfaces of the airplane). Because air has viscosity, this layer of air tends to adhere to the wing. As the wing moves forward through the air, the boundary layer at first flows smoothly over the streamlined shape of the airfoil. Here the flow is called *laminar* and the boundary layer is a laminar layer.

As the boundary layer approaches the centre of the wing, it begins to lose speed due to skin friction and becomes thicker and turbulent. Here it is a turbulent layer. The process of a laminar boundary layer becoming turbulent is known as boundary layer transition. The point at which the boundary layer changes from laminar to turbulent is called the transition point. Where the boundary layer becomes turbulent, drag, due to skin friction, is relatively high. As speed increases, the transition point tends to move forward. As the angle of attack increases the transition point also tends to move forward. One way to limit the size and effect of the turbulent region is to use swept-back delta wings. This is particularly important in supersonic aircraft.

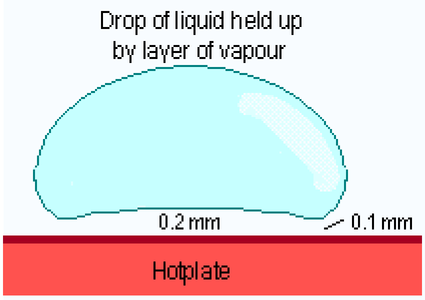
Experiments

A famous experiment involving laminar flow uses two concentric glass cylinders with the gap filled with glycerin. A drop of ink is placed in the fluid. When the outer cylinder is turned, the drop is drawn out into a thread that eventually becomes so thin that it disappears from view. At this point the ink molecules are said to be "enfolded" in the glycerin. If the cylinder is then turned in the opposite direction, the thread reforms and then becomes a drop. This experiment is typically used to show implicit order, but also nicely demonstrates the properties of laminar flow.

Leidenfrost effect

The **Leidenfrost effect** is a phenomenon in which a liquid, in near contact with a mass significantly hotter than its boiling point, produces an insulating vapor layer which keeps that liquid from boiling rapidly. This is most commonly seen when cooking; one sprinkles drops of water in a skillet to gauge its temperature—if the skillet's temperature is at or above the *Leidenfrost point*, the water skitters across the metal and takes *longer* to evaporate than it would in a skillet that is hot, but at a temperature below the *Leidenfrost point.* It has also been used in some dangerous demonstrations, such as dipping a *wet* finger in molten lead and blowing out a mouthful of liquid nitrogen, both enacted without injury to the demonstrator. The effect is also responsible for the ability of liquid nitrogen to skitter across lab floors, collecting dust in the process.

It is named after Johann Gottlob Leidenfrost, who discussed it in *A Tract About Some Qualities of Common Water* in 1756.

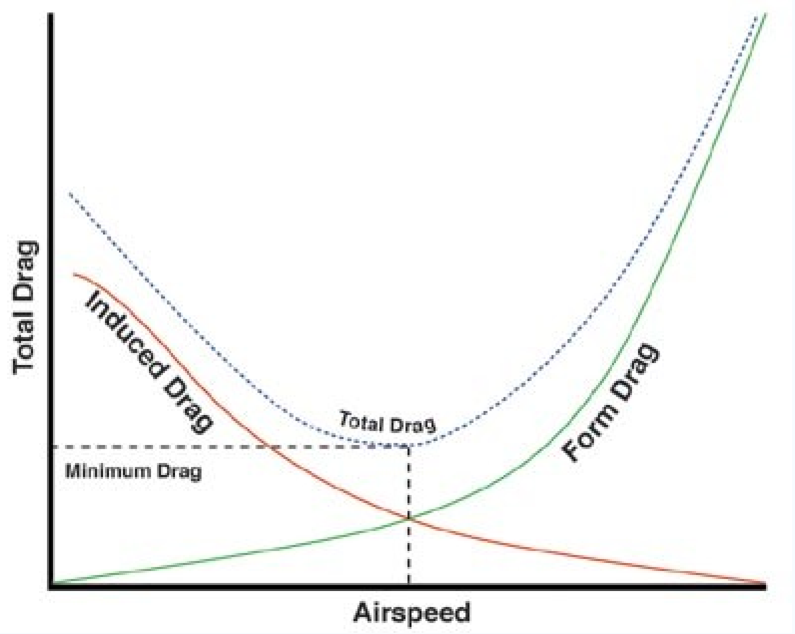


To demonstrate the Leidenfrost effect at home, just take a normal (not nonstick) clean frying pan and heat over a gas stove. Have a bowl of clean water handy. Every now and then, dip your fingers into the bowl of water and sprinkle a few drops onto the pan. Initially, as the temperature of the pan is below 100 °C, the water just flattens out and slowly evaporates. As the temperature of the pan goes above 100 °C, the water drops hiss on touching the pan and evaporate relatively quickly. Later, as the temperature goes past 220°C / 428 °F, the Leidenfrost effect comes into play. On contact the droplets of water do not evaporate away so quickly. This time, they bunch up into small balls of water and skitter around, lasting much longer than when the temperature of the pan was much lower. This effect lasts pretty much until a much higher temperature causes any further drops of water to evaporate too quickly to cause this effect.

This works because, at temperatures above the *Leidenfrost point* (about 220 °C / 428 °F for water), when water touches the hot plate, the bottom part of the water vapourizes immediately on contact. The resulting gas actually suspends the rest of the water droplet just above it, preventing any further direct contact between the liquid water and the hot plate and dramatically slowing down further heat transfer between them. This also results in the drop being able to skid around the pan on the layer of gas just under it .

(The *Leidenfrost point* may also be taken to be the temperature for which the hovering droplet lasts longest.)

Parasitic drag



Parasitic drag (also called **parasite drag**) is drag caused by moving a solid object through a fluid. Parasitic drag is made up of many components, the most prominent being **form drag**. **Skin friction** and **interference drag** are also major components of parasitic drag.

In aviation, induced drag tends to be greater at lower speeds because a high angle of attack is required to maintain lift. However, as speed increases the induced drag becomes much less, but parasitic drag necessarily increases because the fluid is flowing faster. At even higher speeds in the transonic, wave drag enters the picture. Each of these forms of drag changes in proportion to the others based on speed. The combined overall drag curve therefore shows a minimum at some airspeed - an aircraft flying at this speed will be at or close to its optimal efficiency. Pilots will use this speed to maximize endurance (minimum fuel consumption), or maximise gliding range in the event of an engine failure.

Form drag

Form drag, **profile drag**, or **pressure drag**, arises because of the form of the object. The general size and shape of the body is the most important factor in form drag - bodies with a larger apparent cross-section will have a higher drag than thinner bodies. Sleak designs, or designs that are streamlined and change cross-sectional area gradually are also critical for achieving minimum form drag. In some cases, cooling systems can be a serious source of drag, and Evaporative cooling was developed to remedy that. Form drag follows the drag equation, meaning that it rises with the square of speed, and thus becomes more important for high speed aircraft.

Profile Drag (Pxp): depends on the longitudinal section of the body. A diligent choice of body profile is more than essential for low drag coefficient. Streamlines should be continuous and separation of the boundary layer with its attendant vortices should be avoided.

Interference drag

Interference drag arises from vortices. Whenever two surfaces meet at a sharp angle on an airplane, the airflow has a tendency to form a vortex. Accelerating the air into this vortex causes drag on the plane, and the resulting low pressure area behind the plane also contributes. Thus, the primary method of reducing interference drag is eliminating sharp angles by adding fairings which smooth out any sharp angles on the aircraft by forming fillets. Interference drag is also created by closly spaced parallel surfaces such as the wings of a biplane or triplane, or the facing surfaces of an external load (such as an external fuel tank or weapon) and the fuselage or wing. As with other components of parasitic drag, interference drag follows the drag equation and rises with the square of the velocity.

Skin friction

Skin friction arises from the friction of the fluid against the "skin" of the object that is moving through it. Skin friction is a function of the interaction between the fluid and the skin of the body, as well as the wetted surface, or the area of the surface of the body that would become wet if sprayed with water flowing in the wind. As with other components of parasitic drag, skin friction follows the drag equation and rises with the square of the velocity.

Turbulence

Turbulent flow around an obstacle; the flow further upstream is laminar

Laminar and turbulent water flow over the hull of a submarine

Turbulence in the tip vortex from an airplane wing

For the Hyundai Tiburon Turbulence car, see Hyundai Tiburon
For the film of the same name, see Turbulence (film)
For the record label, see Turbulence Records

In fluid dynamics, **turbulence** or **turbulent flow** is a flow regime characterized by chaotic, stochastic property changes. This includes low momentum diffusion, high momentum convection, and rapid variation of pressure and velocity in space and time. Flow that is not turbulent is called laminar flow. The (dimensionless) Reynolds number characterizes whether flow conditions lead to laminar or turbulent flow; e.g. for pipe flow, a Reynolds number above about 2300 will be turbulent.

Explanation

Consider the flow of water over a simple smooth object, such as a sphere. At very low speeds the flow is laminar, i.e., the flow is smooth (though it may involve vortices on a large scale). As the speed increases, at some point the transition is made to turbulent ("chaotic") flow. In turbulent flow, unsteady vortices appear on many scales and interact with each other. Drag due to boundary layer skin friction increases. The structure and location of boundary layer separation often changes, sometimes resulting in a reduction of overall drag. Because laminar-turbulent transition is governed by Reynolds number, the same transition occurs if the size of the object is gradually increased, or the viscosity of the fluid is decreased, or if the density of the fluid is increased.

Turbulence causes the formation of eddies of many different length scales. Most of the kinetic energy of the turbulent motion is contained in the large scale structures. The energy "cascades" from these large scale structures to smaller scale structures by an inertial and essentially inviscid mechanism. This process continues, creating smaller and smaller structures which produces a hierarchy of eddies. Eventually this process creates structures that are small enough that molecular diffusion becomes important and viscous dissipation of energy finally takes place. The scale at which this happens is the Kolmogorov length scale.

Turbulent diffusion is usually described by a turbulent diffusion coefficient. This turbulent diffusion coefficient is defined in a phenomenological sense, by analogy with the molecular diffusivities, but it does not have a true physical meaning, being dependent on the flow conditions, and not a property of the fluid, itself. In addition, the turbulent diffusivity concept assumes a constitutive relation between a turbulent flux and the gradient of a mean variable similar to the relation between flux and gradient that exists for molecular transport. In the best case, this assumption is only an approximation. Nevertheless, the turbulent diffusivity is the simplest approach for quantitative analysis of turbulent flows, and many models have been postulated to calculate it. For instance, in large bodies of water like oceans this coefficient can be found using Richardson's four-third power law and is governed by the random walk principle. In rivers and large ocean currents, the diffusion coefficient is given by variations of Elder's formula.

When designing piping systems, turbulent flow requires a higher input of energy from a pump (or fan) than laminar flow. However, for applications such as heat exchangers and reaction vessels, turbulent flow is essential for good heat transfer and mixing.

While it is possible to find some particular solutions of the Navier-Stokes equations governing fluid motion, all such solutions are unstable at large Reynolds numbers. Sensitive dependence on the initial and boundary conditions makes fluid flow irregular both in time and in space so that a statistical description is needed. Russian mathematician Andrey Kolmogorov proposed the first statistical theory of turbulence, based on the notion of the cascade (energy flow through scales) and self-similarity. As a result, the Kolmogorov microscales were named after him. It is now known that the self-similarity is broken so the statistical description is presently modified [1]. Still, the complete description of turbulence remains one of the unsolved problems in physics. According to an apocryphal story Werner Heisenberg was asked what he would ask God, given the opportunity. His reply was: "When I meet God, I am going to ask him two questions: Why relativity? And why turbulence? I really believe he will have an answer for the first."[2] A similar witticism has been attributed to Horace Lamb (who had published a noted text book on Hydrodynamics)—his choice being quantum mechanics (instead of relativity) and turbulence. Lamb was quoted as saying in a speech to the British Association for the Advancement of Science, "I am an old man now, and when I die and go to heaven there are two matters on which I hope for enlightenment. One is quantum electrodynamics, and the other is the turbulent motion of fluids. And about the former I am rather optimistic."[3]

Examples of turbulence

- Smoke rising from a cigarette. For the first few centimetres it remains laminar, and then becomes unstable and turbulent. Similarly, the dispersion of pollutants in the atmosphere is governed by turbulent processes.
- Flow over a golf ball. (This can be best understood by considering the golf ball to be stationary, with air flowing over it.) If the golf ball were smooth, the boundary layer flow over the front of the sphere would be laminar at typical conditions. However, the boundary layer would separate early, as the pressure gradient switched from favorable (pressure decreasing in the flow direction) to unfavorable (pressure increasing in the flow direction), creating a large region of low pressure behind the ball that creates high form drag. To prevent this from happening, the surface is dimpled to perturb the boundary layer and promote transition to turbulence. This results in higher skin friction, but moves the point of boundary layer separation further along, resulting in lower form drag and lower overall drag.
- The mixing of warm and cold air in the atmosphere by wind, which causes Clear-air turbulence experienced during airplane flight, as well as poor astronomical seeing (the blurring of images seen through the atmosphere.)
- Most of the terrestrial atmospheric circulation
- The oceanic and atmospheric mixed layers and intense oceanic currents.
- The flow conditions in many industrial equipment (such as pipes, ducts, precipitators, gas scrubbers, etc.) and machines (for instance, internal combustion engines and gas turbines).
- The external flow over all kind of vehicles such as cars, airplanes, ships and submarines.
- The motions of matter in stellar atmospheres.
- A jet exhausting from a nozzle into a quiescent fluid. As the flow emerges into this external fluid, shear layers originating at the lips of the nozzle are created. These layers separate the fast moving jet from the external fluid, and at a certain critical Reynolds number they become unstable and break down to turbulence.

Unsolved problems in physics: *Is it possible to make a theoretical model to describe the behavior of a turbulent flow — in particular, its internal structures?*

- Race cars unable to follow each other through fast corners due to turbulence created by the leading car causing understeer.

Viscosity

Continuum mechanics

General

- Classical mechanics

- Stress / Strain
- Tensor
- Conservation of mass

- Conservation of momentum

Solid mechanics

- Solids
- Elasticity
- Plasticity
- Hooke's law
- Poisson's ratio
- Rheology

- Viscoelasticity

Fluid mechanics

- Fluids
- Fluid statics
- Fluid dynamics
- Navier-Stokes equations
- **Viscosity**
- Newtonian fluids

- Non-Newtonian fluids

Viscosity is a measure of the resistance of a fluid to deform under shear stress. It is commonly perceived as "thickness", or resistance to flow. Viscosity describes a fluid's internal resistance to flow and may be thought of as a measure of fluid friction. Thus, water is "thin", having a lower viscosity, while vegetable oil is "thick" having a higher viscosity. All real fluids (except superfluids) have some resistance to shear stress, but a fluid which has no resistance to shear stress is known as an **ideal fluid** or **inviscid fluid** (Symon 1971).

When looking at a value for viscosity the number that one most often sees is the coefficients of viscosity, simply put this is the ratio between the pressure exerted on the surface of a fluid, in the lateral or horizontal direction, to the change in velocity of the fluid as you move down in the fluid (this is what is referred to as a speed gradient). For example water has a viscosity of 1.0 x 10^{-3} Pa·s and motor oil has a viscosity of 250 x 10^{-3} Pa·s. (Serway 1996, p. 440)

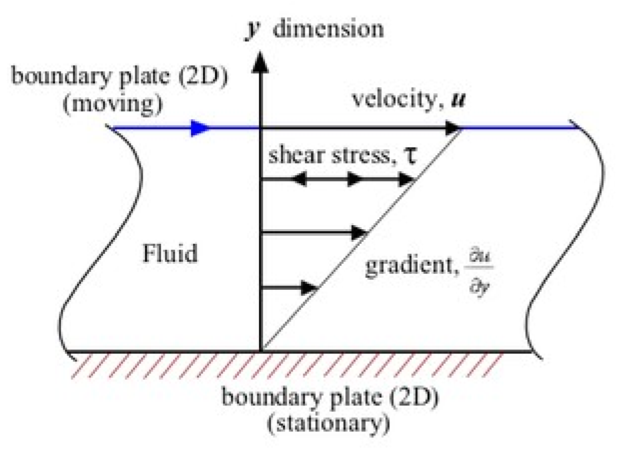
Contents

[hide]

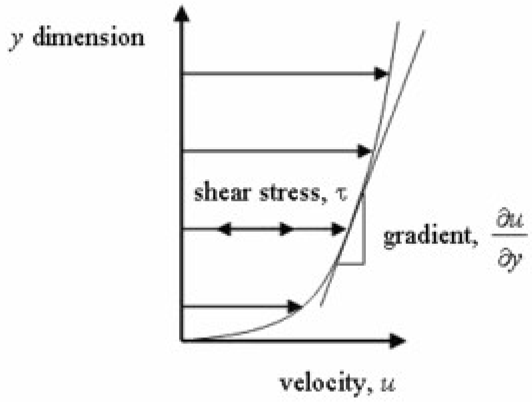
Etymology

The word "viscosity" derives from the Latin word "viscum" for mistletoe. A viscous glue was made from mistletoe berries and used for lime-twigs to catch birds.[citation needed]

Newton's theory



Laminar shear of fluid between two plates. Friction between the fluid and the moving boundaries causes the fluid to shear. The force required for this action is a measure of the fluid's viscosity. This type of flow is known as a Couette flow.



Laminar shear, the non-linear gradient, is a result of the geometry the fluid is flowing through (e.g. a pipe).

In general, in any flow, layers move at different velocities and the fluid's viscosity arises from the shear stress between the layers that ultimately opposes any applied force.

Isaac Newton postulated that, for straight, parallel and uniform flow, the shear stress, τ, between layers is proportional to the velocity gradient, ∂*u*/∂*y*, in the direction perpendicular to the layers, in other words, the relative motion of the layers.

$$\tau = \eta \frac{\partial u}{\partial y}.$$

Here, the constant η is known as the *coefficient of viscosity*, the *viscosity*, or the *dynamic viscosity*. Many fluids, such as water and most gases, satisfy Newton's criterion and are known as Newtonian fluids. Non-Newtonian fluids exhibit a more complicated relationship between shear stress and velocity gradient than simple linearity.

The relationship between the shear stress and the velocity gradient can also be obtained by considering two plates closely spaced apart at a distance *y*, and separated by a homogeneous substance. Assuming that the plates are very large, with a large area *A*, such that edge effects may be ignored, and that the lower plate is fixed, let a force *F* be applied to the upper plate. If this force causes the substance between the plates to undergo shear flow (as opposed to just shearing elastically until the shear stress in the substance balances the applied force), the substance is called a fluid. The applied force is proportional to the area and velocity of the plate and inversely proportional to the distance between the plates. Combining these three relations results in the equation $F = \eta(Au/y)$, where η is the proportionality factor called the *absolute viscosity* (with units Pa·s = kg/(m·s) or slugs/(ft·s)). The absolute viscosity is also known as the *dynamic viscosity*, and is often shortened to simply *viscosity*. The equation can be expressed in terms of shear stress; $\tau = F/A = \eta(u/y)$. The rate of shear deformation is u / y and can be also written as a shear velocity, du/dy. Hence, through this method, the relation between the shear stress and the velocity gradient can be obtained.

In many situations, we are concerned with the ratio of the viscous force to the inertial force, the latter characterised by the fluid density ρ. This ratio is characterised by the *kinematic viscosity*, defined as follows:

$$\nu = \frac{\eta}{\rho}.$$

James Clerk Maxwell called viscosity *fugitive elasticity* because of the analogy that elastic deformation opposes shear stress in solids, while in viscous fluids, shear stress is opposed by *rate* of deformation.

Measuring viscosity

Viscosity is measured with various types of viscometer, typically at 20 °C (standard state). For some fluids, it is a constant over a wide range of shear rates. The fluids without a constant viscosity are called Non-Newtonian fluids.

In paint industries, viscosity is commonly measured with a Zahn cup, in which the efflux time is determined and given to customers. The efflux time can also be converted to kinematic viscosities (cSt) through the conversion equations.

Also used in paint, a Stormer viscometer uses load-based rotation in order to determine viscosity. It uses units, Krebs units (KU), unique to this viscometer.

Units

Viscosity (dynamic/absolute viscosity): η or μ

The IUPAC symbol for viscosity is the Greek symbol eta (η), and dynamic viscosity is also commonly referred to using the Greek symbol mu (μ). The SI physical unit of dynamic viscosity is the pascal-second (Pa·s), which is identical to 1 $kg \cdot m^{-1} \cdot s^{-1}$. If a fluid with a viscosity of one Pa·s is placed between two plates, and one plate is pushed sideways with a shear stress of one pascal, it moves a distance equal to the thickness of the layer between the plates in one second. The name *poiseuille* (Pl) was proposed for this unit (after Jean Louis Marie Poiseuille who formulated Poiseuille's law of viscous flow), but not accepted internationally. Care must be taken in not confusing the poiseuille with the *poise* named after the same person!

The cgs physical unit for dynamic viscosity is the *poise*[1] (P; IPA: [pwaz])) named after Jean Louis Marie Poiseuille. It is more commonly expressed, particularly in ASTM standards, as *centipoise* (cP). The centipoise is commonly used because water has a viscosity of 1.0020 cP (at 20 °C; the closeness to one is a convenient coincidence).

$$1\ P = 1\ g \cdot cm^{-1} \cdot s^{-1}$$

The relation between Poise and Pascal-second is:

$$10\ P = 1\ kg \cdot m^{-1} \cdot s^{-1} = 1\ Pa \cdot s$$
$$1\ cP = 0.001\ Pa \cdot s = 1\ mPa \cdot s$$

Kinematic viscosity: ν

Kinematic viscosity (Greek symbol: ν) has SI units ($m^2 \cdot s^{-1}$). The cgs physical unit for kinematic viscosity is the *stokes* (abbreviated S or St), named after George Gabriel Stokes. It is sometimes expressed in terms of *centistokes* (cS or cSt). In U.S. usage, *stoke* is sometimes used as the singular form.

$$1\ \text{stokes} = 100\ \text{centistokes} = 1\ cm^2 \cdot s^{-1} = 0.0001\ m^2 \cdot s^{-1}.$$
$$1\ \text{centistokes} = 1\ mm^2/s$$

Dynamic versus kinematic viscosity

Conversion between kinematic and dynamic viscosity, is given by $\nu\rho = \eta$. Note that the parameters must be given in SI units not in P, cP or St.

For example, if ν = 1 St (=0.0001 $m^2 \cdot s^{-1}$) and ρ = 1000 $kg \cdot m^{-3}$ then $\eta = \nu\rho$ = 0.1 $kg \cdot m^{-1} \cdot s^{-1}$ = 0.1 Pa·s [1].

For a plot of kinematic viscosity of air as a function of absolute temperature, see James Ierardi's Fire Protection Engineering Site

Molecular origins

The viscosity of a system is determined by how molecules constituting the system interact. There are no simple but correct expressions for the viscosity of a fluid. The simplest exact expressions are the Green-Kubo relations for the linear shear viscosity or the Transient Time Correlation Function expressions derived by Evans and Morriss in 1985. Although these expressions are each exact in order to calculate the viscosity of a dense fluid, using these relations requires the use of molecular dynamics computer simulation.

Gases

Viscosity in gases arises principally from the molecular diffusion that transports momentum between layers of flow. The kinetic theory of gases allows accurate prediction of the behaviour of gaseous viscosity, in particular that, within the regime where the theory is applicable:

- Viscosity is independent of pressure; and
- Viscosity increases as temperature increases.

Liquids

In liquids, the additional forces between molecules become important. This leads to an additional contribution to the shear stress though the exact mechanics of this are still controversial.[citation needed] Thus, in liquids:

- Viscosity is independent of pressure (except at very high pressure); and
- Viscosity tends to fall as temperature increases (for example, water viscosity goes from 1.79 cP to 0.28 cP in the temperature range from 0 °C to 100 °C); see temperature dependence of liquid viscosity for more details.

The dynamic viscosities of liquids are typically several orders of magnitude higher than dynamic viscosities of gases.

Viscosity of materials

The viscosity of air and water are by far the two most important materials for aviation aerodynamics and shipping fluid dynamics. Temperature plays the main role in determining viscosity.

Viscosity of air

The viscosity of air depends mostly on the temperature. At 15.0 °C, the viscosity of air is 1.78 $\times 10^{-5}$ kg/(m·s). You can get the viscosity of air as a function of altitude from the eXtreme High Altitude Calculator

Viscosity of water

The viscosity of water is 8.90×10^{-4} Pa·s or 8.90×10^{-3} dyn·s/cm^{2}at about 25 °C. as a function of temperature: $\mu = A \times 10^{B/(T-C)}$
Where A=2.414 × 10-5 N*s/m2 ; B = 247.8 Kelvin ; C = 140 Kelvin

Viscosity of various materials

Example of the viscosity of milk and water. Liquids with higher viscosities will not make such a splash.

The Sutherland's formula can be used to derive the dynamic viscosity as a function of the temperature:

$$\eta = \eta_0 \frac{T_0 + C}{T + C} \left(\frac{T}{T_0}\right)^{3/2}$$

where:

- η = viscosity in (Pa·s) at input temperature T
- η_0 = reference viscosity in (Pa·s) at reference temperature T_0
- T = input temperature in kelvin
- T_0 = reference temperature in kelvin
- C = Sutherland's constant

Valid for temperatures between $0 < T < 555$ K with an error due to pressure less than 10% below 3.45 MPa

Sutherland's constant and reference temperature for some gases

Gas	***C*** **[K]**	***T*₀** **[K]**	**η₀** **[10^{-6} Pa s]**
air	120	291.15	18.27

nitrogen	111	300.55	17.81
oxygen	127	292.25	20.18
carbon dioxide	240	293.15	14.8
carbon monoxide	118	288.15	17.2
hydrogen	72	293.85	8.76
ammonia	370	293.15	9.82
sulphur dioxide	416	293.65	12.54

Some dynamic viscosities of Newtonian fluids are listed below:

Gases (at 0 °C):

	viscosity [Pa·s]
hydrogen	8.4×10^{-6}
air	17.4×10^{-6}
xenon	21.2×10^{-6}

Liquids (at 25 °C):

	viscosity [Pa·s]
acetone	[a] 0.306×10^{-3}
methanol	[a] 0.544×10^{-3}
benzene	[a] 0.604×10^{-3}
ethanol	[a] 1.074×10^{-3}
mercury	[a] 1.526×10^{-3}
nitrobenzene	[a] 1.863×10^{-3}
propanol	[a] 1.945×10^{-3}
sulfuric acid	[a] 24.2×10^{-3}
olive oil	81×10^{-3}
glycerol	[a] 934×10^{-3}
castor oil	985×10^{-3}
HFO-380	2022×10^{-3}
pitch	2.3×10^{8}

[a] Data from CRC Handbook of Chemistry and Physics, 73rd edition, 1992-1993.

Fluids with variable compositions, such as honey, can have a wide range of viscosities.

A more complete table can be found here

Can solids have a viscosity?

If on the basis that all solids flow to a small extent in response to shear stress then yes, substances know as Amorphous solids, such as glass, may be considered to have viscosity. This has led some to the view that solids are simply liquids with a very high viscosity, typically greater than 10^{12} Pa•s. This position is often adopted by supporters of the widely held misconception that glass flow can be observed in old buildings. This distortion is more likely the result of glass making process rather than the viscosity of glass.

However, others argue that solids are, in general, elastic for small stresses while fluids are not. Even if solids flow at higher stresses, they are characterized by their low-stress behavior. Viscosity may be an appropriate characteristic for solids in a plastic regime. The situation becomes somewhat confused as the term *viscosity* is sometimes used for solid materials, for example Maxwell materials, to describe the relationship between stress and the rate of change of strain, rather than rate of shear.

These distinctions may be largely resolved by considering the constitutive equations of the material in question, which take into account both its viscous and elastic behaviors. Materials for which both their viscosity and their elasticity are important in a particular range of deformation and deformation rate are called *viscoelastic*. In geology, earth materials that exhibit viscous deformation at least three times greater than their elastic deformation are sometimes called rheids.

One example of solids flowing which has been observed since 1930 is the Pitch drop experiment.

Bulk viscosity

The trace of the stress tensor is often identified with the negative-one-third of the thermodynamic pressure,

$$T_a^a = -\frac{1}{3}p,$$

which only depends upon the equilibrium state potentials like temperature and density (equation of state). In general, the trace of the stress tensor is the sum of thermodynamic pressure contribution plus another contribution which is proportional to the divergence of the velocity field. This constant of proportionality is called the **bulk viscosity**.

Eddy viscosity

In the study of turbulence in fluids, a common practical strategy for calculation is to ignore the small-scale *vortices* (or *eddies*) in the motion and to calculate a large-scale motion with an *eddy viscosity* that characterizes the transport and dissipation of energy in the smaller-scale flow. Values of eddy viscosity used in modeling ocean circulation may be from 5×10^4 to 10^6 Pa·s depending upon the resolution of the numerical grid.

Fluidity

The reciprocal of viscosity is *fluidity*, usually symbolized by $\varphi = 1 / \eta$ or $F = 1 / \eta$, depending on the convention used, measured in *reciprocal poise* (cm·s·g^{-1}), sometimes called the *rhe*. *Fluidity* is seldom used in engineering practice.

The concept of fluidity can be used to determine the viscosity of an ideal solution. For two components *a* and *b*, the fluidity when *a* and *b* are mixed is

$$F \approx \chi_a F_a + \chi_b F_b$$

which is only slightly simpler than the equivalent equation in terms of viscosity:

$$\eta \approx \frac{1}{\chi_a/\eta_a + \chi_b/\eta_b}$$

where χ_a and χ_b is the mole fraction of component *a* and *b* respectively, and η_a and η_b are the components pure viscosities.

The linear viscous stress tensor

(See *Hooke's law* and *strain tensor* for an analogous development for linearly elastic materials.)

Viscous forces in a fluid are a function of the rate at which the fluid velocity is changing over distance. The velocity at any point $\mathbf{r}$ is specified by the velocity field $\mathbf{v}(\mathbf{r})$. The velocity at a small distance $d\mathbf{r}$ from point $\mathbf{r}$ may be written as a Taylor series:

$$\mathbf{v}(\mathbf{r} + d\mathbf{r}) = \mathbf{v}(\mathbf{r}) + \frac{d\mathbf{v}}{d\mathbf{r}} d\mathbf{r} + \ldots$$

where $\frac{d\mathbf{v}}{d\mathbf{r}}$ is shorthand for the dyadic product of the del operator and the velocity:

$$\frac{d\mathbf{v}}{d\mathbf{r}} = \begin{bmatrix} \frac{\partial v_x}{\partial x} & \frac{\partial v_x}{\partial y} & \frac{\partial v_x}{\partial z} \\ \frac{\partial v_y}{\partial x} & \frac{\partial v_y}{\partial y} & \frac{\partial v_y}{\partial z} \\ \frac{\partial v_z}{\partial x} & \frac{\partial v_z}{\partial y} & \frac{\partial v_z}{\partial z} \end{bmatrix}$$

This is just the Jacobian of the velocity field. Viscous forces are the result of relative motion between elements of the fluid, and so are expressible as a function of the velocity field. In other words, the forces at $\mathbf{r}$ are a function of $\mathbf{v}(\mathbf{r})$ and all derivatives of $\mathbf{v}(\mathbf{r})$ at that point. In the case of linear viscosity, the viscous force will be a function of the Jacobian tensor alone. For almost all practical situations, the linear approximation is sufficient.

If we represent *x*, *y*, and *z* by indices 1, 2, and 3 respectively, the *i,j* component of the Jacobian may be written as $\partial_i v_j$ where ∂_i is shorthand for $\partial/\partial x_i$. Note that when the first and higher derivative terms are zero, the velocity of all fluid elements is parallel, and there are no viscous forces.

Any matrix may be written as the sum of an antisymmetric matrix and a symmetric matrix, and this decomposition is independent of coordinate system, and so has physical significance. The velocity field may be approximated as:

$$v_i(\mathbf{r} + d\mathbf{r}) = v_i(\mathbf{r}) + \frac{1}{2}\left(\partial_i v_j - \partial_j v_i\right) dr_i + \frac{1}{2}\left(\partial_i v_j + \partial_j v_i\right) dr_i$$

where Einstein notation is now being used in which repeated indices in a product are implicitly summed. The second term on the left is the asymmetric part of the first derivative term, and it represents a rigid rotation of the fluid about $\mathbf{r}$ with angular velocity ω where:

$$\omega = \nabla \times \mathbf{v} = \frac{1}{2}\begin{bmatrix} \partial_2 v_3 - \partial_3 v_2 \\ \partial_3 v_1 - \partial_1 v_3 \\ \partial_1 v_2 - \partial_2 v_1 \end{bmatrix}$$

For such a rigid rotation, there is no change in the relative positions of the fluid elements, and so there is no viscous force associated with this term. The remaining symmetric term is responsible for the viscous forces in the fluid. Assuming the fluid is isotropic (i.e. its properties are the same in all directions), then the most general way that the symmetric term (the rate-of-strain tensor) can be broken down in a coordinate-independent (and therefore physically real) way is as the sum of a constant tensor (the rate-of-expansion tensor) and a traceless symmetric tensor (the rate-of-shear tensor):

$$\frac{1}{2}\left(\partial_i v_j + \partial_j v_i\right) = \frac{1}{3}\partial_k v_k \delta_{ij} + \left(\frac{1}{2}\left(\partial_i v_j + \partial_j v_i\right) - \frac{1}{3}\partial_k v_k \delta_{ij}\right)$$

where δ_{ij} is the unit tensor. The most general linear relationship between the stress tensor σ and the rate-of-strain tensor is then a linear combination of these two tensors (Landau & Lifshitz 1997):

$$\sigma_{visc;ij} = \zeta \partial_k v_k \delta_{ij} + \eta\left(\partial_i v_j + \partial_j v_i - \frac{2}{3}\partial_k v_k \delta_{ij}\right)$$

where ζ is the coefficient of bulk viscosity (or "second viscosity") and η is the coefficient of (shear) viscosity.

The forces in the fluid are due to the velocities of the individual molecules. The velocity of a molecule may be thought of as the sum of the fluid velocity and the thermal velocity. The viscous stress tensor described above gives the force due to the fluid velocity only. The force on an area element in the fluid due to the thermal velocities of the molecules is just the hydrostatic pressure. This pressure term ($p\delta_{ij}$) must be added to the viscous stress tensor to obtain the total stress tensor for the fluid.

$$\sigma_{ij} = p\delta_{ij} + \sigma_{visc;ij}$$

The infinitesimal force dF_i on an infinitesimal area dA_i is then given by the usual relationship:

$$dF_i = \sigma_{ij} dA_j$$

Vortex generator

1967 Model Cessna 182K in flight showing after-market vortex generators on the wing leading edge

After-market Micro Dynamics vortex generators mounted on the wing of a Cessna 182K

The Symphony SA-160 has two unique **vortex generators** on its wing to ensure aileron effectiveness through the stall

A **vortex generator** is an aerodynamic surface, consisting of a small vane that creates a vortex. They can be found in many devices, but the term is most often used in aircraft design.

Vortex generators are added to the leading edge of a swept wing in order to maintain steady airflow over the control surfaces at the rear of the wing. They are typically rectangular or triangular, tall enough to protrude above the boundary layer, and run in spanwise lines near the thickest part of the wing. They can be seen on the wings and vertical tails of many airliners. Vortex generators are positioned in such a way that they have an angle of attack with respect to the local airflow.

A vortex generator creates a tip vortex which draws energetic, rapidly-moving air from outside the slow-moving boundary layer into contact with the aircraft skin. The boundary layer normally thickens as it moves along the aircraft surface, reducing the effectiveness of trailing-edge control surfaces; vortex generators can be used to remedy this problem, among others, by *re-energizing the boundary layer*. Vortex generators delay flow separation and aerodynamic stalling; they improve the effectiveness of control surfaces (e.g Embraer 170 and Symphony SA-160); and, for swept-wing transonic designs, they alleviate potential shock-stall problems (e.g. Harrier, Blackburn Buccaneer, Gloster Javelin).

Many aircraft carry vane vortex generators from time of manufacture, but there are also after-market suppliers who sell VG kits to improve the STOL performance of some light aircraft.

Air jet vortex generators work on a different principle. They direct a jet of air into the boundary layer, thereby re-energising it.

Aircraft engine

The term **aircraft engine**, for the purposes of this article, refers to reciprocating and rotary internal combustion engines used in aircraft. Jet engines and turboprops are the other common aviation powerplants; while operation differs substantially, the basics here apply to all types.

Contents

Engine design

Engines must be:

- *lightweight*, as a heavy engine decreases the amount of excess power available.
- *small* and *easily streamlined*; large engines with substantial surface area, when installed, create too much drag, wasting fuel and reducing power output.
- *powerful*, to overcome the weight and drag of the aircraft.
- *reliable*, as losing power in an airplane is a substantially greater problem than an automobile engine seizing. Aircraft engines operate at temperature, pressure, and speed extremes, and therefore need to operate reliably and safely under all these conditions.
- *repairable*, to keep the cost of replacement down. Minor repairs are relatively inexpensive.

Power

Unlike automobile engines, aircraft engines run at high power settings for extended periods of time. In general, the engine runs at maximum power for a few minutes during taking off, then power is slightly reduced for climb, and then spends the majority of its time at a cruise setting

—typically 65% to 75% of full power. In contrast, a car engine might spend 20% of its time at 65% power accelerating, followed by 80% of its time at 20% power while cruising.

Performance of a propeller motor is measured in units of power (typically horsepower), whereas for a jet engine performance is measured in terms of thrust (typically pounds). In both cases, the performance is relatively constant across the operating envelope. Power is equal to thrust times speed, which means as a propeller-driven aircraft flies faster, it actually develops less and less thrust.

Reliability

The design of aircraft engines tends to favor reliability over performance. It took many years before the reliability was established to fly over the Atlantic or the Pacific Ocean. Engine failure at all stages in flight is a part of flight lessons for student pilots.[1] Forced landings without power are practiced extensively over rural areas until the new pilot is proficient enough to handle such situation during a solo flight.

Long engine operation times and high power settings, combined with the requirement for high-reliability means that engines must be constructed to support this type of operation with ease. The engine, as well as the aircraft, needs to be lifted into the air, meaning it has to overcome lots of weight. The thrust to weight ratio is one of the most important characteristics for an aircraft engine. A typical 250 hp engine weighs just 15% of the total aircraft weight when installed into a 3000 lb (1,400 kg) aircraft.

Aircraft engines also tend to use the simplest parts and include two sets of anything needed for reliability, including ignition system (spark plugs and magnetos) and fuel pumps. Independence of function lessens the likelihood of a single malfunction causing an entire engine to fail. Thus magnetos are used because they do not rely on a battery. Two magnetos were originally installed so a pilot can switch off a faulty magneto and continue the flight on the other—but, later, dual ignition was found to offer some detonation protection too. Similarly, a mechanical engine-driven fuel pump is often backed-up by an electric one.

Another difference between cars and aircraft is that the aircraft spend the vast majority of their time travelling at high speed. This allows aircraft engines to be air cooled, as opposed to requiring a radiator. In the absence of a radiator aircraft engines can boast lower weight and less complexity. The amount of air flow an engine receives is usually carefully designed according to expected speed and altitude of the aircraft in order to maintain the engine at the optimal temperature. Just like overheating, too much cooling can be a bad thing for an engine as well. Some aircraft employ controls that allow a pilot to manually adjust the airflow into the engine compartment.

Aircraft operate at higher altitudes where the air is less dense than at ground level. As engines need oxygen to burn fuel, a forced induction system such as turbocharger or supercharger is especially appropriate for aircraft use. This does bring along the usual drawbacks of additional cost, weight and complexity.

Multi-engine debate

While some countries require twin-engined airplanes for commercial passenger transport, many, such as Canada, Australia, and the United States, allow the use of single-engine aircraft for some commercial services, including charter and sometimes scheduled commuter airline flights (the latter typically use turbine- rather than piston-powered singles).

A second engine adds redundancy so that the aircraft can stay in the air (or at least, descend more slowly) if one engine fails, providing an important safety margin during cruise flight over water or mountainous terrain; however, an engine failure on a twin-engine piston aircraft can also cause serious handling difficulties, especially right after takeoff, due to asymmetrical thrust.

A study of accidents in Australian air charter operations from 1986 to 1996 found that the overall fatal accident rate per hour for multi-engine aircraft was more than triple that for single-engine aircraft, though it did not isolate the accidents specifically caused by engine failure and the multi-engine aircraft did not fly under identical conditions.[2] According to the U.S. Air Safety Foundation, when an engine failure leads to an incident (e.g. some damage or injuries), it has a 10% chance of causing fatalities in a single-engine aircraft, but a 50% chance in a twin.[3]

Size

At one time all engine designs were new and there was no particular difference in design between aircraft and automobile engines. This changed by the start of World War I, however, when a particular class of air-cooled rotary engines became popular. These had a short lifespan, but by the 1920s a large number of engine designs were moving to the similar radial engine design. This combined air-cooled simplicity with large displacements and they were among the most powerful small engines in the world.

Both the rotary and radial engine have the drawback of a very large frontal area (see drag equation). As aircraft increased in speed and demanded better streamlining, designers turned to water-cooled inline engines. Throughout WWII the two designs were generally similar in terms of power and overall performance but some mature-design radials tended to be more reliable. After the war, in the USA, the water-cooled designs rapidly disappeared.

Repairability

For the smaller application, notably in general aviation, a hybrid design in the form an air-cooled inline, almost always 4 or 6 cylinders horizontally opposed, is most common. These combine small frontal area with air-cooled simplicity, although they required careful installation in order to be effectively cooled, notably the rearmost cylinders. To make repairs practical, each cylinder is individually replaceable, as are each of the accessories (pumps, generator and magnetos).

Fuel

Aircraft piston engines are typically designed to run on avgas. As well as being produced to higher quality standards (to avoid fuel-related engine failures), avgas has a higher octane rating, allowing a higher compression ratio and thus more power out of the engine with the same engine design. Avgas uses Tetraethyl lead (TEL) to achieve these high octane ratings, a practice banned in automobile fuel. The shrinking supply of TEL, and the possibility of environmental legislation banning its use, has made a search for replacement fuels for GA planes a priority for pilot's organizations.[4].

New designs

Economics of new designs

Throughout most of the history of aircraft engine design, they tended to be more advanced than their automobile counterparts. High-strength aluminum alloys were used in these engines decades before they became common in cars. Likewise, those engines adopted fuel injection instead of carburetion quite early. Similarly, overhead cams were introduced, while automobile engines continued to use pushrods.

Today the piston-engine aviation market is so small that there is essentially no commercial money for new design work. Most aviation engines flying are based on a design from the 1960s, or before, using original materials, tooling and parts. Meanwhile the financial power of the automobile industry has continued improvement. A new car design is likely to use an engine designed no more than a few years ago, built with the latest alloys and advanced electronic engine controls. Modern car engines require no maintenance at all (other than adding fuel and oil) for over 100,000 km, aircraft engines are now, in comparison and paradoxically, rather heavy, dirty and unreliable.

Much of the innovation (and most newly constructed planes flying) in the past two decades in private aviation has been in ultralights and homebuilt aircraft, and so has innovation in powerplants. Rotax, amongst others, has introduced a number of new small production engine designs for this type of craft. The smallest of these mostly use two-stroke designs, but the larger models are four-strokes. For the reasons discussed above, some hobbyists and experimenters prefer to adapt automotive engines for their home-built aircraft, instead of using certified aircraft engines.

Over the history of the development of aircraft engines, the Otto cycle, that is, conventional gasoline powered, reciprocating-piston engines have been by far the most common type. That is not because they are the best but simply because they were there first and type-certification of new designs is an expensive, time-consuming process.

Powerplant from a Schleicher ASH 26e self-launching motor glider, removed from the glider and mounted on a test stand for maintenance at the Alexander Schleicher GmbH & Co in Poppenhausen, Germany. Counter-clockwise from top left: propeller hub, mast with belt guide, radiator, Wankel engine, muffler shroud.

Wankel engine

Another promising design for aircraft use was the Wankel rotary engine. The Wankel engine is about one half the weight and size of a traditional four stroke cycle piston engine of equal power output, and much lower in complexity. In an aircraft application, the power to weight ratio is very important, making the Wankel engine a good choice. Because the engine is typically constructed with an aluminium housing and a steel rotor, and aluminium expands more than steel when heated, unlike a piston engine, a Wankel engine will not seize when overheated. This is an important safety factor for aeronautical use. Considerable development of these designs started after World War II, but at the time the aircraft industry favored the use of turbine engines. It was believed that turbojet or turboprop engines, could power all aircraft, from the largest to smallest designs. The Wankel engine did not find many applications in aircraft, but was used by Mazda in a popular line of sports cars. Recently, the Wankel engine has been developed for use in motor gliders where the small size, light weight, and low vibration are especially important.[5]

Wankel engines are becoming increasingly popular in homebuilt experimental aircraft, due to a number of factors. Most are Mazda 12A and 13B engines, removed from automobiles and converted to aviation use. This is a very cost-effective alternative to certified aircraft engines, providing engines ranging from 100 to 300 horsepower at a fraction of the cost of traditional engines. These conversions first took place in the early 1970s, and with hundreds or even thousands of these engines mounted on aircraft, as of 10 December 2006 the National Transportation Safety Board has only 7 reports of incidents involving aircraft with Mazda engines, and none of these is of a failure due to design or manufacturing flaws. During the same time frame, they have reports of several thousand reports of broken crankshafts and connecting rods, failed pistons and incidents caused by other components which are not found in the Wankel engines. Rotary engine enthusiasts refer to piston aircraft engines as "Reciprosaurs," and point out that their designs are essentially unchanged since the 1930s, with only minor differences in manufacturing processes and variation in engine displacement.

Peter Garrison, contributing editor for Flying magazine, has said that "the most promising engine for aviation use is the Mazda rotary." Garrison lost an airplane which he had designed and built (and missed death literally by inches), when a piston-powered plane had engine failure and crashed into Garrison's plane, which was waiting to take off.

Diesel engine

The diesel engine is another engine design that has been examined for aviation use. In general diesel engines are more reliable and much better suited to running for long periods of time at medium power settings—this is why they are widely used in trucks for instance. Several attempts to produce diesel aircraft engines were made in the 1930s but, at the time, the alloys were not up to the task of handling the much higher compression ratios used in these designs. They generally had poor power-to-weight ratios and were uncommon for that reason. Improvements in diesel technology in automobiles (leading to much better power-weight ratios), the diesel's much better fuel efficiency (particularly compared to the old designs currently being used in light aircraft) and the high relative taxation of gasoline compared to diesel in Europe have all seen a revival of interest in the concept. As of May 2004 one manufacturer, Thielert Aircraft Engines, is already selling certified diesel aircraft engines for light aircraft, and other companies have alternative designs under development. It remains to be seen whether these new designs will succeed in the marketplace but they potentially represent the biggest change in light aircraft engines in decades.

Angle of attack

In this diagram, the black lines represent the flow of the wind. The wing is shown end on. The angle α is the angle of attack.

Angle of attack (**AOA**, α, Greek letter alpha) is a term used in aerodynamics to describe the angle between the airfoil's chord line and the direction of airflow wind, effectively the direction in which the aircraft is currently moving. It can be described as the angle between where the wing is *pointing* and where it is *going*.

The amount of lift generated by a wing is directly related to the angle of attack, with greater angles generating more lift (and more drag). This remains true up to the stall point, where lift starts to decrease again because of flow separation.

Planes flying at high angles of attack can suddenly enter a stall if, for example, a strong wind gust changes the direction of the relative wind. Also, to maintain a given amount of lift, the angle of attack must be increased as speed through the air decreases. This is why stalling is an effect that occurs more frequently at low speeds.

Nonetheless, a wing (or any other airfoil) can stall at any speed. Planes that already have a high angle of attack, for example because they are pulling g or a heavy payload, will stall at speed well above the normal stall speed, since only a small increase in the angle of attack will take the wing above the critical angle.

The critical angle is typically around 15° for most airfoils. Using a variety of additional aerodynamic surfaces — known as high-lift devices — like leading edge extensions(leading edge wing root extensions), fighter aircraft have increased the potential flyable alpha from about 20° to over 45°, and in some designs, 90° or more. That is, the plane remains flyable when the wing's chord is perpendicular to the direction of motion.

Some aircraft are equipped with a built-in flight computer that automatically prevents the plane from lifting its nose any further when the maximum angle of attack is reached, in spite of pilot input. This is called the *angle of attack limiter* or *alpha limiter*. The pilot may disengage the alpha limiter at any time, thus allowing the plane to perform tighter turns (but with considerably higher risk of going into a stall). A famous military example of this is Pugachev's Cobra, a maneuver which has only been performed by the MiG-29, the Su-27/Su-33 family and some prototype Western aircraft, although some consider the F-15/F-16 to be capable if really pushed. Modern airliners which limit the angle of attack by means of computers include the Airbus 320, 330, 340 and 380 series. Currently, the highest angle of attack recorded for a duration of more than 10 seconds is 89.8, performed in the Russian Su-35 (Flanker-E)/Su-37 (Flanker-F) family. The F-35 is believed to be able to perform in even higher angles of attack for prolonged periods of time.

Sailing

In sailing, the **angle of attack** is the angle between a mid-sail and the direction of the wind. The physical principles involved are the same as for aircraft. See points of sail.

Aspect ratio (wing)

In aerodynamics, the **aspect ratio** is an airplane's wing's span divided by its standard mean chord (SMC). It can be calculated more easily as span squared divided by wing area:

Informally, a "high" aspect ratio indicates long, narrow wings, whereas a "low" aspect ratio indicates short, stubby wings.

Aspect ratio is a powerful indicator of the general performance of a wing. Wingtip vortices greatly deteriorate the performance of a wing, and by reducing the amount of wing tip area, making it skinny or pointed for instance, you reduce the amount of energy lost to this process, induced drag. This is why high performance gliders have very long, skinny wings; with no engine power, they must be as efficient as possible in every respect in order to stay aloft.

Why don't all aircraft have high aspect-ratio wings? There are several reasons:

- **Structural**: the deflection along a high aspect-ratio wing tends to be much higher than for one of low aspect ratio, thus the stresses and consequent risk of fatigue failures are higher - particularly with swept-wing designs.
- **Maneuverability**: a high aspect-ratio wing will have a lower roll rate than one of low aspect ratio, due to the higher moment arm on the drag and greater moment of inertia, thus rendering them unsuitable for fighter aircraft.
- **Pitch Stability** - low aspect ratio wings tend to be more naturally stable than high-aspect ratios. This confers handling advantages, especially at slow speeds. You can try this at home with paper planes - make one with the paper in landscape and one in

portrait - you will see how the higher AR plane will tend to pitch up or down much more than the low AR one.

- **Parasite Drag** - While high aspect wings create less induced drag, they create more parasite drag, (drag due to shape, frontal area, and friction). This is why aircraft that fly at high speed and low altitudes (fighter-bombers and racing planes) typically have low aspect wings. Some aircraft use a swing-wing to gain the advantages of both high and low aspect wings at the cost of increased weight and complexity.
- **Practicality** - low aspect ratios have a greater useful internal volume, which can be used to house the fuel tanks, retractable landing gear and other systems.

The high aspect ratio wing of a USAF B52 bomber

High aspect ratio wings abound in nature. Most birds which have to cover long distances e.g. on migratory routes have a high aspect ratio, and with tapered or elliptical wingtips. This is particularly noticeable on soaring birds such as albatrosses and eagles. In addition, the V-formation (echelon) often seen in flights of geese, ducks and other migratory birds can be considered to act as a single swept wing with a very high aspect ratio - the vortices shed by the lead bird are smoothly transferred to the next and so on. This confers a huge efficiency advantage to the flight as a whole - perhaps as much as a 100% improvement compared to a single bird in flight. Note that the usual common explanation of the V-formation - that following birds are "shielded" from air resistance by the bird in front - may be misleading. While birds do "take turns" at being the lead bird, it is probably to give those at the tips a rest - they are the ones that will experience the most drag when the vortices are finally shed. However, the full explanation of this behaviour is still the subject of research and debate; scientists do not claim to have fully understood the phenomenon yet.

Navier-Stokes equations

The **Navier-Stokes equations**, named after Claude-Louis Navier and George Gabriel Stokes, are a set of equations which describe the motion of fluid substances such as liquids and gases. These equations establish that changes in momentum in infinitesimal volumes of fluid are simply the sum of dissipative viscous forces (similar to friction), changes in pressure, gravity, and other forces acting inside the fluid: an application of Newton's second law.

They are one of the most useful sets of equations because they describe the physics of a large number of phenomena of academic and economic interest. They may be used to model weather, ocean currents, water flow in a pipe, flow around an airfoil (wing), and motion of stars inside a galaxy. As such, these equations in both full and simplified forms, are used in the design of aircraft and cars, the study of blood flow, the design of power stations, the analysis of the effects of pollution, etc. Coupled with Maxwell's equations they can be used to model and study magnetohydrodynamics.

The Navier-Stokes equations are differential equations which, unlike algebraic equations, do not explicitly establish a relation among the variables of interest (e.g. velocity and pressure), rather they establish relations among the rates of change. For example, the Navier-Stokes equations for simple case of an ideal fluid (inviscid) can state that acceleration (the rate of change of velocity) is proportional to the gradient (a type of multivariate derivative) of pressure.

Contrary to what is normally seen in classical mechanics, the Navier-Stokes equations dictate not position but rather velocity. A solution of the Navier-Stokes equations is called a velocity field or flow field, which is a description of the velocity of the fluid at a given point in space and time. Once the velocity field is solved for, other quantities of interest (such as flow rate, drag force, or the path a "particle" of fluid will take) may be found.

Properties

Nonlinearity

The Navier-Stokes equations are nonlinear partial differential equations in almost any real situation (an exception is creeping flow). The nonlinearity makes most problems difficult or impossible to solve and is part of the cause of turbulence.

The nonlinearity is due to convective acceleration, which is an acceleration associated with the change in velocity over position. Hence, any convective flow, whether turbulent or not, will involve nonlinearity, an example of convective but laminar (nonturbulent) flow would be the passage of a viscous fluid (for example, oil) through a small converging nozzle. Such flows, whether exactly solvable or not, can often be thoroughly studied and understood.

Turbulence

Turbulence is the time dependent chaotic behavior seen in many fluid flows. It is generally believed that it is due to the inertia of the fluid as a whole: the culmination of time dependent and convective acceleration; hence flows where inertial effects are small tend to be laminar (the Reynolds number quantifies how much the flow is affected by inertia). It is believed, though not known with certainty, that the Navier-Stokes model turbulence properly.

For essentially all turbulent situations, such as global weather systems like El Niño or the aerodynamics of a wing, solutions of the Navier-Stokes equations must be found with the help of computers. A variety of computer programs (both commercial and academic) have been developed to solve the Navier-Stokes equations using various numerical methods. This approach is collectively called computational fluid dynamics or CFD. Though CFD will theoretically work on any flow, in practice **many** common flows (such as flow over a wing) contain so much detail that no computer can handle in any reasonable amount of time.

Even though turbulence is an everyday experience, it is extremely difficult to find solutions, quantify, or in general characterize. A $1,000,000 prize was offered in May 2000 by the Clay Mathematics Institute to whoever makes preliminary progress toward a mathematical theory which will help in the understanding of this phenomenon.

Applicability

Together with supplemental equations (for example, conservation of mass) and well formulated boundary conditions, the Navier-Stokes equations seem to model fluid motion accurately; even turbulent flows seem (on average) to agree with real world observations.

The Navier-Stokes equations assume that the fluid being studied is a continuum. At very small scales or under extreme conditions, real fluids made out of discrete molecules will produce results different from the continuous fluids modeled by the Navier-Stokes equations. Depending on the Knudsen number of the problem, statistical mechanics or possibly even molecular dynamics may be a more appropriate approach.

Another limitation is very simply the complicated nature of the equations. Time tested formulations exist for common fluid families, but the application of the Navier-Stokes equations to less common families tends to result in very complicated formulations which are an area of current research. For this reason, the Navier-Stokes equations are usually written for Newtonian fluids.

Derivation and description

The derivation of the Navier-Stokes equations begins with the conservation of mass, momentum, and energy being written for an arbitrary control volume. The most general form of the Navier-Stokes equations ends up being:

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -\nabla p + \nabla\cdot\mathbb{T} + \mathbf{f}$$

This is a statement of the conservation of momentum in a fluid, it is an application of Newton's second law to a continuum. This equation is often written using the substantive derivative, making it more apparent that this is a statement of Newton's law:

$$\rho\frac{D\mathbf{v}}{Dt} = -\nabla p + \nabla\cdot\mathbb{T} + \mathbf{f}$$

The right side of the equation is a summation of body forces. ∇p is a pressure gradient and arises from normal stresses that turn up in any flow. $\nabla\cdot\mathbb{T}$ is representative of shear forces in the fluid, normally viscous effects. $\mathbf{f}$ represents "other" forces, such as gravity.

The shear stress term $\nabla\cdot\mathbb{T}$ contains too many unknowns, hence the general form above isn't directly applicable to any problem. For this reason, assumptions on the specific shear stress behavior of a fluid are made (based on natural observations) and applied in order to specify this quantity in terms of familiar variables, such as velocity. For example, this term becomes the useable quantity $\mu\nabla^2\mathbf{v}$ when the fluid is assumed incompressible and Newtonian.

The Navier-Stokes equations are strictly a statement of the conservation of momentum. In order to fully describe fluid flow, more information is needed (how much depends on the assumptions made), this may include the conservation of mass, the conservation of energy, or an equation of state.

Regardless of the flow assumptions, a statement of the conservation of mass is nearly always necessary. This is achieved through the continuity equation, given in its most general form as:

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0$$

Incompressible flow of Newtonian fluids

The vast majority of work on the Navier-Stokes equations is done under an incompressible flow assumption for Newtonian fluids. The incompressible flow assumption typically holds well even when dealing with a "compressible" fluid, such as air at room temperature (even when flowing up to about Mach 0.3). Taking the incompressible flow assumption into account and assuming constant viscosity, the Navier-Stokes equations will read (in vector form):

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -\nabla p + \mu\nabla^2\mathbf{v} + \mathbf{f}$$

f represents "other" body forces, such as gravity or centrifugal force. It's well worth observing the meaning of each term:

$$\overbrace{\rho\Big(\underbrace{\frac{\partial \mathbf{v}}{\partial t}}_{\text{Unsteady acceleration}} + \underbrace{\mathbf{v}\cdot\nabla\mathbf{v}}_{\text{Convective acceleration}}\Big)}^{\text{Inertia}} = \underbrace{-\nabla p}_{\text{Pressure gradient}} + \underbrace{\mu\nabla^2\mathbf{v}}_{\text{Viscosity}} + \underbrace{\mathbf{f}}_{\text{Other forces}}$$

Note that only the convective terms are nonlinear for incompressible Newtonian flow. The convective acceleration is an acceleration caused by a (possibly steady) change in velocity over *position*, for example the speeding up of fluid entering a converging nozzle. Though individual fluid particles are being accelerated and thus are under unsteady motion, the flow field (a velocity distribution) will not necessarily be time dependent.

Another important observation is that the viscosity is represented by the laplacian of the velocity field. This implies that Newtonian viscosity is **diffusion of momentum**, this works in much the same way as the diffusion of heat seen in the heat equation (which also involves the Laplacian).

If temperature effects are also neglected, the only "other" equation needed is the continuity equation. Under the incompressible assumption, density is a constant and it follows that the equation will simplify to:

$$\nabla\cdot\mathbf{v} = 0$$

This is more specifically a statement of the conservation of volume (see divergence). If temperature effects are additionally assumed negligible, this is the only supplemental equation needed.

These equations are commonly used in 3 coordinates systems: Cartesian, cylindrical, and spherical. The Cartesian equations follow directly from the vector equation above, obtaining equations in other coordinate systems will require a change of variables.

Cartesian coordinates

Writing the vector equation explicitly,

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right) = -\frac{\partial p}{\partial x} + \mu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) + \rho g_x$$
$$\rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right) = -\frac{\partial p}{\partial y} + \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right) + \rho g_y$$
$$\rho\left(\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right) = -\frac{\partial p}{\partial z} + \mu\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right) + \rho g_z$$

Note that gravity has been accounted for as a body force, and the values of g_x,g_y,g_z will depend on the orientation of gravity with respect to the chosen set of coordinates.

The continuity equation reads:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0$$

Note that the velocity components (the dependent variables to be solved for) are u, v, w. This system of four equations compromises the most commonly used and studied form. Though comparatively more compact then other representations, this is a nonlinear system of partial differential equations for which solutions are difficult to obtain.

Cylindrical coordinates

A change of variables on the Cartesian equations will yield the following momentum equations for *r*, θ, and *z*:

$$\rho\left(\frac{\partial u_r}{\partial t} + u_r\frac{\partial u_r}{\partial r} + \frac{u_\theta}{r}\frac{\partial u_r}{\partial \theta} + u_z\frac{\partial u_r}{\partial z} - \frac{u_\theta^2}{r}\right) = -\frac{\partial p}{\partial r} + \mu\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_r}{\partial r}\right) + \frac{1}{r}\right.$$

$$\rho\left(\frac{\partial u_\theta}{\partial t} + u_r\frac{\partial u_\theta}{\partial r} + \frac{u_\theta}{r}\frac{\partial u_\theta}{\partial \theta} + u_z\frac{\partial u_\theta}{\partial z} + \frac{u_r u_\theta}{r}\right) = -\frac{1}{r}\frac{\partial p}{\partial \theta} + \mu\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_\theta}{\partial r}\right)\right.$$

$$\rho\left(\frac{\partial u_z}{\partial t} + u_r\frac{\partial u_z}{\partial r} + \frac{u_\theta}{r}\frac{\partial u_z}{\partial \theta} + u_z\frac{\partial u_z}{\partial z}\right) = -\frac{\partial p}{\partial z} + \mu\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_z}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}\right.$$

The gravity components will generally not be constants, however for most applications either the coordinates are chosen so that the gravity components are constant or else it is assumed that gravity is counteracted by a pressure field (for example, flow in horizontal pipe is treated normally without gravity and without a vertical pressure gradient). The continuity equation is:

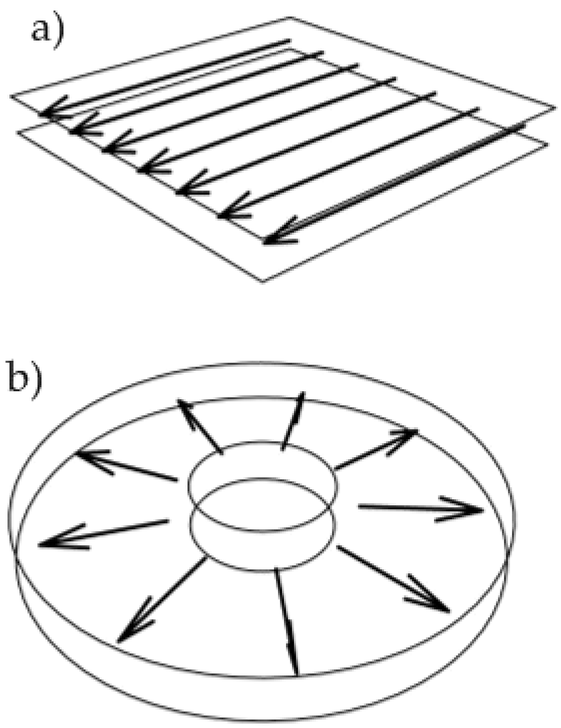
$$\frac{1}{r}\frac{\partial}{\partial r}(r u_r) + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z} = 0.$$

This cylindrical representation of the incompressible Navier-Stokes equations is the second most commonly seen (the first being Cartesian above).

Application to specific problems

The Navier-Stokes equations, even when written explicitly for specific fluids, are rather generic in nature and their proper application to specific problems can be very diverse. This is partly because there is an enormous variety of problems that may be modeled, ranging from as simple as the distribution of static pressure to as complicated as multiphase flow driven by surface tension.

Generally, application to specific problems begins with some flow assumptions and initial/boundary condition formulation, this may be followed by scale analysis to further simplify the problem. For example, after assuming steady, parallel, one dimensional, nonconvective pressure driven flow between parallel plates, the resulting scaled (dimensionless) boundary value problem is:



Visualization of a) parallel flow and b) radial flow.

$$\frac{d^2u}{dy^2} = -1 \quad ; \quad u(0) = u(1) = 0$$

The boundary condition is the no slip condition. This problem is easily solved for the flow field:

$$u(y) = \frac{1}{2}(y - y^2)$$

From this point onward more quantities of interest can be easily obtained, such as viscous drag force or net flow rate.

Difficulties may arise when the problem becomes slightly more complicated. A seemingly modest twist on the parallel flow above would be the *radial* flow between parallel plates; this involves convection and thus nonlinearity. The velocity field may be represented by a function $f(z)$ that must satisfy:

$$\frac{d^2f}{dz^2} + Rf^2 = -1 \quad ; \quad f(0) = f(1) = 0$$

This ordinary differential equation is what is obtained when the Navier-Stokes equations are written and the flow assumptions applied (additionally, the pressure gradient is solved for). The nonlinear term makes this a very difficult problem to solve analytically (a lengthy implicit solution may be found which involves elliptic integrals and roots of cubic polynomials). Issues with the actual existence of solutions arise for $R > 22.609$ (approximately), the parameter R being similar to the Reynolds number. This is an example of flow assumptions losing their applicability, and an example of the difficulty in "high" Reynolds number flows.

Wake turbulence

Wake turbulence is turbulence that forms behind an aircraft as it passes through the air. This turbulence includes various components, the most important of which are wingtip vortices and jetwash. Jetwash refers simply to the rapidly moving air expelled from a jet engine; it is extremely turbulent, but of short duration. Wingtip vortices, on the other hand, are much more stable and can remain in the air for up to two minutes after the passage of an aircraft. Wingtip vortices make up the primary and most dangerous component of wake turbulence.

Wake turbulence is especially hazardous during the landing and take off phases of flight, for two reasons. The first is that during take-off and landing, aircraft operate at low speeds and high angle of attack. This flight attitude maximizes the formation of dangerous wingtip vortices. Secondly, takeoff and landing are the times when a plane is operating closest to its stall speed and to the ground - meaning there is little margin for recovery in the event of encountering a different aircraft's wake turbulence.

This picture from a NASA study on wingtip vortices clearly illustrates the power of this wake turbulence component.

Fixed wing - Level flight

At altitude, vortices sink at a rate of 300 to 500 feet per minute and stabilize about 500 to 900 feet below the flight level of the generating aircraft. For this reason, aircraft operating greater than 2,000 feet above the terrain, are not considered at risk.

Helicopters

Helicopters also produce wake turbulence. Helicopter wakes may be of significantly greater strength than those from a fixed wing aircraft of the same weight. The strongest wake can occur when the helicopter is operating at lower speeds (20 - 50 knots). Some mid-size or executive class helicopters produce wake as strong as that of heavier helicopters. This is because two blade main rotor systems, typical of lighter helicopters, produce stronger wake than rotor systems with more blades.

Parallel or crossing runways

During takeoff and landing, an aircraft's wake sinks toward the ground and moves laterally away from the runway when the wind is calm. A 3 to 5 knot crosswind will tend to keep the upwind side of the wake in the runway area and may cause the downwind side to drift toward another runway. Since the wingtip vortices exist at the outer edge of an airplane's wake, this can be dangerous.

Hazard avoidance

Wake vortex separation

ICAO mandates separation minima based upon wake vortex categories that are, in turn, based upon the Maximum Take Off Mass (MTOM) of the aircraft.

These minima are categorised are as follows:

- Light - MTOM of 7,000kg or less;

- Medium - MTOM of greater than 7,000kg, but less than 136,000kg;
- Heavy - MTOM of 136,000kg or greater.

There are a number of separation criteria for take-off, landing and en-route phases of flight based upon these categories. Air Traffic Controllers will sequence aircraft making instrument approaches with regard to these minima. Aircraft making a visual approach are advised of the relevant recommended spacing and are expected to maintain their own separation.

Common minima are:

Take-off

An aircraft of a lower wake vortex category must not be allowed to take off less than two minutes behind an aircraft of a higher wake vortex category. If the following aircraft does not start its take off roll from the same point as the preceding aircraft, this is increased to three minutes.

Landing

Preceding aircraft	**Following aircraft**	**Minimum radar separation**
A380-800	A380-800	4 nm
	Non-A380-800 Heavy	6 nm
	Medium	8 nm
	Light	10 nm
Heavy	Heavy	4 nm
	Medium	5 nm
	Light	6 nm
Medium	Light	5 nm

Staying on or above leader's glide path

Incident data shows that the greatest potential for a wake vortex incident occurs when a light aircraft is turning from base to final behind a heavy aircraft flying a straight-in approach. Light aircraft pilots must use extreme caution and intercept their final approach path above or well behind the heavier aircraft's path. When a visual approach following a preceding aircraft is issued and accepted, the pilot is required to establish a safe landing interval behind the aircraft s/he was instructed to follow. The pilot is responsible for wake turbulence separation. Pilots must not decrease the separation that existed when the visual approach was issued unless they can remain on or above the flight path of the preceding aircraft.

Warning signs

Any uncommanded aircraft movements (e.g., wing rocking) may be caused by wake. This is why maintaining situation awareness is so critical. Ordinary turbulence is not unusual, particularly in the approach phase. A pilot who suspects wake turbulence is affecting his or her aircraft should get away from the wake, execute a missed approach or go-around and be prepared for a stronger wake encounter. The onset of wake can be insidious and even surprisingly gentle. There have been serious accidents where pilots have attempted to salvage a landing after encountering moderate wake only to encounter severe wake turbulence that they were unable to overcome. Pilots should not depend on any aerodynamic warning, but if the onset of wake is occurring, immediate evasive action is vital.

Accidents/incidents due to wake turbulence

- On June 8, 1966 an XB-70, collided with an F-104. Though the true cause of the collision is unknown, it is believed that due to the XB-70 being designed to have an enhanced wake turbulence to increase lift, the F-104 moved too close, therefore getting caught in the vortex and colliding the wing.
- Delta Air Lines Flight 9570 crashed at the Greater Southwest International Airport on 30 May 1972 while performing "touch and go" landings behind a DC-10. This crash prompted the FAA to create news rules for minimum following separation from "heavy" aircraft.
- A chartered aircraft with 5 onboard, including In-N-Out Burger's president, Rich Snyder, at John Wayne International Airport on December 15, 1993. The aircraft followed in a Boeing 757 for landing, became caught in its wake turbulence, rolled into a deep descent and crashed.
- USAir Flight 427 crashed near Pittsburgh, Pennsylvania in 1994. This accident was believed to involve wake turbulence, though the primary cause was a defective rudder control component (see main article).
- American Airlines Flight 587 crashed into the Belle Harbor neighborhood of Queens, New York shortly after takeoff from John F. Kennedy International Airport on November 12, 2001. This accident was attributed to pilot error in the presence of wake turbulence that resulted in rudder failure and subsequent separation of the vertical stabilizer.

Measurement

Wake turbulence can be measured using several techniques. A high-resolution technique is doppler lidar, a solution now commercially available. Techniques using optics can use the effect of turbulence on refractive index (optical turbulence) to measure the distortion of light that passes through the turbulent area and indicate the strength of that turbulence.

Audibility

Wake turbulence can occasionally, under the right conditions, be heard by ground observers. On a still day, heavy jets flying low and slow on landing approach may produce wake turbulence that is heard as a dull roar/whistle. Often, it is first noticed some seconds after the direct noise of the passing aircraft has diminished. The sound then gets louder, sometimes becoming as loud as was the original direct sound of the aircraft. Nevertheless, being highly directional, wake turbulence sound is easily perceived as originating a considerable distance behind the aircraft, its apparent source moving across the sky just as the aircraft did. It can persist for 30 seconds or more, continually changing timbre, sometimes with swishing and cracking notes, until it finally dies away.

In Popular Culture

In the movie *Top Gun*, Lieutenant Pete Mitchell, played by Tom Cruise, suffers 2 "jet washes" or Wake Turbulences. The 1st one being during a training mission, and he is caught in Tom

Kazansky's jet turbulence. (Kazansky is played by Val Kilmer.) In the 1st jet-wash, Mitchell loses his RIO and best friend, "Goose" as they eject out of the plane. In the second jet-wash, he is with "Merlin" and they are caught in a bogey's jet wash. Mitchell recovers from the turbulence but is shaken up.

Wing

A **wing** is a surface used to produce lift and therefore flight, for travel in the air or another gaseous medium. The wing shape is usually an airfoil. The first use of the word was for the foremost limbs of birds, but has been extended to include the wings of insects, bats and pterosaurs and also man-made devices.

A wing is a device for generating lift. Its *aerodynamic quality*, expressed as a Lift-to-drag ratio, can be up to 60 on some gliders. This means that a significantly smaller thrust force can be applied to propel the wing through the air in order to obtain a specified lift.

Use

A common use of wings is in flight, using forward motion to create vertical lift, but wings are also used to produce downforce holding racing cars on the ground. A sail boat moves by using sails and a keel like a vertical wings to produce lift (in the horizontal plane).

Artificial wings

Terms used to describe aircraft wings

- Leading edge: the front edge of the wing
- Trailing edge: the back edge of the wing
- Span: distance from wing tip to wing tip
- Chord: distance from wing leading edge to wing trailing edge, usually measured parallel to the long axis of the fuselage
- Aspect ratio: ratio of span to standard mean chord
- Aerofoil (or Airfoil in US English): the shape of the top and bottom surfaces when viewed as cross sections cut from leading edge to trailing edge.
- Sweep angle: the angle between the perpendicular to the design centreline of the wing in the wing plane, and either the leading edge or ¼ chord line.
- Twist: gradual change of the airfoil (*aerodynamic twist*) and/or angle of incidence of the wing cross-sections (*geometrical twist*) along the span.

Design features

Aeroplane wings may feature some of the following:

- A rounded (rarely sharp) leading edge cross-section
- A sharp trailing edge cross-section
- Leading-edge devices such as slats, slots, or extensions
- Trailing-edge devices such as flaps
- Ailerons (usually near the wingtips) to provide roll control
- Spoilers on the upper surface to disrupt lift and additional roll control

- Vortex generators to help prevent flow separation
- Wing fences to keep flow attached to the wing
- Dihedral, or a positive wing angle to the horizontal. This gives inherent stability in roll. Anhedral, or a negative wing angle to the horizontal, has a destabilising effect
- Folding wings allow more aircraft to be carried in the confined space of the hangar of an aircraft carrier.

The wings of a Boeing 737-800 equipped with performance-enhancing winglet

The wing of a landing BMI Airbus A319-100. The slats at the leading edge and the flaps at the trailing edge are extended.

Wing types

- Swept wings are wings that are bent back at an angle, instead of sticking straight out from the fuselage.

- Forward-swept wings are bent forward, the reverse of a traditional swept wing. Forward swept wings have been used in some two seat gliders, and in the experimental X-29.
- Elliptical wings (technically wings with an elliptical lift distribution) are theoretically optimum for efficiency at subsonic speeds. A good example of this wing type can be seen on the British Supermarine Spitfire World War II fighter aircraft.
- Delta wings have reasonable performance at subsonic and supersonic speeds and are good at high angles of attack. For examples see the F-102, F-106, Avro Vulcan and B-58.
- Waveriders are efficient supersonic wings that take advantage of shock waves. For an example, see the XB-70.
- Rogallo wings are two partial cone sections arranged with the apexes together and the convex side up. One of the simplest wings to construct using cloth or other membrane material and a frame.
- Swing-wings (or variable geometry wings) are able to move in flight to give the benefits of dihedral and delta wing. Although they were originally proposed by German aerodynamicists during the 1940s, they are now only found on military aircraft such as the Grumman F-14, Panavia Tornado, General Dynamics F-111, B-1 Lancer, Tupolev Tu-160, MiG-23 and Sukhoi Su-24.
- Ring wings are optimally loaded closed lifting surfaces with higher aerodynamic efficiency than planar wings having the same aspect-ratios. Other nonplanar wing systems display an aerodynamic efficiency intermediate between ring wings and planar wings.
- Oblique wing

Science of wings

The science of wings is one of the principal applications of the science of aerodynamics. However, at the simplest level, a wing operates by generating a greater pressure below the wing than above it. Pressure and force are directly related; higher pressure equals higher force. When enough force is applied below the wing, flight can take place. This, however is not due to the unequal path explanation commonly given. Despite popular belief, flat plates at angles and curved airfoils do not generate lift in different ways. They both, in fact, use Bernoulli's principle to generate lift. On a flat plate, there is a dividing point on the bottom surface, forward of which, the air must curl back and over the leading edge of the airfoil. This, in incompressible flow (a good approximation below Mach .6) results in the airflow over the top moving faster, at the same density, inducing a lower relative pressure. In aerodynamics the actual lift generated by an airfoil is found by integrating the area between the upper surface Cp and lower surface Cp from the front of the wing to the back. Each Cp is the relative pressure (pressure at some infinitely small point) minus the pressure of the free stream air (the condition before the air is affected by the airfoil) divided by dynamic pressure. [1]

It was at one point believed that lift over a wing would be produced through the molecules of air colliding with the surface, and imparting some of their momentum to it, having been deflected somewhat downward from their initial path. This theory, however, also proves, conclusively, that flight is impossible, as the lift created would be insignificant when compared to the drag induced. The equation arrived at through a Newtonian model of lift is $L=\rho*V^2*\sin^2\alpha*\cos\alpha$. (see Newton's Third Law). The fundamental flaw with this being that it does not account for upstream effects seen in airflows.

The science of wings applies in other areas beyond conventional fixed-wing aircraft, including:

- Helicopters which use a rotating wing with a variable pitch or angle to provide a directional force
- The space shuttle which uses its wings only for lift during its descent
- Some racing cars, especially Formula One cars, which use upside-down wings to give cars greater adhesion at high speeds
- Sailing boats which use sails as vertical wings with variable fullness and direction to move across water.

Structures with the same purpose as wings, but designed to operate in liquid media, are generally called fins or hydroplanes, with hydrodynamics as the governing science. Applications arise in craft such as hydrofoils and submarines. Sailing boats use both fins and wings.

Animal wings

Biologists believe that animal wings evolved at least four separate times, an example of convergent evolution.

- insect wings are believed to have evolved between 300 and 400 million years ago
- pterosaur wings at least 225 million years ago
- bird wings at least 150 million years ago
- bat wings about 55 million years ago.

Wings in these groups are analogous structures because they evolved independently rather than being passed from a common ancestor.

Wing loading

In aerodynamics, **wing loading** is the loaded weight of the aircraft divided by the area of the wing. It is broadly reflective of the aircraft's lift-to-mass ratio, which affects its rate of climb, load-carrying ability, and turn performance.

Typical wing loadings range from 20 lb/ft² (100 kg/m²) for general aviation aircraft, to 80 to 120 lb/ft² (390 to 585 kg/m²) for high-speed designs like modern fighter aircraft. The critical limit for bird flight is about 5 lb/ft² (25 kg/m²) (Meunier, 1951).

Wings generate lift owing to the motion of air over the wing surface. Larger wings move more air, so an aircraft with a large wing area relative to its mass (i.e., low wing loading) will have more lift at any given speed. Therefore, an aircraft with lower wing loading will be able to take off and land at a lower speed (or be able to take off with a greater load). It will also tend to have a superior rate of climb because less additional forward speed is necessary to generate the additional lift to increase altitude. It may also be capable of more efficient cruising performance because less thrust is required to maintain the lift for sustained flight.

Wing loading is also a useful measure of the general maneuvering performance of an aircraft. To turn, an aircraft must roll in the direction of the turn (i.e., in a right turn the pilot rolls to right wing low, left wing high), increasing the aircraft's bank angle. Banks lower the wing's lift against gravity and the nose moves toward the earth, so a control must be moved into the air stream (often the rudder) to keep the nose level. This increases drag. Also, turning is 'climbing around a circle' (wing lift is diverted to turning the aircraft) so the increase in wing angle of attack creates even more drag. The harder the turn attempted, the more drag. All this requires that power (thrust) be added to overcome the drag. The maximum rate of turn possible for a given aircraft design is limited by its wing size and available engine power: the maximum turn the aircraft can achieve and hold is its *sustained turn performance*. Aircraft with low wing loading tend to have superior sustained turn performance because they can generate more lift for a given quantity of engine thrust. Conversely, however, a large, lightly loaded wing will tend to have greater mass and inertia and create greater induced drag when the bank angle or angle of attack increases. The immediate turn position an aircraft can get into before drag seriously bleeds off speed is its *instantaneous turn performance*, its ability to rapidly change direction. An aircraft with a small, highly loaded wing may have superior instantaneous turn performance, but poor sustained turn performance: it reacts quickly to control input, but its ability to sustain a tight turn is limited. A classic example is the F-104 Starfighter, which has a very small wing. At the opposite end of the spectrum was the gigantic Convair B-36. Its large wings resulted in a low wing loading, and there are disputed claims that this made the bomber more agile than contemporary jet fighters at high altitude. A blended wing-fuselage design often helps to reduce wing loading; in such a design (such as that found on the F-16 Fighting Falcon or MiG-29 Fulcrum), the shape of the fuselage generates some aerodynamic lift itself, improving wing loading while maintaining high performance.

All else being equal, a larger wing generates more drag than a small one. The construction of a large wing also tends to be thicker, which further increases drag. This drag reduces the aircraft's acceleration, particularly at supersonic speeds. A smaller, thinner wing will (all else being equal) have less drag, making it more suitable for high-speed flight (albeit at the cost of higher take-off speeds and reduced turning performance).

Wing loading also affects *gust response*, the degree to which the aircraft is affected by turbulence and variations in air density. A highly loaded wing has more inertia and a small wing has less area on which a gust can act, both of which serve to smooth the ride. For high-speed, low-level flight (such as a fast low-level bombing run in an attack aircraft), a small, thin, highly loaded wing is preferable: aircraft with low wing loading are often subject to a rough, punishing ride in this flight regime. The F-15E Strike Eagle has been criticized for its ride quality, as have most delta wing aircraft (such as the Dassault Mirage III), which tend to have large wings and low wing loading.

A further complication with wing loading is that it is difficult to substantially alter the wing area of an existing aircraft design (although modest improvements are possible). As aircraft are developed they are prone to "*weight growth*" -- the addition of equipment and features that substantially increase the operating mass of the aircraft. An aircraft whose wing loading is moderate in its original design may end up with very heavy wing loading as new gear is added. Although engines can be replaced or upgraded for additional thrust, the effects on turning and takeoff performance resulting from higher wing loading are not so easily reconciled. This was a major reason for the well-known disparity between the World War II-vintage Supermarine Spitfire and Messerschmitt Bf 109. Earlier marks of the Messerschmitt design were significantly lighter than later ones as armament, armor, and equipment increased, and while improved engine power maintained the power-to-weight ratio, later models had such heavily loaded wings that their maneuverability suffered badly, eventually tilting the balance in favor of the Spitfire.

Wingtip vortices

Wingtip vortices stream from an F-15E as it disengages from a KC-10 Extender following midair refueling.

Wingtip vortices are regions of high vorticity which develop at the tip of a wing as it flies through the air (or potentially another fluid). Wingtip vortices are a form of induced drag, an essentially unavoidable side-effect of the wing generating lift. Designing a wing with a vortex of preferable shape is critically important in aerospace engineering. Wingtip vortices also form the major component of wake turbulence.

Cause and effects

As a wing flies through the air, it generates aerodynamic lift by creating a region of higher air pressure beneath the wing than above it, among other factors like air deflection for instance. It must be kept in mind that lift is a sum of forces not a single force. Fluids are forced to flow from high to low pressure and the relatively high pressure air below the wing tends to escape to the top of the wing. The air does not escape around the leading or trailing edge of the wing due to airspeed, but it can flow around the tip. Consequently, air flows from below the wing and out around the tip to the top of the wing in a circular fashion. This leakage will raise the pressure on top of the wing and lower the overall lift that the wing can produce. It also produces an emergent flow pattern with low pressure in the center surrounded by fast moving air with curved streamlines.

Wingtip vortices only affect the portion of the wing closest to the end. Thus, the longer a wing is, the smaller the affected fraction of it will be. As well, the shorter the chord of the wing, the less opportunity air will have to form vortices. This means that for an aircraft to be most efficient, it should have a very high aspect ratio. This is evident in the design of long-range airliners and gliders, where fuel efficiency is of critical importance. However, increasing the wingspan reduces the maneuverability of the aircraft, which is why combat and aerobatic planes usually feature short, stubby wings despite the efficiency losses.

Another method of reducing fuel consumption is use of winglets, as seen on a number of modern airliners such as the Airbus A340. Winglets work by forcing the vortex to move to the very tip of the wing and allowing the entire span to produce lift, thereby effectively increasing the aspect ratio of the wing. Winglets also change the pattern of vorticity in the core of the vortex pattern; spreading it out and reducing the kinetic energy in the circular air flow, which reduces the amount of fuel expended to perform work by the wing upon the spinning air. Winglets can yield very worthwhile economy improvements on long distance flights.

Since vortices cause a low-pressure area at their centre, sometimes water precipitates out to form clouds in the vortex cores, allowing wingtip vortices to be seen. This is most common on aircraft flying at high angles of attack, such as fighter aircraft pulling high g manoeuvres, or airliners landing.

Hazards

A NASA study on wingtip vortices produced these pictures of smoke in the wake of an aircraft, clearly illustrating the size and power of the vortices produced.

Wingtip vortices can also pose a severe hazard to light aircraft, especially during the landing and take off phases of flight. The intensity or strength of the vortex is a function of aircraft size, speed, and configuration (flap setting, etc.). The strongest vortices are produced by heavy aircraft, flying slowly, in a clean configuration; large jet aircraft can generate vortices which are larger than an entire small plane. These vortices can persist for several minutes, drifting with the prevailing wind. If a small plane is immediately preceded by a large aircraft on the runway, there is a high risk that the winds in a vortex will cause uncontrollable and sudden variations in altitude, possibly violently slamming the aircraft into the ground without warning. Worse, the circular nature of vortices can flip a small plane upside down. At the low altitudes involved with landing and takeoff, this is completely unrecoverable. The hazardous aspects of wingtip vortices are most often discussed in the context of wake turbulence.

Air-to-air missile

A US Navy VF-103 Jolly Rogers F-14 Tomcat fighter launches an AIM-54 Phoenix long-range air-to-air missile. Photo courtesy U.S. Navy Atlantic Fleet.

An **air-to-air missile** (**AAM**) is a guided missile fired from an aircraft for the purpose of destroying another aircraft. It is typically powered by one or more rocket motors, usually solid fuelled but sometimes liquid fuelled. Ramjet engines, as used on the MBDA Meteor (currently in development), are emerging as propulsion that will enable future medium-range missiles to maintain higher average speed across their engagement envelope.

Air-to-air missiles are broadly grouped into short-range missiles (also called "dogfight" or "within visual range" (WVR)) and medium- or long-range missiles (beyond visual range (BVR)). Short-range missiles tend to use infrared guidance, while medium- and long-range missiles rely upon some type of radar guidance (and sometimes inertial guidance).

Guidance

Guided missiles operate by detecting their target (usually by either radar or infrared methods, although rarely others such as laser guidance or optical tracking), and then "homing" in on the target on a collision course.

The target is usually destroyed or damaged by means of an explosive warhead, often throwing out fragments to increase the lethal radius, typically detonated by a proximity fuze (or impact fuze if it scores a direct hit).

Note that although the missile may use radar or infra-red guidance to home on the target, this does not necessarily mean that the same means is used by the launching aircraft to detect and track the target before launch. Infra-red guided missiles can be "slaved" to an attack radar in order to find the target and radar-guided missiles can be launched at targets detected visually or via an infra-red search and track (IRST) system, although they may require the attack radar to illuminate the target during part or all of the missile interception itself.

Radar guidance

Radar guidance is normally used for medium or long range missiles, where the infra-red signature of the target would be too faint for an infra-red detector to track. There are two major types of radar-guided missile - active and semi-active.

Radar guided missiles can be countered by rapid maneuvering (which may result in them "breaking lock", or may cause them to overshoot), deploying chaff or using electronic counter-measures.

Active radar homing

Active radar (AR)-guided missiles carry their own radar system to detect and track their target. However, the size of the radar antenna is limited by the small diameter of missiles, limiting its range which typically means such missiles have to use another method to get close to the target before turning their radar set on, often relying on inertial guidance systems.

Semi-active radar homing

Semi-active radar (SAR)-guided missiles are simpler and more common. They function by detecting the radar energy reflected from the target, the radar energy is emitted from the launch aircraft's own radar signal. However, this means the launch aircraft has to maintain a "lock" on the target (keep illuminating the target aircraft with its' own radar) until the missile makes the interception, limiting the attacking aircraft's ability to maneuver, which may be necessary should threats to the attacking aircraft appear. It also makes jamming the missile lock easier because the launching aircraft is further from the target than the missile, so the radar signal has to travel further and is greatly attenuated over the distance.

Beam riding

An early form of radar guidance was "beam-riding" (BR). In this method the attacking aircraft directed a narrow beam of radar energy at the target. The air-to-air missile was launched into the beam where sensors on the aft of the missile controlled the missile, keeping it within the beam. So long as the beam was kept on the target aircraft, the missile would ride the beam until making the interception. While simple in concept, the difficulty of simultaneously keeping the beam solidly on the target (which couldn't be relied upon to cooperate by flying straight and level), continuing to fly one's own aircraft, all the while keeping an eye out for enemy countermeasures, can be readily appreciated.

Although radar beam-riding air-to-air missiles are obsolete, the technology has since evolved toward laser-beam guided air-to-ground munitions, such as laser-guided bombs (LGB). These precision-strike munitions are sometimes called "smart weapons" by the press.

Infrared guidance

Infrared guided (IR) missiles home on the heat produced by an aircraft. Early infra-red detectors had poor sensitivity, so could only track the hot exhaust pipes of an aircraft. This meant an attacking aircraft had to maneuver to a position behind its target before it could fire an infra-red guided missile. This also limited the range of the missile as the infra-red signature soon become too small to detect with increasing distance and after launch the missile was playing "catch-up" with its target.

More modern infra-red guided missiles can detect the heat of an aircraft's skin, warmed by the friction of airflow, in addition to the fainter heat signature of the engine when the aircraft is seen from the side or head-on. This, combined with greater maneuverability, gives them an "all-aspect" capability, and an attacking aircraft no longer had to be behind its target to fire. Although launching from behind the target increases the probability of a hit, the launching aircraft usually has to be closer to the target in a tail-chase engagement.

An aircraft can defend against infra-red missiles by dropping flares that are hotter than the aircraft, so the missile homes in on the brighter, hotter target. Towed decoys and infra-red jammers can also be used. Some large aircraft and many combat helicopters make use of so called "hot brick" infra-red jammers, typically mounted near the engines. Current research is developing laser devices which can spoof or destroy the guidance systems of infra-redguided missiles.

However, the latest missiles such as the ASRAAM use an "imaging" infra-red seeker which "sees" the target (much like a digital video camera), and can distinguish between an aircraft and a point heat source such as a flare. They also feature a very wide detection angle, so the attacking aircraft does not have to be pointing straight at the target for the missile to lock on. The pilot can use a helmet mounted sight (HMS) and target another aircraft by looking at it, and then firing. This is called "off-boresight" launch. The Russian Su-27 is equipped with an infrared search and track (IRST) system with laser rangefinder for its HMS-guided missiles.

In order to maneuver sufficiently from a poor launch angle at short ranges to hit its target, missiles are now employing gas-dynamic flight control methods such as vectored thrust, which allow the missile to start turning "off the rail", before its motor has accelerated it up to high enough speeds for its small aerodynamic surfaces to be useful.

Electro-optical

A recent advancement in missile guidance is **electro-optical** imaging. The Israeli Python-5 has an electro-optical seeker that scans designated area for targets via optical imaging. Once a target is acquired, the missile will lock-on to it for the kill. Electro-optical seekers can be programmed to target vital area of an aircraft, such as the cockpit. Since it doesn't depend on the target aircraft's heat signature, it can be used against low-heat targets such as UAV's and cruise missiles.

Design

Air-to-air missiles are typically long, thin cylinders in order to reduce their cross section and thus minimize drag at the high speeds at which they travel.

At the front is the seeker, either a radar system, radar homer, or infra-red detector. Behind that lies the avionics which control the missile. Typically after that, in the centre of the missile, is the warhead, usually several kilogrammes of high explosive surrounded by metal that fragments on detonation (or in some cases, pre-fragmented metal).

The rear part of the missile contains the propulsion system, usually a rocket of some type. Dual-thrust solid-fuel rockets are common, but some longer-range missiles use liquid-fuel motors that can "throttle" to extend their range and preserve fuel for energy-intensive final maneuvering. Some solid-fuelled missiles mimic this technique with a second rocket motor which burns during the terminal homing phase. There are missiles in development, such as the MBDA Meteor, that "breathe" air (using a ramjet, similar to a jet engine) in order to extend their range.

Modern missiles use "low-smoke" motors - early missiles produced thick smoke trails, which were easily seen by the crew of the target aircraft alerting them to the attack and helping them determine how to evade it.

Missile range

Missiles are often cited with their maximum engagement range, which is very misleading. A missile's effective range is dependent on factors such as altitude, speed, position, and direction of target aircraft. For example the Vympel R-77 has stated range of 100 km. That's only true for a head-on, non-evading target at high altitude. At low altitude, the effective range is reduced by as much as 75%-80% to 20-25 km. If the target is taking evasive action, or in stern-chase position, the effective range is further reduced. See Air-to-Air missile non-comparison table for more information. The effective range of an air-to-air missile is known as the 'no-escape zone', noting the range at which the target can not evade the missile once launched.

Poorly-trained pilots and mercenary pilots more interested in collecting their pay and going home alive than dead, are known to fire their missiles at maximum-range engagement with poor results. In the 1998-2000 Eritrean-Ethiopian War, fighters from both sides shot over a dozen medium-range R-27 (*AA-10 Alamo*) missiles at distance with little effect. But when better-trained Ethiopian Su-27 pilots gave chase and attacked with short-range R-73 (*AA-11 Archer*) missiles, the results were often deadly to the Eritrean aircraft. [1]

Performance

A number of terms frequently crop up in discussions of air to air missile performance.

Launch success zone
: The Launch Success Zone is the range within which there is a high (defined) kill probability against a target that remains unaware of its engagement until the final moment. When alerted visually or by a warning system the target attempts a last ditch manoeuvre sequence.

F-Pole
: A closely related term is the F-Pole. This is the slant range between the launch aircraft and target, at the time of interception. The greater the F-Pole, the greater the confidence that the launch aircraft will achieve air superiority with that missile.

No-Escape Zone
: The No-Escape Zone is the zone within which there is a high (defined) kill probability against a target even if it has been alerted. This zone is defined as a conical shape with the tip at the missile launch. The cone's length and width are determined by the missile and seeker performance. A missile's speed, range and seeker sensitivity will mostly determine the length of this imaginary cone, while its agility (turn rate) and seeker complexity (speed of detection and ability to detect off axis targets) will determine the width of the cone.

Dogfight

Short-range air-to-air missiles used in "dogfighting" are usually classified into five "generations" according to the historical technological advances. Most of these advances were in infrared seeker technology (later combined with digital signal processing).

First generation

Early short-range missiles such as the early Sidewinders and Vympel K-13 (*AA-2 Atol*) had infrared seekers with a narrow (30 degree) field of views and required the attacker to position them self behind the target (rear aspect engagement). This meant the target aircraft only had to perform a slight turn to move outside the missile seeker field of view and cause the missile to lose track of the target ("break lock").[1]

Second generation

Second generation missiles utilized better seekers that improved the field of view to 45 degrees.

Third generation

This generation introduced "all aspect" missiles, because more sensitive seekers allowed the attacker to fire at a target which was side-on to itself, i.e. from *all aspects* not just the rear. This meant that while the field-of-view was still restricted to a fairly narrow cone, the attack at least did not have to be behind the target.[1]

Fourth generation

The Vympel R-73 (*AA-11 Archer*) entered service in 1985 and marked a new generation of dogfight missile. These missiles employed more advanced seeker technologies such focal plane arrays that improved resistance to infrared countermeasures (IRCM) such as flares and increased off-bore sight capability to in excess of 60 degrees, i.e. a 120 degree field of view.

To take advantage of the increased field-of-view that now exceeded the capabilities of most aircraft radars also meant that helmet mounted sights gained popularity.[2] Popularly know as "look-down-shoot-down", because the pilot would turn their head towards the target and the missile use this directional information in inertial guidance until the target was in the infrared seekers' field of view. The Israeli Python 4 was one example of a missile integrated with such a system.

These missiles are also much more agile, some by employing thrust vectoring (typically gimballed thrust).

Fifth generation

The latest generation of short-range missiles again defined by advances in seeker technologies, this time electro-optical imaging infrared (IIR) seekers that allow the missiles to "see" images rather than single "points" of infrared radiation (heat). The sensors combined with more powerful digital signal processing provide the following benefits:[2]

- greater infrared counter countermeasures (IRCCM) ability, by being able to distinguish aircraft from infrared countermeasures (IRCM) such as flares.
- greater sensitivity means greater range and ability to identify smaller low flying targets such as UAVs.
- more detailed target image allows targeting of more vulnerable parts of instead of just homing in on the brightest infrared source (aircraft exhaust).

Examples of fifth generation missiles include:

- AIM-132 ASRAAM – Britain (1998–)

- AIM-9X Sidewinder – USA (2003–)
- IRIS-T – German lead consortium (2005–)
- Python 5 – Israeli
- A-Darter (under development) – South Africa

List of missiles by country

For each missile, short notes are given, including an indication of its range and guidance mechanism.

Brazil

- Mectron MAA-1 Piranha - short range IR

France

- Matra R550 Magic - short-range, IR guided
- Matra Magic II - IR guided missile.
- Magic Super 530F/Super 530D - medium-range, radar-guided
- MBDA MICA - medium-range, IR or radar guided

Germany

- R4M rocket - first practical anti-aircraft rocket, used at the end of WW2
- Ruhrstahl X-4 - World War II design, first practical anti-aircraft missile, MCLOS, never saw service
- Henschel Hs 298 - World War II design, MCLOS, never saw service
- MBDA Meteor
- IRIS-T

European

- MBDA Meteor - medium range, active radar homing; design to replace AMRAAM
- IRIS-T - short range infrared homing; replacement for AIM-9 Sidewinder

India

- Astra missile (Undergoing developmental trials) - long range

Iraq

- Al Humurrabi- Long range, semi active radar

Israel

- Rafael Shafrir - first Israeli domestic AAM

- Rafael Shafrir 2 - improved Shafrir missile
- Rafael Python 3 - medium range IR-homing missile with all aspect capability [3]
- Rafael Python 4 - medium range IR-homing missile with HMS-guidance capability [4]
- Rafael Python 5 - improved Python 4 with electro-optical imaging seeker [5]
- Rafael Derby - Also known as the Alto, this is a medium-range, BVR active radar-homing missile [6]

Italy

- Alenia Aspide - Italian manufactured version of the AIM-7 Sparrow, based on the AIM-7E.

Japan

- AAM-3 - short-range Type 90 air-to-air missile
- AAM-4 - middle-range Type 99 air-to-air missile
- AAM-5 - short-range Type 04 air-to-air missile

Pakistan

- Sarab 1 - Pakistani version of Matra Magic Missile, Short Range Missile Project Cancelled due to unsatisfactory results (Ahmad Usman).
- SD 10 - Jointly Developed by China and Pakistan
- PL-9 - Jointly Developed by Pakistan and China.

People's Republic of China

- PL-1 - PRC version of the Soviet Kaliningrad K-5 (AA-1 Alkali), retired.
- PL-2 - PRC version of the Soviet Vympel K-13 (AA-2 Atoll), which was based on AIM-9B Sidewinder. [7] Retired & replaced by PL-5 in PLAAF service.
- PL-3 - updated version of the PL-2, did not enter service.
- PL-5 - updated version of the PL-2, known versions include: [8]
    - PL-5A - semi-active radar-homing AAM intended to replace the PL-2, did not enter service. Resembles AIM-9G in appearance.
    - PL-5B - IR version, entered service in 1990s to replace the PL-2 SRAAM. Limited off-boresight
    - PL-5C - Improved version comparable to AIM-9H or AIM-9L in performance
    - PL-5E - All-aspect attack version, resembles AIM-9P in appearance.
- PL-7 - PRC version of the IR-homing French R550 Magic AAM, did not enter service. [9]
- PL-8 - PRC version of the Israeli Rafael Python 3 [10]
- PL-9 - short range IR guided missile, marketed for export. One known improved version (PL-9C). [11]
- PL-10 - semi-active radar-homing medium-range missile based on the HQ-61 SAM, [12] often confused with PL-11. Did not enter service.
- PL-11 - medium-range air-to-air missile (MRAAM), based on the HQ-61C & Italian Aspide (AIM-7) technology. Limited service with J-8-B/D/H fighters. Known versions include: [13]
    - PL-11 - MRAAM with semi-active radar homing, based on the HQ-61C SAM and Aspide seeker technology, exported as FD-60 [14]
    - PL-11A - Improved PL-11 with better range, warhead, and seeker. The new seeker only requires fire-control radar guidance during the terminal stage, providing a basic LOAL (lock-on after launch) capability.

    - PL-11B - Also known as PL-11 AMR, improved PL-11 with AMR-1 active radar-homing seeker.
    - LY-60 - PL-11 adopted for navy ships for air-defense, sold to Pakistan but does not appear to be in service with the Chinese Navy. [15]
- PL-12 (SD-10) - medium-range active radar missile [16]
- TY-90 - light IR-homing air-to-air missile designed for helicopters [17]

Russia/Soviet

- Kaliningrad K-5 (NATO reporting name **AA-1 'Alkali'**) - beam-riding
- Vympel K-13 (NATO reporting name **AA-2 'Atoll'**) - short-range IR or SARH
- Kaliningrad K-8 (NATO reporting name **AA-3 'Anab'**) - IR or SARH
- Raduga K-9 (NATO reporting name **AA-4 'Awl'**) - IR or SARH
- Bisnovat R-4 (NATO reporting name **AA-5 'Ash'**) - IR or SARH
- Bisnovat R-40 (NATO reporting name **AA-6 'Acrid'**) - long-range IR or SARH
- Vympel R-23 (NATO reporting name **AA-7 'Apex'**) - medium-range SARAH or IR
- Molniya R-60 (NATO reporting name **AA-8 'Aphid'**) - short-range IR
- Vympel R-33 (NATO reporting name **AA-9 'Amos'**) - long range active radar
- Vympel R-27 (NATO reporting name **AA-10 'Alamo'**) - medium-range SARH or IR
- Vympel R-73 (NATO reporting name **AA-11 'Archer'**) - short-range IR
- Vympel R-77 (NATO reporting name **AA-12 'Adder'**) - medium-range active radar
- Vympel R-37 (NATO reporting name **AA-X-13 'Arrow'**) - long-range SARH or active radar
- Novator KS-172 AAM-L - extreme long range, inertial navigation with active radar for terminal homing

South Africa

- A-Darter - short range IR
- R-Darter - beyond visual range (BVR) radar-guided missile

United Kingdom

- Fireflash - short range beam-riding
- Firestreak - short range IR
- Red Top - short range IR
- Skyflash - medium-range radar-guided missile based on the AIM-7E2, said to have quick warm-up times of 1 to 2 seconds.
- AIM-132 ASRAAM - short range IR

United States

- AIM-4 Falcon - radar (later IR) guided
- AIM-7 Sparrow - medium range semi-active radar
- AIM-9 Sidewinder - short range IR
- AIM-54 Phoenix - long range, semi-active and active radar
- AIM-120 AMRAAM - medium range, active radar; replaces AIM-7 Sparrow

Air-to-surface missile

A pilot inspects an AGM-65 Maverick missile on his A-10 Thunderbolt II.

An **air-to-surface missile** (also, **air-to-ground missile**, **ASM** or **AGM**) is a missile designed to be launched from military aircraft (bombers, attack aircraft, fighter aircraft or other kinds) and strike ground targets on land, at sea, or both. They are similar to guided glide bombs but to be considered a missile, they usually contain some form of propulsion system. The two most common propulsion systems for air-to-surface missiles are rocket motors and jet engines. These also tend to correspond to the range of the missiles - short and long, respectively. Some Soviet air-to-surface missiles are powered by ramjets, giving them both long range and high speed.

Guidance for air-to-surface missiles is typically via laser guidance, infrared guidance, optical guidance or via GPS signals. The type of guidance depends on the type of target. Ships, for example, may be detected via passive or active radar, while this wouldn't work very well against land targets which typically don't contain such a large mass of metal surrounded by empty space.

There is some cross-over between air-to-surface missiles and surface-to-surface missiles. For example, there was an air-launched version of the Tomahawk missile, although this has been superseded by the AGM-86 ALCM. Other missiles used in both roles include the Penguin anti-ship missile and AGM-84 Harpoon anti-ship missile. Many air-to-surface missiles can be used against both ships and land targets, although some of them have to be modified to perform both roles effectively. For example, the Standoff Land Attack Missile is a land-attack version of the Harpoon.

One of the major advantages of air-to-surface missiles over other weapons available for aircraft to use to attack ground targets is the standoff distance they provide. This allows them to launch the weapons outside the most intense air defences around the target site. Most air-to-surface missiles are fire-and-forget in order to take most advantage of the standoff distance - they allow the launching platform to turn away after launch. Some missiles have enough range to be launched over the horizon. These missiles (typically either cruise or anti-ship missiles) need to be able to find and home in on the target autonomously.

Sub-categories of air-to-surface missiles include:

- air-launched anti-tank guided missiles (typically launched from helicopters)
- air-launched cruise missiles
- air-launched anti-ship missiles
- anti-radiation missiles

Typically, the higher and faster the launching aircraft is flying, the further away the missile's target can be. For long range missiles this difference can be small, but short range missiles (like the AGM-65 Maverick) often dramatically increase in range when launched at altitude.

These have been examples of air-launched ballistic missiles, but they are rare. Sometimes air-to-surface missiles are divided into the categories of *tactical* and *strategic*. Typically this indicates conventional explosive or small nuclear warhead (tactical) and large nuclear warhead (strategic).

List of air-to-surface missiles

Argentina

- MP-1000 Martín Pescador
- AS-25K

France

- MBDA AS 30
- MBDA Apache
- MBDA Exocet

France/UK

- Brimstone
- Storm Shadow

Germany

- Taurus KEPD 350
- AGM Armiger
- PARS-3
- Euromissile HOT

India

- BrahMos
- Nag

Iraq

- Al Quds

Norway

- Penguin missile
- Joint Strike Missile

Pakistan

- H2 Missile
- H4 Missile

- Baktar-Shikan
- Babur_missile
- Hafr

South Africa

- Mokopa

Sweden

- RBS 15

UK

- Blue Steel missile
- Brimstone missile
- Green Cheese missile
- AGM-48 Skybolt
- ALARM

USA

- AGM-12 Bullpup
- AGM-22
- AGM-28 Hound Dog
- AGM-45 Shrike
- AGM-48 Skybolt
- AGM-53 Condor
- AGM-62 Walleye
- AGM-63
- AGM-64 Hornet
- AGM-65 Maverick
- AGM-69 SRAM
- AGM-76 Falcon
- AGM-78 Standard ARM
- AGM-79 Blue Eye
- AGM-80 Viper
- AGM-83 Bulldog
- AGM-84 Harpoon
- AGM-86 CALCM
- AGM-87 Focus
- AGM-88 HARM
- AGM-112
- AGM-114 Hellfire
- AGM-122 Sidearm
- AGM-123 Skipper
- AGM-124 Wasp
- AGM-129 ACM
- AGM-130
- AGM-131 SRAM II
- AGM-136 Tacit Rainbow
- AGM-137 TSSAM
- AGM-142 Have Nap

- AGM-153
- AGM-154 JSOW
- AGM-158 JASSM
- AGM-159 JASSM

USSR/Russia

- AS-1 'Kennel' (KS-1 *Kometa*)
- AS-2 'Kipper' (K-10S *Yen*)
- AS-3 'Kangaroo' (H-20)
- AS-4 'Kitchen' (H-22 Burya)
- AS-5 'Kelt' (H-11/KSR-2)
- AS-6 'Kingfish' (H-26/KSR-5)
- AS-7 'Kerry' (H-66, H-23 *Grom*)
- AS-8 (9M114V *Sturm-V*)
- AS-9 'Kyle' (H-28)
- AS-10 'Karen' (H-25)
- AS-11 'Kilter' (H-58 *Izdeliye*)
- AS-12 'Kegler' (H-25MP, H-27PS)
- AS-13 'Kingbolt' (H-59 *Ovod*)
- AS-14 'Kedge' (H-29)
- AS-15 'Kent' (H-55/H-65S *Izdeliye*)
- AS-16 'Kickback' (H-15)
- AS-17 'Krypton' (H-31)
- AS-18 Kazoo (H-59M *Ovod-M*)
- AS-19 'Koala' (P-750 *Grom*)
- AS-X-19 'Koala' (3M25A *Meteorit-A*)
- AS-20 'Kayak' (H-35/H-37 *Uran*)
- AS-X-21 (Kh-90 *Gela*)

Chaff (radar countermeasure)

Modern US Navy RR-129 and RR-124 chaff countermeasures and containers

Chaff, originally called **Window** by the British, is a radar countermeasure in which aircraft or other targets spread a cloud of small, thin pieces of aluminium, metallised glass fibre or plastic, which either appears as a cluster of secondary targets on radar screens or swamps the screen with multiple returns.

Modern armed forces use chaff (in naval applications,for instance, using short-range SRBOC rockets) to distract radar-guided missiles from their targets. Most military aircraft and warships have chaff dispensing systems for self-defense. An intercontinental ballistic missile may release in its midcourse phase several independent warheads, a large number of decoys, and chaff.

Chaff can also be used to signal distress by an aircraft when communications are not functional. This has the same effect as an SOS, and can be picked up on radar. It is done by dropping chaff every 2 minutes.

World War II

The idea of using chaff was independently developed in the UK, Germany, and the United States.

As far back as 1937, R. V. Jones had suggested that a piece of metal foil falling through the air might create radar echoes. In early 1942, a TRE researcher named Joan Curran had investigated the idea and come up with a scheme for dumping packets of aluminum strips from aircraft to generate a cloud of false echoes.[1] The British referred to the idea as Window. Meanwhile in Germany, similar research had led to the development of Düppel. In the US, Fred Whipple developed a similar system (according to Harvard Gazette Archives) for the USAAF.

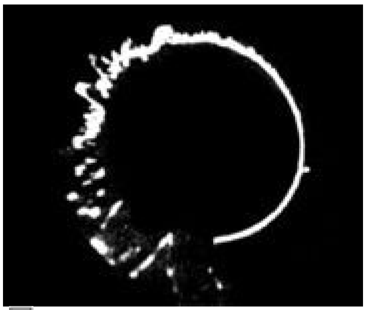
A Lancaster dropping chaff (the crescent-shaped white cloud on the left of the picture)

The systems were all essentially identical in concept, small aluminum strips cut to one-half of the target radar's wavelength. When dropped, the strips would give a strong echo, appearing as a bomber on radar screens. Opposing defenses would find it almost impossible to pick out the "real" bombers from the false echos. Other radar-confusing techniques included Mandrel, Piperack, and Jostle.

Then something odd happened: no one used it. Unaware of the opposing air force's knowledge of the chaff concept, planners felt that using it was even more dangerous than not: as soon as it was used the enemy could easily duplicate it and use it against them. In particular the British government's leading scientific adviser, Professor Lindemann, balefully pointed out that if the RAF used it against the Germans, the Luftwaffe would quickly copy it and could launch a new Blitz. This caused panic in Fighter Command and Anti-Aircraft Command, who managed to suppress the use of Window until July 1943.

Examination of the Würzburg radar equipment brought back to the UK during Operation Biting and subsequent reconnaissance, revealed to the British that all German radars were operating in no more than three major frequency ranges, and thus were prone to jamming. "*Bomber*" Harris, Commander-in-Chief (C-in-C) of RAF Bomber Command, finally got approval to use Window as part of Operation Gomorrah, the raids against Hamburg.

The first to use it were 76 squadron. Twenty-four crews were briefed onto how to drop the bundles of aluminised-paper strips (treated-paper was used to minimise the weight and maximise the time that the strips would remain in the air, prolonging the effect), one every minute through the flare chute, using a stopwatch to time them. The results were spectacular. The radar guided master searchlights wandered aimlessly across the sky. The AA guns fired randomly or not at all and the night fighters utterly failed to find the bomber stream. A vast area of Hamburg was devastated with the loss of only 12 bombers. Squadron Commanders quickly had special chutes fitted to their bombers to make the deployment even easier. Seeing this as a development that made it safer to go on ops, many crews got in as many trips as they could before the Germans found a countermeasure.



The effect of chaff on the display of a Giant Wurzburg

Although the metal strips puzzled the German civilians at first (many thought they were radioactive or carrying anthrax, or some other disease), the German scientists knew exactly what they were because they had developed Düppel themselves and refrained from using it for exactly the same reasons as Lindemann had pointed out to the British.

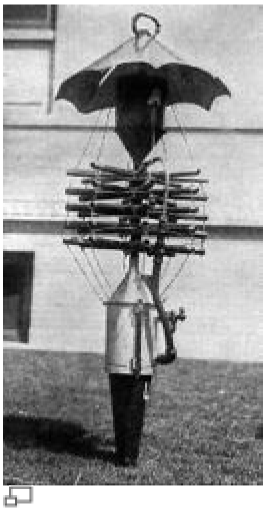
The use of Window rendered the ground-controlled 'Himmelbett' fighters of the Kammhuber Line redundant overnight but the Germans responded quickly, using non-radar equipped free-ranging 'Wild Boar' day fighters to attack visually. Some argue that, by using Window, the British forced the Germans to devise a more effective night fighter defense and had they left well alone then Allied bomber losses may have been ultimately smaller, and not worth the momentary advantage Window gave.

A lesser known fact is that Luftwaffe used this technology just six weeks after the above mentioned Hamburg raid. In a series of raids in 1943, and a larger series known as Operation Steinbock between February and May 1944, Düppel allowed German bombers to once again return to London. Although theoretically effective, the small number of bombers, notably in relation to the RAF's now-large night fighter force, doomed the effort from the start. The British fighters were able to go aloft in large numbers and often found the German bombers in spite of their Düppel.

Falklands War

Chaff was heavily used by ships in the Falklands War. The absence of chaff launchers on the *Atlantic Conveyor*, while used by all other Royal Navy ships in the group, may have led to the ship's sinking by an Exocet missile — although given the vessel's large radar cross section, it is unlikely that chaff would have been effective.

Flare (pyrotechnic)



A World War I-era parachute flare dropped from aircraft for illumination.

A **flare** is a type of pyrotechnic that produces a brilliant light or intense heat without an explosion. Flares are used for signaling, illumination, or defensive countermeasures in civilian and military applications.

Delivery and composition

A IHB brakeman uses a fusee to demonstrate a hand signal indicating "stop".

Flares generally produce their light through the combustion of magnesium metal, sometimes colored by the inclusion of other metals. Calcium flares are used underwater for illumination.

Flares may be ground pyrotechnics, projectile pyrotechnics, or parachute-suspended to provide maximum illumination time over a large area. Projectile pyrotechnics may be dropped from aircraft, fired from rocket or artillery, or deployed by handheld percussive tubes. Flares may also be dropped in the water to illuminate submerged objects.

Civilian use

In the civilian world, flares are commonly used as distress signals, and may be ignited on the ground or fired as an aerial signal from a pistol-like flare gun. Flare guns are commonly found in marine survival kits.

Flares in a football match between Real Zaragoza and RCD Espanyol (Copa del Rey Final 2006)

Another type of flare is the *fusee*, which burns for 15-60 minutes with a bright red light. Fusees are commonly used to indicate obstacles or advise caution on roadways at night; in this usage they are also called *highway flares* or *ground flares*. They are commonly found in roadside emergency kits.

In forestry and firefighting, fusees are sometimes used in wildland fire suppression and in the ignition of controlled burns. They are especially effective in igniting burnouts or backburns in very dry conditions, but not so effective when fuel conditions are moist. Since controlled burns are often done during relatively high humidity levels (on the grounds that they could not be safely contained during periods of very low humidity), the driptorch is more effective and more often used. Fusees are also commonly carried by wildland firefighters for emergency use, to ignite an escape fire in surrounding fuels in case of being overrun by a fire if no other escape routes are available.

Fusees are also known as *railroad flares* and are used to perform hand signals in rail transport applications. Since they can be used only once, fusees nowadays are usually intended for emergency use (as opposed to the incandescent lanterns typically used during normal operating conditions). However, in the days before train radio communications, fusees were used to keep trains apart on un-signaled lines. A railroad fusee was timed to burn for 5 minutes and quantities were dropped behind a train to ensure a safe spacing. If a following train encountered a burning fusee it was not to pass until the fusee burned out.

Military use

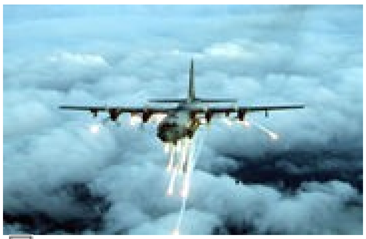
Land

Ground military forces in need of a large-area illumination for artilleries or for an attack, often request the delivery of parachute-flares. Ground forces may also deploy hand-held flares for aerial or ground signaling to indicate the correct area for releasing ordnance, deploying paratroopers, or landing an aircraft. In World War II, clusters of coloured flares were deployed by reconnaissance aircraft or pathfinders to mark targets for bomber missions and supply drops.

Sea

Naval flares may be employed by naval forces to illuminate undersea targets such as submarines at depth. Naval flares are also launched from anti-submarine aircraft from fixed, multi-barrel, ejectors on the sides of the fuselage.

Air



An AC-130H releases decoy flares

A special variety of flare is used in military aircraft as a defensive countermeasure against heat-seeking missiles. These flares are usually discharged individually or in salvoes by the pilot or automatically by tail-warning devices, and are accompanied by vigorous evasive maneuvering. Since they are intended to deceive infrared missiles, these flares burn at temperatures of thousands of degrees, incandescing in the visible spectrum as well. Soids are floating flares that are effective only in the terminal phase of missiles with infrared signature seeker heads.

Fuse (explosives)

In an explosive, pyrotechnic device or military munition, a **fuse** (or **fuze**) is the part of the device that initiates function. In common usage, the word fuse is used indiscriminately. However, when being specific (and in particular in a military context), the term *fuse* describes a simple pyrotechnic detonating device, like the cord on a firecracker, whereas the term *fuze* is used to describe a more complicated ignition device incorporating mechanical and/or electronic components eg a proximity fuze for an artillery shell.

Burning/Safety fuse

A burning fuse.

The simplest form of fuse is the burning fuse, believed to date back to the 10th Century and originating in China, this simple fuse consisted of lightweight paper filled with loose gunpowder, and served as a means of delaying ignition in fireworks. This simple form of burning fuse can still be found today in many modern pyrotechnics. The commercial and military version of a burning fuse, is often referred to as a safety fuse. The safety fuse is a simply length of cord either filled with combustible material, or itself made from combustible material. Safety fuse is often coded by its burn time for 30cms i.e. 60secs, which means that a length of fuse 30cm long will take 60 seconds to burn.

Modern day safety fuses are often used in mining and military operations, to provide a time-delay before ignition, and they more often than not are used to initiate an explosive detonator, thereby starting an explosive chain reaction to detonate a larger more stable main charge.

Safety fuse is typically colored black to distinguish it from detonating cords such as Primacord, which is brightly colored or transparent.

Fuses are found in fireworks, model cannons, antique smoothbore firearms, some improvised explosive devices and many forms of pyrotechnics.

Burning fuse may take many forms:

- Some safety fuse has an outer layer of plastic around the cloth and black powder. This fuse can burn reliably underwater once lit, since the explosive provides both its own fuel and oxidant.
- Slow match is a very slow-burning fuse consisting of a hemp or cotton rope saturated with an oxidizer such as potassium nitrate. Slow match is used as a source of fire for manually lighting other devices, such as matchlock guns, or fuses on black powder cannons. Before percussion caps, slow match was most suitable for use around black powder weapons because it could be roughly handled without going out, and only presented a small glowing tip instead of a large flame that risked igniting powder supplies nearby.
- Today's punks (wood splints covered with ground plant pith saturated with nitrate) used for lighting consumer fireworks are a type of slow match.
- Black match is a type of fuse consisting of cotton string coated with a dried slurry of black powder and glue. This acts as a simple pass-fire, and was used to fire ancient cannons. It is used today in fireworks construction.
- Quick match or piped match is a type of black powder fuse that burns very quickly, some hundreds of feet per second. It consists of black match covered with a loose paper wrap (pipe). When lit, the flame propagates quickly down the paper pipe from the hot gases produced by the burning powder. Quick match is used in professional fireworks displays to pass fire nearly instantly between devices that must be physically separated while firing simultaneously, such as a finale rack. Devices which should fire in sequence can be branched from a single master fuse, consisting of quick match spliced onto Visco fuses of various length for time delays.
- An Ignitor Safety Fuse Electric (ISFE) lights a main fuse or device when activated by an electrical current. They typically consist of a pair of wires leading to a thin resistance wire that heats when current is applied. The resistance wire is covered by a bit of pyrotechnic composition that ignites from the wire heating, providing enough fire to reliably ignite the main fuse via a mechanical connection, or the device directly. Estes model rocket motors are lit by a type of electric match. Large fireworks displays are launched with complex timing sequences using a computer that energizes electric matches connected to the individual device fuses.
- Flying fish fuse is an unusual type of component for fireworks. It is made like Visco fuse, but contains a metallic spark composition or other effect instead of black powder. Flying fish can thus perform as a main effect instead of just an initiator. For example, simply lighting a short piece of flying fish on the ground makes it fly through the air, seeming to swim in random directions, while emitting sparks and noise. A aerial shell loaded with many such pieces results in a beautiful myriad of pieces flying and sparking high in the air.
- A **spoolette** is a delay fuse consisting of a hollow wooden dowel rammed full of black powder. A spoolette is glued into the wall of a fireworks shell and ignited by the lift charge that launches the shell into the air. The spoolette, after a delay that allows the shell to reach its top of trajectory, ignites the shell's main effect(s). The tough wood

construction ensures that the fuse burns reliably despite the explosive force and acceleration of the launch.
- The saucisson was an early form of fuse.

Munition fuzes

Also known as fuse, some countries use the z spelling to distinguish between burning fuses and more complicated munition fuzes. Examples of both spelling can be found.

A fuze refers to a device used in munitions which is designed to detonate, or to set forces into action to ignite, detonate or deflagrate, the charge (or primer) under specified conditions.

Types of fuzes include:

- time fuzes detonate after a set period of time by using mechanical, electronic, igniferious or chemical timers.
- contact detonators or point detonating fuzes explode on impact.
- proximity fuzes cause a missile or other munition to explode when it comes within a certain distance of the target. Some proximity fuses utilize radar, sonar, Infrared, photo-electric or television cameras.
- remote detonators use wires or radio waves to remotely command the explosive to function.
- altitude fuzes cause a bomb to explode at a certain altitude above sea level by means of an infrared rangefinder, radar, or barometric altimeter

Many weapons have fuzing systems to ensure that they do not initiate (explode) prematurely. In most cases the munition has to travel some distance or wait for a period of time before it can detonate.

In modern artillery shells, most fuzes incorporate several safety features to prevent a fuze arming before it leaves the gun barrel. These safety features may include arming on "set-back" or by centrifugal force, and often both operating together. Set-back arming uses the inertia of the accelerating artillery shell to remove a safety feature as the projectile accelerates from rest to its in-flight speed. Rotational arming requires that the artillery shell reach a certain rpm before centrifugal forces cause a safety feature to disengage or move an arming mechanism to its armed position (artillery shells are fired through a rifled barrel, and so spin during flight).

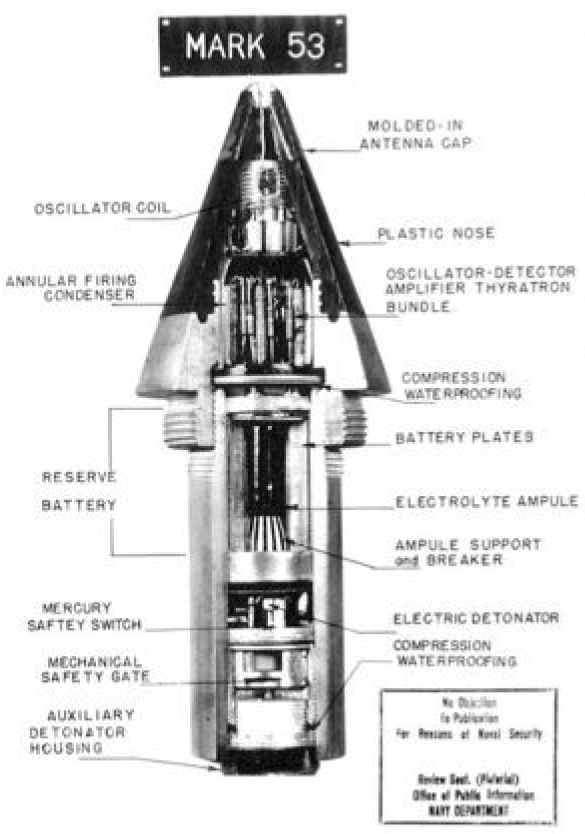
In other cases the bomb, mine or projectile has a fuze that prevents accidental initiation (for example, stopping turning of a small propellor so that the striker-pin cannot touch the detonator even if the weapon is dropped on the ground). These sorts of fuze operate with aircraft weapons where the weapon may have to be jetisoned over *friendly* territory to allow a damaged aircraft to continue to fly. The crew can choose to jettison the weapons *safe* by dropping the devices with safety pins still attached, or drop them *live* by retaining the safety pins as the weapons leave the aircraft.

Aerial bombs and depth charges can be *nose* and *tail* fuzed using different detonator/initiator characteristics so that the crew can choose which effect fuze will suit target conditions that may not have been known before the flight. The arming switch is set to one of *safe*, *nose*, or *tail* at the crew's choice.

Base fuzes are also used by artillery and tanks for shells of the 'squash head' type. Some types of armour piercing shells have also used base fuzes, as have nuclear artillery shells.

There also exist more complicated integrated fuzing and firing systems, typically used in missiles and special weapons.

Proximity fuze



A **proximity fuze** (also called a VT fuze, for "variable time") is a fuze that is designed to detonate an explosive automatically when the distance to target becomes smaller than a predetermined value or when the target passes through a given plane. There are different sensing principles:

- radio frequency sensing
- optical sensing
- acoustic sensing
- magnetic sensing

Radio frequency sensing

Radio frequency sensing is the main sensing principle for shells and this is mostly in mind when one speaks of "proximity fuzes".

The WW2 patent works as follows: The shell contains a radio transmitter which uses the shell body as antenna and emits a continuous wave of roughly 180 - 220 MHz. As the shell approaches a reflecting object, an interference pattern is created. This pattern changes with shrinking distance: every half wavelength in distance, the transmitter is in or out of resonance. This causes a small oscillation of the radiated power and consecutively the oscillator supply current of about 200 - 800 Hz, the Doppler frequency. This signal is sent through a band pass filter, amplified, and triggers the detonation when it exceeds a given amplitude.

Before the fuze's invention, detonation had to be induced by direct contact, a timer set at launch, or an altimeter. All of these have disadvantages. Getting direct contact with a relatively small moving target is hard (even ignoring the effect of wind); to set a time- or height-triggered fuze one must measure the height of the target (or even predict the height of the target at the time one will be able to get a shell or missile in its neighbourhood). With a proximity fuze, all one has to worry about is getting a shell or missile on a trajectory that, at some time, will pass close by the target. This is still not a trivial task, but it is much easier than previous methods were.

Use of timing to produce air bursts against ground targets requires observers to provide information for adjusting the timing. This is not practical in all situations and is slow in any event. Proximity fuzes remove these problems.

The proximity fuze was invented in the UK in 1940, but developed mainly by the U.S. (with British collaboration) during World War II. Vannevar Bush, head of the U.S. Office of Scientific Research and Development (OSRD) during this war, credits it with three significant effects. It was important in defense from Japanese Kamikaze attacks in the Pacific. It was an important part of the radar-controlled anti-aircraft batteries that finally neutralized the German V-1 bomb attacks on England. Third, it was released for use in land warfare just in time for use in the Battle of the Bulge, where it decimated German divisions caught in the open. The Germans felt safe from timed fire because the weather prevented accurate observation. Bush cites an estimated seven times increase in the effect of artillery with this innovation.

Optical sensing

Optical sensing was also developed first in WW2, mainly for anti-aircraft missiles. It used then a toroidal lens, that concentrated all light out of a plane perpendicular to the missile's main axis onto a photo cell. When the cell current changed a certain amount in a certain time interval, the detonation was triggered.

Some modern air-to-air missiles make use of lasers. They project narrow beams of laser light perpendicular to the flight of the missile. As the missile cruises towards the target the laser energy simply beams out into space. However, as the missile passes its target some of the laser energy strikes the target and is reflected back towards the missile where detectors sense the reflected laser energy and trigger the missile warhead.

Acoustic sensing

Acoustic sensing used a microphone in a missile. The characteristic frequency of an aircraft engine was filtered and triggered the detonation. This principle was applied in German anti-aircraft missiles, which were mostly still in development when the war ended.

Magnetic sensing

Magnetic sensing can only be applied to detect huge masses of iron such as ships. It is being used in mines and torpedoes.

Guided missile

A **guided missile** is a military rocket that can be directed in flight to change its flight path. In typical usage the term "missile" refers to guided rockets, and "rockets" to unguided ones. The differences between the two may be fairly minor other than the guidance system.

The first missiles to be used operationally were a series of German missiles of WW2. Most famous of these are the V1 and V2, both of which used a simple mechanical autopilot to keep the missile flying along a pre-chosen route. Less well known were a series of anti-shipping and anti-aircraft missiles, typically based on a simple radio control system directed by the operator.

Basic roles

Ballistic missiles

After the boost-stage ballistic missiles follow a trajectory mainly determined by ballistics, the guidance is for relatively small deviations from that.

The V2 had demonstrated that a ballistic missile could deliver a warhead to a target city with no possibility of interception, and the introduction of nuclear weapons meant it could do useful damage when it arrived. The accuracy of these systems was fairly poor, but post-war development by most military forces improved the basic inertial platform concept to the point where it could be used as the guidance system on ICBMs flying thousands of miles. Today the ballistic missile represents the only strategic deterrent in most military forces; the USAFs continued support of manned bombers is considered by some to be entirely political in nature.

Cruise missiles

The V1 had been successfully intercepted during the war, but this did not make the cruise missile concept entirely useless. After the war the US deployed a small number of nuclear armed cruise missiles in Germany, but these were considered to be of limited usefulness. Continued research into much longer ranged and faster versions led to the US's Navaho missile, and its Soviet counterparts, the Burya and Buran cruise missile. However these were rendered largely obsolete by the ICBM, and none was used operationally. Instead shorter-range developments have become widely used as highly accurate attack systems, such as the US Tomahawk missile.

Anti-shipping

Another major German missile development project was the anti-shipping class (such as the Fritz X and Henschel Hs 293), intended to stop any attempt at a cross-channel invasion.

However the British were able to render their systems useless by jamming their radios, and missiles with wire guidance were not ready by D-Day. After the war the anti-shipping class slowly developed, and became a major class in the 1960s with the introduction of the low-flying turbojet powered cruise missiles known as "sea-skimmers". These became famous during the Falklands War when an Argentine Exocet missile sank a Royal Navy destroyer.

Anti-aircraft

The Stinger shoulder-launched surface-to-air missile system.

By 1944 US and British airforces were sending huge airfleets over occupied Europe, increasing the pressure on the Luftwaffe day and night fighter forces. The Germans were keen to get some sort of useful ground-based anti-aircraft system into operation. Several systems were under development, but none had reached operational status before the war's end. The US Navy also started missile research to deal with the Kamikaze threat. By 1950 systems based on this early research started to reach operational service, including the US Army's Nike Ajax, the Navy's "3T's" (Talos, Terrier, Tartar), and soon followed by the Soviet SA-1 and SA-2 and French and British systems.

Air-to-air

German experience in WWII demonstrated that destroying a large aircraft was quite difficult, and they had invested considerable effort into air-to-air missile systems to do this. This gave birth to the Me262's R4M rockets. It was developed by the Luftwaffe during World War II, and used operationally for a very brief time just prior to the end of the war. In the post-war period the R4M served as the pattern for a number of similar systems, used by almost all interceptor aircraft during the 1940s and '50s. The US Navy and USAF used their superior electronics to deliver a number of such designs in the early 1950s, most famous being the US Navy's AIM-9 Sidewinder and USAF's AIM-4 Falcon. These systems have continued to advance, and modern air warfare consists almost entirely of missile firing.

Anti-tank

By the end of WWII all forces had widely introduced unguided rockets using HEAT warheads as their major anti-tank weapon. However these had a limited useful range of a 100 m or so, and the Germans were looking to extend this with the use of a missile using wire guidance, the X-7. After the war this became a major design class in the later 1950s, and by the 1960s had developed into practically the only non-tank anti-tank system in general use.

Anti-ballistic

Like most missiles, the Arrow missile and MIM-104 Patriot for defense against short-range missiles, carry explosives.

However, in the case of a large closing speed, a projectile without explosives is used, just a collision is sufficient to destroy the target. See Missile Defense Agency for the following systems being developed:

- Kinetic Energy Interceptor (KEI)
- Aegis Ballistic Missile Defense System (Aegis BMD) - a SM-3 missile with Lightweight Exo-Atmospheric Projectile (LEAP) Kinetic Warhead (KW)

Anti-satellite weapon (ASAT)

Also the proposed Brilliant Pebbles defense system would use kinetic energy collisions without explosives.

Guidance systems

Missile guidance systems generally fall into a number of basic classes, each one associated with a particular role. Modern electronics has allowed systems to be mixed on a single airframe, dramatically increasing the capabilities of the missiles.

Inertial guidance system

An **inertial guidance system** consists of an Inertial Measurement Unit (IMU) combined with a set of guidance algorithms and control mechanisms, allowing the path of a vehicle to be controlled according to the position determined by the inertial navigation system. These systems are also referred to as an **inertial platform**.

An **inertial navigation system** (INS) provides the position, velocity, orientation, and angular velocity of a vehicle by measuring the linear and angular accelerations applied to the system in an inertial reference frame. It is widely used because it refers to no real-world item beyond itself (other than the earth's magnetic field). It is therefore immune to jamming and deception. (See the principle of relativity and Mach's principle for some background in the physics involved.)

Overview

Inertial guidance systems were originally developed for navigating rockets. American rocket pioneer Robert Goddard experimented with rudimentary gyroscopic systems. Dr. Goddard's systems were of great interest to contemporary German pioneers including Wernher von Braun.

A typical inertial navigation system integrates the information gathered from a combination of gyroscopes and accelerometers in order to determine the current state of the system.

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

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

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

All inertial navigation systems suffer from integration drift. As small errors in the measurement of acceleration and angular velocity are integrated into progressively larger errors in velocity, which is compounded into errors in position. This is a problem that is inherent in every open loop control system. The inaccuracy of a good-quality navigational system is normally less than 0.6 nautical miles per hour in position and on the order of tenths of a degree per hour orientation.

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

Control theory in general and Kalman filtering in particular provide a theoretical framework for combining information from various sensors. One of the most common alternative sensors is a satellite navigation radio such as GPS. By properly combining the information from an INS and the GPS system, the errors in position and velocity are stable.

Inertial navigation systems in detail

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

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

Linear accelerometers measure how the vehicle is moving in space. Since it can move in three axes (up & down, left & right, forward & back), there is a linear accelerometer for each axis.

A computer continually calculates the vehicle's current position. First, for each of the six degrees of freedom (x,y,z and theta x, theta y and theta z), it integrates the sensed amount of acceleration over time to figure the current velocity. Then it integrates the velocity to figure the current position.

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

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

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

Basic schemes

Gimbaled Gyrostabilized platforms

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

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

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

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

Fluidically Suspended Gyrostabilized Platforms

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

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

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

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

Strapdown systems

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

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

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

GPS Align in Motion

Honeywell has developed a new initialization process called Align in Motion. Strapdown inertial navigation systems require an initialization process that establishes the relationship between the aircraft body frame and the local geographic reference. This process, called alignment, generally requires the device to remain stationary for some period of time in order to establish this initial state. This paper[[1]] describes an alignment process where the initialization occurs while the device is moving. This is possible because an accurate determination of the aircraft motion is available based on measurements obtained from GPS.

Align In Motion allows initialization of a Strapdown Inertial Navigation System while an aircraft is moving, in the air or on the ground. This is accomplished using Civilian grade GPS and an inertial reasonableness test, thereby allowing commercial data integrity requirements to be met. Align In Motion has been FAA certified to recover pure INS performance equivalent to stationary align procedures for civilian flight times up to 18 hours.

This Align In Motion capability allows the removal of dedicated backup batteries on aircraft resulting in weight, cost, and reliability improvements. Align In Motion also has benefits for aircraft operations on the ground, on board ship, and in the air such as reduced turn backs, quicker dispatch, and world-wide alignment including polar regions.

Types of sensors

Laser gyros

Laser gyroscopes were supposed to eliminate the bearings in the gyroscopes, and thus the last bastion of precision machining and moving parts.

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

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

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

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

Vibrating gyros

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

Hemispherical Resonator Gyros ("Brandy Snifter Gyros")

If a standing wave is induced in a globular brandy snifter, and then the snifter is tilted, the waves tend to continue in the same plane of movement. They don't fully tilt with the snifter. This trick is used to measure angles. Instead of brandy snifters, the system uses hollow globes machined from piezoelectric materials such as quartz. The electrodes to start and sense the waves are evaporated directly onto the quartz.

This system has almost no moving parts, and is very accurate. However it is still relatively expensive due to the cost of the precision ground and polished hollow quartz spheres.

Although successful systems were constructed, and an HRG's kinematics appear capable of greater accuracy, they never really caught on. Laser gyros were just more popular.

The classic system is the Delco 130Y Hemispherical Resonator Gyro, developed about 1986. See also [2] for a picture of an HRG resonator.

Quartz rate sensors

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

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

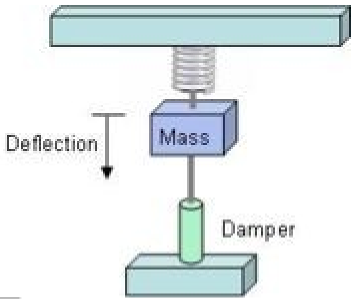
Current state of the art non-military technology (2005) can build small solid state sensors that can measure human body movements. These devices have no moving parts, and weigh about 50 grams.

Solid state devices using the same physical principles are used to stabilize images taken with small cameras or camcorders. These can be extremely small (≈5 mm) and are built with MEMS (Microelectromechanical Systems) technologies.

MHD sensor

Sensors based on magnetohydrodynamic principles can be used to measure angular velocities and are described in "MHD sensor".

Pendular accelerometers



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

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

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

Infrared

Image of two girls in mid-infrared ("thermal") light (false-color)

Infrared (**IR**) radiation is electromagnetic radiation of a wavelength longer than that of visible light, but shorter than that of radio waves. The name means "below red" (from the Latin *infra*, "below"), red being the color of visible light of longest wavelength. Infrared radiation has wavelengths between approximately 750 nm and 1 mm, spanning three orders of magnitude.[1]

The infrared portion of the spectrum has a number of technological uses, including target acquisition and tracking by the military; remote temperature sensing; short-ranged wireless communication; spectroscopy, and weather forecasting. Telescopes equipped with infrared sensors are used in infrared astronomy to penetrate dusty regions of space, such as molecular clouds; detect low temperature objects such as planets orbiting stars, and to view highly red-shifted objects from the early history of the universe.[2]

At the atomic level, infrared energy elicits vibrational modes in a molecule through a change in the dipole moment, making it a useful frequency range for study of these energy states. Infrared spectroscopy is the examination of absorption and transmission of photons in the infrared energy range, based on their frequency and intensity.[3]

Different regions in the infrared

This article or section does not adequately cite its references or sources. Please help improve this article by adding citations to reliable sources. (help, get involved!) Any material not supported by sources may be challenged and removed at any time. This article has been tagged since **July 2006**.

The infrared band is often subdivided into smaller sections but the divisions are not precise, and are used differently by different authors.

One such scheme is:[citation needed]

- near infrared (**NIR**, IR-A *DIN*): 0.75–1.4 µm in wavelength, defined by the water absorption, and commonly used in fiber optic telecommunication because of low attenuation losses in the SiO_2 glass (silica) medium.
- short wavelength IR (**SWIR**, IR-B *DIN*): 1.4–3 µm, water absorption increases significantly at 1450 nm. The 1530 to 1560 nm range is the dominant spectral region for long-distance telecommunications.

- mid wavelength IR (**MWIR**, IR-C *DIN*) also intermediate-IR (IIR): 3–8 µm
- long wavelength IR (**LWIR**, IR-C *DIN*): 8–15 µm
- far infrared (**FIR**): 15–1,000 µm (see also far infrared laser)

NIR and SWIR is sometimes called *Reflected Infrared* while MWIR and LWIR is sometimes referred to as *Thermal Infrared*.

Astronomers typically divide the infrared spectrum as follows:[4]

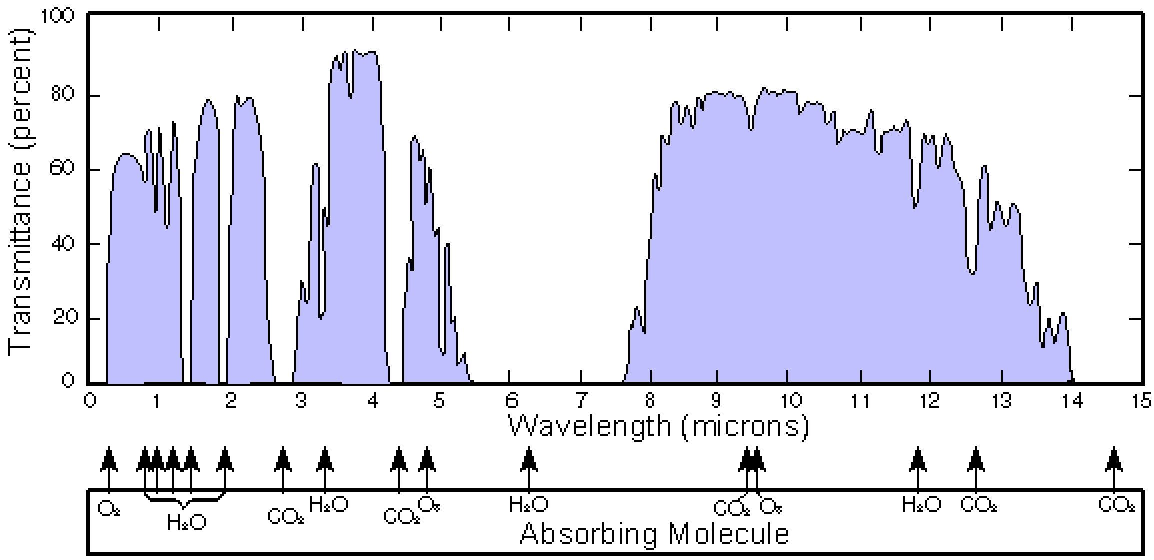
- **near**: (0.7-1) to 5 µm
- **mid**: 5 to (25-40) µm
- **long**: (25-40) to (200-350) µm

These divisions are not precise and can vary depending on the publication. The three regions are used for observation of different temperature ranges, and hence different environments in space.

A third scheme divides up the band based on the response of various detectors:[5]

- Near IR (**NIR**): from 0.7 to 1.0 micrometers (from the approximate end of the response of the human eye to that of silicon)
- Short-wave infrared (**SWIR**): 1.0 to 3 micrometers (from the cut off of silicon to that of the MWIR atmospheric window. InGaAs covers to about 1.8 micrometers; the less sensitive lead salts cover this region
- Mid-wave infrared (**MWIR**): 3 to 5 micrometers (defined by the atmospheric window and covered by InSb and HgCdTe and partially PbSe)
- Long-wave infrared (**LWIR**): 8 to 12, or 7 to 14 micrometers: the atmospheric window (Covered by HgCdTe and microbolometers)
- Very-long wave infrared (**VLWIR**): 12 to about 30 micrometers, covered by doped silicon

These divisions are justified by the different human response to this radiation: near infrared is the region closest in wavelength to the radiation detectable by the human eye, mid and far infrared are progressively further from the visible regime. Other definitions follow different physical mechanisms (emission peaks, vs. bands, water absorption) and the newest follow technical reasons (The common silicon detectors are sensitive to about 1,050 nm, while InGaAs' sensitivity starts around 950 nm and ends between 1,700 and 2,600 nm, depending on the specific configuration). Unfortunately, international standards for these specifications are not currently available.



Plot of atmospheric transmittance in part of the infrared region.

The boundary between visible and infrared light is not precisely defined. The human eye is markedly less sensitive to light above 700 nm wavelength, so shorter frequencies make insignificant contributions to scenes illuminated by common light sources. But particularly intense light (e.g., from lasers, or from bright daylight with the visible light removed by colored gels[1]) can be detected up to approximately 780 nm, and will be perceived as red light. The onset of infrared is defined (according to different standards) at various values typically between 700 nm and 800 nm.

Telecommunication bands in the infrared

In optical communications, the part of the infrared spectrum that is used is divided into several bands based on availability of light sources, transmitting/absorbing materials (fibers) and detectors:[6]

Band	Descriptor	Wavelength range
O band	Original	1260–1360 nm
E band	Extended	1360–1460 nm
S band	Short wavelength	1460–1530 nm
C band	Conventional	1530–1565 nm
L band	Long wavelength	1565–1625 nm
U band	Ultralong wavelength	1625–1675 nm

The C-band is the dominant band for long-distance telecommunication networks. The S and L bands are based on less well established technology, and are not as widely deployed.

"Heat"

Main article: Thermal radiation

Infrared radiation is popularly known as "heat" or sometimes "heat radiation," since many people attribute all radiant heating to infrared light. This is a widespread misconception, since light and electromagnetic waves of any frequency will heat surfaces that absorb them. Infrared light from the Sun only accounts for 50%[citation needed] of the heating of the Earth, the rest being caused by visible light that is absorbed then re-radiated at longer wavelengths. Visible light or ultraviolet-emitting lasers can char paper and incandescently hot objects emit visible radiation. It is true that objects at room temperature will emit radiation mostly concentrated in the 8–12 micron band, but this is not distinct from the emission of visible light by incandescent objects and ultraviolet by even hotter objects (see black body and Wien's displacement law).[7]

Heat is energy in transient form that flows due to temperature difference. Unlike heat transmitted by thermal conduction or thermal convection, radiation can propagate through a vacuum.

Applications

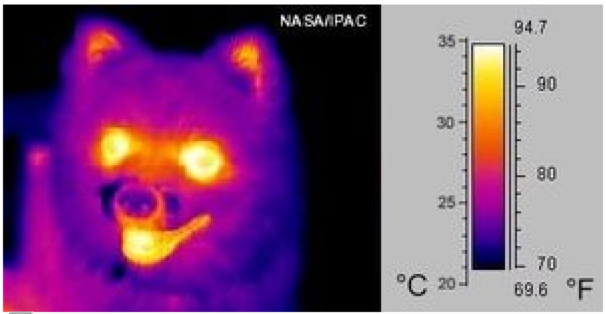
Night vision

Infrared is used in night-vision equipment when there is insufficient visible light to see an object. The radiation is detected and turned into an image on a screen, hotter objects showing up in different shades than cooler objects, enabling the police and military to distinguish warm targets, such as human beings and automobiles. *Also see Forward looking infrared*. IR radiation is a secondary effect of heat; it is not heat itself. Heat itself is a measure of the translational energy of an amount of matter. "Thermal" detectors do not actually detect heat directly but the difference in IR radiation from objects. The device itself that detects the radiation is known as a photocathode. Military gunnery ranges sometimes use special materials that reflect IR radiation to simulate enemy vehicles with running engines. The targets can be at the exact same temperature as the surrounding terrain, but they emit (reflect) much more IR radiation. Different materials emit more or less IR radiation as temperature increases or decreases, depending on the composition of the material. Infrared imagery is usually formed as a result of the integrated inband intensity of the radiation, based on temperate and emissivity.

Simple infrared sensors were used by British, American and German forces in the Second World War as night vision aids for snipers.

Smoke is more transparent to infrared than to visible light, so firefighters use infrared imaging equipment when working in smoke-filled areas.

Thermography



A thermographic image of a dog

Infrared thermography is a non-contact, non-destructive test method that utilizes a thermal imager to detect, display and record thermal patterns and temperatures across the surface of an object. Infrared thermography may be applied to any situation where knowledge of thermal profiles and temperatures will provide meaningful data about a system, object or process. Thermography is widely used in industry for predictive maintenance, condition assessment, quality assurance, and forensic investigations of electrical, mechanical and structural systems. Other applications include, but are not limited to: law enforcement, firefighting, search and rescue, and medical and veterinary sciences.

Aside from test equipment, training is the most important investment a company will make in an infrared inspection program. Advances in technology have provided infrared equipment that is user-friendly; however, infrared thermography is not a “simply point and shoot” technology. In addition to understanding the object or system being inspected, thermographers must also understand common error sources that can influence observed thermal data. Typically,infrared training courses should cover the topics of infrared theory, heat transfer concepts, equipment selection and operation, how to eliminate or overcome common error sources, and specific applications. Training courses from independent training companies are preferred since they are not biased toward a single brand or type of equipment.

Other imaging

Infrared light from the LED of a remote control as seen by a digital camera.

In infrared photography, infrared filters are used to capture the near-infrared spectrum. Digital cameras often use infrared blockers. Cheaper digital cameras and some camera phones which do not have appropriate filters can "see" near-infrared, appearing as a bright white colour (try pointing a TV remote at your digital camera). This is especially pronounced when taking pictures of subjects near IR-bright areas (such as near a lamp), where the resulting infrared interference can wash out the image. There is also a technique called 'T-ray' imaging, which is imaging using far infrared or terahertz radiation. Lack of bright sources makes terahertz photography technically more challenging than most other infrared imaging techniques. Recently T-ray imaging has been of considerable interest due to a number of new developments such as terahertz time-domain spectroscopy.

Heating

Infrared radiation is used in infrared saunas to heat the occupants, and to remove ice from the wings of aircraft (de-icing). It is also gaining popularity as a method of heating asphalt pavements in place during new construction or in repair of damaged asphalt. Infrared can be used in cooking and heating food as it heats only opaque, absorbent objects and not the air around them, if there are no particles in it.

Infrared heating is also becoming more popular in industrial manufacturing processes, e.g. curing of coatings, forming of plastics, annealing, plastic welding, print drying. In these applications, infrared heaters replace convection ovens and contact heating. If the wavelength of the infrared heater is matched to the absorption characteristics of the material, significant gains in energy efficiency are possible.

Communications

IR data transmission is also employed in short-range communication among computer peripherals and personal digital assistants. These devices usually conform to standards published by IrDA, the Infrared Data Association. Remote controls and IrDA devices use infrared light-emitting diodes (LEDs) to emit infrared radiation which is focused by a plastic lens into a narrow beam. The beam is modulated, i.e. switched on and off, to encode the data. The receiver uses a silicon photodiode to convert the infrared radiation to an electric current. It responds only to the rapidly pulsing signal created by the transmitter, and filters out slowly changing infrared radiation from ambient light. Infrared communications are useful for indoor use in areas of high population density. IR does not penetrate walls and so does not interfere with other devices in adjoining rooms. Infrared is the most common way for remote controls to command appliances.

Free space optical communication using infrared lasers can be a relatively inexpensive way to install a communications link in an urban area operating at up to 4 gigabit/s, compared to the cost of burying fiber optic cable.

Infrared lasers are used to provide the light for optical fiber communications systems. Infrared light with a wavelength around 1,330 nm (least dispersion) or 1,550 nm (best transmission) are the best choices for standard silica fibers.

Spectroscopy

Infrared radiation spectroscopy (see also near infrared spectroscopy) is the study of the composition of (usually) organic compounds, finding out a compound's structure and composition based on the percentage transmittance of IR radiation through a sample. Different frequencies are absorbed by different stretches and bends in the molecular bonds occurring inside the sample. Carbon dioxide, for example, has a strong absorption band at 4.2 µm.

Meteorology

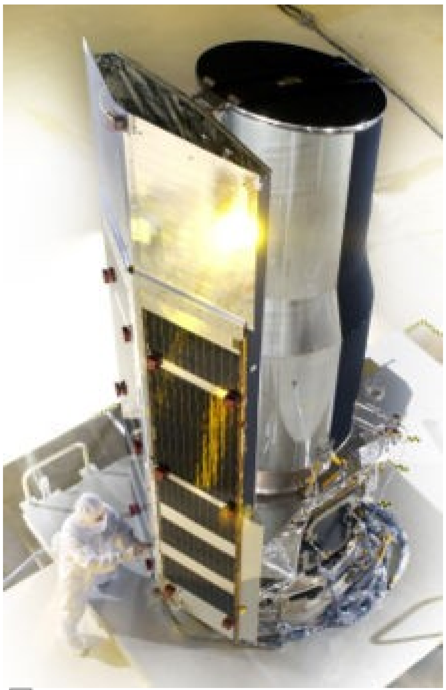
IR Satellite picture taken 1315 Z on 15th October 2006. A frontal system can be seen in the Gulf of Mexico with embedded Cumulonimbus cloud. Shallower Cumulus and Stratocumulus can be seen off the Eastern Seaboard.

Weather satellites equipped with scanning radiometers produce thermal or infrared images which can then enable a trained analyst to determine cloud heights and types, to calculate land and surface water temperatures, and to locate ocean surface features. The scanning is typically in the range 10.3-12.5 µm (IR4 and IR5 channels).

High, cold ice cloud such as Cirrus or Cumulonimbus show up bright white, lower warmer cloud such as Stratus or Stratocumulus show up as grey with intermediate clouds shaded accordingly. Hot land surfaces will show up as dark grey or black. One disadvantage of infrared imagery is that low cloud such as stratus or fog can be a similar temperature to the surrounding land or sea surface does not show up. However using the difference in brightness of the IR4 channel (10.3-11.5 µm) and the near-infrared channel (1.58-1.64 µm), low cloud can be distinguished, producing a *fog* satellite picture. The main advantage of infrared is that images can be produced at night, allowing a continuous sequence of weather to be studied.

These infrared pictures can depict ocean eddies or vortices and map currents such as the Gulf Stream which are valuable to the shipping industry. Fishermen and farmers are interested in knowing land and water temperatures to protect their crops against frost or increase their catch from the sea. Even El Niño phenomena can be spotted. Using color-digitized techniques, the gray shaded thermal images can be converted to color for easier identification of desired information.

Astronomy



The Spitzer Space Telescope is a dedicated infrared space observatory currently in orbit around the Earth. *NASA image.*

Astronomers observe objects in the infrared portion of the electromagnetic spectrum using optical components, including mirrors, lenses and solid state digital detectors. For this reason it is classified as part of optical astronomy. To form an image, the components of an infrared telescope need to be carefully shielded from heat sources, and the detectors are chilled using liquid helium.

The sensitivity of Earth-based infrared telescopes is significantly limited by water vapor in the atmosphere, which absorbs a portion of the infrared radiation arriving from space outside of selected atmospheric windows. This limitation can be partially alleviated by placing the telescope observatory at a high altitude, or by carrying the telescope aloft with a balloon or an aircraft. Space telescopes do not suffer from this handicap, and so outer space is considered the ideal location for infrared astronomy.

The infrared portion of the spectrum has several useful benefits for astronomers. Cold, dark molecular clouds of gas and dust in our galaxy will glow with radiated heat as they are irradiated by imbedded stars. Infrared can also be used to detect protostars before they begin to emit visible light. Stars emit a smaller portion of their energy in the infrared spectrum, so nearby cool objects such as planets can be more readily detected. (In the visible light spectrum, the glare from the star will drown out the reflected light from a planet.)

Infrared light is also useful for observing the cores of active galaxies which are often cloaked in gas and dust. Distant galaxies with a high redshift will have the peak portion of their spectrum shifted toward longer wavelengths, so they are more readily observed in the infrared.[2]

Art history and Archaeology

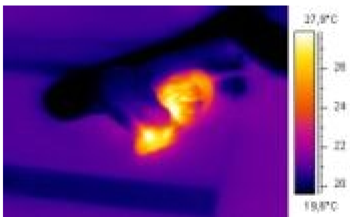
The Arnolfini Portrait by Jan van Eyck, National Gallery, London

Infra-red (as art historians call them) reflectograms are taken of paintings to reveal underlying layers, in particular the underdrawing or outline drawn to by the artist as a guide. This often uses carbon black which shows up well in reflectograms, so long as it has not also been used in the ground underlying the whole painting. Art historians are looking to see if the visible layers of paint differ from the under-drawing or layers in between - such alterations are called pentimenti when made by the original artist. This is very useful information in deciding whether a painting is the prime version by the original artist or a copy, and whether it has been altered by over-enthusiatic restoration work. Generally the more pentimenti, the more likely a painting is to be the prime version. It also gives useful insights into working practices. [2]

Among many other changes in the Arnolfini Portrait of 1434 (right), his face was higher by about the height of his eye, hers was higher, and her eyes looked more to the front. Each of his feet was underdrawn in one position, painted in another, and then overpainted in a third. These alterations are seen in infra-red reflectograms.[8]

Similar uses of infrared are made by archaeologists on various types of objects, especially very old written documents such as the Dead Sea Scrolls, the Roman works in the Villa of the Papyri, and the Silk Road texts found in the Dunhuang Caves.[9] Carbon black used in ink can show up extremely well.

Biological systems



Thermographic image of a snake eating a mouse

The pit viper is known to have two infrared sensory pits on its head. There is controversy over the exact thermal sensitivity of this biological infrared detection system.[10][11]

Other organisms that actively employ thermo-receptors are rattlesnakes (Crotalinae subfamily) and boas (Boidae family), the Common Vampire Bat (*Desmodus rotundus*), a variety of jewel beetles (*Melanophila acuminata*), darkly pigmented butterflies (*Pachliopta aristolochiae* and *Troides rhadamathus plateni*), and possibly blood-sucking bugs (*Triatoma infestans*).[12]

The Earth as an infrared emitter

This article or section does not adequately cite its references or sources. Please help improve this article by adding citations to reliable sources. (help, get involved!) Any material not supported by sources may be challenged and removed at any time. This article has been tagged since **July 2006**.

The Earth's surface and the clouds absorb visible and invisible radiation from the sun and re-emit much of the energy as infrared back to the atmosphere. Certain substances in the atmosphere, chiefly cloud droplets and water vapor, but also carbon dioxide, methane, nitrous oxide, sulfur hexafluoride, and chlorofluorocarbons, absorb this infrared, and re-radiate it in all directions including back to Earth. Thus the greenhouse effect keeps the atmosphere and surface much warmer than if the infrared absorbers were absent from the atmosphere.

History of infrared science

The discovery of infrared radiation is ascribed to William Herschel, the astronomer, in the early 19th century. Herschel published his results in 1800 before the UK Royal Society. Herschel used a prism to refract light from the sun and detected the infrared, beyond the red part of the spectrum, through an increase in the temperature recorded on a thermometer. He was surprised at the result and called them "Calorific Rays". The term 'Infrared' did not appear until late in the 19th century. Incidentally, Herschel is buried in Westminster Abbey between Darwin and Newton.

Other important dates include:[5]

- 1835: Macedonio Melloni makes the first thermopile IR detector;
- 1859: Gustav Kirchhoff formulates the blackbody theorem $E = J(T,n)$;
- 1873: Willoughby Smith discovers the photoconductivity of selenium;
- 1879: Stefan-Boltzmann law formulated empirically ω_T^4
- 1880s & 1890s: Lord Rayleigh and Wilhelm Wien both solve part of the blackbody equation, but both solutions are approximations that "blow up" out of their useful ranges. This problem was called the "UV Catastrophe and Infrared Catastrophe".
- 1901: Max Planck published the blackbody equation and theorem. He solved the problem by quantizing the allowable energy transitions.
- Early 1900s: Albert Einstein develops the theory of the photoelectric effect, determining the photon. Also William Coblentz in spectroscopy and radiometry.
- 1917: Case develops thallous sulfide detector; British develop the first infra-red search and track (IRST) in World War I and detect aircraft at a range of one mile;
- 1935: Lead salts—early missile guidance in World War II;
- 1938: Teau Ta—predicted that the pyroelectric effect could be used to detect infrared radiation.
- 1952: H. Welker discovers InSb;
- 1950s: Paul Kruse (at Honeywell) and Texas Instruments form infrared images before 1955;

- 1950s and 1960s: Nomenclature and radiometric units defined by Fred Nicodemenus, G.J. Zissis and R. Clark, Jones defines *D**;
- 1958: W.D. Lawson (Royal Radar Establishment in Malvern) discovers IR detection properties of HgCdTe;
- 1958: Falcon & Sidewinder missiles developed using infrared and the first textbook on infrared sensors appears by Paul Kruse, et al.
- 1962: J. Cooper demonstrated pyroelectric detection;
- 1962: Kruse and ? Rodat advance HgCdTe; Signal Element and Linear Arrays available;
- 1965: First IR Handbook; first commercial imagers (Barnes, Agema {now part of FLIR Systems Inc.}; Richard Hudson's landmark text; F4 TRAM FLIR by Hughes; phenomenology pioneered by Fred Simmons and A.T. Stair; U.S. Army's night vision lab formed (now Night Vision and Electronic Sensors Directorate (NVESD), and Rachets develops detection, recognition and identification modeling there;
- 1970: ? Boyle & ? Smith propose CCD at Bell Labs for picture phone;
- 1972: Common module program started by NVESD;
- 1978: Pommernig & ? Francis fabricate IRCCDs; US Common Module leads to a proliferation of IR Sensors in the U.S. military; commercial IR companies formed (Inframetrics in Boston, MA and FLIR Systems Inc. in Portland OR); Infrared imaging astronomy comes of age, observatories planned, IRTF on Mauna Kea opened; 32 by 32 and 64 by 64 arrays are produced in InSb, HgCdTe and other materials.

Laser guidance

Laser guidance is a technique of guiding a missile or other projectile or vehicle to a target by means of a laser beam. Some laser guided systems utilise beam riding guidance, but most operate more similarly to semi-active radar homing (SARH). This technique is sometimes called **SALH**, for **Semi-Active Laser Homing**. With this technique, a laser is kept pointed at the target and the laser radiation bounces off the target and is scattered in all directions. The missile, bomb, etc. is launched or dropped somewhere near the target. When it is close enough that some of the reflected laser energy from the target reaches it, a laser seeker notices which direction this energy is coming from and aims the projectile towards the source. As long as the projectile is in the right general area and the laser is kept aimed at the target, the projectile should be guided accurately to the target.

Note that laser guidance isn't useful against targets that don't reflect much laser energy, including those coated in special paint which absorbs laser energy. This is likely to be widely used by advanced military vehicles in order to make it harder to use laser range-finders against them and harder to hit them with laser-guided munitions.

Types of missile

Air-to-air missile (AAM) · Air-to-surface missile (ASM) · Surface-to-air missile (SAM) · Surface-to-surface missile (SSM) · Ballistic missile · Intercontinental ballistic missile (ICBM) · Submarine-launched ballistic missile (SLBM) · Anti-ballistic missile (ABM) · Cruise missile · Anti-ship missile (AShM) · Anti-submarine Rocket (ASROC) · Anti-tank guided missile (ATGM) · Anti-satellite weapon (ASAT)
List of missiles

Laser-guided bomb

A **laser-guided bomb (LGB)** is a precision-guided munition (PGM) that uses semi-active laser homing to strike a designated target with greater accuracy than a free-fall bomb. LGBs are one of the most common and widespread PGMs, used by a large number of the world's air forces.

Overview

BOLT-117 laser guided bomb

Laser-guided munitions use a laser designator to mark (illuminate) a target. The reflected laser light ("sparkle") from the target is then detected by the seeker head of the weapon, which sends signals to the weapon's control surfaces to guide it toward the designated point. Laser-guided bombs are generally unpowered, using small wings to glide towards their targets. Powered laser-guided missiles, such as some variants of the US AGM-65 Maverick and the French AS.30L, use the same guidance system, but have greater range and maneuverability because they are not limited to unpowered flight. Some LGBs have been fitted with strap-on rocket motors to increase their range; one such weapon is the USAF AGM-123 Skipper.

The earliest laser guidance seekers measured the intensity of the reflected laser light at four corners of the seeker window. The seeker then actuated the control fins to steer the weapon in the direction of the strongest signal return, thereby keeping the weapon centered on the laser sparkle. Later weapons have more sensitive seekers and more sophisticated control systems that waste less energy with course corrections, improving accuracy and range, but the principle remains essentially the same. The first such weapon to be developed was the Texas Instruments BOLT-117.

Most laser-guided bombs are produced in the form of strap-on kits: seeker heads, steering fins, and wings that can be attached to a standard general-purpose bomb or penetration bomb. Such kits are modular, allowing relatively easy upgrades, and are considerably cheaper than purpose-built weapons.

Development

Laser-guided weapons were first developed in the United States in the early 1960s. The USAF issued the first development contracts in 1964, leading to the development of the Paveway™ series, which was used operationally in Vietnam starting in 1968. Although there were a variety of technical and operational problems, the results were generally positive. LGBs proved to offer a much higher degree of accuracy than unguided weapons, but without the expense, complexity, and limitations of guided air-to-ground missiles like the AGM-12 Bullpup. The LGB proved particularly effective against difficult fixed targets like bridges, which previously had required huge loads of "dumb" ordnance, and large numbers of sorties, to destroy.

It was determined that 48 percent of Paveways dropped during 1972–73 around Hanoi and Haiphong achieved direct hits, compared with only 5.5 percent of unguided bombs dropped on the same area a few years earlier.[1] The average Paveway landed within 23 feet of its target, as opposed to 447 feet for gravity bombs.[1] The leap in accuracy brought about primarily by laser guidance made it possible to take out heavily defended, point objectives that had eluded earlier air raids. The most dramatic example was the Thanh Hoa Bridge, 70 miles south of Hanoi, a critical part crossing point over the Red River. Starting in 1965, U.S. pilots had flown 871 sorties against it, losing 11 planes without managing to put it out of commission. In 1972 the "Dragon's Jaw" bridge was attacked with Paveway bombs, and 14 jets managed to do what the previous 871 had not: drop the span, and cut a critical North Vietnamese supply artery.[1]

In the wake of this success, other nations, specifically the Soviet Union, France, and Great Britain, began developing similar weapons in the late 1960s and early 1970s, while US weapons were refined based on combat experience.

The USAF and other air forces are now seeking to upgrade their LGBs with GPS guidance as a back-up. These weapons, such as the USAF Enhanced Guided Bomb Unit (part of the Paveway™ family), use laser designation for precision attacks, but contain an inertial navigation system with GPS receiver for back-up, so that if the target illumination is lost or broken, the weapon will continue to home in on the GPS coordinates of the original target.

Paveway™ is a trademark of Raytheon Company.

Problems and Limitations

While LGBs are highly accurate under ideal conditions, they present several challenges for successful use, making them somewhat less than the "silver bullet" sometimes suggested.

The first problem is designation. To ensure accurate guidance, the target must be illuminated for several seconds before launch, allowing the weapon's seeker to obtain a positive lock, and the target must remain illuminated during much of the weapon's transit time. If the designator's "sparkle" is turned off, blocked, or moved, the weapon's accuracy will be greatly reduced.

For an accurate attack against a small target, uninterrupted designation is essential. But, the guidance controls of many LGBs (such as the American Paveway™ II) cause large deflections (visible as a noticeable wobble) which reduce the bomb's range. To compensate, crews will often release their weapons in an unguided, ballistic arc, activating the designator only to refine the bomb's final impact point. This is more demanding of crew and aircraft, requiring a high standard of basic, unguided bombing accuracy and more attention to the bomb's flight.

Laser designation is very sensitive and vulnerable to weather conditions. Cloud cover, rain, and smoke often make reliable designation impossible. In war conditions, many attacks have been aborted due to poor visibility.

In the 1970s and 1980s it was common for aircraft to rely on a separate designator, either carried by ground forces, operated by the forward air controller, or carried by another aircraft in the strike group. It was often deemed more practical for one aircraft to designate for its comrades. Modern conflicts and a growing emphasis on precision-guided weapons have pointed to the need for autonomous designation, and many fighter-bomber aircraft are now being fitted with designator pods to self-designate for laser-guided munitions.

Even if the launch aircraft is capable of autonomous designation, problems remain. Laser illumination can be interrupted by smoke, fog, or clouds, limiting the usefulness of LGBs in poor weather or very dusty conditions. In desert warfare, such as the 1991 Gulf War, laser designation sometimes reflected off the sand, causing weapons to home on false targets. Furthermore, the need to provide designation may leave the aircraft dangerously exposed to ground fire or enemy air support.

An additional concern is the limited "launch envelope" of an unguided weapon. The reflected laser "sparkle" can be described as a basket into which the weapon must be steered to hit the target. If the weapon is released too low or to far from the target, or in a trajectory that puts the weapon outside the seeker's field of view, it is likely to miss. Optimum altitude for an effective LGB attack is relatively high, increasing the aircraft's vulnerability to surface-to-air missile (SAM) attacks.

For these reasons, while all modern air forces have put an increasing emphasis on LGBs and other precision-guided munitions, some tacticians still see an important role for the accurate delivery of unguided bombs. During their 1981 raid on the Iraqi nuclear reactor at Osirak, the Israeli Air Force chose to use unguided Mark 84 bombs rather than laser-guided weapons because they felt the need to designate the target would leave the attackers unacceptably vulnerable.

Missile guidance

A guided bomb strikes an underground facility

Missile guidance technologies of missile systems use a variety of methods to guide a missile to its intended target. These can generally be classified into a number of categories, with the broadest categories being active vs. passive vs. preset.

Passive systems use signals generated by the target itself as a signal on which to "home in". A number of such systems have been developed, but by far the most common are sound in the case of torpedoes and infrared in the case of air-to-air missiles.

Active systems use some "input" signal instead. One common sort of signal is a controller who watches the missile and sends corrections to its flight path. Another common system is to use radar signals or radio control. The semi-active radar homing is a crossover, homing passively on a reflected active radar signal generated by some other system.

Preset systems are used to attack targets at fixed locations, such as military bases and cities.

- command guided, an active system in which signals are sent to the missile using radio control or some similar system. more specifically the term is typically used to describe anti-aircraft systems in which the tracking and guidance systems are all ground-based.

- MCLOS, manually command to line of sight, the operator watches the missile flight and uses some sort of signaling system to command the missile back into the straight line between the operator and the target (the "line of sight"). Typically useful only for slower targets where significant "lead" is not required. MCLOS is a subtype of command guided systems. In the case of glide bombs missiles against ships or the supersonic Wasserfall against slow-moving B-17 Flying Fortress bombers this system worked fine, but as speeds increased MCLOS was quickly rendered useless for most roles.

- SACLOS, semi-automatic command to line of sight, is similar to MCLOS but some automatic system positions the missile in the line of sight while the operator simply tracks the target. SACLOS has the advantage of allowing the missile to start in a position invisible to the user, as well as generally being considerably easier to operate. SACLOS is the most common form of guidance against ground targets such as tanks and bunkers.

- beam riding, in which a "beam" of some sort, typically radio or laser, is pointed at the target and detectors on the rear of the missile keep it centered in the beam. Beam riding systems are often SACLOS, but don't have to be, in other systems the beam is part of an automated radar tracking system.

- active radar homing uses a radar on the missile to provide a guidance signal. Typically electronics in the missile keep the radar pointed directly at the target, and the missile then looks at this "angle off" its own centerline to guide itself. Radar resolution is based on the size of the antenna, so in a smaller missile these systems are useful for attacking only large targets, ships or large bombers for instance. Active radar systems remain in widespread use in anti-shipping missiles, and in "fire and forget" air-to-air missile systems such as AMRAAM and R-77

- semi-active radar homing (SARH) systems combine a radar receiver on the missile with a radar broadcaster located "elsewhere". Since the missile is typically being launched after the target was detected using a powerful radar system, it makes sense to use that same radar system to track the target, thereby avoiding problems with resolution or power. SARH is by far the most common "all weather" guidance solution for anti-aircraft systems, both ground and air launched. SALH is a similar system using a laser as a signal.

- track-via-missile (TVM) is like a hybrid between command guidance, semi-active radar homing and active radar homing. The missile picks up radiation broadcast by the tracking radar which bounces off the target and relays it to the tracking station, which relays commands back to the missile.

- Infrared homing, a passive system in which heat generated by the target is detected and homed on. Typically used in the anti-aircraft role to track the heat of jet engines, it has also been used in the anti-vehicle role with some success. This means of guidance is sometimes also referred to as "heatseeking".

- inertial guidance uses sensitive measurement devices to calculate the location of the missile due to the acceleration put on it after leaving a known position. Early mechanical systems were not very accurate, and required some sort of external adjustment to allow them to hit targets even the size of a city. Modern systems use solid state ring gyros that are accurate to within metres over ranges of 10,000km, and no longer require additional inputs. Gyroscope development has culminated in the AIRS found on the MX missile, allowing for an accuracy of less than 100m at intercontinental ranges. Many civilian aircraft use inertial guidance using the ring laser gyroscope, which is less accurate than the mechanical systems found in ICBMs, but which provide an inexpensive means of attaining a fairly accurate fix on location (when most airliners such as Boeing's 707 and 747 were designed, GPS was not the widely commercially available means of tracking that it is today). Today guided weapons can use a combination of INS, GPS and radar terrain mapping to achieve extremely high levels of accuracy such as that found in modern cruise missiles.

- Stellar-inertial guidance was first used in the American Poseidon missile and uses star positioning to fine-tune the accuracy of the inertial guidance system after launch. As the accuracy of a ballistic missile is dependent upon the guidance system knowing the exact position of the rocket at any given moment during its boost phase, the fact that stars are a fixed reference point from which to calculate that position makes this a potentially very effective means of improving accuracy. In the Polaris system this was achieved by a single camera that was trained to spot just one star in its expected position (it is believed that the missiles from Soviet submarines would track two separate stars to achieve this), if it was not quite aligned to where it should be then this would indicate that the inertial system was not precisely on target and a correction would be made. Apparently this system is sufficiently sensitive to detect stars in daylight.

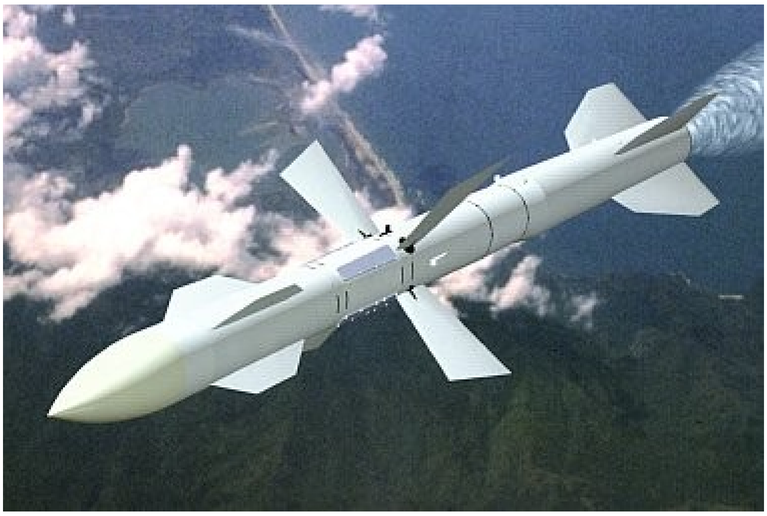
- TERCOM, for "terrain contour matching", uses altitude maps of the strip of land from the launch site to the target, and compares them with information from a radar altimeter onboard. More sophisticated TERCOM systems allow the missile to fly a complex route over a full 3D map, instead of flying directly to the target. TERCOM is the typical system for cruise missile guidance, but is being supplanted by GPS systems and by DSMAC, Digital Scene-Matching Area Correlator, which employs a camera to view an area of land, digitizes the view, and compares it to stored scenes in an onboard computer to guide the missile to its target.

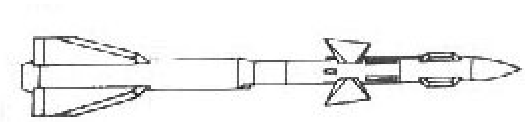
- Contrast seekers use a television camera, typically black and white, to image a field of view in front of the missile, which is presented to the operator. When launched, the electronics in the missile look for the spot on the image where the contrast changes the fastest, both vertically and horizontally, and then attempts to keep that spot at a constant location in its view. Contrast seekers have been used for air-to-ground missiles, including the famous AGM-65 Maverick, because most ground targets can be distinguished only by visual means. However they rely on there being strong contrast changes to track, and even traditional camouflage can render them unable to "lock on".

Medium-Range Air-to-Air Missile

R - 27

NATO Designation AA-10 "Alamo"





Model	R-27R	R-27T	R-27RE	R-27TE	R-27AE	R-27EM
Entered service in	1983 (?)	1984	1985 (?)	1985 (?)	1990	1990
Range against closing target	80 km	70 km	130 km	120 km	130 km	110 - 170 km
Range against receding target	50 km	45 km	60 km	55 km	60 km	?
Weight	253 kg	254 kg	350 kg	343 kg	350 kg	350 kg
Weight of warhead	39 kg					
Type of warhead	expanding rod					
Speed	Mach 2.5 - 4.5					
Guidance	semi-active radarhoming	infrared	semi-active radarhoming	infrared	active radarhoming	semi-active radarhoming
Hit probability	0.7					
Length	4.08 m	3.79 m	4.78 m	4.5 m	4.78 m	4.78 m
Diameter	0.23 m	0.23 m	0.26 m	0.26 m	0.26 m	0.26 m

Fin span	0.77 m	0.77 m	0.8 m	0.8 m	0.8 m	0.8 m
Carried by	MiG-29, MiG-31, Su-27, Su-30, Su-33, Su-34, Su-35, Su-37, Yak-141					

The R-27 is a medium-range air-to-air missile, designated as AA-10 "Alamo" with NATO countries. This missile was designed for the Soviet Union's forth-generation fighters such as the MiG-29 and Su-27 featuring exceptional maneuverability replacing in service MiG-23 fighters armed with the R-23 missiles. Furthermore this missile is carried by improved versions of the MiG-21, MiG-23 and MiG-25. This missile was also designed as a counterweight for the United States F-15 fighters armed with the AIM-7F "Sparrow" missiles. Suggestions of the new missile conception were made in 1972 - 1973 and at the end of 1973 development of this missile begins. There were made two competeting projects by "Molniya" and "Vympel" design bureaus and early in the 1980s the "Vympel" design was selected for further development. Series production of the R-27 began in 1986.

The R-27 missiles are intended to intercept and defeat aircraft and helicopters of all types, unmanned reconnaissance aircraft and cruise missiles under active enemy electronic jamming, counteractions and maneuvering. There are produced some variants of the AA-10 "Alamo" with two different seeker types - semi-active radarhoming and infrared, and two types of engines - with standard and extended range engine.

The R-27 missiles have a modular design, thus missile can be easily converted from semi-active radarhoming to infrared just replacing the seeker module. Furthermore such design allowed to use the same missile both with the MiG-29 light frontline fighter and mounting extended range module - with the Su-27 long-range interceptor.

"Alamo" missiles are capable to intercept an air target traveling at speed up to 3 500 km/h. Interception altitude varies from 20 meters to 27 kilometers. Maximum altitude difference between target and missile carrying aircraft is 10 kilometers. All R-27 missiles have a minimum range of fire in 0.5 - 1 km and carry 39 kg weight expanding rod warheads.

Initially, as almost all soviet air-to-air missiles, the R-27 came in two variants, having either semi-active radarhoming or an infrared homing seeker. The semi-active radarhoming variant was designated as the R-27R (AA-10A "Alamo-A"), while the infrared variant was designated as R-27T (AA-10B "Alamo-B"). The standard soviet tactic for interceptions is based on firing two missiles with a various seeker types at the same target to maximize kill probability. There are also designed downgraded export versions of these missiles, designated as R-27R1 (AA-10A "Alamo-A") and R-27T1 (AA-10B "Alamo-B") respectively.

A few years later "Vympel" developed extended range of the R-27 variants featuring a larger engine and longer range of fire. However internal changes done to the missiles were insignificant. Extended range version of the R-27R received designation as R-27RE (AA-10C "Alamo-C") and version of the R-27T - the R-27TE (AA-10D "Alamo-D") respectively. Their export versions are R-27RE1 and R-27TE1 with similar NATO designations.

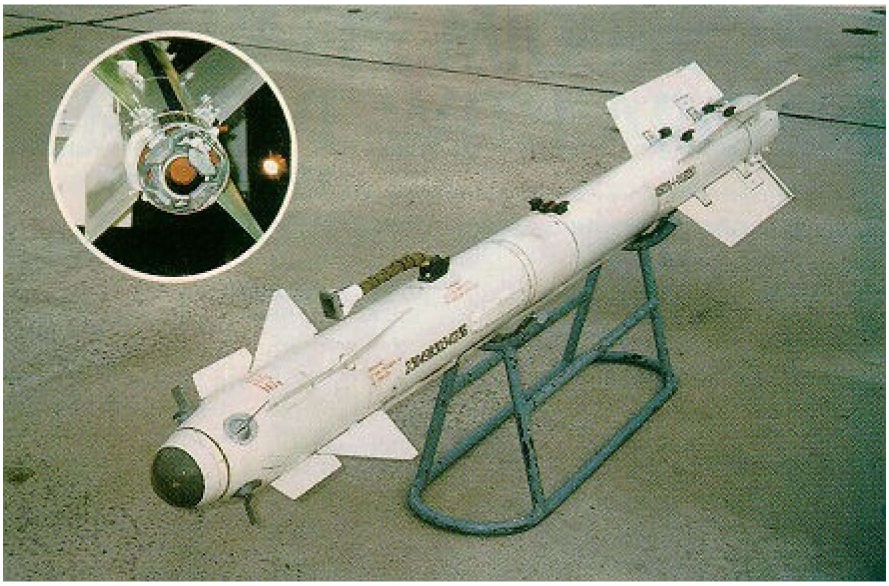
The R-27RE missile became a base developing a R-27AE and R-27EM missiles. The R-27AE (AA-10C "Alamo-C") is a medium-range missile, featuring an active radarhoming seeker. The R-27EM (AA-10C "Alamo-C") is optimized for a long-range low-level interception on water. Current status of this missile is unknown. Both missiles entered service in 1990.

One more variant of the R-27 is an R-27P (AA-10D "Alamo-D") missile with a passive seeker. This is an anti-radar homing missile launched against aircraft using active radars, such as AWACS and jamming aircraft. Missile's long-range variant with an extended range engine is the R-27PE. Both missile operational status is unknown.

Short-Range Air-to-Air Missile

R – 73

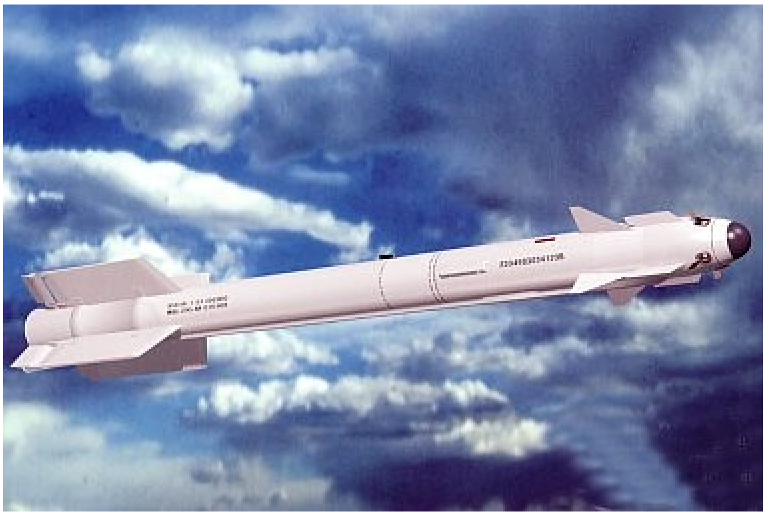
NATO Designation AA-11 "Archer"



Model	**R-73E**	**R-73M1**	**R-73M2**
Entered service in	?	1982	?
Range against closing target	< 30 km	30 km	40 km

Range against receding target	< 15 km	15 km	?
Weight	115 kg	105 kg	110 kg
Weight of warhead	7.4 kg	8 kg	?
Type of warhead	expanding rod		
Speed	Mach 2.5		
Guidance	infrared		
Kill probability	?	0.6	?
Length	2.9 m		
Diameter	0.17 m		
Fin span	0.51 m		
Carried by	Ka-50, Ka-52, MiG-29, MiG-31, Su-25, Su-27, Su-30, Su-33, Su-34, Su-35, Su-37, Su-39, Yak-141		

The R-73 short-range air-to-air missile was developed by "Molniya" (recently the special design bureau Nr.4) design bureau. It's team at the beginning of the 1970s developed the R-60 missile and the R-73 was intended to replace it. It is known as the AA-11 "Archer" with NATO countries.



Missile features a wide angle infrared seeker and extreme maneuverability. The R-73 is completed with a vectored trust system to make very tight turns. It's minimum range of fire is 0.3 km and missile is intended as a dogfight weapon in close air combats. Furthermore the R-73 is considered to be the most dangerous weapon system in close visual combat.

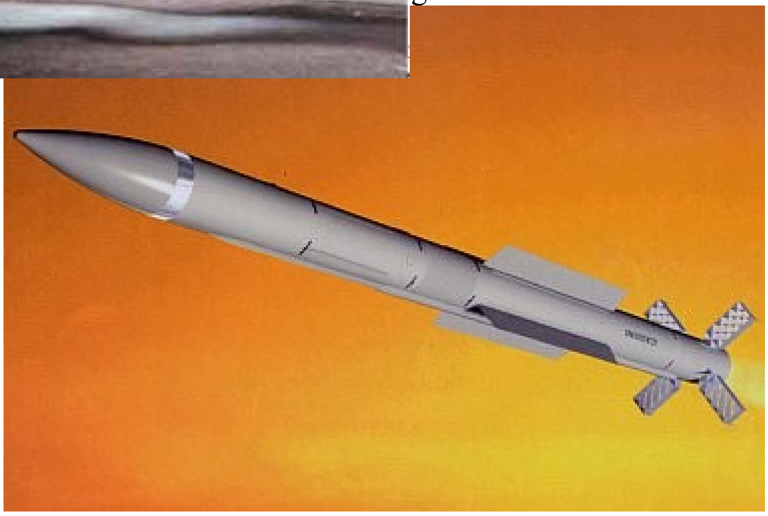
There were developed later variants of the R-73 missile:

- The R-73E missile features extended range;
- The R-73M1 (sometimes designated as R-73 RDM-1) features improved overall performance;
- The R-73M2 (R-73 RDM-2) has even better performance characteristics than it's predecessor - the R-73M1;
- The K-74ME.

All these missiles have the same AA-11 "Archer" NATO designation.

ssile

gnation AA-12 "Adder"



Entered service in	1994
Range against closing target	150 km
Range against receding target	50 - 90 km
Weight	175 kg
Weight of warhead	30 kg
Type of warhead	expanding rod
Speed	Mach 4+
Guidance	active radarhoming
Length	3.6 m
Diameter	0.2 m
Fin span	0.7 m
Carried by	Ka-50, Ka-52, MiG-29, MiG-31, Su-27, Su-30, Su-33, Su-34, Su-35, Su-37, Yak-141

The R-77 is a modern medium-range air-to-air missile. It was developed by "Vympel" design bureau since 1980s and entered service with Russian army in 1994. It is believed that it will be the standard Russian fighter aircraft missile. NATO countries designated it as the AA-12 "Adder".

It has an active radarhoming seeker and alike the R-27 medium-range missile it is guided to a certain point with the help of a data link. But then it uses an onboard radar to illuminate the target and steers towards it. Similar to the AIM-120 AMRAAM it gives the pilot a certain "fire and forget" capability. One more interesting detail is that it's NATO unofficial designation is "AMRAAMski". Missile has a range of fire in 0.5 - 150 kilometers.

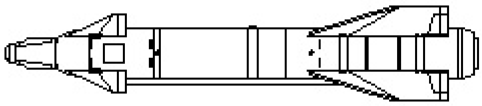
The R-77 became a base for the R-77E featuring extended range and even more improved R-77M. The last mentioned is completed with a different engine and has a greater weight. Relying on official sources the R-77 has a range of fire in 200 kilometers. This missile is expected to be in service not earlier than 2007.

Recently besides Russian air forces the R-77 missile were exported and are operational in India, Malaysia and Peru.

Air-to-Surface Missile

Kh - 29T

NATO Designation AS-14A "Kedge"

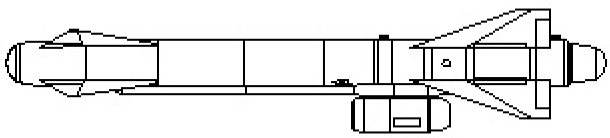


Entered service in	1980
Range	12 - 30 km
Weight	660 kg
Weight of warhead	320 kg
Speed	2 880 km/h
Guidance	TV
Deflection from the target	3.88 m
Carried by	MiG-23, MiG-29, Su-30MK, Su-35, Su-37 fighters; MiG-27, Su-17, Su-25, Su-39 ground attack aircraft; Su-24, Su-34 frontline bombers

Air-to-Surface Missile

Kh - 59M "Ovod-M"

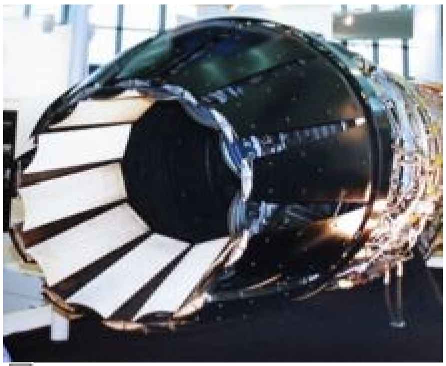
NATO Designation AS-18 "Kazoo"



Entered service in	1982
Range	120 km
Weight	920 kg
Weight of warhead	320 kg
Speed	1 020 km/h
Length	5.62 m
Carried by	Su-24, Su-34 frontline bombers

Thrust vectoring

Harrier AV-8A - world's first operational fighter jet with thrust vectoring



A thrust-vectoring jet engine nozzle

Thrust vectoring is the ability of an aircraft or other vehicle to direct the thrust from its main engine(s) in a direction other than parallel to the vehicle's longitudinal axis. The technique was originally envisaged to provide upward vertical thrust as a means to give aircraft vertical (VTOL) or short (STOL) takeoff and landing ability. Subsequently, it was realized that using vectored thrust in combat situations enabled aircraft to perform various maneuvers not available to conventional-engined planes. To perform turns, aircraft that use no thrust vectoring must rely on only ailerons or flaps; craft with vectoring still must use ailerons, but to a lesser extent.

Principles

Most currently operational vectored thrust aircraft use turbofans with rotating nozzles or vanes to deflect the exhaust stream. This method can successfully deflect thrust through as much as 90 degrees, relative to the aircraft centerline. However, the engine must be sized for vertical lift, rather than normal flight, which results in a weight penalty. Afterburning (or Plenum Chamber Burning, PCB, in the bypass stream) is difficult to incorporate and is impractical for take-off and landing, because the very hot exhaust can damage runway surfaces. Without afterburning it is hard to reach supersonic flight speeds. A PCB engine, the Bristol Siddeley BS100, was cancelled in 1965.

Tiltrotor aircraft vector thrust via rotating turboprop engine nacelles. The mechanical complexities of this design are quite troublesome, including twisting flexible internal components and driveshaft power transfer between engines.

Most current tiltrotor designs feature 2 rotors in a side-by-side configuration. If such a craft is flown in a way where it enters a vortex ring state, one of the rotors will always enter slightly before the other, causing the aircraft to perform a drastic and unplanned roll.

Thrust vectoring is also used as a control mechanism for airships, particularly modern non-rigid airships. In this use, most of the load is usually supported by buoyancy and vectored thrust is used to control the motion of the aircraft. But, designs have recently been proposed, especially for Project WALRUS, where a significant portion of the weight of the craft is supported by vectored thrust. The first airship that used a control system based on pressurized air was the Forlanini's *Omnia Dir* in 1930s.

Now being researched, fluidic injection nozzles divert thrust via fluid effects[1][2][3]. Tests show that air forced into a jet engine exhaust stream can deflect thrust up to 15 degrees. Such nozzles are desirable for their lower: mass and cost (up to 50% less), inertia (for faster, stronger control response), complexity (mechanically simpler, no moving parts or surfaces), and radar cross section for Stealth. This will likely be used in many unmanned aircraft, and 6th generation fighter aircraft.

Operational examples

The best known example of thrust vectoring in an engine is the Rolls-Royce Pegasus engine of the Hawker Siddeley Harrier brother to the BS100. (with variants built by McDonnell Douglas). Contrary to popular belief, the practice of using sudden changes as a combat maneuver was not applied against conventional Argentine fighters in the Falklands War. Thrust vectoring with the Harrier, also called "Vectoring In Forward Flight" or VIFFing, is actively discouraged by the Royal Air Force and Royal Navy, but encouraged and actively practiced by the United States Marine Corps. The technique has been used in various experimental and development planes, some with vectored thrust in directions other than downwards. Widespread use of thrust vectoring for maneuverability in a Western fighter aircraft would have to wait for the 21st century, and the deployment of the Lockheed Martin F-22 Raptor fifth-generation jet fighter, with its afterburning, thrust-vectoring Pratt & Whitney F119 turbofan.

Lockheed Martin F-35 Lightning II is currently in the pre-production test and development stage. Although this aircraft incorporates a conventional afterburning turbofan (F135 or F136) which facilitates supersonic operation, the variant for the US Marine Corp and RAF also incorporates a vertically mounted, Low pressure shaft-driven remote fan, which is driven through a clutch during landing from the engine. The exhaust from this fan is deflected by a thrust vectoring nozzle, as is the main engine exhaust, to provide the appropriate combination of lift and propulsive thrust during transition.

The F-18 HARV, X-31, and F-16 MATV in flight.

Rockets or rocket-powered aircraft can also use thrust vectoring. In particular, many missiles use this technique since at launch they are moving too slowly to be able to steer effectively without massive fins (and their accompanying drag penalty at high speeds). In addition, rockets often go near or beyond the edge of the atmosphere, where aerodynamic surfaces are useless, making gas-dynamic steering necessary. Examples of rockets and missiles which use thrust vectoring include both large systems such as the Space Shuttle SRB, S-300P (SA-10) surface-to-air missile, UGM-27 Polaris nuclear ballistic missile and RT-23 (SS-24) ballistic missile and smaller battlefield weapons such as Swingfire.

List of vectored thrust aircraft

Thrust vectoring can convey two main benefits: VTOL/STOL, and higher maneuverability. Aircraft are usually optimized to maximally exploit one benefit, though will gain in the other.

For VTOL ability

- Harrier Jump Jet
  - Hawker Siddeley Harrier
  - British Aerospace Sea Harrier
  - Boeing/BAE Systems AV-8B Harrier II
  - BAE Systems/Boeing Harrier II
- Boeing V-22 Osprey (Turboprop)
- Boeing X-32
- Lockheed Martin F-35 Lightning II (B model)
- Moller Skycar
- Dornier Do 31
- Armstrong Whitworth AW.681
- Yakovlev Yak-38
- Yakovlev Yak-141

For higher maneuverability

Two dimension vectoring (pitch axis)

- Lockheed Martin F-22 Raptor
- McDonnell Douglas F-15S/MTD
- McDonnell Douglas X-36
- Sukhoi Su-30MKI
- Super-10
- JF-17 (speculated future addition)

Three dimension vectoring (pitch and yaw axes)

- Lockheed F-16 MATV
- McDonnell Douglas F-15 ACTIVE
- McDonnell Douglas F-18 HARV
- Mikoyan-Gurevich MiG-35 MFI
- Mikoyan Project 1.44
- Rockwell-MBB X-31
- Sukhoi Su-37
- Sukhoi Su-47
- X-44 MANTA

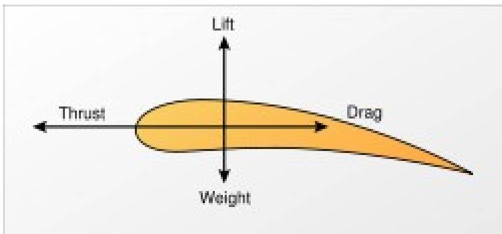
Other

- Zeppelin NT modern thrust vectoring airship

Thrust

Thrust is a reaction force described quantitatively by Newton's Second and Third Laws. When a system expels or accelerates mass in one direction the accelerated mass will cause a proportional but opposite force on that system.

Examples



Forces on an aircraft

A fixed-wing aircraft generates forward thrust when a spinning propeller moves air, or gases are ejected from a jet engine (or rocket engine), opposite the direction of flight. The forward thrust is proportional to the (mass of the air) multiplied by (average velocity of the airstream). Reverse thrust can be generated to aid braking after landing by reversing the pitch of variable pitch propeller blades, or using a thrust reverser on a jet engine. Rotary wing aircraft and thrust vectoring V/STOL aircraft use engine thrust to support the weight of the aircraft, and vector some of this thrust fore and aft to control forward speed.

A motorboat generates thrust (or reverse thrust) when the propellers are turned to accelerate water backwards (or forwards). The resulting thrust pushes the boat in the equal and opposite direction to the sum of the momentum change in the water flowing through the propeller.

A rocket's mass is propelled forward by a thrust force equal to, and opposite of, the time-rate of momentum change of the exhaust mass accelerated from the combustion chamber through the rocket engine nozzle. This is the exhaust velocity with respect to the rocket, times the time-rate at which the mass is expelled, or in mathematical terms:

$$T = v\frac{dm}{dt}$$

where:

T = thrust generated (force),
$\frac{dm}{dt} =$ rate of change of mass with respect to time (fuel burn rate).
v = exhaust velocity.

Of course, for a launch the thrust at lift-off should be more than the weight, and with a fair margin, because a "slow launch" would be very inefficient.

Each of the three Space shuttle main engines can produce a thrust of 1.8 MN, and each of its two Solid Rocket Boosters 14.7 MN, together 34.8 MN. Compare with the mass at lift-off of 2,040,000 kg, hence a weight of 20 MN.

By contrast, the simplified Aid for EVA Rescue (SAFER) has 24 thrusters of 3.56 N each.

In the air breathing category, the AMT-USA AT-180 jet engine developed for radio-controlled aircraft produce 90 N (20 Lb_f) of thrust.[1] The GE90-115B engines fitted on the Boeing 777-300ER, recognized by the Guinness Book of World Records as the "World's Most Powerful Commercial Jet Engine," have a tested thrust of 569 kN (127,900 lb_f).

VTOL

The **Hawker Harrier**, one of the famous examples of a plane with VTOL capability.

VTOL is an abbreviaton for **Vertical Take-Off and Landing**. **VTOL** describes fixed-wing aircraft that can lift off vertically. This classification includes only a very few aircraft; helicopters, autogyros; balloons and airships are not normally considered VTOL[citation needed]. Some aircraft can operate in VTOL mode in addition to others, such as CTOL (Conventional Take-off and Landing) and/or STOL (Short Take-Off and Landing). Others can only operate by VTOL, due to the aircraft lacking landing gear that can handle horizontal motion.

Currently there are two types of practical VTOL aircraft in service:

- Using a tiltrotor — the Bell Boeing V-22 Osprey
- Using directed jet thrust — the Harrier Jump Jet family.

History

In 1928, Nikola Tesla received patents for an apparatus for aerial transportation. Tesla called it the "*Flivver*". It is one of the earliest examples of VTOL aircraft.

In late World War II, Germany designers studied the possibility of a VTOL aircraft, the Heinkel Lerche, but the plan never got off the drawing board.

An early contribution to VTOL was Rolls-Royce's Thrust Measuring Rig ("flying bedstead") of 1953. This led to the first VTOL engines as used in the first British VTOL aircraft, the Short SC.1 (1957) which used a 4 vertical lift engines with a horizontal one for forward thrust.

The idea of using the same engine for vertical and horizontal flight by altering the path of the thrust led to the Bristol Siddeley Pegasus engine which used rotating ducts to direct thrust over a range of angles. This was developed side by side with an airframe, the Hawker P.1127, which became subsequently the Kestrel and then entered production as the Hawker Siddeley Harrier though the supersonic Hawker Siddeley P.1154 was cancelled in 1965.

The Harrier is often flown in STOVL mode which enables it to carry a higher fuel or weapon load over a given distance. The Indian Navy operate Sea Harriers, mainly from its aircraft carrier. The United States Marine Corps, and the Italian and Spanish Navies use the AV-8 Harrier II, an advanced derivative of the Harrier. The Harrier II will be replaced in the air arms of the US and UK by a STOVL variant of the F-35 Joint Strike Fighter.

NASA has flown other VTOL craft such as the XV-15 research craft (1977), as have the Soviet Navy and *Luftwaffe*. Sikorsky tested an aircraft dubbed the X-Wing, which took off in the manner of a helicopter. The rotors would become stationary in mid-flight, and function as wings, providing lift in addition to the static wings. Boeing X-50 is a Canard Rotor/Wing prototype that utilizes a similar concept.

In the 1960s France developed a version of the Dassault Mirage III capable of attaining Mach 1. The Dassault Mirage III - V *Balzac* (not to be confused with the Mirage 5) achieved transition from vertical to horizontal flight in March of 1966 and reached Mach 1.3 in level flight a short time later.

The Soviet Yak-38 Forger was the Soviet Navy's VTOL aircraft for their light carriers, cargoships, and capital ships. It was developed from the Yak-36 Freehand experimental aircraft. Before the Soviet Union collapsed, a supersonic VTOL aircraft was developed as the Yak-38's successor, the Yak-141, which never went into production. The Yak-141, also called Yak-41 was further developed into the Yak-43.

In the 1960s and early 70s Germany used the F-104 as a base for research for a V/STOL aircraft. Although two models (X1 and X2) were built, the project was canceled due to high costs and political problems as well as changed needs in the Luftwaffe and NATO. The EWR VJ 101C did perform free VTOL take-offs and landings, as well as test flights beyond mach 1 in the mid- and late 60s. One of the test-aircraft is preserved in the Deutsches Museum in Munich, Germany.

The Moller Skycar is a prototype personal VTOL aircraft -- literally, a "flying air vehicle" (PAV). It has, as of this date, never made the transition to level flight, nor has it ever flown with anybody on board.

Aircraft designed to operate in extraterrestrial environments often utilize VTOL. An example of this type of aircraft is the LLRV. Spacecraft typically operate in environments where runways or even a suitably flat surface for skids is nonexistent.

STOL

STOL is an acronym for *Short Take-Off and Landing*, a term used in the aircraft industry to describe aeroplanes with very short runway requirements.

The formal NATO definition (since 1964) is:

> *Short Take-Off and Landing (décollage et atterrissage courts) is the ability of an aircraft to clear a 15 m (50 ft) obstacle within 450 m (1,500 ft) of commencing take-off or, in landing, to stop within 450 m (1,500 ft) after passing over a 15 m (50 ft) obstacle.*

Many STOL aircraft are bush planes, though some, like the de Havilland Dash-7, are designed for use on prepared airstrips; likewise, many STOL aircraft are taildraggers, though there are exceptions like the de Havilland Twin Otter, the Cessna 208, the Yakovlev Yak-40, and the Peterson 260SE.

Runway length requirement is a function of the square of the minimum flying speed (stall speed), and most design effort is spent on reducing this number. For takeoff, large power/weight ratios and low drag help the plane to accelerate for flight. The landing run is minimized by strong brakes, low landing speed or spoilers (less common). Overall STOL performance is set by the length of runway needed to land or take off, whichever is longer.

Of equal importance to short ground run is the ability to clear obstacles, such as trees, on both take off and landing. For takeoff, large power/weight ratios and low drag result in a high rate of climb required to clear obstacles. For landing, high drag allows the aeroplane to descend steeply to the runway without building excess speed resulting in a longer ground run. Drag is increased by use of flaps (devices on the wings) and by a forward slip (causing the aeroplane to fly somewhat sideways though the air to increase drag).

Normally, a STOL plane will have a large wing for its weight. These wings often use aerodynamic devices like flaps, slots, slats, and vortex generators. Typically, designing an aeroplane for excellent STOL performance reduces maximum speed, but does not reduce payload lifting ability. The payload is critical, because many small, isolated communities rely on STOL aircraft as their only transportation link to the outside world for passengers or cargo; examples include many communities in the Canadian north and Alaska.

A Zenair CH701 **STOL** light aircraft

Most STOL aeroplanes can land either on- or off-airport. Typical off-airport landing areas include snow or ice (using skis), fields or gravel riverbanks (often using special fat, low-pressure tundra tires), and water (using floats): these areas are often extremely short and obstructed by tall trees or hills. Wheel skis and amphibious floats combine wheels with skis or floats, allowing the choice of landing on snow/water or a prepared runway. A STOLport is an airport designed with STOL operations in mind, normally having a short single runway. These are not common but can be found, for example, at London City Airport in England.

List of some STOL aircraft

- ATR 42/72
- Alenia G.222
- American Champion/Bellanca/Champion Citabria
- Cessna 208
- 8CKAB Decathlon
- American Champion/Bellanca Scout
- Antonov An-72
- Aviat Husky
- BAC TSR-2
- BAe 146
- Boeing C-17
- Crouch-Bolas Dragonfly
- Dornier Do 228
- Fieseler Fi 156
- De Havilland DHA-3 Drover
- De Havilland Canada DHC-2 Beaver
- De Havilland Canada DHC-4 Caribou
- De Havilland Canada DHC-5 Buffalo
- De Havilland Canada DHC-6 Twin Otter
- De Havilland Canada Dash 7

- De Havilland Canada Dash 8
- Fairchild-Dornier 328 family
- Eurofighter Typhoon
- Helio Courier
- Hunting H126
- Ilyushin Il-76
- IAI Arava
- Jetpod
- Lockheed C-130
- Maule
- Wren 460 and Peterson 260SE
- Pilatus PC-6
- Piper Cub
- Piper PA-20 Pacer family
- PZL-104 Wilga
- Saab Gripen
- Saab Viggen
- Westland Lysander
- Kitfox
- Yakovlev Yak-40
- Zenith STOL CH 701

Radar

This long range radar antenna, known as ALTAIR, is used to detect and track space objects in conjunction with ABM testing at the Ronald Reagan Test Site on the Kwajalein atoll.[1]

Radar is a system that uses electromagnetic waves to identify the range, altitude, direction, or speed of both moving and fixed objects such as aircraft, ships, motor vehicles, weather formations, and terrain. A transmitter emits radio waves, which are reflected by the target and detected by a receiver, typically in the same location as the transmitter. Although the radio signal returned is usually very weak, radio signals can easily be amplified. This enables a radar to detect objects at ranges where other emissions, such as sound or visible light, would be too weak to detect. Radar is used in many contexts, including meteorological detection of precipitation, air traffic control, police detection of speeding traffic, and by the military. The term *RADAR* was coined in 1941 as an acronym for **Ra**dio **D**etection **a**nd **R**anging. This acronym of American origin replaced the previously used British abbreviation *RDF* (which stands for *Radio Direction Finding*). The term has since entered the English language as a standard word, *radar*, losing the capitalization in the process.

History

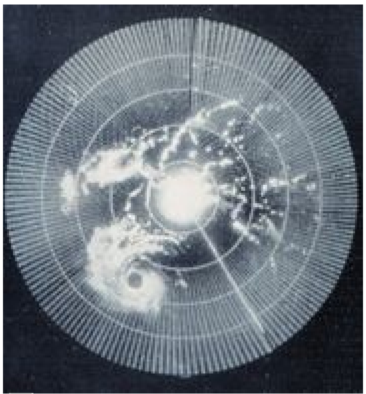
Several inventors, scientists, and engineers contributed to the development of radar. The use of radio waves to detect "the presence of distant metallic objects via radio waves" was first implemented in 1904 by Christian Hülsmeyer,[2][3] who demonstrated the feasibility of detecting the presence of a ship in dense fog, but not its distance. He received a Reichspatent patent Nr. 165546 for his pre-radar device in April and on November 11, 1904 the patent 169154 an amendment of his patent for ranging that is indirectly related to his device. He received a patent (GB13170) in England for his *telemobiloscope* on September 22 1904.[2][4]

Prior to the Second World War, developments by the Americans (Dr. Robert M. Page tested the first monopulse radar in 1934),[5] the Germans, the French (French Patent n° 788795 in 1934),[6] and the British (British Patent GB593017 by Robert Watson-Watt in 1935),[6][7] led to the first real radars. Hungarian Zoltán Bay produced a working model by 1936 at the Tungsram laboratory in the same vein.

The war precipitated the research to find better resolution, more portability, more features for that new defensive weapon. Post-war years have seen the use of radar in fields as diverse as air traffic control, weather monitoring, astrometry and road speed control.

Principles

Reflection



Brightness can indicate reflectivity as in this 1960 weather radar image. The radar's frequency, pulse form, and antenna largely determine what it can observe.

Electromagnetic waves reflect (scatter) from any large change in the dielectric or diamagnetic constants. This means that a solid object in air or a vacuum, or other significant change in atomic density between the object and what's surrounding it, will usually scatter radar (radio) waves. This is particularly true for electrically conductive materials, such as metal and carbon fibre, making radar particularly well suited to the detection of aircraft and ships. Radar absorbing material, containing resistive and sometimes magnetic substances, is used on military vehicles to reduce radar reflection. This is the radio equivalent of painting something a dark color.

Radar waves scatter in a variety of ways depending on the size (wavelength) of the radio wave and the shape of the target. If the wavelength is much shorter than the target's size, the wave will bounce off in a way similar to the way light is reflected by a mirror. If the wavelength is much longer than the size of the target, the target is polarized (positive and negative charges are separated), like a dipole antenna. This is described by Rayleigh scattering, an effect that creates the Earth's blue sky and red sunsets. When the two length scales are comparable, there may be resonances. Early radars used very long wavelengths that were larger than the targets and received a vague signal, whereas some modern systems use shorter wavelengths (a few centimetres or shorter) that can image objects as small as a loaf of bread.

Short radio waves reflect from curves and corners, in a way similar to glint from a rounded piece of glass. The most reflective targets for short wavelengths have 90° angles between the reflective surfaces. A structure consisting of three flat surfaces meeting at a single corner, like the corner on a box, will always reflect waves entering its opening directly back at the source. These so-called corner reflectors are commonly used as radar reflectors to make otherwise difficult-to-detect objects easier to detect, and are often found on boats in order to improve their detection in a rescue situation and to reduce collisions. For similar reasons, objects attempting to avoid detection will angle their surfaces in a way to eliminate inside corners and avoid surfaces and edges perpendicular to likely detection directions, which leads to "odd" looking stealth aircraft. These precautions do not completely eliminate reflection because of diffraction, especially at longer wavelengths. Half wavelength long wires or strips of conducting material, such as chaff, are very reflective but do not direct the scattered energy back toward the source. The extent to which an object reflects or scatters radio waves is called its radar cross section.

Radar equation

The amount of power P_r returning to the receiving antenna is given by the radar equation:

$$P_r = \frac{P_t G_t A_r \sigma F^4}{(4\pi)^2 R_t^2 R_r^2}$$

where*'

- P_t = transmitter power
- G_t = gain of the transmitting antenna
- A_r = effective aperture (area) of the receiving antenna
- σ = radar cross section, or scattering coefficient, of the target
- F = pattern propagation factor
- R_t = distance from the transmitter to the target
- R_r = distance from the target to the receiver.

In the common case where the transmitter and the receiver are at the same location, $R_t = R_r$ and the term $R_t^2 R_r^2$ can be replaced by R^4, where R is the range. This yields:

$$P_r = \frac{P_t G_t A_r \sigma F^4}{(4\pi)^2 R^4}$$

This shows that the received power declines as the fourth power of the range, which means that the reflected power from distant targets is very, very small.

The equation above with $F = 1$ is a simplification for vacuum without interference. The propagation factor accounts for the effects of multipath and shadowing and depends on the details of the environment. In a real-world situation, pathloss effects should also be considered.

Other mathematical developments in radar signal processing include time-frequency analysis (Weyl Heisenberg or wavelet), as well as the chirplet transform which makes use of the fact that radar returns from moving targets typically "chirp" (change their frequency as a function of time, as does the sound of a bird or bat).

Polarization

In the transmitted radar signal, the electric field is perpendicular to the direction of propagation, and this direction of the electric field is the Polarization of the wave. Radars use horizontal, vertical, linear and circular polarization to detect different types of reflections. For example, circular polarization is used to minimize the interference caused by rain. Linear polarization returns usually indicate metal surfaces. Random polarization returns usually indicate a fractal surface, such as rocks or soil, and are used by navigational radars.

Interference

Radar systems must overcome several different sources of unwanted signals in order to focus only on the actual targets of interest. These unwanted signals may originate from internal and external sources, both passive and active. The ability of the radar system to overcome these unwanted signals defines its signal-to-noise ratio (SNR): the higher a system's SNR, the better it is in isolating actual targets from the surrounding noise signals.

Noise

Signal noise is an internal source of random variations in the signal, which is inherently generated to some degree by all electronic components. Noise typically appears as random variations superimposed on the desired echo signal received in the radar receiver. The lower the power of the desired signal, the more difficult it is to discern it from the noise (similar to trying to hear a whisper while standing near a busy road). Therefore, the most important noise sources appear in the receiver and much effort is made to minimize these factors. Noise figure is a measure of the noise produced by a receiver compared to an ideal receiver, and this needs to be minimized.

Noise is also generated by external sources, most importantly the natural thermal radiation of the background scene surrounding the target of interest. In modern radar systems, due to the high performance of their receivers, the internal noise is typically about equal to or lower than the external scene noise. An exception is if the radar is aimed upwards at clear sky, where the scene is so cold that it generates very little thermal noise.

Clutter

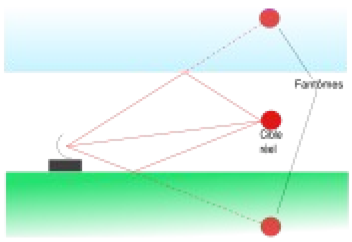
Clutter refers to actual radio frequency (RF) echoes returned from targets which are by definition uninteresting (i.e unwanted targets)to the radar operators in general. Such targets mostly include natural objects such as ground, sea, precipitation (such as rain, snow or hail), sand storms, animals (especially birds), atmospheric turbulence, and other atmospheric effects (such as ionosphere reflections and meteor trails). Clutter may also be returned from man-made objects such as buildings and, intentionally, by radar countermeasures such as chaff.

Some clutter may also be caused by a long waveguide between the radar transceiver and the antenna. In a typical plan position indicator (PPI) radar with a rotating antenna, this will usually be seen as a "sun" or "sunburst" in the centre of the display as the receiver responds to echoes from dust particles and misguided RF in the waveguide. Adjusting the timing between when the transmitter sends a pulse and when the receiver stage is enabled will generally reduce the sunburst without affecting the accuracy of the range, since most sunburst is caused by diffused transmit pulse reflected before it leaves the antenna.

While some clutter sources may be undesirable for some radar applications (such as storm clouds for air-defence radars), they may be desirable for others (meteorological radars in this example). Clutter is considered a passive interference source, since it only appears in response to radar signals sent by the radar.

There are several methods of detecting and neutralizing clutter. Many of these methods rely on the fact that clutter tends to appear static between radar scans. Therefore, when comparing subsequent scans echoes, desirable targets will appear to move and all stationary echoes can be eliminated. Sea clutter can be reduced by using horizontal polarization, while rain is reduced with circular polarization (note that meteorological radars wish for the opposite effect, therefore using linear polarization the better to detect precipitation). Other methods attempt to increase the signal-to-clutter ratio.

CFAR (Constant False-Alarm Rate, a form of Automatic Gain Control, or AGC) is a method relying on the fact that clutter returns far outnumber echoes from targets of interest. The receiver's gain is automatically adjusted to maintain a constant level of overall visible clutter. While this does not help detect targets masked by stronger surrounding clutter, it does help to distinguish strong target sources. In the past, radar AGC was electronically controlled and affected the gain of the entire radar receiver. As radars evolved, AGC became computer-software controlled, and affected the gain with greater granularity, in specific detection cells.



Radar multipath echoes from an actual target cause ghosts to appear.

Clutter may also originate from multipath echoes from valid targets due to ground reflection, atmospheric ducting or ionospheric reflection/refraction. This specific clutter type is especially bothersome, since it appears to move and behave like other normal (point) targets of interest, thereby creating a ghost. In a typical scenario, an aircraft echo is multipath-reflected from the ground below, appearing to the receiver as an identical target below the correct one. The radar may try to unify the targets, reporting the target at an incorrect height, or - worse - eliminating it on the basis of jitter or a physical impossibility. These problems can be overcome by incorporating a ground map of the radar's surroundings and eliminating all echoes which appear to originate below ground or above a certain height.

Jamming

[Radar jamming](#) refers to RF signals originating from sources outside the radar, transmitting in the radar's frequency and thereby masking targets of interest. Jamming may be intentional, as with an anti-radar electronic warfare (EW) tactic, or unintentional, as with friendly forces operating equipment that transmits using the same frequency range. Jamming is considered an active interference source, since it is initiated by elements outside the radar and in general unrelated to the radar signals.

Jamming is problematic to radar since the jamming signal only needs to travel one-way (from the jammer to the radar receiver) whereas the radar echoes travel two-ways (radar-target-radar) and are therefore significantly reduced in power by the time they return to the radar receiver. Jammers therefore can be much less powerful than their jammed radars and still effectively mask targets along the line of sight from the jammer to the radar (*Mainlobe Jamming*). Jammers have an added effect of affecting radars along other line-of-sights, due to the radar receiver's sidelobes (*Sidelobe Jamming*).

Mainlobe jamming can generally only be reduced by narrowing the mainlobe solid angle, and can never fully be eliminated when directly facing a jammer which uses the same frequency and polarization as the radar. Sidelobe jamming can be overcome by reducing receiving sidelobes in the radar antenna design and by using an omnidirectional antenna to detect and disregard non-mainlobe signals. Other anti-jamming techniques are frequency hopping and polarization. See Electronic counter-counter-measures for details.

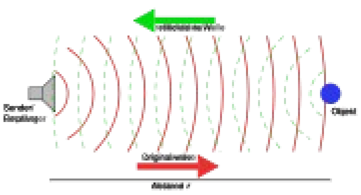
Interference has recently become a problem for C-band (5.66 GHz) meteorological radars with the proliferation of 5.4 GHz band WiFi equipment.[8]

Radar signal processing

Distance measurement

Transit time

Pulse radar



One way to measure the distance to an object is to transmit a short pulse of radio signal (electromagnetic radiation), and measure the time it takes for the reflection to return. The distance is one-half the product of round trip time (because the signal has to travel to the target and then back to the receiver) and the speed of the signal. Since radio waves travel at the speed of light (186,000 miles per second or 300,000,000 meters per second),accurate distance measurement requires high-performance electronics.

In most cases, the receiver does not detect the return while the signal is being transmitted. Through the use of a device called a *diplexer*, the radar switches between transmitting and receiving at a predetermined rate. The minimum range is calculated by measuring the length of the pulse multiplied by the speed of light, divided by two. In order to detect closer targets one must use a shorter pulse length.

A similar effect imposes a maximum range as well. If the return from the target comes in when the next pulse is being sent out, once again the receiver cannot tell the difference. In order to maximize range, one wants to use longer times between pulses, the inter-pulse time.

These two effects tend to be at odds with each other, and it is not easy to combine both good short range and good long range in a single radar. This is because the short pulses needed for a good minimum range broadcast have less total energy, making the returns much smaller and the target harder to detect. This could be offset by using more pulses, but this would shorten the maximum range again. So each radar uses a particular type of signal. Long range radars tend to use long pulses with long delays between them, and short range radars use smaller pulses with less time between them. This pattern of pulses and pauses is known as the pulse repetition frequency (or PRF), and is one of the main ways to characterize a radar. As electronics have improved many radars now can change their PRF thereby changing their range.

The distance resolution and the characteristics of the received signal as compared to noise depends heavily on the shape of the pulse. The pulse is often modulated to achieve better performance thanks to a technique known as pulse compression.

Frequency modulation

Another form of distance measuring radar is based on frequency modulation. Frequency comparison between two signals is considerably more accurate, even with older electronics, than timing the signal. By changing the frequency of the returned signal and comparing that with the original, the difference can be easily measured.

This technique can be used in continuous wave radar, and is often found in aircraft radar altimeters. In these systems a "carrier" radar signal is frequency modulated in a predictable way, typically varying up and down with a sine wave or sawtooth pattern at audio frequencies. The signal is then sent out from one antenna and received on another, typically located on the bottom of the aircraft, and the signal can be continuously compared.

Since the signal frequency is changing, by the time the signal returns to the aircraft the broadcast has shifted to some other frequency. The amount of that shift is greater over longer times, so greater frequency differences mean a longer distance, the exact amount being the "ramp speed" selected by the electronics. The amount of shift is therefore directly related to the distance traveled, and can be displayed on an instrument. This signal processing is similar to that used in speed detecting Doppler radar.

Speed measurement

Speed is the change in distance to an object with respect to time. Thus the existing system for measuring distance, combined with a memory capacity to see where the target last was, is enough to measure speed. At one time the memory consisted of a user making grease-pencil marks on the radar screen, and then calculating the speed using a slide rule. Modern radar systems perform the equivalent operation faster and more accurately using computers.

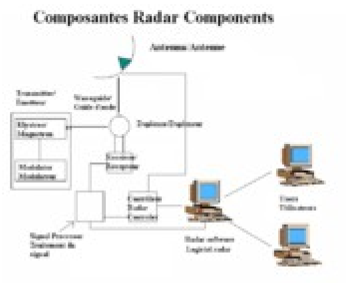
However, if the transmitter's output is coherent (phase synchronized), there is another effect that can be used to make almost instant speed measurements (no memory is required), known as the Doppler effect. Most modern radar systems use this principle in the pulse-doppler radar system. Return signals from targets are shifted away from this base frequency via the Doppler effect enabling the calculation of the speed of the object relative to the radar. The Doppler effect is only able to determine the relative speed of the target along the line of sight from the radar to the target. Any component of target velocity perpendicular to this line of sight cannot be determined by Doppler alone tracking the target's azimuth over time must be used. Additional information of the nature of the Doppler returns may be found in the Radar signal characteristics article.

It is also possible to make a radar without any pulsing, known as a continuous-wave radar (CW radar), by sending out a very pure signal of a known frequency. CW radar is ideal for determining the radial component of a target's velocity, but it cannot determine the target's range. CW radar is typically used by traffic enforcement to measure vehicle speed quickly and accurately where range is not important.

Reduction of interference effects

Signal processing is employed in radar systems to reduce the interference effects. Signal processing techniques include moving target indication (MTI), pulse doppler, moving target detection (MTD) processors, correlation with secondary surveillance radar (SSR) targets and space-time adaptive processing (STAP). Constant false alarm rate (CFAR) and digital terrain model (DTM) processing are also used in clutter environments.

Radar engineering



Radar components

A radar has different components:

- A transmitter that generates the radio signal with an oscillator such as a klystron or a magnetron and controls its duration by a modulator.
- A waveguide that links the transmitter and the antenna.
- A duplexer that serves as a switch between the antenna and the transmitter or the receiver for the signal when the antenna is used in both situations.
- A receiver.
- An electronic section that controls all those devices and the antenna to perform the radar scan ordered by a software.
- A link to end users.

Antenna design

Radio signals broadcast from a single antenna will spread out in all directions, and likewise a single antenna will receive signals equally from all directions. This leaves the radar with the problem of deciding where the target object is located.

Early systems tended to use omni-directional broadcast antennas, with directional receiver antennas which were pointed in various directions. For instance the first system to be deployed, Chain Home, used two straight antennas at right angles for reception, each on a different display. The maximum return would be detected with an antenna at right angles to the target, and a minimum with the antenna pointed directly at it (end on). The operator could determine the direction to a target by rotating the antenna so one display showed a maximum while the other shows a minimum.

One serious limitation with this type of solution is that the broadcast is sent out in all directions, so the amount of energy in the direction being examined is a small part of that transmitted. To get a reasonable amount of power on the "target", the transmitting aerial should also be directional.

Parabolic reflector

More modern systems used a steerable parabolic "dish" to create a tight broadcast beam, typically using the same dish as the receiver. Such systems often combine two radar frequencies in the same antenna in order to allow automatic steering, or **radar lock**.

Types of scan

Primary Scan – A scanning technique where the main antenna aerial is moved to produce a scanning beam, examples include circular scan, sector scan etc

Secondary Scan – A scanning technique where the antenna feed is moved to produce a scanning beam, example include conical scan, unidirectional sector scan, loge switching etc.

Palmer Scan – A scanning technique that produces a scanning beam by moving the main antenna and its feed. A Palmer Scan is a combination of a Primary Scan and a Secondary Scan.

Phased array: Not all radar antennas must rotate to scan the sky.

Slotted waveguide

Applied similarly to the parabolic reflector the slotted waveguide is moved mechanically to scan and is particularly suitable for non-tracking surface scan systems, where the vertical pattern may remain constant. Owing to lower cost and less wind exposure, shipboard, airport surface, and harbour surveillance radars now use this in preference to the parabolic antenna.

Phased array

Another method of steering is used in a phased array radar. This uses an array of similar aerials suitably spaced, the phase of the signal to each individual aerial being controlled so that the signal is reinforced in the desired direction and cancels in other directions. If the individual aerials are in one plane and the signal is fed to each aerial in phase with all others then the signal will reinforce in a direction perpendicular to that plane. By altering the relative phase of the signal fed to each aerial the direction of the beam can be moved because the direction of constructive interference will move. Because phased array radars require no physical movement the beam can scan at thousands of degrees per second, fast enough to irradiate and track many individual targets, and still run a wide-ranging search periodically. By simply turning some of the antennas on or off, the beam can be spread for searching, narrowed for tracking, or even split into two or more virtual radars. However, the beam cannot be effectively steered at small angles to the plane of the array, so for full coverage multiple arrays are required, typically disposed on the faces of a triangular pyramid (see picture).

Phased array radars have been in use since the earliest years of radar use in World War II, but limitations of the electronics led to fairly poor accuracy. Phased array radars were originally used for missile defense. They are the heart of the ship-borne Aegis combat system, and the Patriot Missile System, and are increasingly used in other areas because the lack of moving parts makes them more reliable, and sometimes permits a much larger effective antenna, useful in fighter aircraft applications that offer only confined space for mechanical scanning.

As the price of electronics has fallen, phased array radars have become more and more common. Almost all modern military radar systems are based on phased arrays, where the small additional cost is far offset by the improved reliability of a system with no moving parts. Traditional moving-antenna designs are still widely used in roles where cost is a significant factor such as air traffic surveillance, weather radars and similar systems.

Phased array radars are also valued for use in aircraft, since they can track multiple targets. The first aircraft to use a phased array radar is the B-1B Lancer. The first aircraft fighter to use phased array radar was the Mikoyan MiG-31. The MiG-31M's SBI-16 Zaslon phased array radar is considered to be the world's most powerful fighter radar [1].

Frequency bands

The traditional band names originated as code-names during World War II and are still in military and aviation use throughout the world in the 21st century. They have been adopted in the United States by the IEEE, and internationally by the ITU. Most countries have additional regulations to control which parts of each band are available for civilian or military use.

Other users of the radio spectrum, such as the broadcasting and electronic countermeasures (ECM) industries, have replaced the traditional military designations with their own systems.

Radar frequency bands

Band Name	Frequency Range	Wavelength Range	Notes
HF	3-30 MHz	10-100 m	coastal radar systems, over-the-horizon (OTH) radars; 'high frequency'
P	< 300 MHz	1 m+	'P' for 'previous', applied retrospectively to early radar systems
VHF	50-330 MHz	0.9-6 m	very long range, ground penetrating; 'very high frequency'
UHF	300-1000 MHz	0.3-1 m	very long range (e.g. ballistic missile early warning), ground penetrating, foliage penetrating; 'ultra high

			frequency'
L	1-2 GHz	15-30 cm	long range air traffic control and surveillance; 'L' for 'long'
S	2-4 GHz	7.5-15 cm	terminal air traffic control, long range weather, marine radar; 'S' for 'short'
C	4-8 GHz	3.75-7.5 cm	Satellite transponders; a compromise (hence 'C') between X and S bands; weather
X	8-12 GHz	2.5-3.75 cm	missile guidance, marine radar, weather, medium-resolution mapping and ground surveillance; in the USA the narrow range 10.525 GHz ±25 MHz is used for airport radar. Named X band because the frequency was a secret during WW2.
K_u	12-18 GHz	1.67-2.5 cm	high-resolution mapping, satellite altimetry; frequency just under K band (hence 'u')
K	18-27 GHz	1.11-1.67 cm	from German *kurz*, meaning 'short'; limited use due to absorption by water vapour, so K_u and K_a were used instead for surveillance. K-band is used for detecting clouds by meteorologists, and by police for detecting speeding motorists. K-band radar guns operate at 24.150 ± 0.100 GHz.
K_a	27-40 GHz	0.75-1.11 cm	mapping, short range, airport surveillance; frequency just above K band (hence 'a') Photo radar, used to trigger cameras which take pictures of license plates of cars running red lights, operates at 34.300 ± 0.100 GHz.
mm	40-300 GHz	7.5 mm - 1 mm	millimetre band, subdivided as below. The letter designators appear to be random, and the frequency ranges dependent on waveguide size. Multiple letters are assigned to these bands by different groups. These are from Baytron, a now defunct company that made test equipment.
Q	40-60 GHz	7.5 mm - 5 mm	Used for Military communication.
V	50-75 GHz	6.0 - 4 mm	Very strongly absorbed by the atmosphere.
E	60-90 GHz	6.0 - 3.33 mm	
W	75-110 GHz	2.7 - 4.0 mm	used as a visual sensor for experimental autonomous vehicles, high-resolution meteorological observation, and imaging.

Radar modulators

Modulators are sometimes called pulsers and act to provide the short pulses of power to the magnetron. This technology is known as Pulsed power. In this way, the transmitted pulse of RF radiation is kept to a defined, and usually very short, duration. Modulators consist of a high voltage pulse generator formed from a HV supply, a pulse forming network or line (PFN) and a high voltage switch such as a thyratron.

A klystron tube is an amplifier, so it can be modulated by its low power input signal.

Radar coolant

Coolanol and PAO (poly-alpha olefin) are the two main coolants used to cool airborne radar equipment today.[citation needed]

The U.S. Navy has instituted a program named Pollution Prevention (P2) to reduce or eliminate the volume and toxicity of waste, air emissions, and effluent discharges. Because of this Coolanol is used less often today.

PAO is a synthetic lubricant composition is a blend of a polyol ester admixed with effective amounts of an antioxidant, yellow metal pacifier and rust inhibitors. The polyol ester blend includes a major proportion of poly(neopentyl polyol) ester blend formed by reacting poly(pentaerythritol) partial esters with at least one C7 to C12 carboxylic acid mixed with an ester formed by reacting a polyol having at least two hydroxyl groups and at least one C8-C10 carboxylic acid. Preferably, the acids are linear and avoid those which can cause odours during use. Effective additives include secondary arylamine antioxidants, triazole derivative yellow metal pacifier and an amino acid derivative and substituted primary and secondary amine and/or diamine rust inhibitor.

A synthetic coolant/lubricant composition, comprising an ester mixture of 50 to 80 weight percent of poly(neopentyl polyol) ester formed by reacting a poly(neopentyl polyol) partial ester and at least one linear monocarboxylic acid having from 6 to 12 carbon atoms, and 20 to 50 weight percent of a polyol ester formed by reacting a polyol having 5 to 8 carbon atoms and at least two hydroxyl groups with at least one linear monocarboxylic acid having from 7 to 12 carbon atoms, the weight percents based on the total weight of the composition.

Radar functions and roles

Surface search radar display commonly found on ships

Detection and search radars

- Early Warning (EW) Radar Systems
  - Early Warning Radar
  - Ground Control Intercept (GCI) Radar
  - Airborne Early Warning (AEW)
  - Over-the-Horizon (OTH) Radar
- Target Acquisition (TA) Radar Systems
  - Surface-to-Air Missile (SAM) Systems
  - Anti-Aircraft Artillery (AAA) Systems
- Surface Search (SS) Radar Systems

    - Surface Search Radar
    - Coastal Surveillance Radar
    - Harbour Surveillance Radar
    - Antisubmarine Warfare (ASW) Radar
- Height Finder (HF) Radar Systems
- Gap Filler Radar Systems

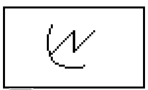
Threat radars

- Target Tracking (TT) Systems
    - AAA Systems
    - SAM Systems
    - Precision Approach Radar (PAR) Systems
- Multi-Function Systems
    - Fire Control (FC) Systems
        - Acquisition Mode
        - Semiautomatic Tracking Mode
        - Manual Tracking Mode
    - Airborne Intercept (AI) Radars
        - Search Mode
        - TA Mode
        - TT Mode
        - Target Illumination (TI) Mode
        - Missile Guidance (MG) Mode

Missile guidance systems

- Air-to-Air Missile (AAM)
- Air-to-Surface Missile (ASM)
- SAM Systems
- Surface-to-Surface Missiles (SSM) Systems

Battlefield and reconnaissance radar



Military map marking symbol *Radar* as of NATO standard APP-6a

- Battlefield Surveillance Systems
    - Battlefield Surveillance Radars
- Countermortar/Counterbattery Systems
    - Shell Tracking Radars
- Air Mapping Systems
    - Side Looking Airborne Radar (SLAR)
    - Synthetic Aperture Radar (SAR)
    - Perimeter Surveillance Radar (PSR)

Air Traffic Control and navigation

Air traffic control radar at London Heathrow Airport

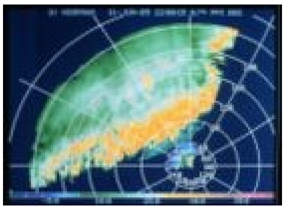
- Air Traffic Control Systems
  - Air Traffic Control (ATC) Radars
  - Secondary Surveillance Radar (SSR) (Airport Surveillance Radar)
  - Ground Control Approach (GCA) Radars
  - PAR Systems
- Distance Measuring Equipment (DME)
- Radio Beacons
- Identification Friend or Foe (IFF) Systems
  - IFF Interrogator
  - IFF Transponder
- Altimeter (AL) Radar Systems
- Terrain-Following Radar (TFR) Systems

Space and range instrumentation radar systems

- Space (SP) Tracking Systems
- Range Instrumentation (RI) Systems
- Video Relay/Downlink Systems
- Space-Based Radar

Weather-sensing Radar systems

- Weather radar
- Wind profilers



Storm front reflectivities on a Weather radar screen (NOAA) Wind profiling radar

Radars for biological research

- Bird radar
- Insect radar

- Surveillance radar (mostly X and S band, i.e. primary ATC Radars)
- Tracking radar (mostly X band, i.e. Fire Control Systems)

Video games

Many video games dedicate a small portion of the screen to a subsidiary display that indicates the position of the player relative to other objects and players. The games sometimes refer to this sub-display as the 'radar', although it is not usually meant to represent a real radar system.

Rocket engine

A **rocket engine** is a reaction engine that can be used for spacecraft propulsion as well as terrestrial uses, such as missiles. Rocket engines take their reaction mass from within the vehicle and form it into a high speed jet, obtaining thrust in accordance with Newton's third law. Most rocket engines are internal combustion engines, although non combusting forms exist.

Principle of operation

Rocket engines give part of their thrust due to unopposed pressure on the combustion chamber

Classic rocket engines produce a high temperature, hypersonic gaseous exhaust. This is most often achieved by the combustion of solid, liquid or gaseous propellant, containing oxidiser and a fuel, within a combustion chamber at high pressure. Alternatively, a chemically inert reaction mass can be heated to high temperature using a high energy power source.

The hot gas produced is then allowed to escape through a narrow hole (the 'throat'), into a high-expansion ratio nozzle. The effect of the nozzle is to dramatically accelerate the mass, converting most of the thermal energy into kinetic energy. The large bell or cone shaped expansion nozzle gives a rocket engine its characteristic shape. Exhaust speeds as high as 10 times the speed of sound at sea level are not uncommon.

Part of the rocket engine's thrust comes from the gas pressure inside the combustion chamber but the majority comes from the pressure against the inside of the expansion nozzle. Inside the combustion chamber the gas produces a similar force against all the sides of the combustion chamber but the throat gives no force producing an unopposed resultant force from the diametrically opposite end of the chamber. As the gases (adiabatically) expand inside the nozzle they press against the bell's walls forcing the rocket engine in one direction, and accelerating the gases in the opposite direction.

For optimum performance hot gas is used because it maximises the speed of sound at the throat — for aerodynamic reasons the flow goes sonic ("chokes") at the throat, so the highest speed there is desirable. By comparison, at room temperature the speed of sound in air is about 340m/s, the speed of sound in the hot gas of a rocket engine can be over 1700m/s.

The expansion part of the rocket nozzle then multiplies the speed of the flow by a further factor, typically between 1.5 and 4 times, giving a highly collimated exhaust jet. The speed ratio of a rocket nozzle is mostly determined by its area expansion ratio — the ratio of the area of the throat to the area at the exit, but details of the gas properties are also important. Larger ratio nozzles are more massive and bulkier, but they are able to extract more heat from the combustion gases, which become lower in pressure and colder, but also faster.

A significant complication arises when launching a vehicle from the Earth's surface as the ambient atmospheric pressure changes with altitude. For maximum performance it turns out that the pressure of the gas leaving a rocket nozzle should be the same as ambient pressure; if lower the vehicle will be slowed by the difference in pressure between the top of the engine and the exit, if higher then this represents pressure that the bell has not turned into thrust. To achieve this ideal, the diameter of the nozzle would need to increase with altitude, which is difficult to arrange. A compromise nozzle is generally used and some percentage reduction in performance occurs. To improve on this, various exotic nozzle designs such as the plug nozzle, stepped nozzles, the expanding nozzle and the aerospike have been proposed, each having some way to adapt to changing ambient air pressure and each allowing the gas to expand further against the nozzle giving extra thrust at higher altitude.

Thermal issues

The reaction mass's combustion temperatures can fairly typically reach ~3500 K (~5800 F) which is often far higher than the melting point of the nozzle and combustion chamber materials (~1200K for copper). Indeed many construction materials can make perfectly acceptable propellants in their own right. It is important that these materials be prevented from combusting, melting or vapourising to the point of failure. Materials technology could potentially place an upper limit on the exhaust temperature of chemical rockets.

To avoid this problem rockets can use ablative materials that erode in a controlled fashion, or very high temperature materials, such as graphite, ceramics or certain exotic metals.

Alternatively, rockets may use more common construction materials such as aluminum, steel, nickel or copper alloys and employ cooling systems that prevent the construction material itself becoming too hot. Regenerative cooling, where the propellant is passed through tubes around the combustion chamber or nozzle, and other techniques such as curtain cooling or film cooling, may be employed to give essentially unlimited nozzle and chamber life. These techniques ensure that the gas boundary layer touching the material is kept below the point where the material would fail.

Mechanical issues

The combustion chamber is often under substantial pressure, typically 10-200 bar, higher pressures giving better performance. This causes the outermost part of the chamber to be under very large hoop stresses.

Worse, due to the high temperatures created in rocket engines the materials used tend to have a significantly lowered working tensile strength.

Safety

Rocket engines are tested at a test facility before being put into production.

Rockets have a reputation for unreliability and danger; especially catastrophic failures.

Contrary to this reputation, carefully designed rockets can be made arbitrarily reliable. In military use, rockets are not unreliable. However, one of the main non-military uses of rockets is for orbital launch. In this application, the premium is on minimum weight, and it is difficult to achieve high reliability and low weight simultaneously. In addition, if the number of flights launched is low, there is a very high chance of a design, operations or manufacturing error causing destruction of the vehicle. Essentially, as of 2006 all launch vehicles are test vehicles by normal aerospace standards.

The X-15 rocket plane achieved a 0.5% failure rate, with a single catastrophic failure during ground test, and the SSME has managed to avoid catastrophic failures in over 300 engine-flights.

Noise

The Saturn V launch was detectable on seismometers a considerable distance from the launch site. As the hypersonic exhaust mixes with the ambient air, shock waves are formed. The sound intensity from these shock waves depends on the size of the rocket, and on large rockets can actually kill. The Space Shuttle generates over 200 dB(A) of noise around its base.

Generally speaking noise is most intense when a rocket is close to the ground, since the noise from the engines radiate up away from the plume, as well as reflecting off the ground. This noise can be reduced somewhat by flame trenches with roofs, by water injection around the plume and by deflecting the plume at an angle.

Chemistry

Contrary to popular belief, while rocket propellants require reasonably high energy per kilogram, many common materials are more energetic; for example petrol/gasoline or paraffin has as much energy as a rocket fuel and far more than the fuel/oxidiser mix used for rocket fuels. This is due to the necessity of the propellant containing large amounts of oxidiser, normal propellants used on earth for say, Turbojet engines, are reacted with the atmosphere and hence can have several times higher energy density.

Good rocket propellants require large quantities of hydrogen in the propellant, as this gives the highest exhaust speeds primarily due to the low molecular mass; but this is not the whole story.[1]

Programs exist to predict the performance of propellants in rocket engines.(e.g.[2]).

Ignition

With liquid propellants immediate ignition of the propellants as they first enter the combustion chamber is essential.

Failure to ignite within milliseconds causes too much liquid propellant to be within the chamber, and if/when ignition occurs the amount of hot gas created will often exceed the maximum design pressure of the chamber. The pressure vessel will often fail catastrophically. This is sometimes called a *hard start*.

Ignition can be achieved by a number of different methods; a pyrotechnic charge can be used, the propellants can ignite spontaneously on contact (hypergolic), a plasma torch can be used, or electric spark plugs may be employed.

Gaseous propellants generally will not cause hard starts, with rockets the total injector area is less than the throat thus the chamber pressure tends to ambient prior to ignition and high pressures cannot form even if the entire chamber is full of flammable gas at ignition.

Solid propellants are usually ignited with one-shot pyrotechnic devices.

Once ignited, rocket chambers are self sustaining and igniters are not needed. Indeed chambers often spontaneously reignite if they are restarted after being shut down for a few seconds.

However, when cooled, many rockets cannot be started more than once without minor maintenance, such as replacement of the pyrotechnic igniter.

Types of rocket engines

Type	Description	Advantages	Disadvantages
water rocket	Partially filled pressurised carbonated drinks container with tail and nose weighting	Very simple to build	altitude limited to a few hundred feet or so
cold gas thruster	A non combusting form, used for attitude jets	Non contaminating exhaust	Low performance
Solid rocket	Ignitable, self sustaining solid fuel/oxidiser mixture ("grain") with central hole and nozzle	Simple, often no moving parts, reasonably good mass fraction, reasonable I_{sp}. A thrust schedule can be designed into the grain.	Once lit, extinguishing it is difficult although often possible, cannot be throttled in real time; handling issues from ignitable mixture, lower performance than liquid rockets, if grain cracks it can block nozzle with disastrous results, cracks burn and widen during burn. Refuelling grain harder than simply filling tanks.
Hybrid rocket	Separate oxidiser/fuel, typically oxidiser is liquid and kept in a tank, the other solid with central hole	Quite simple, solid fuel is essentially inert without oxidiser, safer; cracks do not escalate, throttleable and easy to switch off.	Some oxidisers are monopropellants, can explode in own right; mechanical failure of solid propellant can block nozzle, central hole widens over burn and negatively affects mixture ratio. Replacing grain harder than simply refuelling tanks.
Monopropellant rocket	Propellant such as Hydrazine, Hydrogen Peroxide or Nitrous Oxide, flows over catalyst and exothermically decomposes and hot gases are emitted through nozzle	Simple in concept, throttleable, low temperatures in combustion chamber	catalysts can be easily contaminated, monopropellants can detonate if contaminated or provoked, I_{sp} is perhaps 1/3 of best liquids
Bipropellant rocket	Two fluid (typically liquid) propellants are introduced through injectors into combustion chamber and burnt	Up to ~99% efficient combustion with excellent mixture control, throttleable, can be used with turbopumps which permits incredibly lightweight tanks, can be safe with extreme care	Pumps needed for high performance are expensive to design, huge thermal fluxes across combustion chamber wall can impact reuse, failure modes include major explosions, a lot of plumbing is needed.
Dual mode	Rocket takes off as a	Simplicity and ease of	Lower performance than

propulsion rocket	bipropellant rocket, then turns to using just one propellant as a monopropellant	control	bipropellants
Tripropellant rocket	Three different propellants (usually hydrogen, hydrocarbon and liquid oxygen) are introduced into a combustion chamber in variable mixture ratios, or multiple engines are used with fixed propellant mixture ratios and throttled or shut down	Reduces take-off weight, since hydrogen is lighter; combines good thrust to weight with high average I_{sp}, improves payload for launching from Earth by a sizeable percentage	Similar issues to bipropellant, but with more plumbing, more R&D
Air-augmented rocket	Essentially a ramjet where intake air is compressed and burnt with the exhaust from a rocket	Mach 0 to Mach 4.5+ (can also run exoatmospheric), good efficiency at Mach 2 to 4	Similar efficiency to rockets at low speed or exoatmospheric, inlet difficulties, a relatively undeveloped and unexplored type, cooling difficulties, very noisy, thrust/weight ratio is similar to ramjets.
Turborocket	A combined cycle turbojet/rocket where an additional oxidizer such as oxygen is added to the airstream to increase maximum altitude	Very close to existing designs, operates in very high altitude, wide range of altitude and airspeed	Atmospheric airspeed limited to same range as turbojet engine, carrying oxidizer like LOX can be dangerous. Much heavier than simple rockets.
Precooled jets / LACE (combined cycle with rocket)	Intake air is chilled to very low temperatures at inlet before passing through a ramjet or turbojet engine. Can be combined with a rocket engine for orbital insertion.	Easily tested on ground. High thrust/weight ratios are possible (~14) together with good fuel efficiency over a wide range of airspeeds, mach 0-5.5+; this combination of efficiencies may permit launching to orbit, single stage, or very rapid intercontinental travel.	Exists only at the lab prototyping stage. Examples include RB545, SABRE, ATREX

Electric heating

Type	**Description**	**Advantages**	**Disadvantages**
Resistojet rocket (electric heating)	A monopropellant is electrically heated by a filament for extra performance	Higher I_{sp} than monopropellant alone, about 40% higher.	Uses a lot of power and hence gives typically low thrust

Arcjet rocket (chemical burning aided by electrical discharge)	Similar to resistojet in concept but with inert propellant, except an arc is used which allows higher temperatures	1600 seconds I_{sp}	Very low thrust and high power, performance is similar to Ion drive.
Pulsed plasma thruster (electric arc heating; emits plasma)	Plasma is used to erode a solid propellant	High I_{sp}, can be pulsed on and off for attitude control	Low energetic efficiency
Variable specific impulse magnetoplasma rocket	Microwave heated plasma with magnetic throat/nozzle	Variable I_{sp} from 1000 seconds to 10,000 seconds	similar thrust/weight ratio with ion drives (worse), thermal issues, as with ion drives very high power requirements for significant thrust, really needs advanced nuclear reactors, never flown, requires low temperatures for superconductors to work

Solar heating

The Solar thermal rocket would make use of solar power to directly heat reaction mass, and therefore does not require an electrical generator as most other forms of solar-powered propulsion do. A solar thermal rocket only has to carry the means of capturing solar energy, such as concentrators and mirrors. The heated propellant is fed through a conventional rocket nozzle to produce thrust. The engine thrust is directly related to the surface area of the solar collector and to the local intensity of the solar radiation.

Type	Description	Advantages	Disadvantages
Solar thermal rocket	Propellant is heated by solar collector	Reasonably simple, good performance with liquid hydrogen propellant, adequate performance with in-situ water for short-range interplanetary flight	only useful once in space, as thrust is fairly low, but hydrogen is not easily stored in space, otherwise moderate/low I_{sp} if higher molecular mass propellants are used

Nuclear heating

Nuclear propulsion includes a wide variety of propulsion methods that use some form of nuclear reaction as their primary power source. Various types of nuclear propulsion have been proposed, and some of them tested, for spacecraft applications:

Type	Description	Advantages	Disadvantages
Radioisotope rocket/"Poodle thruster" (radioactive decay energy)	Heat from radioactive decay is used to heat hydrogen	about 700-800 seconds, almost no moving parts	low thrust/weight ratio
Nuclear thermal rocket (nuclear fission energy)	propellant (typ. hydrogen) is passed through a nuclear reactor to heat to high	I_{sp} can be high, perhaps 900 seconds or more, above unity thrust/weight ratio with some designs	Maximum temperature is limited by materials technology, some radioactive particles can be present in exhaust in some designs, nuclear reactor

	temperature		shielding is heavy, unlikely to be permitted from surface of the Earth, thrust/weight ratio is not high
Gas core reactor rocket (nuclear fission energy)	Nuclear reaction using a gaseous state fission reactor in intimate contact with propellant	Very hot propellant, not limited by keeping reactor solid, I_{sp} between 1500 and 3000 seconds but with very high thrust	difficulties in heating propellant without losing fissionables in exhaust, exhaust inherently highly radioactive, massive thermal issues particularly for nozzle/throat region
Fission-fragment rocket (nuclear fission energy)	Fission products are directly exhausted to give thrust		Theoretical only at this point
Fission sail (nuclear fission energy)	A sail material is coated with fissionable material on one side	No moving parts, works in deep space	
Nuclear salt-water rocket (nuclear fission energy)	Nuclear salts are held in solution, caused to react at nozzle	Very high I_{sp}, very high thrust	Thermal issues in nozzle, propellant could be unstable, highly radioactive exhaust
Nuclear pulse propulsion (exploding fission/fusion bombs)	Shaped nuclear bombs are detonated behind vehicle and blast is caught by a 'pusher plate'	Very high I_{sp}, very high thrust/weight ratio, no show stoppers are known for this technology	Never been tested, pusher plate may throw off fragments due to shock, minimum size for nuclear bombs is still pretty big, expensive at small scales, nuclear treaty issues
Antimatter catalyzed nuclear pulse propulsion (fission and/or fusion energy)	Nuclear pulse propulsion with antimatter assist for smaller bombs	Smaller sized vehicle might be possible	Containment of antimatter, production of antimatter in macroscopic quantities isn't currently feasible
Fusion rocket (nuclear fusion energy)	Fusion is used to heat propellant	Very high exhaust velocity	Largely beyond current state of the art
Antimatter rocket (annihilation energy)	Antimatter reaction is used to heat propellant	Extremely energetic, very high exhaust velocity is possible on paper	Antimatter containment issues, thermal issues, beyond current state of the art.

S-13 Aircraft Weapon System

S-13 AIRCRAFT WEAPON SYSTEM: CURRENT STATUS AND PROSPECTS FOR GROWTH

S-13 rocket with different warheads

The weapon system had been built to help tackle a major task facing front-line and army-level aviation that of destroying hostile aircraft kept in a variety of concrete shelters, as well as destroy runways, command posts, communications nodes and other fortified facilities.

In the mid-1980s the S-13 airbcraft rocket pod system (NARV), designed to include the B-13L five-tube launcher and a complement of 122mm S-13T, S-13OF and S-13D rockets, entered service with the Soviet Air Force. The weapon system had been built to help tackle a major task facing fron-line and army-level aviation, that of destroying hostile aircraft kept in a variety of concrete shelters, as well as destroy runways, command posts, communications centers and other fortified facilities. The S-13T baseline rocket was developed for that purpose. It featured an extended high-power, solid-propellant rocket engine and a double-module enetrator-type warhead. Later the product was expanded to produce the S-13OF and S-13D rockets, carrying high-explosive warheads designed to neutralize light-armored and vulnerable materiel and personnel.

Given the ingenious engineering solutions employed in developing the munition (an optimized TNT charge design boasting no armor shielding, the S-13T warhead's self-contained modules in a tandem configuration, casting technology for producing S-13OF warhead casings that uses a diamond-shaped pattern to achieve the desired fragmentation effect, etc.), we have managed to build highly-lethal weapons and maintain the most affordable per-unit costs (the latter traditionally being a crucial requirement for unguided aircraft weapons). For example, the fire power of a Su-25 attack aircraft, armed with high-explosive rockets magnified by its inherent agility in the air, is equal to salvo firings of several Grad MLRS systems. The warhead's principal characteristics are given in the Table below.

Basic Characteristics of the S-13 Rockets

S-13T 	Double-module high-explosive warhead. Penetrates 6 meters of soil plus 1 meter of concrete. Craters 20 square meters of runway. Warhead weight - 37 kg Maximum effective range - 4,000 m
S-130F 	High-explosive with 450 precut fragments, each weighing between 25- 35 grams. Warhead weight - 33 kg
S-13D 	Vacuum-type. Warhead weight - 32 kg TNT equivalent - 35 kg

At the same time, given general trends in the development of armaments and military hardware, the upgrading of the S-13 weapon system to still higher performance levels has become a pressing challenge. To expand the tactical uses and combat capabilities of the existing aircraft, the first task is to extend effective ranges and augment firing accuracy rates at maximum engagement distances. Given the results of tests done at the Institute of Applied Physics, these goals could be achieved by the use of a semiactive laser homing head with terminal corrector. A semiactive seeker is chosen because of the low-contrast nature of likely targets in the IR and millimeter-wave bands. In addition, semiactive laser seeker heads are generally simple in design, affordable, can be used with the baseline S-13 core components (launcher and sight system) without heavy reengineering, and allow for refitting of in-service unguided aircraft rockets with a smart-munition capability.

S-13 Rockets

To increase the rocket's sighting launch range, a more powerful engine needs to be installed. This increase is achievable via the use of a composite solid propellant. This option would not only enable the rockets to fly faster, which would obviously translate into greater penetrating ability, but also increase the rocket's size and weight reserve, thereby making the standard product upgradable into a smart munition. Should these solutions be applied, the penetrator-type rocket's lethality would be enhanced at least 5-7 times.

Later the S-13 system's modernization should be focused on the B-13L launcher in order to secure aircraft-to-rocket interface, in order to:
- activate the rocket's fuse option depending on the type of target;
- assure the use of a mixed complement of rockets;
- indicate on the cockpit display the types and quantity of rockets available in the pod.

These affordable improvements would allow S-13 aircraft rocket pods to keep their proven strengths while radically boosting lethality.

Semi-active radar homing

Semi-active radar homing, or **SARH**, is a common type of missile guidance system, perhaps the most common type for longer range air-to-air and surface-to-air missile systems. The name refers to the fact that the missile itself is only a passive detector of a radar signal – provided by an external ("offboard") source – as it reflects off the target.

NATO brevity code for a semi-active radar homing missile launch is **Fox One**.

Concept

The basic concept of SARH is that almost all detection and tracking systems consist of a radar system, so duplicating this hardware on the missile itself is wasted. In addition, the resolution of a radar is strongly related to the physical size of the antenna, in the small nose cone of a missile there isn't enough room to provide the sort of accuracy needed for guidance. Instead the larger radar dish on the ground or launch aircraft will provide the needed signal and tracking logic, and the missile simply has to listen to the target reflected signal and point itself in the right direction. Additionally, the missile will listen rearward to the the launch platform's transmitted signal as a reference, enabling it to avoid some kinds of radar jamming distractions offered by the target.

Contrast this with beam riding systems, in which the radar is pointed at the target and the missile keeps itself centered in the beam by listening to the signal at the rear of the missile body. In the SARH system the missile listens for the reflected signal at the nose, and is still responsible for providing some sort of "lead" guidance. The advantages are twofold. One is that a radar signal is "fan shaped" growing larger, and therefore less accurate, with distance. This means that the beam riding system is not terribly accurate at long ranges, while SARH is largely independent of range and grows more accurate as it approaches the target—the "source" of the signal it listens for. Another addition is that a beam riding system must accurately track the target at high speeds, typically requiring one radar for tracking and another "tighter" beam for guidance. The SARH system needs only one radar set to a wider pattern.

Continuous-wave radar

Modern SARH systems use continuous-wave (CW) radar for guidance. Even though most modern fighter radars are pulse Doppler sets, most have a CW function to guide radar missiles. A few Soviet aircraft, such as some versions of the MiG-23 and MiG-27, used an auxiliary guidance pod or aerial to provide a CW signal. Vympel R-33 AA missile for MiG-31 interceptor uses SARH as the main type of guidance (with supplement of inertial guidance on initial stage).

SARH missiles require the tracking radar to lock on to the target and then illuminate it for the entire duration of the missile's flight. This could leave the launch aircraft vulnerable to counter attack, as well giving the target's electronic warning systems time to detect the attack and engage countermeasures. Because most SARH missiles require guidance during their entire flight, only one target per radar emitter can be engaged at a time.

Electronic counter-countermeasure (ECCM)

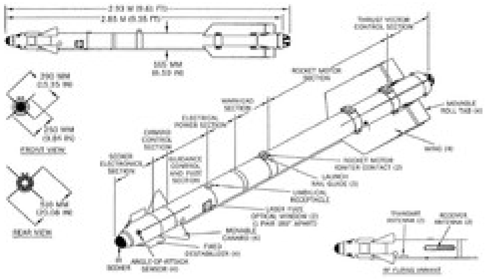
Recent-generation SARH weapons have superior electronic counter-countermeasure (**ECCM**) capability, but the system still has fundamental limitations. Some newer missiles, such as the Standard SM-2 incorporate terminal semi-active radar homing (TSARH). TSARH missiles use inertial guidance for most of their flight, only activating their SARH system for the final attack. This can keep the target from realising it is under attack until shortly before the missile strikes. Since the missile only requires guidance during the terminal phase, each radar emitter can be used to engage more targets. Some of these weapons, like the SM-2, allow the firing aircraft to update the missile with mid-course updates via datalink.

Combat record

The combat record of SARH missiles was unimpressive during the Vietnam War. USAF and US Navy fighters armed with AIM-7 Sparrow attained a success rate of barely 10%, which tended to amplify the effect of deleting the gun on most F-4 Phantoms, which carried 4 Sparrows. Some of the failures were attributable to mechanical failure of 1960s era electronics which could be disturbed by pulling a cart over uneven pavement, or pilot error; the intrinsic accuracy of these weapons was relatively low relative to Sidewinder and guns. However, since Desert Storm, most F-15 Eagle combat victories have been scored with the Sparrow at BVR ranges. Also notable is the combat victory of an Israeli Eagle on one fast, high-flying MiG-25 Foxbat.

Vympel R-73

Vympel R-73	



Vympel R-73

Basic data	
Function	Short-range Air to Air Missile
Manufacturer	Vympel NPO
Entered service	
General characteristics	
Engine	solid-fuel rocket engine
Launch mass	105 kg (231 lb)
Length	2900 mm (9 ft 6 in)
Diameter	170 mm (6.7 in)
Wingspan	510 mm (20 in)
Speed	Mach 2.5
Range	30 km (18.75 mi)
Warhead	7.4 kg (16.3 lb)
Guidance	infrared homing
Launch platform	• MiG-29, Su-27, Su-35 • MiG-21, MiG-23, Sukhoi Su-24, Su-25 • Mil Mi-24, Mil Mi-28, Kamov Ka-50

The **Vympel R-73** (NATO reporting name **AA-11** ***Archer***) developed by Vympel machine Building Design Bureau, is the most modern Russian short-range air-to-air missile. It is generally considered one of the most formidable modern dogfight weapons.

Development

The **R-73** was developed to replace the earlier R-60 (AA-8 'Aphid') weapon for short-range use by Soviet fighter aircraft. Work began in 1973, and the first missiles entered service in 1985.

The R-73 is an infrared-guided (heat-seeking) missile with a sensitive, cryogenic cooled seeker with a substantial "off-boresight" capability: the seeker can "see" targets up to 60° off the missile's centerline. It can be targeted by a helmet-mounted sight (HMS) allowing pilots to designate targets by looking at them. Minimum engagement range is about 300 meters, with maximum aerodynamic range of nearly 30 km (18.75 mi) at altitude.

The R-73 is a highly maneuverable missile that in most respects is believed to be superior to the United States AIM-9M Sidewinder a fact demonstrated by the reunified Luftwaffe winning all dogfight engagements with their Mig 29/R-73 combine against the F-16/Aim 9M combination fielded by the USAF, prompting the development of Sidewinder and other SRM successors like AIM-132 ASRAAM, IRIS-T, MICA IR, Python IV and the latest Sidewinder variant, AIM-9X, that entered squadron service in 2003.

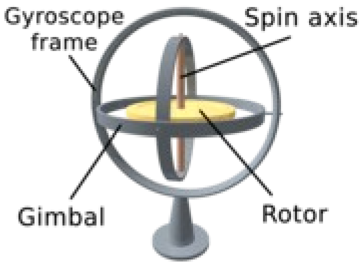
From 1994 the R-73 has been upgraded in production to **R-74EM** standard (originally **R-73M**), which entered CIS service in 1997. The R-74EM has greater range and a wider seeker angle (to 60° off-boresight), as well as improved IRCCM (InfraRed Counter-Counter Measures).

The weapon is used by the MiG-29, Su-27, Su-32 and Su-35, and can be carried by newer versions of the MiG-21, MiG-23, Sukhoi Su-24, and Su-25 aircraft. It can also be carried by Russian attack helicopters, including the Mil Mi-24, Mil Mi-28, and Kamov Ka-50.

R-73 before R-77

R-73Ae, R-27R1(AeR1), R-27T1(AeT1), and Kh-59MAe at MACS, Zhukovski, 1999.

Gyroscope



A gyroscope

For other uses, see Gyroscope (disambiguation).

A **gyroscope** is a device for measuring or maintaining orientation, based on the principle of conservation of angular momentum. The essence of the device is a spinning wheel on an axle. The device, once spinning, tends to resist changes to its orientation due to the angular momentum of the wheel. In physics this phenomenon is also known as **gyroscopic inertia** or rigidity in space.

Description and diagram

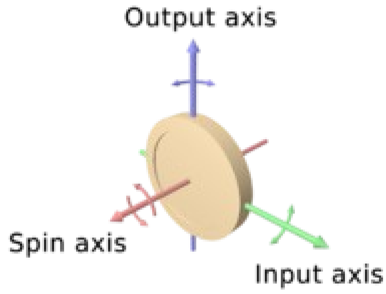


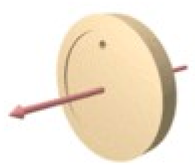
Diagram of a gyro wheel. Reaction arrows about the output axis (blue) correspond to forces applied about the input axis (green), and vice versa.

Within mechanical combinations or devices constituting portions of machines, a conventional *gyroscope* is a mechanism comprising a rotor journaled to spin about one axis, the journals of the rotor being mounted in an inner gimbal or ring, the inner gimbal being journaled for oscillation in an outer gimbal which in turn is journaled for oscillation relative to a support. The outer gimbal or ring is mounted so as to pivot about an axis in its own plane determined by the support. The outer gimbal possesses one degree of rotational freedom and its axis possesses none. The inner gimbal is mounted in the outer gimbal so as to pivot about an axis in its own plane, which axis is always perpendicular to the pivotal axis of the outer gimbal.

The axle of the spinning wheel defines the spin axis. The inner gimbal possesses two degrees of rotational freedom and its axis possesses one. The rotor is journaled to spin about an axis which is always perpendicular to the axis of the inner gimbal. Hence the rotor possesses three degrees of rotational freedom and its axis possesses two. The wheel responds to a force applied about the input axis by a reaction force about the output axis. The 3 axes are perpendicular, and this cross-axis response is the simple essence of the gyroscopic effect.

The behaviour of a gyroscope can be most easily appreciated by consideration of the front wheel of a bicycle. If the wheel is leaned away from the vertical so that the top of the wheel moves to the left, the forward rim of the wheel also turns to the left. In other words, rotation on one axis of the turning wheel produces rotation of the third axis.

A **gyroscope flywheel** will roll or resist about the output axis depending upon whether the output gimbals are of a free- or fixed- configuration. Examples of some free-output-gimbal devices would be the attitude reference gyroscopes used to sense or measure the pitch, roll and yaw attitude angles in a spacecraft or aircraft.



Animation of a gyro wheel in action

The center of gravity of the rotor can be in a fixed position. The rotor simultaneously spins about one axis and is capable of oscillating about the two other axes, and thus, except for its inherent resistance due to rotor spin, it is free to turn in any direction about the fixed point. Some gyroscopes have mechanical equivalents substituted for one or more of the elements, e.g., the spinning rotor may be suspended in a fluid, instead of being pivotally mounted in gimbals. A control moment gyroscope (CMG) is an example of a fixed-output-gimbal device that is used on spacecraft to hold or maintain a desired attitude angle or pointing direction using the gyroscopic resistance force.

In some special cases, the outer gimbal (or its equivalent) may be omitted so that the rotor has only two degrees of freedom. In other cases, the center of gravity of the rotor may be offset from the axis of oscillation, and thus the center of gravity of the rotor and the center of suspension of the rotor may not coincide.

History

The gyroscope effect was discovered in 1817 by Johann Bohnenberger; the gyroscope was invented, and the effect named after it, in 1852 by Léon Foucault for an experiment involving the rotation of the Earth. Foucault's experiment to see (*skopeein*, to see) the Earth's rotation (*gyros*, circle or rotation) was unsuccessful due to friction, which effectively limited each trial to 8 to 10 minutes, too short a time to observe significant movement. In the 1860s, however, electric motors made the concept feasible, leading to the first prototype gyrocompasses; the first functional marine gyrocompass was developed between 1905 and 1908 by German inventor Hermann Anschütz-Kaempfe. The American Elmer Sperry followed with his own design in 1910, and other nations soon realized the military importance of the invention—in an age in which naval might was the most significant measure of military power—and created their own gyroscope industries. The Sperry Gyroscope Company quickly expanded to provide aircraft and naval stabilizers as well, and other gyroscope developers followed suit.[1]

In the first several decades of the 20th century, other inventors attempted (unsuccessfully) to use gyroscopes as the basis for early black box navigational systems by creating a stable platform from which accurate acceleration measurements could be performed (in order to bypass the need for star sightings to calculate position). Similar principles were later employed in the development of inertial guidance systems for ballistic missiles.[2]

Properties

A gyroscope in operation with freedom in all three axis. The rotor will maintain its spin axis direction regardless of the orientation of the outer frame.

A gyroscope exhibits a number of behaviours including precession and nutation. Gyroscopes can be used to construct gyrocompasses which complement or replace magnetic compasses (in ships, aircraft and spacecraft, vehicles in general), to assist in stability (bicycle, Hubble Space Telescope, ships, vehicles in general) or be used as part of an Inertial guidance system. Gyroscopic effects are used in toys like yo-yos and Powerballs. Many other rotating devices, such as flywheels, behave gyroscopically although the gyroscopic effect is not used.

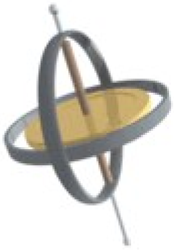
The fundamental equation describing the behavior of the gyroscope is:

$$\boldsymbol{\tau} = \frac{d\mathbf{L}}{dt} = \frac{d(I\boldsymbol{\omega})}{dt} = I\boldsymbol{\alpha}$$

where the vectors $\boldsymbol{\tau}$ and $\mathbf{L}$ are, respectively, the torque on the gyroscope and its angular momentum, the scalar I is its moment of inertia, the vector $\boldsymbol{\omega}$ is its angular velocity, and the vector $\boldsymbol{\alpha}$ is its angular acceleration.

It follows from this that a torque $\boldsymbol{\tau}$ applied perpendicular to the axis of rotation, and therefore perpendicular to $\mathbf{L}$, results in a motion perpendicular to both $\boldsymbol{\tau}$ and $\mathbf{L}$. This motion is called *precession*. The angular velocity of precession $\boldsymbol{\Omega}_P$ is given by the cross product:

$$\boldsymbol{\tau} = \boldsymbol{\Omega}_P \times \mathbf{L}$$



Precession on a gyroscope

Precession can be demonstrated by placing a spinning gyroscope with its axis horizontal and supported loosely (frictionless toward precession) at one end. Instead of falling, as might be expected, the gyroscope appears to defy gravity by remaining with its axis horizontal, when the other end of the axis is left unsupported and the free end of the axis slowly describes a circle in a horizontal plane, the resulting precession turning. This effect is explained by the above equations. The torque on the gyroscope is supplied by a couple of forces: gravity acting downwards on the device's centre of mass, and an equal force acting upwards to support one end of the device. The motion resulting from this torque is not downwards, as might be intuitively expected, causing the device to fall, but perpendicular to both the gravitational torque (downwards) and the axis of rotation (outwards from the point of support), i.e. in a forward horizontal direction, causing the device to rotate slowly about the supporting point.

As the second equation shows, under a constant torque due to gravity or not, the gyroscope's speed of precession is inversely proportional to its angular momentum. This means that, for instance, if friction causes the gyroscope's spin to slow down, the rate of precession increases. This continues until the device is unable to rotate fast enough to support its own weight, when it stops precessing and falls off its support, mostly because friction against precession cause another precession that goes to cause the fall.

By convention, these three vectors, torque, spin, and precession, are all oriented with respect to each other according to the right-hand rule.

To easily ascertain the direction of gyro effect, simply remember that a rolling wheel tends, when entering a corner, to turn over to the inside.

Gyrostat

A **gyrostat** is a variant of the gyroscope. The first gyrostat was designed by Lord Kelvin to illustrate the more complicated state of motion of a spinning body when free to wander about on a horizontal plane, like a top spun on the pavement, or a hoop or bicycle on the road. It consists essentially of a massive flywheel concealed in a solid casing. Its behaviour on a table, or with various modes of suspension or support, serves to illustrate the curious reversal of the ordinary laws of static equilibrium due to the gyrostatic behaviour of the interior invisible flywheel when rotated rapidly.

Small manually-spun gyrostats are sold as childrens' toys under the brand name Wizzer.